Hands - On Machine Learning with Scikit - Learn Keras and TensorFlow 3rd edition GÃ©ron Test Bank

SKU: 9781098125974-TEST-BANK

Hands-On Machine Learning with Scikit-Learn and TensorFlow

ISBN 9781098125974 | Chapter 1: The Machine Learning Landscape

Total Questions: 150

Multiple Choice

Q1.

Which definition best matches Machine Learning as introduced in the chapter?

A) The field of storing large volumes of data for later retrieval
B) The science and art of programming computers so they can learn from data ✓ Correct
C) The process of writing explicit rules for every possible input
D) The study of faster computer hardware for automation

Rationale: The chapter defines Machine Learning as the science and art of programming computers so they can learn from data.

Q2.

According to Arthur Samuel's 1959 definition, Machine Learning gives computers the ability to learn without being what?

A) Being connected to the internet
B) Being supervised by humans
C) Being explicitly programmed ✓ Correct
D) Being evaluated with data

Rationale: Samuel's definition emphasizes learning without being explicitly programmed.

Q3.

In Tom Mitchell's definition, a program learns if its performance on task T improves with experience E as measured by what quantity?

A) A utility report
B) A performance measure P ✓ Correct
C) A feature vector F
D) A model parameter θ

Rationale: Mitchell's formulation states that learning is judged by improvement on task T as measured by performance measure P.

Q4.

In the spam filter example, what is the training set?

A) The collection of rules written by the programmer
B) The archive of all incoming emails after deployment
C) The examples of spam and ham used by the system to learn ✓ Correct
D) The list of words most associated with spam

Rationale: The training set is the set of examples the system uses to learn, such as spam and ham emails.

Q5.
What is another term used in the chapter for a training instance?

A) Policy
B) Sample ✓ Correct
C) Hyperparameter
D) Cluster
Rationale: Each training instance is also called a sample.

Q6.
In the spam filter example, which performance measure is specifically named as the ratio of correctly classified emails?

A) Recall
B) Precision
C) Accuracy ✓ Correct
D) Specificity
Rationale: The chapter identifies accuracy as the ratio of correctly classified emails.

Q7.
Why does downloading a copy of Wikipedia not count as Machine Learning in the chapter's discussion?

A) The data is unlabeled
B) The computer gains data but does not improve at a task ✓ Correct
C) Wikipedia is too large for training
D) The data is not numerical
Rationale: Simply acquiring data without improving task performance does not satisfy the definition of learning.

Q8.
Which application is described as the first mainstream Machine Learning application that improved the lives of hundreds of millions of people?

A) Optical character recognition
B) Autonomous driving
C) Spam filtering ✓ Correct
D) Speech synthesis
Rationale: The chapter highlights the spam filter as the first mainstream ML application.

Q9.
Which of the following is identified as a specialized application in which Machine Learning had existed for decades?

A) Optical Character Recognition ✓ Correct
B) Stock price prediction
C) Customer segmentation
D) Robotic surgery
Rationale: OCR is specifically mentioned as an earlier specialized ML application.

Q10.

Which of the following is listed as a broad criterion for classifying Machine Learning systems?

A) Whether they use cloud computing or edge computing
B) Whether they are trained with human supervision ✓ Correct
C) Whether they use Python or R
D) Whether they rely on relational databases

Rationale: One major classification criterion is the amount and type of human supervision during training.

Q11.

How many major categories are named under classification by amount and type of supervision?

A) Two
B) Three
C) Four ✓ Correct
D) Five

Rationale: The chapter lists four: supervised, unsupervised, semisupervised, and reinforcement learning.

Q12.

In supervised learning, the desired solutions included in the training data are called what?

A) Features
B) Labels ✓ Correct
C) Policies
D) Rewards

Rationale: Supervised learning uses labeled training data, where desired solutions are called labels.

Q13.

Which supervised learning task involves assigning inputs to categories such as spam or ham?

A) Regression
B) Clustering
C) Classification ✓ Correct
D) Dimensionality reduction

Rationale: Classification is the supervised task of assigning examples to classes like spam or ham.

Q14.

Which supervised learning task predicts a target numeric value such as the price of a car?

A) Regression ✓ Correct
B) Association rule learning
C) Anomaly detection
D) Clustering

Rationale: Regression predicts numeric target values such as price.

Q15.

According to the chapter, in Machine Learning an attribute is best described as what?

A) A data type such as Mileage ✓ Correct
B) A learned decision boundary
C) A reward signal from the environment
D) A subset of the test set

Rationale: An attribute is described as a data type, for example Mileage.

Q16.

Which example best matches the chapter's general meaning of a feature?

A) A model's slope parameter
B) An attribute plus its value, such as Mileage = 15,000 ✓ Correct
C) A dataset split reserved for testing
D) A sequence of rewards received by an agent

Rationale: The chapter notes that a feature generally means an attribute together with its value.

Q17.

Which algorithm is specifically noted as commonly used for classification even though its name suggests regression?

A) Linear Regression
B) Principal Component Analysis
C) Logistic Regression ✓ Correct
D) K-Means

Rationale: Logistic Regression is commonly used for classification because it can output class probabilities.

Q18.

Which of the following is listed as a supervised learning algorithm in the chapter?

A) Apriori
B) DBSCAN
C) Decision Trees and Random Forests ✓ Correct
D) t-SNE

Rationale: Decision Trees and Random Forests are listed among the main supervised learning algorithms.

Q19.

In unsupervised learning, the training data is typically what?

A) Labeled with desired outputs
B) Unlabeled ✓ Correct
C) Balanced across classes
D) Restricted to numeric variables only

Rationale: Unsupervised learning uses unlabeled training data.

Q20.

Which of the following is presented as an unsupervised learning task?

A) Classification
B) Regression
C) Clustering ✓ Correct
D) Calibration

Rationale: Clustering is one of the major unsupervised learning tasks listed in the chapter.

Q21.

Which algorithm is listed under clustering methods?

A) K-Means ✓ Correct
B) Linear Regression
C) Logistic Regression
D) Support Vector Machines

Rationale: K-Means is explicitly listed as a clustering algorithm.

Q22.

Which algorithm is listed under anomaly detection and novelty detection?

A) Apriori
B) Isolation Forest ✓ Correct
C) Hierarchical Cluster Analysis
D) Kernel PCA

Rationale: Isolation Forest appears in the chapter under anomaly and novelty detection methods.

Q23.

Which method is listed as a visualization and dimensionality reduction algorithm?

A) Eclat
B) One-class SVM
C) t-SNE ✓ Correct
D) Random Forest

Rationale: t-SNE is listed as a visualization and dimensionality reduction technique.

Q24.

Which pair of algorithms is listed under association rule learning?

A) Apriori and Eclat ✓ Correct
B) PCA and Kernel PCA
C) K-Means and DBSCAN
D) Linear Regression and Logistic Regression

Rationale: Apriori and Eclat are the association rule learning algorithms named in the chapter.

Q25.

What is the goal of dimensionality reduction as described in the chapter?

A) To increase the number of labels available for training

B) To simplify the data without losing too much information ✓ Correct

C) To convert supervised learning problems into reinforcement learning problems

D) To force all features to become categorical

Rationale: Dimensionality reduction aims to simplify data while retaining as much useful information as possible.

Q26.

Merging correlated features such as car age and mileage into one feature representing wear and tear is called what?

A) Feature extraction ✓ Correct

B) Sampling bias

C) Regularization

D) Cross-validation

Rationale: Combining correlated features into a more useful one is described as feature extraction.

Q27.

What key distinction does the chapter make between novelty detection and anomaly detection?

A) Novelty detection uses labels, whereas anomaly detection does not

B) Novelty detection expects only normal data during training, whereas anomaly detection can tolerate some outliers ✓ Correct

C) Novelty detection is supervised, whereas anomaly detection is reinforcement learning

D) Novelty detection predicts numeric values, whereas anomaly detection predicts classes

Rationale: Novelty detection assumes only normal instances in training, while anomaly detection is usually more tolerant of some outliers.

Q28.

Semisupervised learning usually involves what combination of data?

A) Only labeled data from two classes

B) Only unlabeled data from many sources

C) A lot of unlabeled data and a little labeled data ✓ Correct

D) Equal amounts of training, validation, and test data

Rationale: Semisupervised learning is described as using mostly unlabeled data with a smaller amount of labeled data.

Q29.

In the chapter's photo-hosting example, what unsupervised technique helps recognize that the same person appears in multiple photos?

A) Regression

B) Clustering ✓ Correct

C) Regularization

D) Reinforcement learning

Rationale: The unsupervised part of the photo example is clustering similar faces together.

Q30.

In Reinforcement Learning, the learning system is called what?

A) Estimator
B) Sampler
C) Agent ✓ Correct
D) Validator

Rationale: The chapter states that the learning system in reinforcement learning is called an agent.

Q31.

In Reinforcement Learning, the strategy that specifies what action the agent should choose in a given situation is called a what?

A) Feature map
B) Policy ✓ Correct
C) Label
D) Cost function

Rationale: A policy defines which action the agent should take in each situation.

Q32.

Which of the following is given as an example of Reinforcement Learning?

A) AlphaGo learning a winning strategy for Go ✓ Correct
B) PCA reducing the number of dimensions
C) A spam filter classifying email as spam or ham
D) Apriori finding co-purchased items

Rationale: AlphaGo is presented as a reinforcement learning example that learned a policy for Go.

Q33.

What is another name given in the chapter for batch learning?

A) Similarity learning
B) Offline learning ✓ Correct
C) Outlier learning
D) Policy learning

Rationale: Batch learning is described as offline learning because training occurs before deployment.

Q34.

Which statement best describes batch learning?

A) The system updates itself continuously from each new instance in production
B) The system must be trained on all available data and cannot learn incrementally ✓ Correct
C) The system uses only unlabeled data and clusters it automatically
D) The system stores only summary statistics and discards all raw data

Rationale: Batch learning systems are trained using all available data and are not capable of incremental learning.

Q35.
If a batch learning system must incorporate new data such as a new type of spam, what does the chapter say is typically required?

A) Only updating the similarity measure
B) Training a new version from scratch on the full dataset ✓ Correct
C) Relabeling only the new instances
D) Increasing the validation set size

Rationale: Batch systems typically require retraining a new model from scratch on both old and new data.

Q36.
In online learning, data is commonly fed to the system in small groups called what?

A) Validation folds
B) Mini-batches ✓ Correct
C) Policies
D) Clusters

Rationale: Online learning can process instances individually or in small groups called mini-batches.

Q37.
What is out-of-core learning?

A) Using only external validation data for model selection
B) Training on data that fits entirely in cache memory
C) Using incremental algorithms to train on datasets too large to fit in main memory ✓ Correct
D) Performing reinforcement learning outside the production system

Rationale: Out-of-core learning handles huge datasets by loading parts of the data sequentially because the full set cannot fit in main memory.

Q38.
In online learning, what does the learning rate control?

A) The percentage of data reserved for testing
B) How fast the system adapts to changing data ✓ Correct
C) The number of features extracted from raw input
D) The similarity threshold used for nearest neighbors

Rationale: The learning rate determines how quickly an online system adapts to new data.

Q39.
A high learning rate in online learning tends to produce which trade-off?

A) Slower adaptation but stronger resistance to noise
B) Faster adaptation but quicker forgetting of old data ✓ Correct
C) Lower memory use but higher sampling bias
D) Better test performance but worse training performance

Rationale: A high learning rate makes the system adapt quickly, but it also tends to forget old data faster.

Q40.

What is a major risk if bad data is fed into an online learning system?

A) The model will become perfectly regularized

B) The system's performance may gradually decline ✓ Correct

C) The validation set will become too large

D) The task will automatically convert to unsupervised learning

Rationale: The chapter warns that bad incoming data can gradually degrade an online system's performance.

Q41.

Which type of learning algorithm generalizes to new cases by comparing them with learned examples using a similarity measure?

A) Model-based learning

B) Instance-based learning ✓ Correct

C) Reinforcement learning

D) Semisupervised learning

Rationale: Instance-based learning memorizes examples and compares new instances to them using similarity.

Q42.

Which type of learning algorithm builds a model from examples and then uses that model to make predictions?

A) Instance-based learning

B) Association-based learning

C) Model-based learning ✓ Correct

D) Rule-free learning

Rationale: Model-based learning generalizes by fitting a model to the training data and then using it for prediction.

Q43.

Selecting a linear function to relate life satisfaction to GDP per capita is an example of what step?

A) Model selection ✓ Correct

B) Cross-validation

C) Anomaly detection

D) Feature scaling

Rationale: Choosing the form of the model, such as a linear model, is called model selection.

Q44.

In the simple linear model life_satisfaction = $\theta 0 + \theta 1 \times$ GDP_per_capita, what do $\theta 0$ and $\theta 1$ represent?

A) Features extracted from the data

B) Model parameters ✓ Correct

C) Validation metrics

D) Learning rates

Rationale: The chapter identifies $\theta 0$ and $\theta 1$ as the model's parameters.

Q45.

For linear regression problems, what objective is typically used during training?

A) Maximizing the number of clusters

B) Minimizing a cost function measuring prediction error on training examples ✓ Correct

C) Maximizing the number of features retained

D) Minimizing the number of labels required

Rationale: Linear regression typically minimizes a cost function based on the distance between predictions and training examples.

Q46.

What term does the chapter use for applying a trained model to make predictions on new cases?

A) Sampling

B) Inference ✓ Correct

C) Regularization

D) Segmentation

Rationale: Applying the trained model to new cases is called inference.

Q47.

Which of the following is identified as a main challenge involving bad data?

A) Insufficient quantity of training data ✓ Correct

B) Too many epochs of gradient descent

C) Too few hidden layers in every model

D) Too much cross-validation

Rationale: Insufficient training data is one of the main data-related challenges discussed in the chapter.

Q48.

What is sampling bias?

A) Error caused by using too many model parameters

B) Nonrepresentative data caused by a flawed sampling method ✓ Correct

C) A model's inability to learn from unlabeled data

D) Performance decline caused by a high learning rate

Rationale: Sampling bias occurs when the sampling method produces a nonrepresentative dataset.

Q49.

If a model performs well on the training data but generalizes poorly to new instances, what problem is most likely occurring?

A) Underfitting

B) Feature extraction

C) Overfitting ✓ Correct

D) Dimensionality reduction

Rationale: Good training performance combined with poor generalization is the hallmark of overfitting.

Q50.

What is regularization in the context of Machine Learning?

A) Adding more labels to the training set

B) Constraining a model to make it simpler and reduce the risk of overfitting ✓ Correct

C) Splitting data into training and test sets

D) Converting batch learning into online learning

Rationale: Regularization simplifies the model by imposing constraints, which helps reduce overfitting.

Q51.

A hospital compliance team wants to automate email spam detection for staff inboxes. The current rule-based filter requires weekly manual updates because spammers keep changing wording. Which reason best supports replacing it with a machine learning approach?

A) Machine learning automatically adapts when new spam patterns become frequent in flagged messages ✓ Correct

B) Machine learning eliminates the need for any training data once deployed

C) Machine learning guarantees perfect accuracy on all future emails

D) Machine learning works only when spam messages are identical to prior examples

Rationale: ML-based spam filters can learn new predictive patterns from updated examples, reducing ongoing manual rule writing.

Q52.

A health system downloads a massive archive of public medical articles onto a server but does not run any learning algorithm. According to the chapter's definition, why is this not machine learning?

A) The server lacks labeled examples for classification only

B) The data was acquired from external rather than internal sources

C) The system has more data but has not improved at a task through experience ✓ Correct

D) Machine learning requires online incremental updates in all cases

Rationale: Simply storing data is not ML unless performance on a task improves through experience with that data.

Q53.

A clinical analytics manager defines task T as predicting appointment no-shows, experience E as historical scheduling records, and performance measure P as the proportion of correct predictions. Which performance measure is being used?

A) Novelty score

B) Accuracy ✓ Correct

C) Dimensionality

D) Learning rate

Rationale: The ratio of correctly classified cases is accuracy, a common classification metric.

Q54.
A coding education platform wants to predict a student's numeric exam score from study hours, prior GPA, and quiz performance. Which supervised learning task is this?

A) Clustering
B) Association rule learning
C) Regression ✓ Correct
D) Anomaly detection

Rationale: Predicting a continuous numeric target such as an exam score is a regression task.

Q55.
A rehabilitation robotics team needs a system that learns how to choose movements on uneven surfaces by receiving rewards for stable steps and penalties for falls. Which machine learning paradigm is most appropriate?

A) Supervised learning
B) Reinforcement learning ✓ Correct
C) Association rule learning
D) Dimensionality reduction

Rationale: Reinforcement learning is designed for agents that learn policies from actions, rewards, and penalties.

Q56.
A hospital marketing department has years of unlabeled website behavior data and wants to identify naturally occurring groups of visitors to tailor outreach. Which approach best fits this objective?

A) Regression
B) Clustering ✓ Correct
C) Logistic calibration
D) Holdout validation

Rationale: Clustering is an unsupervised method used to discover similar groups in unlabeled data.

Q57.
A pathology image team has 500 labeled slides and 50,000 unlabeled slides. They want a method that can use both sources of data together. Which learning setup is most appropriate?

A) Batch-only supervised learning
B) Semisupervised learning ✓ Correct
C) Pure anomaly detection
D) Association rule learning

Rationale: Semisupervised learning is intended for problems with a small labeled set and a much larger unlabeled set.

Q58.

A photo management app clusters faces automatically, then asks the user to identify one example of each person so it can name the remaining photos. This workflow is best described as:

A) Batch reinforcement learning

B) Semisupervised learning ✓ Correct

C) Pure supervised regression

D) Instance-based novelty detection

Rationale: The system combines unsupervised grouping with a small amount of labeling, which is semisupervised learning.

Q59.

A clinical operations dashboard receives streaming vital-sign data every second and must update predictions continuously with limited memory. Which learning strategy is most suitable?

A) Batch learning with weekly retraining only

B) Online learning ✓ Correct

C) Hierarchical clustering

D) Association rule mining

Rationale: Online learning updates incrementally from a data stream and is well suited to limited-resource, rapidly changing settings.

Q60.

A healthcare startup has a dataset too large to fit into one machine's RAM, so it loads chunks sequentially, updates the model on each chunk, and repeats until all data are processed. What is this approach called?

A) Out-of-core learning ✓ Correct

B) Transfer learning

C) Sampling bias correction

D) Cross-sectional validation

Rationale: Out-of-core learning trains incrementally on data subsets when the full dataset cannot fit in memory.

Q61.

A payer's fraud model is updated after every small batch of new claims. Leadership raises the learning rate to make the model respond faster to new fraud tactics. What trade-off should they expect?

A) Faster adaptation but quicker forgetting of older patterns ✓ Correct

B) Lower sensitivity to outliers with greater inertia

C) Guaranteed improvement in generalization error

D) Automatic conversion from model-based to instance-based learning

Rationale: A high learning rate helps the system adapt quickly but also makes it forget old data more readily.

Q62.
An online triage model begins performing worse after a faulty device starts sending corrupted measurements into the live update stream. What is the most appropriate immediate operational response?

A) Increase the learning rate so the model adapts faster to the bad data
B) Switch learning off and investigate performance and input anomalies ✓ Correct
C) Discard the test set and retrain only on the corrupted stream
D) Convert the model to unsupervised clustering without monitoring

Rationale: For online systems, degraded performance from bad incoming data calls for close monitoring and prompt suspension of learning if needed.

Q63.
A document retrieval tool predicts whether a new note belongs to a category by comparing it to stored examples using a similarity score based on shared terms. This is primarily what type of learning?

A) Model-based learning
B) Instance-based learning ✓ Correct
C) Reinforcement learning
D) Association rule learning

Rationale: Instance-based learning generalizes by comparing new cases to memorized examples using a similarity measure.

Q64.
A public health analyst builds a linear equation to predict life satisfaction from GDP per capita and then estimates values for countries not in the dataset. This exemplifies:

A) Model-based learning ✓ Correct
B) Novelty detection
C) Hierarchical clustering
D) Batch labeling

Rationale: Model-based learning fits a model to training data and uses the fitted model to make predictions on new cases.

Q65.
A data scientist selects a linear model with parameters $\theta 0$ and $\theta 1$ for a healthcare cost prediction problem. During training, what is the learning algorithm primarily searching for?

A) The labels most similar to the test instances
B) The parameter values that minimize a cost function on the training data ✓ Correct
C) The smallest possible validation set
D) The highest number of irrelevant features

Rationale: Model training typically searches for parameter values that minimize a defined cost function.

Q66.
A quality improvement team notices that patient age and years since diagnosis are highly correlated. They combine them into a single variable representing disease duration burden before modeling. This is an example of:

A) Feature extraction ✓ Correct
B) Sampling bias
C) Regularization
D) Nonresponse bias

Rationale: Combining correlated features into a more useful derived feature is feature extraction.

Q67.
A biomedical researcher uses PCA before fitting a supervised model because the original dataset has thousands of variables. Which operational benefit is most directly expected from dimensionality reduction?

A) It guarantees elimination of all noise
B) It usually makes training faster and can reduce storage needs ✓ Correct
C) It converts supervised tasks into reinforcement tasks
D) It ensures perfect representativeness of the sample

Rationale: Dimensionality reduction often speeds training and reduces disk and memory use, sometimes improving performance as well.

Q68.
A hospital fraud unit trains a model mostly on legitimate transactions so it can flag unusual future claims. Which unsupervised task best matches this use case?

A) Regression
B) Anomaly detection ✓ Correct
C) Association rule learning
D) Classification with labels

Rationale: Anomaly detection learns patterns of normal cases and identifies unusual new instances.

Q69.
A data engineer explains that the model was trained on mostly normal sensor readings but may tolerate a small proportion of outliers in training. This description aligns more closely with:

A) Novelty detection
B) Anomaly detection ✓ Correct
C) Pure supervised classification
D) Instance memorization only

Rationale: Anomaly detection is generally more tolerant of some outliers in the training data than novelty detection.

Q70.
A retail pharmacy analyzes purchase logs and discovers that customers who buy cough drops and tissues also often buy tea. Which machine learning task produced this finding?

A) Association rule learning ✓ Correct
B) Regression
C) Reinforcement learning
D) Dimensionality reduction

Rationale: Association rule learning discovers interesting co-occurrence relationships among attributes or items.

Q71.
A machine learning engineer reports that a new model fits the training data extremely well but performs poorly on unseen patient records. What problem is most likely occurring?

A) Underfitting
B) Overfitting ✓ Correct
C) Dimensionality reduction
D) Sampling without replacement

Rationale: Strong training performance with weak performance on new data is the hallmark of overfitting.

Q72.
A hospital readmission model overfits because the training set is small and noisy. Which change would best reduce overfitting according to the chapter?

A) Add more parameters and more irrelevant features
B) Simplify the model or regularize it ✓ Correct
C) Tune hyperparameters on the test set repeatedly
D) Reduce the amount of training data further

Rationale: Overfitting can often be reduced by simplifying the model, reducing features, or applying regularization.

Q73.
A team constrains a linear model so the slope remains small even if the unconstrained fit would be steeper. What is the main purpose of this constraint?

A) To increase variance and fit noise more closely
B) To reduce the risk of overfitting by simplifying the model ✓ Correct
C) To transform the task from regression to clustering
D) To ensure the training and test sets are identical

Rationale: Regularization constrains model flexibility so it is less likely to fit noise in the training data.

Q74.
In a machine learning pipeline, the regularization strength is chosen before training and remains fixed while the algorithm learns model weights. This regularization strength is a:

A) Training instance
B) Model parameter
C) Hyperparameter ✓ Correct
D) Label

Rationale: A hyperparameter belongs to the learning algorithm setup and is set prior to training rather than learned from the data.

Q75.
A simple linear model gives poor predictions even on the training examples because the real relationship is clearly more complex. Which issue is most likely present?

A) Overfitting
B) Underfitting ✓ Correct
C) Nonresponse bias
D) Data leakage from test to train

Rationale: Poor performance even on training data suggests the model is too simple and is underfitting.

Q76.
A nurse informaticist suspects underfitting in a staffing forecast model. Which action is most appropriate?

A) Use a more powerful model or engineer better features ✓ Correct
B) Increase constraints until the model becomes nearly flat
C) Remove informative predictors to simplify the task
D) Evaluate only on the training set

Rationale: Underfitting is addressed by increasing model capacity, improving features, or reducing excessive constraints.

Q77.
A health insurer has only 400 training examples for a complex image-recognition task. Which challenge from the chapter is most directly illustrated?

A) Insufficient quantity of training data ✓ Correct
B) Association rule conflict
C) Excessive online learning rate
D) No Free Lunch violation

Rationale: Complex tasks such as image recognition often require large amounts of training data to perform well.

Q78.
A medical device company trains a model using records from one highly specialized tertiary center and then deploys it across community clinics with very different patient populations. What data problem is most likely?

A) Regularization bias
B) Nonrepresentative training data ✓ Correct
C) Underfitting due to low slope
D) Feature extraction error

Rationale: Training data that do not reflect the target population are nonrepresentative and may generalize poorly.

Q79.
A survey-based wellness prediction model is built from a huge sample, but the sampling frame excluded people without internet access. This is best described as:

A) Sampling bias ✓ Correct
B) Regularization
C) Cross-validation
D) Feature extraction

Rationale: Even a large sample can be nonrepresentative if the sampling method systematically excludes key groups.

Q80.
A hospital patient experience survey has a low response rate, and respondents differ systematically from nonrespondents. Which specific bias is this?

A) Dimensionality bias
B) Nonresponse bias ✓ Correct
C) Model variance bias
D) Learning-rate bias

Rationale: When the people who respond differ systematically from those who do not, the sample suffers from nonresponse bias.

Q81.
A claims dataset contains incorrect values, missing fields, and obvious outliers. Before model development, what action is most justified?

A) Ignore data quality because algorithms can always infer the truth
B) Spend time cleaning the data and deciding how to handle missing values ✓ Correct
C) Use the test set to impute all missing fields
D) Increase model complexity to absorb the errors

Rationale: Poor-quality data make pattern detection harder, so cleaning errors, outliers, and missingness is often essential.

Q82.
A sepsis prediction model includes many variables, but several are unrelated administrative codes that add noise. According to the chapter, this problem reflects:

A) Irrelevant features ✓ Correct
B) Reinforcement penalties
C) Out-of-core processing
D) Holdout validation

Rationale: Too many irrelevant features can impair learning because the model may focus on unhelpful inputs.

Q83.
A data science lead says, 'Garbage in, garbage out,' while reviewing a failed model. Which competency is most directly implicated?

A) Feature engineering and data quality management ✓ Correct
B) Increasing the number of test sets
C) Replacing labels with rewards
D) Switching every task to online learning

Rationale: Model success depends heavily on having relevant, high-quality features, which is the focus of feature engineering and data preparation.

Q84.
A team splits its dataset so that 80% is used to fit the model and 20% is reserved for final evaluation on unseen cases. The reserved portion is the:

A) Validation set
B) Training set
C) Test set ✓ Correct
D) Mini-batch set

Rationale: The test set is held out from training and used to estimate performance on new data.

Q85.
After training, a model shows very low training error but high error on the held-out test set. What does this difference estimate?

A) The model's novelty score
B) Its generalization problem on unseen data ✓ Correct
C) Its clustering purity
D) Its reward function

Rationale: High test error despite low training error indicates poor generalization, typically due to overfitting.

Q86.
A researcher compares 50 hyperparameter settings and always chooses the one with the best performance on the test set. Why is this problematic?

A) The test set becomes part of model selection, leading to overly optimistic estimates ✓ Correct
B) The model can no longer be trained on the training set
C) The validation set automatically becomes unnecessary and harmful
D) The algorithm changes from supervised to unsupervised

Rationale: Repeatedly tuning to the test set leaks information from it into model selection and biases the estimated generalization error.

Q87.
A team withholds part of the training data to compare candidate models and hyperparameters, then retrains the selected model on the full training data before final testing. The withheld subset is the:

A) Test set
B) Validation set ✓ Correct
C) Inference set
D) Production set

Rationale: A validation set is used during model selection before the final evaluation on the test set.

Q88.
A small healthcare dataset makes model comparisons unstable because a single validation split gives highly variable results. Which method best addresses this?

A) Repeated cross-validation ✓ Correct
B) Increasing regularization to the maximum
C) Using the test set for all tuning decisions
D) Removing the training set entirely

Rationale: Repeated cross-validation averages performance over multiple validation splits to obtain a more reliable estimate.

Q89.
What is the main drawback of repeated cross-validation in an operational analytics project?

A) It prevents any estimate of generalization error
B) It multiplies training time because each model is trained repeatedly ✓ Correct
C) It can only be used for unsupervised learning
D) It eliminates the need for a final test set in all circumstances

Rationale: Repeated cross-validation improves estimate stability at the cost of substantially more computation.

Q90.
A mobile dermatology app is trained mostly on web images, but validation and test sets are built only from photos taken within the app. Why is this split strategy appropriate?

A) Because validation and test data should match the production data as closely as possible ✓ Correct
B) Because web images should never be used for training
C) Because test sets must always be larger than training sets
D) Because representative data should be excluded from final evaluation

Rationale: Validation and test sets should be highly representative of the data the model will face in production.

Q91.
A flower-classification model trained on web images performs well on a held-out subset of web images but poorly on mobile-app validation images. What does this pattern most strongly suggest?

A) The model is underfitting the web training data
B) The main issue is likely data mismatch between training and production data ✓ Correct
C) The labels in the validation set are unnecessary
D) The model must be instance-based rather than model-based

Rationale: Good train-dev performance with poor validation performance points to mismatch between training-source and production-source data.

Q92.
A machine learning team creates an additional holdout from the web-image training source to determine whether poor app performance is due to overfitting or source mismatch. Andrew Ng's term for this extra set is the:

A) Reward set
B) Train-dev set ✓ Correct
C) Mini-batch set
D) Inference set

Rationale: A train-dev set is held out from the training source to help distinguish overfitting from data mismatch.

Q93.
A model performs poorly on the train-dev set after being trained on the remaining training data. What conclusion is most justified?

A) The model is probably overfitting the training data ✓ Correct
B) The validation set is too representative
C) The task should be reframed as clustering
D) The data mismatch problem has been solved

Rationale: Poor performance on a same-source holdout such as a train-dev set suggests the model has overfit the training data.

Q94.
A chief analytics officer asks whether one algorithm can be assumed best for every prediction problem in healthcare. Which principle from the chapter is the best response?

A) The No Free Lunch theorem says no single model is guaranteed best across all datasets ✓ Correct
B) Linear models are always superior when data are noisy
C) Neural networks always outperform simpler algorithms with enough tuning
D) Supervised learning dominates unsupervised learning for every task

Rationale: The No Free Lunch theorem states that no learning algorithm is universally best without assumptions about the data.

Q95.
A machine learning consultant says model choice requires assumptions about the data, such as linearity, before narrowing candidate algorithms. Why?

A) Because assumptions determine what structure is treated as signal versus noise ✓ Correct
B) Because assumptions eliminate the need for evaluation
C) Because assumptions convert hyperparameters into labels
D) Because assumptions guarantee zero generalization error

Rationale: Models embody assumptions about the data that influence which patterns are preserved and which are ignored as noise.

Q96.

A hospital wants to discover hidden patterns in a large operational dataset to better understand patient flow, not necessarily to predict a predefined label. Which benefit of machine learning is most relevant?

A) Reinforcement reward shaping

B) Data mining for insight discovery ✓ Correct

C) Guaranteed causal inference

D) Manual rule optimization

Rationale: ML can be used to inspect large datasets and uncover patterns that help humans learn about the problem domain.

Q97.

A security team needs a system that can identify the latest phishing tactics within hours instead of waiting for a nightly full retrain. Which learning setup is most operationally appropriate?

A) Weekly batch learning only

B) Online learning with close monitoring ✓ Correct

C) Pure dimensionality reduction

D) Static instance memorization without updates

Rationale: Online learning is better suited for rapidly changing environments that require fast adaptation.

Q98.

A model predicts inpatient length of stay from age, comorbidities, and severity score. In this context, which statement best distinguishes an attribute from a feature as described in the chapter?

A) An attribute is the algorithm, while a feature is the hyperparameter

B) An attribute is a data type such as age, while a feature often refers to the attribute together with its value ✓ Correct

C) An attribute is a label, while a feature is an unlabeled instance

D) There is no possible distinction between attribute and feature

Rationale: The chapter describes an attribute as a data type and a feature often as that attribute with a specific value.

Q99.

A patient-risk model is retrained from scratch every night using all historical data and then deployed unchanged during the day. This is best classified as:

A) Online learning

B) Offline batch learning ✓ Correct

C) Reinforcement learning

D) Semisupervised learning

Rationale: Batch learning trains on the full dataset offline and then runs in production without incremental learning.

Q100.

A graduate student proposes selecting a model solely because it achieved the lowest error on the training set. Which recommendation best reflects the chapter's workflow?

A) Choose the model with the lowest training error because it will generalize best

B) Evaluate candidate models using validation data and reserve a test set for final generalization estimation ✓ Correct

C) Skip validation if the dataset is large enough

D) Prefer the most complex model to avoid underfitting

Rationale: Model selection should rely on validation performance, with a separate test set reserved for final estimation of generalization error.

Q101.

A data science team claims its system is performing machine learning because it downloaded a complete copy of Wikipedia onto a server. Based on the chapter's definitions, which evaluation best explains why this claim is insufficient?

A) The system has acquired data, but there is no evidence that performance on a task improved with experience ✓ Correct

B) The system has increased storage capacity, which is the primary criterion for machine learning

C) The system has become unsupervised because Wikipedia lacks labels for every sentence

D) The system has completed feature extraction automatically, which guarantees learning

Rationale: Machine learning requires improved performance on a task from experience, not merely storing more data.

Q102.

A hospital IT team is choosing between a rule-based claims denial detector and an ML-based detector. Denial patterns change weekly as payer edits evolve. Which argument most strongly favors the ML approach?

A) ML guarantees zero maintenance after deployment

B) ML can automatically adapt when new patterns become frequent in incoming data ✓ Correct

C) ML eliminates the need for any training examples

D) ML always uses fewer computational resources than rule-based systems

Rationale: The chapter emphasizes ML's advantage in fluctuating environments because it can adapt to changing patterns.

Q103.

A researcher must classify a project into the appropriate ML category. The algorithm receives email examples labeled spam or ham and learns to classify future emails. Which category best fits?

A) Unsupervised learning

B) Reinforcement learning

C) Supervised learning ✓ Correct

D) Association rule learning

Rationale: Labeled examples with desired outputs define supervised learning.

Q104.

Which scenario is the clearest example of regression rather than classification?

A) Predicting whether a message is spam or ham

B) Predicting a patient's length of stay in days as a numeric value ✓ Correct

C) Grouping patients into utilization clusters

D) Detecting unusual transactions without labels

Rationale: Regression predicts a numeric target value rather than a class label.

Q105.

A team uses clustering to group website visitors with no prior labels. What is the strongest justification for calling this unsupervised learning?

A) The algorithm predicts a probability for each visitor

B) The algorithm learns from rewards and penalties

C) The algorithm is trained on unlabeled data and discovers structure without target outputs ✓ Correct

D) The algorithm uses a cost function with labeled outcomes

Rationale: Unsupervised learning works with unlabeled data to detect patterns or structure.

Q106.

A photo platform groups images by person automatically, then a user provides one name per cluster so the system can label all matching photos. Which classification is most appropriate?

A) Pure supervised learning

B) Semisupervised learning ✓ Correct

C) Pure reinforcement learning

D) Batch anomaly detection

Rationale: The system combines large amounts of unlabeled data with a small amount of labeled data, which is semisupervised learning.

Q107.

An autonomous wheelchair learns navigation by trying actions, observing the environment, and receiving rewards for safe movement. Which concept best characterizes this setup?

A) Reinforcement learning with a learned policy ✓ Correct

B) Supervised learning with labels

C) Association rule learning with support counts

D) Dimensionality reduction with feature extraction

Rationale: Reinforcement learning involves an agent learning a policy from actions and rewards.

Q108.

A financial forecasting model is retrained from scratch every Sunday on all historical data and does not update itself during the week. Which description is most accurate?

A) Online learning because it uses sequential records

B) Offline batch learning because it learns only during periodic full retraining ✓ Correct

C) Reinforcement learning because markets provide rewards

D) Semisupervised learning because new labels arrive later

Rationale: Batch learning is trained on all available data offline and does not learn incrementally in production.

Q109.
A streaming fraud detector updates its parameters after every mini-batch of transactions and discards processed instances to conserve memory. Which approach is being used?

A) Batch learning
B) Online learning ✓ Correct
C) Hierarchical clustering
D) Pure instance memorization

Rationale: Online learning updates incrementally from sequential instances or mini-batches and can discard past data.

Q110.
Which situation most strongly supports using out-of-core learning?

A) The dataset is tiny and fits comfortably in memory
B) The data must be labeled by experts before training
C) The full dataset is too large to fit in main memory, so it must be processed in chunks ✓ Correct
D) The model should never be retrained after deployment

Rationale: Out-of-core learning is used when huge datasets must be trained incrementally because they do not fit in memory.

Q111.
A classifier predicts a new case by comparing it with stored examples using a similarity measure and assigning the majority class among the most similar cases. Which generalization strategy is this?

A) Model-based learning
B) Instance-based learning ✓ Correct
C) Policy-based reinforcement learning
D) Holdout validation

Rationale: Instance-based learning generalizes by comparing new cases to memorized instances using similarity.

Q112.
A linear model for life satisfaction is written as life_satisfaction = $\theta 0 + \theta 1 \times$ GDP_per_capita. What are $\theta 0$ and $\theta 1$ in this context?

A) Validation metrics
B) Hyperparameters chosen after testing
C) Model parameters learned during training ✓ Correct
D) Features extracted from the dataset

Rationale: $\theta 0$ and $\theta 1$ are model parameters adjusted by the learning algorithm to fit the training data.

Q113.
A team adds mileage and vehicle age into a single derived variable representing wear and tear before training a model. Which concept does this illustrate most directly?

A) Feature extraction ✓ Correct
B) Sampling bias
C) Novelty detection
D) Policy optimization

Rationale: Combining existing correlated features into a more useful derived feature is feature extraction.

Q114.

A model performs poorly because many input variables are unrelated to the target and obscure useful signal. Which challenge from the chapter is most directly implicated?

A) Insufficient quantity of data

B) Irrelevant features ✓ Correct

C) Excessive validation size

D) No Free Lunch theorem

Rationale: Too many irrelevant features can prevent the system from learning useful patterns.

Q115.

A polling dataset contains millions of responses, but the sampling frame overrepresents affluent voters and underrepresents others. Which conclusion is most accurate?

A) Large sample size guarantees representativeness

B) The main issue is underfitting due to too few parameters

C) The dataset may still be nonrepresentative because of sampling bias ✓ Correct

D) The problem can only be solved with reinforcement learning

Rationale: Even very large datasets can be nonrepresentative if the sampling method is biased.

Q116.

A team tunes dozens of hyperparameter settings by repeatedly checking performance on the test set and selecting the best one. What is the most likely consequence?

A) The test set becomes a validation set in practice, yielding an optimistically biased estimate of generalization ✓ Correct

B) The training set becomes more representative automatically

C) The model is guaranteed to underfit

D) The cost function is eliminated from training

Rationale: Repeatedly using the test set for tuning adapts the model to that set and compromises its role as an unbiased estimate.

Q117.

A startup has abundant web images for training a flower classifier, but production images will come from mobile phones under different conditions. Which dataset design is most defensible?

A) Use web images for training, and ensure validation and test sets consist only of representative mobile images ✓ Correct

B) Use only web images for all splits because larger datasets always dominate representativeness

C) Use mobile images only for training and web images only for testing

D) Mix duplicates across all splits to stabilize evaluation

Rationale: Validation and test sets should be as representative as possible of production data.

Q118.

A model achieves very low training error and very high error on new cases. Which diagnosis and remedy pair is best aligned with the chapter?

A) Underfitting; increase regularization sharply

B) Overfitting; simplify the model or gather more data ✓ Correct

C) Sampling bias; use the test set for training

D) Data mismatch; remove the validation set

Rationale: Low training error with high generalization error indicates overfitting, often addressed by simplification or more data.

Q119.

A linear model of a complex nonlinear relationship performs poorly even on the training examples. What is the most likely problem?

A) Overfitting caused by too much flexibility

B) Underfitting caused by an overly simple model ✓ Correct

C) Data leakage from the test set

D) Novelty detection failure

Rationale: Poor performance even on the training data suggests the model is too simple and is underfitting.

Q120.

A practitioner increases the regularization hyperparameter to an extreme value. What tradeoff should they expect?

A) Higher risk of overfitting and more complex decision boundaries

B) Less risk of overfitting but greater risk of missing a good fit because the model becomes too constrained ✓ Correct

C) No effect because hyperparameters are learned automatically

D) Guaranteed improvement on both training and test data

Rationale: Strong regularization simplifies the model and reduces overfitting risk, but can make the model too constrained.

Q121.

A team is comparing candidate models using a single small validation set. Which critique is most valid?

A) The validation estimates may be unstable, increasing the chance of selecting a suboptimal model ✓ Correct

B) A small validation set always causes overfitting to the training data

C) Validation sets should always be larger than training sets

D) A validation set should contain all test examples to improve power

Rationale: Small validation sets can yield imprecise estimates and lead to poor model selection decisions.

Q122.

Why might repeated cross-validation be preferred over a single holdout validation set when data are limited?

A) It removes the need for a final test set

B) It yields a more accurate performance estimate by averaging across multiple validation splits ✓ Correct

C) It guarantees the best model under the No Free Lunch theorem

D) It prevents all forms of sampling bias

Rationale: Repeated cross-validation reduces variance in evaluation by averaging performance across many splits.

Q123.

A team notices that its online recommendation model is rapidly changing after every small burst of unusual user behavior. Which parameter should be examined first?

A) The similarity metric only

B) The learning rate because adaptation may be too aggressive ✓ Correct

C) The test set size because it controls forgetting

D) The number of labels because it determines online status

Rationale: In online learning, a high learning rate can make the system adapt too quickly and become sensitive to noise.

Q124.

A search engine's online ranking model is being manipulated by adversarial inputs, causing gradual performance decline. Which response best reflects the chapter's guidance?

A) Increase the learning rate so poisoned data dominate less

B) Monitor performance and input data closely, disable learning if needed, and revert to a prior working state ✓ Correct

C) Replace the validation set with the test set immediately

D) Convert the task to unsupervised clustering

Rationale: The chapter recommends monitoring online systems, switching learning off, and possibly rolling back when bad data degrade performance.

Q125.

A researcher trains a complex model that discovers a pattern: countries with the letter "w" in their names have high life satisfaction. What is the best interpretation?

A) A robust causal relationship has likely been found

B) The model has probably learned noise due to overfitting ✓ Correct

C) The training set is necessarily representative because the pattern is memorable

D) This demonstrates effective feature extraction

Rationale: The chapter uses such spurious patterns as examples of overfitting to noise rather than real structure.

Q126.

Which evaluation of ML for speech recognition is most consistent with the chapter's argument?

A) Traditional hand-coded pitch rules scale better than learned models across languages and speakers

B) Speech recognition is a domain where ML is useful because no simple traditional algorithm scales to real-world complexity ✓ Correct

C) Speech recognition is primarily an anomaly detection problem

D) Speech recognition should avoid data-driven approaches because they obscure causal rules

Rationale: The chapter presents speech recognition as too complex for hand-coded rules and well suited to ML.

Q127.

A team wants to inspect a trained spam filter to identify words most predictive of spam and use those insights to refine business policy. Which ML benefit is being leveraged?

A) Reinforcement learning

B) Data mining for human insight ✓ Correct

C) Sampling correction

D) Out-of-core inference

Rationale: Inspecting learned patterns to uncover trends and relationships is described as data mining that can help humans learn.

Q128.

A model-based learner is being developed. Which sequence best captures the typical workflow described in the chapter?

A) Select a model, define a performance measure, train to find parameter values, then use the model for inference ✓ Correct

B) Tune on the test set, gather labels later, then define a model if needed

C) Cluster the data, discard labels, then optimize rewards

D) Memorize instances, compute similarity, then estimate a cost function only after deployment

Rationale: Model-based learning involves selecting a model, specifying how performance is assessed, training parameters, and then making predictions.

Q129.

A team argues that because one algorithm outperformed another on a previous project, it should always be preferred for future datasets. Which principle most directly challenges this claim?

A) Sampling bias principle

B) No Free Lunch theorem ✓ Correct

C) Feature extraction principle

D) Nonresponse bias principle

Rationale: The No Free Lunch theorem states that no single model is guaranteed to be best across all possible datasets.

Q130.

A company has only 10,000 highly representative mobile images but millions of less representative web images. After training on web images, the model performs well on held-out web images but poorly on mobile validation images. Which interpretation is strongest?

A) The model is definitely underfitting the web training data

B) Data mismatch is a plausible cause and should be distinguished from overfitting using a train-dev set ✓ Correct

C) The test set must be too large

D) The web images should replace the validation set entirely

Rationale: Good performance on held-out training-distribution data but poor performance on representative validation data suggests possible data mismatch.

Q131.

A data scientist removes obvious outliers, imputes missing ages with the median, and corrects known entry errors before model training. What strategic objective is being pursued?

A) Reducing poor-quality data that can obscure underlying patterns ✓ Correct

B) Increasing model complexity to avoid underfitting

C) Transforming supervised learning into unsupervised learning

D) Eliminating the need for a test set

Rationale: Cleaning errors, handling missing values, and addressing outliers improve data quality and help learning.

Q132.

A system groups customers into segments, then a separate supervised model predicts churn using those segment identifiers as inputs. Which reasoning best justifies the first step?

A) Clustering can reveal latent structure that may become useful engineered features for later prediction ✓ Correct

B) Clustering guarantees causal explanations for churn

C) Clustering replaces the need for labels in the churn model

D) Clustering is identical to regression when enough data are available

Rationale: Unsupervised structure discovery can support later supervised learning by creating informative features or insights.

Q133.

A team must decide whether to spend effort on collecting more labeled examples or on marginal algorithmic tweaks for a difficult language task. Which conclusion is most aligned with the chapter's cited evidence?

A) Algorithm choice always dominates data quantity

B) For complex problems, additional data can be at least as important as algorithmic sophistication ✓ Correct

C) Small datasets are always preferable because they reduce noise

D) Labels become unnecessary once enough features are engineered

Rationale: The chapter cites work showing that, for complex tasks, more data can make different algorithms perform similarly well.

Q134.

A classifier is trained on a dataset with many missing countries at the high and low ends of GDP, then used to predict life satisfaction for those extremes. What is the primary methodological concern?

A) The model is guaranteed to be unbiased because linear regression was used

B) The training data are nonrepresentative, so generalization to those extremes is unreliable ✓ Correct

C) The issue is only that the test set is too large

D) The model cannot be overfit if it has only two parameters

Rationale: Missing important parts of the target population makes the training set nonrepresentative and weakens generalization.

Q135.

An operations team wants a Mars rover model that can continue learning with limited onboard memory and computing power. Which learning setup is most suitable?

A) Large weekly batch retraining on all historical data stored locally

B) Incremental online learning that can discard processed instances ✓ Correct

C) Only association rule learning because it uses fewer features

D) Exclusive use of instance-based learning with full memory of all examples

Rationale: Online learning is suited to limited-resource settings because it learns incrementally and need not retain all past data.

Q136.

A practitioner says, “Our model has two parameters, so its regularization strength is the third parameter learned during training.” Which correction is most accurate?

A) Regularization strength is usually a hyperparameter set before training, not a model parameter learned by the algorithm ✓ Correct

B) Regularization strength is part of the test set design

C) Regularization strength is a label attached to each training instance

D) Regularization strength is a similarity measure used only in instance-based learning

Rationale: Hyperparameters belong to the learning algorithm setup and are fixed prior to training rather than learned as model parameters.

Q137.

A team evaluates two candidate models on the test set, chooses the better one, retrains it on the combined training and test data, and reports the same test performance as final. Why is this problematic?

A) The test set has been used for model selection and then contaminated by training, so it no longer provides an unbiased generalization estimate ✓ Correct

B) Retraining on more data always increases overfitting, making any estimate invalid

C) Model selection should be based only on training error

D) The issue is that two models were compared instead of three

Rationale: Using the test set for selection and then including it in training destroys its role as an independent final evaluation set.

Q138.

A hospital analytics team wants a fast preliminary method to visualize thousands of unlabeled patient records in two dimensions to inspect possible groupings before formal modeling. Which approach best matches the chapter?

A) t-SNE or another visualization/dimensionality reduction method ✓ Correct

B) Linear regression with a continuous target

C) Reinforcement learning with reward shaping

D) Logistic regression with probability thresholds

Rationale: The chapter lists t-SNE and related methods as unsupervised tools for 2D or 3D visualization of complex unlabeled data.

Q139.

A fraud team trains on mostly normal transactions but expects a small fraction of outliers in the training data. Which task framing is most appropriate?

A) Novelty detection only, because any outlier in training makes the method invalid

B) Anomaly detection, because it can often tolerate some outliers during training ✓ Correct

C) Regression, because fraud severity is numeric

D) Association rule learning, because rules replace labels

Rationale: The chapter distinguishes anomaly detection as often more tolerant of a small percentage of outliers in training data.

Q140.

A supermarket mines sales logs and discovers that customers buying barbecue sauce and chips often also buy steak. Which unsupervised objective does this illustrate?

A) Classification

B) Association rule learning ✓ Correct

C) Novelty detection

D) Policy learning

Rationale: Discovering co-occurrence relationships among attributes in large datasets is association rule learning.

Q141.

A model with a high-degree polynomial fits the training data extremely closely but makes unstable predictions for unseen cases. Which response best reflects sound model governance?

A) Prefer the model because lower training error is the definitive criterion

B) Question the model's generalization and consider a simpler or regularized alternative ✓ Correct

C) Assume the instability is caused by the test set alone

D) Eliminate the need for validation by increasing polynomial degree further

Rationale: Very low training error with unstable unseen performance is characteristic of overfitting and calls for simplification or regularization.

Q142.

A team is deciding whether to hold out 20% or 1% of a 10-million-record dataset for testing. Which choice is most defensible according to the chapter?

A) Always hold out exactly 20%, regardless of dataset size

B) A 1% test split may be sufficient because 100,000 test cases can still estimate generalization well ✓ Correct

C) Testing is unnecessary when the training set is very large

D) Use the smallest possible test set and tune on it repeatedly

Rationale: The chapter notes that for very large datasets, even a small percentage can produce a sufficiently large test set.

Q143.

A team reports, “Our algorithm detected subtle patterns in the training set, so those patterns are likely real.” Which critique is strongest?

A) Subtle patterns found by complex models may reflect noise, especially with small or noisy datasets ✓ Correct

B) Any detected pattern is valid if the cost function decreased during training

C) Patterns are only questionable in unsupervised learning

D) Noise affects only instance-based methods, not model-based methods

Rationale: Complex models can learn noise as if it were signal, particularly when data are limited or noisy.

Q144.

A scientist compares candidate models trained on a heavily reduced training subset because a very large validation set was reserved. Why can this distort model selection?

A) Candidate models may be judged under conditions unlike the final training regime, making the comparison less informative ✓ Correct

B) Large validation sets eliminate the need for training data

C) Validation data should always include labels from the test set

D) Reduced training sets only affect unsupervised models

Rationale: If candidate models are trained on much less data than the final model will receive, selection may not reflect final performance well.

Q145.

Which statement best distinguishes instance-based from model-based learning in terms of generalization?

A) Instance-based learning generalizes by similarity to stored examples, whereas model-based learning generalizes through a fitted predictive model ✓ Correct

B) Instance-based learning requires labels, whereas model-based learning never uses labels

C) Instance-based learning is always online, whereas model-based learning is always batch

D) Instance-based learning cannot be used for regression, whereas model-based learning cannot be used for classification

Rationale: The core distinction is whether prediction comes from comparing to learned instances or from applying a fitted model.

Q146.

A team chooses a linear model for a problem because it assumes the relationship is approximately linear and deviations are noise. What broader methodological point does this illustrate?

A) Modeling always avoids assumptions about the data

B) Every model selection embeds assumptions about what structure matters and what can be ignored ✓ Correct

C) The No Free Lunch theorem proves linear models are optimal for noisy data

D) Hyperparameters remove the need for assumptions

Rationale: The chapter explains that models are simplifications and therefore necessarily rely on assumptions about the data.

Q147.

A mobile health app team finds poor validation performance after training on web images. They then evaluate on a held-out train-dev set drawn from the same web distribution and observe poor performance there as well. What is the best next conclusion?

A) The main issue is likely overfitting or poor fit to the training distribution rather than only data mismatch ✓ Correct

B) The validation set must be mislabeled because train-dev is poor too

C) The model is necessarily optimal because two sets were used

D) The test set should now be used for hyperparameter tuning

Rationale: Poor train-dev performance indicates the model is not generalizing well even within the training distribution, pointing to overfitting or poor fit.

Q148.

A manager asks why the team cannot simply pick the single best algorithm from a textbook and apply it to every predictive problem. Which response is most accurate?

A) Because algorithm performance depends on the dataset, and no model is guaranteed best a priori across all problems ✓ Correct

B) Because only unsupervised algorithms are reusable across domains

C) Because supervised algorithms cannot generalize

D) Because model parameters cannot be estimated on new data

Rationale: The No Free Lunch theorem implies that no algorithm is universally best for all datasets.

Q149.

A startup with limited engineering capacity wants to reduce a huge hand-maintained rule base for spam detection while improving adaptability. Which recommendation best synthesizes the chapter's reasoning?

A) Replace the system with a supervised ML spam filter trained on labeled spam and ham examples ✓ Correct

B) Replace the system with pure clustering because labels are unnecessary for spam detection

C) Keep the rule base and only add more handcrafted exceptions

D) Use reinforcement learning because email users provide rewards implicitly

Rationale: The chapter uses spam filtering as a prime example where supervised ML can simplify code and adapt better than long rule lists.

Q150.

A team must choose between spending its final budget on more representative data collection, better feature engineering, or a more complex model. Current evidence shows the model underfits, several important predictors are missing, and the existing sample is biased toward one subgroup. Which priority order is most defensible?

A) Increase model complexity first and ignore data issues because underfitting proves the data are adequate

B) Address representativeness and missing relevant features before assuming complexity alone will solve the problem ✓ Correct

C) Tune the test set repeatedly until training and test error align

D) Switch to instance-based learning because it eliminates bias in the sample

Rationale: Biased samples and missing relevant features can fundamentally limit learning, so data representativeness and feature quality should be addressed before relying solely on added complexity.