Lesson 5.1: Machine Learning Foundations for Finance – Quiz

Digital Finance Course

Question 1

A bank wants to predict whether a loan applicant will default (yes/no). Which type of supervised learning task is this?

A. Clustering
B. Anomaly detection
C. Classification
D. Regression

Question 1

A bank wants to predict whether a loan applicant will default (yes/no). Which type of supervised learning task is this?

A. Clustering
B. Anomaly detection
C. Classification
D. Regression

[Answer hidden – compile with \solutionstrue to reveal]
Predicting a categorical outcome (default yes/no) is a classification task. Regression predicts continuous values, clustering is unsupervised, and anomaly detection does not require labeled data.

Question 2

A model achieves 99.5% training accuracy but only 72% test accuracy on credit default data. What problem does this indicate?

A. Class imbalance
B. Underfitting (high bias)
C. Data leakage
D. Overfitting (high variance)

Question 2

A model achieves 99.5% training accuracy but only 72% test accuracy on credit default data. What problem does this indicate?

A. Class imbalance
B. Underfitting (high bias)
C. Data leakage
D. Overfitting (high variance)

[Answer hidden – compile with \solutionstrue to reveal]
The large gap between training accuracy (99.5%) and test accuracy (72%) is the classic signature of overfitting: the model memorized training data instead of learning generalizable patterns.

Question 3

Which component of total prediction error CANNOT be reduced by a better model?

A. Irreducible error (noise)
B. Variance
C. Training error
D. Bias

Question 3

Which component of total prediction error CANNOT be reduced by a better model?

A. Irreducible error (noise)
B. Variance
C. Training error
D. Bias

[Answer hidden – compile with \solutionstrue to reveal]
Irreducible error comes from inherent randomness in the data (noise) and cannot be eliminated by any model. Bias and variance can be managed through model selection and tuning.

Question 4

Your fraud detection model runs standard 5-fold cross-validation on 3 years of transaction data. What critical mistake was made?

A. 5 folds is too few; use 10 folds

B. Cross-validation is unnecessary for fraud detection

C. The dataset is too large for cross-validation

D. Standard k-fold shuffles data, allowing future transactions to leak into training

Question 4

Your fraud detection model runs standard 5-fold cross-validation on 3 years of transaction data. What critical mistake was made?

A. 5 folds is too few; use 10 folds

B. Cross-validation is unnecessary for fraud detection

C. The dataset is too large for cross-validation

D. Standard k-fold shuffles data, allowing future transactions to leak into training

[Answer hidden – compile with \solutionstrue to reveal]
Standard k-fold randomly shuffles data, violating temporal order. Future transactions can leak into training folds (look-ahead bias). Financial data requires time-series split: always train on past, test on future.

Question 5

Which statement about decision trees is FALSE?

A. They are prone to overfitting when grown too deep
B. They require feature scaling (normalization) to work properly
C. They are highly interpretable
D. They handle both numeric and categorical features

Question 5

Which statement about decision trees is FALSE?

A. They are prone to overfitting when grown too deep

B. They require feature scaling (normalization) to work properly

C. They are highly interpretable

D. They handle both numeric and categorical features

[Answer hidden – compile with \solutionstrue to reveal]
Decision trees split on thresholds and do NOT require feature scaling. This is an advantage over distance-based models (e.g., k-NN, SVMs) and neural networks. All other statements are true.

Question 6

How does a random forest reduce the variance problem of individual decision trees?

A. By using a single tree with regularization
B. By training each tree on a different bootstrap sample with random feature subsets, then averaging
C. By removing outliers from the training data
D. By pruning each tree to maximum depth 3

Question 6

How does a random forest reduce the variance problem of individual decision trees?

A. By using a single tree with regularization

B. By training each tree on a different bootstrap sample with random feature subsets, then averaging

C. By removing outliers from the training data

D. By pruning each tree to maximum depth 3

[Answer hidden – compile with \solutionstrue to reveal]
Random forests use bagging (bootstrap aggregating) with random feature selection to create diverse trees. Averaging many diverse, overfitting trees cancels out individual errors, reducing variance.

Question 7

What distinguishes gradient boosting from random forest?

A. Gradient boosting builds trees sequentially, each correcting the errors of previous trees
B. Gradient boosting cannot handle missing values
C. Gradient boosting builds trees in parallel; random forest builds them sequentially
D. Gradient boosting uses neural networks; random forest uses trees

Question 7

What distinguishes gradient boosting from random forest?

A. Gradient boosting builds trees sequentially, each correcting the errors of previous trees

B. Gradient boosting cannot handle missing values

C. Gradient boosting builds trees in parallel; random forest builds them sequentially

D. Gradient boosting uses neural networks; random forest uses trees

[Answer hidden – compile with \solutionstrue to reveal]
Gradient boosting builds trees sequentially—each new tree focuses on the residual errors of the ensemble so far. Random forest builds trees independently in parallel. XGBoost also natively handles missing values.

Question 8

In a confusion matrix for fraud detection with 10,000 transactions, the model produces: TP=240, FP=160, FN=60, TN=9,540. What is the precision?

A. 97.8%

B. $240/10{,}000 = 2.4\%$

C. 80.0%

D. 60.0%

Question 8

In a confusion matrix for fraud detection with 10,000 transactions, the model produces: TP=240, FP=160, FN=60, TN=9,540. What is the precision?

A. 97.8%

B. 240/10,000 = 2.4%

C. 80.0%

D. 60.0%

[Answer hidden – compile with \solutionstrue to reveal]

Precision = $\frac{TP}{TP+FP} = \frac{240}{240+160} = \frac{240}{400} = 60.0\%$. Of every 100 transactions flagged as fraud, 60 actually were fraud and 40 were false alarms.

Question 9

Using the same confusion matrix (TP=240, FP=160, FN=60, TN=9,540), what is the recall?

A. 80.0%
B. 97.8%
C. 60.0%
D. 24.0%

Question 9

Using the same confusion matrix (TP=240, FP=160, FN=60, TN=9,540), what is the recall?

A. 80.0%

B. 97.8%

C. 60.0%

D. 24.0%

[Answer hidden – compile with \solutionstrue to reveal]

Recall $= \frac{TP}{TP+FN} = \frac{240}{240+60} = \frac{240}{300} = 80.0\%$. The model catches 80% of actual fraud cases, but misses 20% (60 fraudulent transactions slip through).

Question 10

A "naïve" model predicts every transaction as legitimate. With 99% legitimate and 1% fraud, what is its accuracy and recall for fraud?

A. Accuracy = 99%, Recall = 0%
B. Accuracy = 1%, Recall = 100%
C. Accuracy = 50%, Recall = 50%
D. Accuracy = 99%, Recall = 99%

Question 10

A "naïve" model predicts every transaction as legitimate. With 99% legitimate and 1% fraud, what is its accuracy and recall for fraud?

A. Accuracy = 99%, Recall = 0%

B. Accuracy = 1%, Recall = 100%

C. Accuracy = 50%, Recall = 50%

D. Accuracy = 99%, Recall = 99%

[Answer hidden – compile with \solutionstrue to reveal]
Predicting "always legitimate" gets 99% accuracy (correct for 99% of data) but 0% recall for fraud (catches zero fraud cases). This demonstrates why accuracy is misleading for imbalanced datasets.

Question 11

The F1-score is the harmonic mean of precision and recall. If precision = 60% and recall = 80%, what is F1?

A. 48.0%
B. 68.6%
C. 70.0%
D. 80.0%

Question 11

The F1-score is the harmonic mean of precision and recall. If precision = 60% and recall = 80%, what is F1?

A. 48.0%

B. 68.6%

C. 70.0%

D. 80.0%

[Answer hidden – compile with \solutionstrue to reveal]

$F_1 = 2 \cdot \frac{0.60 \times 0.80}{0.60 + 0.80} = 2 \cdot \frac{0.48}{1.40} = \frac{0.96}{1.40} \approx 0.686 = 68.6\%$. The harmonic mean penalizes extreme imbalance between precision and recall.

Question 12

What does an AUC (Area Under the ROC Curve) of 0.50 indicate?

A. The model is 50% accurate
B. The model has no discriminative ability (equivalent to random guessing)
C. The model achieves 50% precision
D. Half the predictions are correct

Question 12

What does an AUC (Area Under the ROC Curve) of 0.50 indicate?

A. The model is 50% accurate

B. The model has no discriminative ability (equivalent to random guessing)

C. The model achieves 50% precision

D. Half the predictions are correct

[Answer hidden – compile with \solutionstrue to reveal]
AUC = 0.50 means the model ranks positives and negatives no better than chance. The ROC curve lies on the diagonal. A useful model should have AUC significantly above 0.50.

Question 13

For a credit scoring model, which evaluation metric best captures the model's ability to rank applicants by default risk?

A. Mean Squared Error (MSE)
B. AUC (Area Under ROC Curve)
C. Accuracy
D. Precision

Question 13

For a credit scoring model, which evaluation metric best captures the model's ability to rank applicants by default risk?

A. Mean Squared Error (MSE)

B. AUC (Area Under ROC Curve)

C. Accuracy

D. Precision

[Answer hidden – compile with \solutionstrue to reveal]
AUC measures the model's ranking ability: the probability that it assigns a higher default score to a true defaulter than a non-defaulter. This is exactly what credit scoring requires.

Question 14

Which feature engineering technique is MOST appropriate for detecting money laundering patterns in transaction data?

A. Rolling 30-day statistics (mean, std, max) of transaction amounts
B. Log transformation of account balance
C. Standardizing all features to zero mean
D. One-hot encoding of transaction currency

Question 14

Which feature engineering technique is MOST appropriate for detecting money laundering patterns in transaction data?

A. Rolling 30-day statistics (mean, std, max) of transaction amounts

B. Log transformation of account balance

C. Standardizing all features to zero mean

D. One-hot encoding of transaction currency

[Answer hidden – compile with \solutionstrue to reveal]
Rolling window statistics capture behavioral patterns over time (e.g., sudden spikes in transaction volume). AML detection relies on temporal patterns, making rolling aggregations the most informative feature type.

Question 15

A bank has 10 million labeled transactions (fraud/legitimate) spanning 5 years. Which model is the best starting point?

A. Isolation Forest for anomaly detection
B. Deep neural network with 20 hidden layers
C. Simple decision tree with max_depth=3
D. XGBoost with time-series cross-validation

Question 15

A bank has 10 million labeled transactions (fraud/legitimate) spanning 5 years. Which model is the best starting point?

A. Isolation Forest for anomaly detection

B. Deep neural network with 20 hidden layers

C. Simple decision tree with max_depth=3

D. XGBoost with time-series cross-validation

[Answer hidden – compile with \solutionstrue to reveal]
With abundant labeled data on tabular features, XGBoost with time-series CV is the industry standard. Deep NNs rarely outperform XGBoost on tabular data. Isolation Forest is unsupervised and wastes the labels.

Question 16

When is Isolation Forest preferred over supervised classification for fraud detection?

A. When you need the highest possible accuracy
B. When you have millions of labeled fraud examples
C. When fraud patterns are well-known and stable
D. When labeled fraud data is scarce or fraud patterns evolve rapidly

Question 16

When is Isolation Forest preferred over supervised classification for fraud detection?

A. When you need the highest possible accuracy

B. When you have millions of labeled fraud examples

C. When fraud patterns are well-known and stable

D. When labeled fraud data is scarce or fraud patterns evolve rapidly

[Answer hidden – compile with \solutionstrue to reveal]
Isolation Forest excels when labels are scarce, unreliable, or when new (previously unseen) fraud types emerge. It detects anomalies without needing examples of each fraud type.

Question 17

Which statement about neural networks in finance is CORRECT?

A. Neural networks always outperform tree-based models on tabular financial data
B. Neural networks are easily interpretable, satisfying regulatory requirements
C. Neural networks excel with unstructured data (text, images) but often underperform XGBoost on tabular data
D. Neural networks require less data than decision trees

Question 17

Which statement about neural networks in finance is CORRECT?

A. Neural networks always outperform tree-based models on tabular financial data

B. Neural networks are easily interpretable, satisfying regulatory requirements

C. Neural networks excel with unstructured data (text, images) but often underperform XGBoost on tabular data

D. Neural networks require less data than decision trees

[Answer hidden – compile with \solutionstrue to reveal]
Research consistently shows tree-based methods (XGBoost, LightGBM) outperform neural networks on tabular data. Neural networks shine with unstructured data (NLP for news sentiment, image-based document processing).

Question 18

What is backpropagation?

A. A technique to select features by removing the least important ones
B. A regularization technique to prevent overfitting
C. The algorithm that computes gradients of the loss function and updates neural network weights
D. A method to split data into training and test sets

Question 18

What is backpropagation?

A. A technique to select features by removing the least important ones

B. A regularization technique to prevent overfitting

C. The algorithm that computes gradients of the loss function and updates neural network weights

D. A method to split data into training and test sets

[Answer hidden – compile with \solutionstrue to reveal]
Backpropagation computes how much each weight contributed to the prediction error, then updates weights using gradient descent. It propagates the error signal backward from the output layer to the input layer.

Question 19

A trading firm's ML model performs well in backtesting (2015–2023) but loses money when deployed live in 2024. Which pitfall is most likely?

A. The model used too few features

B. The model was underfitting

C. Non-stationarity: market regimes changed, making historical patterns invalid

D. The training data was too large

Question 19

A trading firm's ML model performs well in backtesting (2015–2023) but loses money when deployed live in 2024. Which pitfall is most likely?

A. The model used too few features

B. The model was underfitting

C. Non-stationarity: market regimes changed, making historical patterns invalid

D. The training data was too large

[Answer hidden – compile with \solutionstrue to reveal]
Financial markets are non-stationary—patterns that held in 2015–2023 may not persist in 2024. This is the fundamental challenge of financial ML: the future distribution differs from the past.

Question 20

EU regulations (e.g., EU AI Act, GDPR) require that financial ML models be explainable. Which model type BEST satisfies this requirement?

A. Ensemble of 1,000 random forest trees
B. Isolation Forest
C. Logistic regression or decision tree with limited depth
D. Deep neural network with 50 layers

Question 20

EU regulations (e.g., EU AI Act, GDPR) require that financial ML models be explainable. Which model type BEST satisfies this requirement?

A. Ensemble of 1,000 random forest trees

B. Isolation Forest

C. Logistic regression or decision tree with limited depth

D. Deep neural network with 50 layers

[Answer hidden – compile with \solutionstrue to reveal]
Logistic regression provides directly interpretable coefficients, and shallow decision trees provide transparent decision paths. Regulators require that affected individuals understand why a decision was made ("right to explanation").