Lesson 2.2 Quiz: Credit Scoring — From FICO to ML

Module 2: The Access Problem

Prof. Dr. Joerg Osterrieder

Digital Finance — BSc Course (v2026.05)

Q1: Purpose of a Credit Score

What is the primary purpose of a credit score?

- A. To verify the borrower's identity
- B. To determine the exact amount a borrower should receive
- C. To calculate the interest rate on a mortgage
- D. To predict the probability that a borrower will default within a defined time horizon

Q1: Purpose of a Credit Score

What is the primary purpose of a credit score?

A. To verify the borrower's identity
B. To determine the exact amount a borrower should receive
C. To calculate the interest rate on a mortgage
D. To predict the probability that a borrower will default within a defined time horizon

[Answer hidden – compile with \solutionstrue to reveal]

Q2: FICO Score Components

Which factor carries the **highest weight** in a traditional FICO credit score?

- A. Payment history (35%)
- B. Amounts owed / credit utilisation (30%)
- C. Length of credit history (15%)
- D. Credit mix (10%)

Q2: FICO Score Components

Which factor carries the **highest weight** in a traditional FICO credit score?

- (A) Payment history (35%)
- (B) Amounts owed / credit utilisation (30%)
- (C) Length of credit history (15%)
- (D) Credit mix (10%)

[Answer hidden – compile with \solutionstrue to reveal]

Q3: WoE Interpretation

A bin in a credit scorecard has a Weight of Evidence (WoE) of $+0.40$. What does this indicate?

- A The bin contains exactly 40% of all defaults
- B The bin has proportionally more bads than goods
- C The bin has proportionally more goods (non-defaulters) than bads (defaulters)
- D The feature is not predictive

Q3: WoE Interpretation

A bin in a credit scorecard has a Weight of Evidence (WoE) of $+0.40$. What does this indicate?

- A The bin contains exactly 40% of all defaults
- B The bin has proportionally more bads than goods
- C The bin has proportionally more goods (non-defaulters) than bads (defaulters)
- D The feature is not predictive

[Answer hidden – compile with \solutionstrue to reveal]

Q4: Information Value Thresholds

An analyst calculates the Information Value (IV) of a candidate feature and obtains IV $= 0.03$. How should this feature be classified?

- A. Weak predictor
- B. Medium predictor
- C. Suspicious — check for data leakage
- D. Strong predictor

Q4: Information Value Thresholds

An analyst calculates the Information Value (IV) of a candidate feature and obtains IV $= 0.03$. How should this feature be classified?

- (A) Weak predictor
- (B) Medium predictor
- (C) Suspicious — check for data leakage
- (D) Strong predictor

[Answer hidden – compile with \solutionstrue to reveal]

Q5: WoE Calculation

A bin contains 20% of all goods and 10% of all bads. What is the WoE of this bin?

- A $0.20 \times 0.10 = 0.02$
- B $0.20 - 0.10 = 0.10$
- C $\ln(0.10/0.20) = -0.693$
- D $\ln(0.20/0.10) = 0.693$

Q5: WoE Calculation

A bin contains 20% of all goods and 10% of all bads. What is the WoE of this bin?

- (A) $0.20 \times 0.10 = 0.02$
- (B) $0.20 - 0.10 = 0.10$
- (C) $\ln(0.10/0.20) = -0.693$
- (D) $\ln(0.20/0.10) = 0.693$

[Answer hidden – compile with \solutionstrue to reveal]

Q6: Expected Loss Calculation

A personal loan has PD $=$ 4%, LGD $=$ 45%, and EAD $=$ \$20,000. What is the expected loss?

- A \$3,600
- B \$36
- C \$900
- D \$360

Q6: Expected Loss Calculation

A personal loan has PD $=$ 4%, LGD $=$ 45%, and EAD $=$ \$20,000. What is the expected loss?

- (A) \$3,600
- (B) \$36
- (C) \$900
- (D) \$360

[Answer hidden – compile with \solutionstrue to reveal]

Q7: Gini from AUC

A credit scoring model achieves an AUC of 0.78 on the validation dataset. What is the corresponding Gini coefficient?

- A. 0.78
- B. 1.56
- C. 0.56
- D. 0.39

Q7: Gini from AUC

A credit scoring model achieves an AUC of 0.78 on the validation dataset. What is the corresponding Gini coefficient?

- (A) 0.78
- (B) 1.56
- (C) 0.56
- (D) 0.39

[Answer hidden – compile with \solutionstrue to reveal]

Q8: ROC Curve Interpretation

A credit model has AUC $= 0.50$ on out-of-time validation data. What does this mean?

- A. The model is perfectly calibrated
- B. The model is overfitting
- C. The model has moderate discrimination
- D. The model has no discriminatory power—equivalent to random guessing

Q8: ROC Curve Interpretation

A credit model has AUC $= 0.50$ on out-of-time validation data. What does this mean?

- A. The model is perfectly calibrated
- B. The model is overfitting
- C. The model has moderate discrimination
- D. The model has no discriminatory power—equivalent to random guessing

[Answer hidden – compile with \solutionstrue to reveal]

Q9: PSI Interpretation

A quarterly monitoring report shows PSI $= 0.18$ for the application scorecard. What action is recommended?

- A. No action — this is within acceptable limits
- B. Investigate the source of the shift
- C. Retire the scorecard
- D. Immediately rebuild the model

Q9: PSI Interpretation

A quarterly monitoring report shows PSI $= 0.18$ for the application scorecard. What action is recommended?

- A No action — this is within acceptable limits
- B Investigate the source of the shift
- C Retire the scorecard
- D Immediately rebuild the model

[Answer hidden – compile with \solutionstrue to reveal]

Q10: IV Calculation

A feature has two bins. Bin 1: Distr. Goods $= 0.60$, Distr. Bads $= 0.40$, WoE $= 0.405$. Bin 2: Distr. Goods $= 0.40$, Distr. Bads $= 0.60$, WoE $= -0.405$. What is the IV?

- A 0.081
- B 0.000
- C 0.162
- D 0.405

Q10: IV Calculation

A feature has two bins. Bin 1: Distr. Goods $= 0.60$, Distr. Bads $= 0.40$, WoE $= 0.405$. Bin 2: Distr. Goods $= 0.40$, Distr. Bads $= 0.60$, WoE $= -0.405$. What is the IV?

- (A) 0.081
- (B) 0.000
- (C) 0.162
- (D) 0.405

[Answer hidden – compile with \solutionstrue to reveal]

Q11: Score-to-PD Conversion

A scorecard uses the formula $\text{Score} = 600 - 40 \times \ln(\text{PD}/(1 - \text{PD}))$. A borrower receives a score of 640. Which statement is correct about their PD?

- A. The score is invalid because it exceeds 600
- B. PD is lower than the borrower with score 600
- C. PD is higher than the borrower with score 600
- D. PD is exactly 50%

Q11: Score-to-PD Conversion

A scorecard uses the formula Score $= 600 - 40 \times \ln(\mathrm{PD}/(1 - \mathrm{PD}))$. A borrower receives a score of 640. Which statement is correct about their PD?

- A. The score is invalid because it exceeds 600
- B. PD is lower than the borrower with score 600
- C. PD is higher than the borrower with score 600
- D. PD is exactly 50%

[Answer hidden – compile with \solutionstrue to reveal]

Q12: Gradient Boosting Hyperparameters

When building a credit scoring model with gradient boosting, a data scientist sets `max_depth = 3` and `learning_rate = 0.05`. What is the rationale?

- A. These settings are only valid for random forests, not gradient boosting
- B. Shallow trees cannot capture non-linear relationships
- C. Deep trees with fast learning maximise training speed
- D. Shallow trees with slow learning reduce overfitting and improve generalisation

Q12: Gradient Boosting Hyperparameters

When building a credit scoring model with gradient boosting, a data scientist sets `max_depth = 3` and `learning_rate = 0.05`. What is the rationale?

- (A) These settings are only valid for random forests, not gradient boosting
- (B) Shallow trees cannot capture non-linear relationships
- (C) Deep trees with fast learning maximise training speed
- (D) Shallow trees with slow learning reduce overfitting and improve generalisation

[Answer hidden – compile with \solutionstrue to reveal]

Q13: Discrimination vs. Calibration

A credit model has AUC $= 0.82$ but consistently predicts PDs that are 3 percentage points lower than observed default rates. What is the diagnosis?

- A. The model has good discrimination but poor calibration
- B. The model has poor discrimination but good calibration
- C. The model has poor discrimination and poor calibration
- D. The model has good discrimination and good calibration

Q13: Discrimination vs. Calibration

A credit model has AUC $= 0.82$ but consistently predicts PDs that are 3 percentage points lower than observed default rates. What is the diagnosis?

- (A) The model has good discrimination but poor calibration
- (B) The model has poor discrimination but good calibration
- (C) The model has poor discrimination and poor calibration
- (D) The model has good discrimination and good calibration

[Answer hidden – compile with \solutionstrue to reveal]

Q14: Alternative Data Trade-off

A FinTech lender in East Africa uses mobile airtime top-up patterns to score thin-file borrowers. Which risk is **most specific** to this alternative data source?

- A The logistic regression model may overfit
- B The FICO score range is too narrow for this population
- C Mobile usage patterns may change rapidly, causing model drift
- D The model's AUC will always be below 0.50

Q14: Alternative Data Trade-off

A FinTech lender in East Africa uses mobile airtime top-up patterns to score thin-file borrowers. Which risk is **most specific** to this alternative data source?

- A The logistic regression model may overfit
- B The FICO score range is too narrow for this population
- C Mobile usage patterns may change rapidly, causing model drift
- D The model's AUC will always be below 0.50

[Answer hidden – compile with \solutionstrue to reveal]

Q15: Proxy Discrimination

A credit model uses zip code as a feature. The zip code has high IV and improves AUC. However, zip codes are highly correlated with race in the US. What is the issue?

- A. Zip codes should always be included because they improve accuracy
- B. Zip code acts as a proxy for a protected characteristic, creating illegal discrimination
- C. This is only a problem if race is directly included in the model
- D. The model will have low AUC because zip code is not predictive

Q15: Proxy Discrimination

A credit model uses zip code as a feature. The zip code has high IV and improves AUC. However, zip codes are highly correlated with race in the US. What is the issue?

- A Zip codes should always be included because they improve accuracy
- B Zip code acts as a proxy for a protected characteristic, creating illegal discrimination
- C This is only a problem if race is directly included in the model
- D The model will have low AUC because zip code is not predictive

[Answer hidden – compile with \solutionstrue to reveal]

Q16: Feedback Loop in Credit Scoring

A bank declines 30% of applicants based on its scorecard. Declined applicants never generate repayment data. Over time, what happens to the model?

- A The model improves because it only trains on successful loans
- B The model becomes biased toward the approved population, potentially reinforcing initial biases
- C The model automatically corrects for missing data
- D Nothing changes because the model was validated at deployment

Q16: Feedback Loop in Credit Scoring

A bank declines 30% of applicants based on its scorecard. Declined applicants never generate repayment data. Over time, what happens to the model?

- A The model improves because it only trains on successful loans
- B The model becomes biased toward the approved population, potentially reinforcing initial biases
- C The model automatically corrects for missing data
- D Nothing changes because the model was validated at deployment

[Answer hidden – compile with \solutionstrue to reveal]

Q17: ML vs. Logistic Regression

Compared to a well-built logistic regression scorecard, gradient boosting typically shows the largest AUC improvement in which scenario?

- A. When hundreds of features including alternative data and interaction effects are available
- B. When all features have linear relationships with log-odds of default
- C. When only 5 traditional bureau features are available
- D. When the training dataset has fewer than 100 observations

Q17: ML vs. Logistic Regression

Compared to a well-built logistic regression scorecard, gradient boosting typically shows the largest AUC improvement in which scenario?

- A. When hundreds of features including alternative data and interaction effects are available
- B. When all features have linear relationships with log-odds of default
- C. When only 5 traditional bureau features are available
- D. When the training dataset has fewer than 100 observations

[Answer hidden – compile with \solutionstrue to reveal]

Q18: Model Validation Dimension

A bank's credit model shows AUC $= 0.75$ at deployment but AUC $= 0.62$ twelve months later. Which validation dimension has failed?

- A. Stability (temporal performance)
- B. Calibration
- C. Concentration
- D. Stress testing

Q18: Model Validation Dimension

A bank's credit model shows AUC $= 0.75$ at deployment but AUC $= 0.62$ twelve months later. Which validation dimension has failed?

- A. Stability (temporal performance)
- B. Calibration
- C. Concentration
- D. Stress testing

[Answer hidden – compile with \solutionstrue to reveal]

Q19: Two-Model Approach Evaluation

A FinTech lender uses a gradient boosting model (AUC $= 0.84$) for decisioning and a logistic regression surrogate (AUC $= 0.76$) for generating adverse action reasons. A regulator questions this approach. Which criticism is **most valid**?

- A. Two models always cost more than one, so this is wasteful
- B. The logistic model is too inaccurate to use at all
- C. The surrogate may give reasons that do not reflect the actual decision logic of the ML model
- D. The gradient boosting model should be abandoned in favour of the logistic model

Q19: Two-Model Approach Evaluation

A FinTech lender uses a gradient boosting model (AUC $= 0.84$) for decisioning and a logistic regression surrogate (AUC $= 0.76$) for generating adverse action reasons. A regulator questions this approach. Which criticism is **most valid**?

- A. Two models always cost more than one, so this is wasteful
- B. The logistic model is too inaccurate to use at all
- C. The surrogate may give reasons that do not reflect the actual decision logic of the ML model
- D. The gradient boosting model should be abandoned in favour of the logistic model

[Answer hidden – compile with \solutionstrue to reveal]

Q20: Policy Recommendation

A developing country's central bank is deciding whether to allow FinTech lenders to use social media data for credit scoring. Which recommendation best balances inclusion and fairness?

- A. Require all lenders to use only FICO-equivalent bureau scores
- B. Permit social media data with mandatory fairness testing, opt-in consent, and regular bias audits
- C. Allow unrestricted use of social media data to maximise financial inclusion
- D. Ban all alternative data to protect consumer privacy

Q20: Policy Recommendation

A developing country's central bank is deciding whether to allow FinTech lenders to use social media data for credit scoring. Which recommendation best balances inclusion and fairness?

- A. Require all lenders to use only FICO-equivalent bureau scores
- B. Permit social media data with mandatory fairness testing, opt-in consent, and regular bias audits
- C. Allow unrestricted use of social media data to maximise financial inclusion
- D. Ban all alternative data to protect consumer privacy

[Answer hidden – compile with \solutionstrue to reveal]