**ES 335: Machine Learning**

# Quiz 1 — Foundations, Trees & Nearest Neighbours

Date: 2026-08-28 Duration: 50 minutes Maximum marks: 30

**SET D**

Name: _____ Roll No.: _____

**Instructions**

- Write your name and roll number before starting.
- This paper has 12 questions in 3 parts, for a total of 30 marks.
- Record your final MCQ answers in the answer grid below — the grid is what gets graded.
- All logarithms are base 2 unless stated otherwise.
- Calculators are permitted; devices with network access are not.

**MCQ Answer Grid**

| Q | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Ans | | | | | | |

## Part A — Multiple Choice

*12 marks*

*Choose the single best answer unless a question says otherwise.*

---

**Q1.** A binary classifier produces the confusion matrix below. *[2 marks]*

| | **Predicted +** | **Predicted −** |
|---|---|---|
| **Actual +** | 40 | 10 |
| **Actual −** | 20 | 30 |

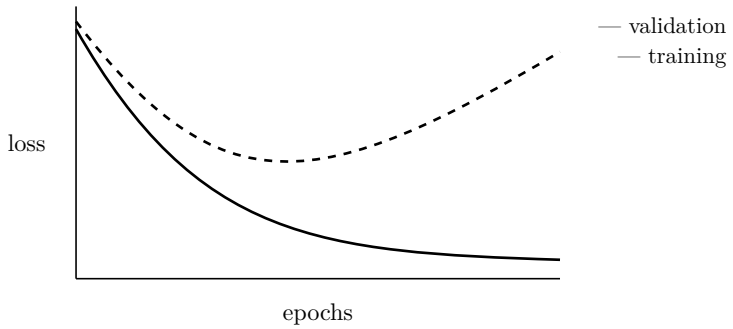
What is its precision?

(A) 40/100
(B) 40/60
(C) 70/100
(D) 40/50

**Q2.** Which of the following objectives are convex in their parameters? *(Select all that apply.)* *[2 marks]*

(A) Logistic regression negative log-likelihood
(B) The $k$-means clustering objective (jointly in assignments and centroids)
(C) Linear regression with squared loss
(D) A two-layer neural network with squared loss

**Q3.** The curves below were logged during training. *[2 marks]*



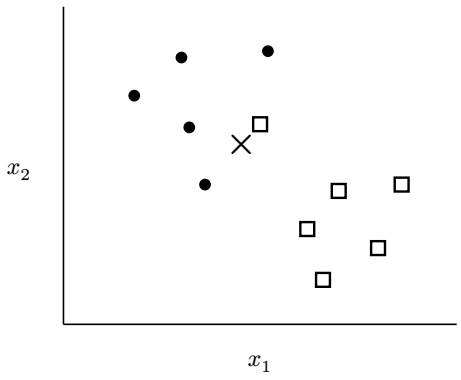

What is the most appropriate response?

(A) Train longer — the model is underfitting
(B) Increase model capacity — both losses are too high
(C) Lower the learning rate — training has diverged
(D) Stop earlier or regularize — the model is overfitting

**Q4.** Increasing $k$ in $k$-NN typically has which effect? *[2 marks]*

(A) Increases both bias and variance
(B) Decreases bias, increases variance
(C) Increases bias, decreases variance
(D) None of the above

**Q5.** The figure shows training points from two classes and a query point $\times$. *[2 marks]*




Using Euclidean distance, how does the query point get classified by 1-NN and by 3-NN respectively?

(A) 1-NN: class B (□), 3-NN: class B (□)
(B) 1-NN: class B (□), 3-NN: class A (●)
(C) 1-NN: class A (●), 3-NN: tie
(D) 1-NN: class A (●), 3-NN: class A (●)

**Q6.** For a split that sends all positives left and all negatives right, the weighted entropy of the children is: *[2 marks]*

(A) $\log_2 3$
(B) 0
(C) 1
(D) 0.5

## Part B — Fill in the Blanks

*4 marks*

**Q7.** For linear regression in matrix form, the least-squares solution is $\hat{\theta} =$ $(X^\top X)^{\text{\_\_\_\_\_}} X^\top$ _____. *[2 marks]*

**Q8.** Compared with L2, L1 regularization tends to produce weight vectors that are _____. *[1 mark]*

**Q9.** Computing inner products in a high-dimensional feature space without ever constructing the features is known as the kernel _____. *[1 mark]*

## Part C — Long Answers

*14 marks*

*Answer in the space provided; show your working.*

---

**Q10.** The table records eight days of data for predicting whether a match is *Played*. *[5 marks]*

| Day | Outlook | Windy | Played |
|---|---|---|---|
| 1 | Sunny | No | Yes |
| 2 | Sunny | Yes | No |
| 3 | Overcast | No | Yes |
| 4 | Overcast | Yes | Yes |
| 5 | Rainy | No | Yes |
| 6 | Rainy | Yes | No |
| 7 | Sunny | No | No |
| 8 | Rainy | No | Yes |

Compute (i) the entropy of *Played*, and (ii) the information gain of splitting on *Windy*. Which single-feature split would a decision tree prefer if the gain for *Outlook* is 0.311?

**Q11.** Starting from $L(\theta) = \|y - X\theta\|_2^2$, derive the normal equation. State the condition under which the solution is unique, and give one practical remedy when it is not. *[5 marks]*

**Q12.** Your colleague reports 99.2% accuracy on a fraud-detection dataset in which 0.8% of transactions are fraudulent. Explain why this number is not impressive, and name two evaluation metrics (or tools) better suited to this setting, justifying each in one line. *[4 marks]*

*— End of paper —*