

CS 4410: Midterm Summer 2026B

Exam Time: 50 Minutes

Total Marks: 30

Write Your section, your Troy ID, and your name on every sheet.

Troy ID:

Name:

Q1 [10 points] Answer any two of the following. Each response is worth 5 points.

- (A) Distinguish the likelihood $p(\mathcal{X} \mid \theta)$ from the posterior $p(\theta \mid \mathcal{X})$. Why is the log-likelihood commonly optimized instead of the original likelihood product?
- (B) Distinguish a validation set from a test set. Explain how high bias and high variance appear in training and validation performance.

Part II: Numerical and Problem-Solving Questions

20 points

Q2 [4 points] A disease has prevalence $P(D) = 0.01$. A diagnostic test has sensitivity

$$P(+ \mid D) = 0.80$$

and false-positive rate

$$P(+ \mid \overline{D}) = 0.096.$$

A randomly selected person receives a positive result. Compute $P(D \mid +)$ using Bayes' rule. Show all steps and briefly interpret the result.

Q3 [4 points] Assume that the observations

$$\mathcal{X} = \{2, 3, 5, 6, 9\}$$

are independently drawn from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$. Compute the maximum-likelihood estimates

$$\hat{\mu}_{ML} \quad \text{and} \quad \hat{\sigma}_{ML}^2$$

CS 4410: Midterm 2026C

Exam Time: 50 Minutes

Total Marks: 30

Write Your section, your Troy ID, and your name on every sheet.

Troy ID:

Name:

Part I: Short-Answer Questions

10 points

Q1 [10 points] Answer following. Each response is worth 5 points.

- (A) Explain the meanings of the specific boundary S , the general boundary G , and the version space in Candidate Elimination. What does an empty version space indicate?
- (B) List the major stages of a complete machine-learning project, from defining the problem to presenting the solution.

Part II: Numerical and Problem-Solving Questions

20 points

Q2 [7 points] Apply the Candidate Elimination algorithm to the training data below. Show the S and G boundaries after every example and state the final version space.

Weather	Temperature	Wind	Class
Sunny	Warm	Weak	+
Rainy	Warm	Weak	+
Cloudy	Cold	Strong	−
Sunny	Cold	Weak	+

Q3 [7 points] Consider two classes C_1 and C_2 with the following loss matrix:

	C_1	C_2
α_1 : choose C_1	0	5
α_2 : choose C_2	2	0

Let

$$P(C_1 \mid x) = p, \quad P(C_2 \mid x) = 1 - p.$$

- (a) Write $R(\alpha_1 \mid x)$ and $R(\alpha_2 \mid x)$.
- (b) Derive the decision threshold without rejection.
- (c) Add a reject action α_R with loss 0.6 for either true class. Determine the ranges of p for choosing C_1 , choosing C_2 , and rejecting.

Q4 [6 points] Consider the feature values

$$X = \{10, 20, 30, 40, 100\}.$$

For $x = 40$, compute:

- (a) Min–Max scaling,

$$x' = \frac{x - X_{\min}}{X_{\max} - X_{\min}};$$

- (b) robust scaling when the median is 30, $Q_1 = 20$, and $Q_3 = 40$,

$$x' = \frac{x - \text{median}(X)}{Q_3 - Q_1};$$

- (c) Max–Abs scaling.

Which of these methods is generally least influenced by the outlier 100, and why?