CS 4410 Final Exam Summer 2026B

Exam time: 55 Minutes ***Total Marks: 35***

Name: ***ID:*** ***Date:***

your name, and your Troy ID on every sheet.

Q1 [15 points] Answer any three of the following. Each answer is worth 5 points. Briefly define and differentiate the concepts, with an example where appropriate.

(a) Feature selection and feature extraction.
(b) PCA and LDA.
(c) Hard-margin and soft-margin Support Vector Machines.
(d) Bagging and boosting.
(e) Prepruning and postpruning in decision trees.

Q2 [5 points] A decision-tree node contains 10 observations:

$$6 \text{ positive}, \qquad 4 \text{ negative}.$$

Attribute A produces

$$A = 0 : \ 4+,\ 0-, \qquad\qquad A = 1 : \ 2+,\ 4- .$$

Using

$$H(S) = -\sum_i p_i \log_2 p_i,$$

calculate the original entropy, the weighted entropy after splitting on A, and the information gain $IG(A)$.

Q3 [5 points] Consider the Bayesian network

$$R \longrightarrow W \longleftarrow S,$$

where R is rain, S is sprinkler, and W is wet grass. Given

$$P(S = 1 \mid R = 1) = 0.10,$$

$$P(W = 1 \mid R = 1, S = 1) = 0.99,$$

$P(W = 1 \mid R = 1, S = 0) = 0.80,$ compute

$$P(S = 1 \mid W = 1, R = 1).$$

Show the normalization step.

Q4 [10 points] Apply one complete iteration of k-means clustering to the following points:

$$A = (1, 1), \qquad B = (1, 2), \qquad C = (2, 1),$$

$$D = (5, 5), \qquad E = (6, 5), \qquad F = (5, 6).$$

Use $k = 2$ with initial centroids

$$m_1^{(0)} = A = (1, 1), \qquad m_2^{(0)} = D = (5, 5).$$

Using Euclidean distance:

(a) Assign each point to its nearest initial centroid.

(b) Compute the updated centroids $m_1^{(1)}$ and $m_2^{(1)}$.

(c) Reassign all points using the updated centroids and state whether the assignments have converged.

(d) Calculate the final within-cluster sum of squared errors:

$$SSE = \sum_{i=1}^{2} \sum_{x \in G_i} \|x - m_i\|^2.$$

CS 4410 Final Exam Summer 2026C

Exam time: 55 Minutes ***Total Marks: 35***

Name: ***ID:*** ***Date:***

your name, and your Troy ID on every sheet.

Q1 [15 points] Answer any three of the following. Each answer is worth 5 points. Briefly define and differentiate the concepts, with an example where appropriate.

(a) Forward selection, backward elimination, and floating feature search.
(b) k-means clustering and Gaussian mixture models.
(c) Online, stochastic, and batch neural-network learning.
(d) Margin, support vectors, and slack variables in an SVM.
(e) Majority voting, weighted voting, and stacking.
(f) Exclusive, overlapping, and fuzzy clustering.

Q3 [5 points] Four observations have coordinates

$$A = (1, 2), \qquad B = (2, 3), \qquad C = (6, 7), \qquad D = (7, 8).$$

Using Euclidean distance and single-link hierarchical clustering, show the complete sequence of merges and the distance at which each merge occurs. Draw the dendrogram.

Q4 [10 points] A perceptron has

$$w_0 = 0.1, \qquad w_1 = -0.3, \qquad w_2 = 0.2.$$

For

$$x_0 = 1, \qquad x_1 = 1, \qquad x_2 = 2, \qquad r = 0,$$

use

$$y = \begin{cases} 1, & w^T x > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad \Delta w_j = \eta (r - y) x_j,$$

with $\eta = 0.1$. Calculate the initial output, update all weights, and calculate the new output.

Q5 [5 points] A Naive Bayes classifier has

$$P(Yes) = \frac{9}{14}, \qquad P(No) = \frac{5}{14}.$$

For $x = (Sunny, Cool, Strong)$, use

	Yes	No
$Sunny$	2/9	3/5
$Cool$	3/9	1/5
$Strong$	3/9	3/5

to compute the two unnormalized posterior scores and predict the class.