CS 273a Final Exam

Intro to Machine Learning: Fall 2023

Monday December 11th, 2023

Your name: **Row/Seat Number:**

Your ID #(e.g., 123456789) **UCINetID (e.g.ucinetid@uci.edu)**

- Please put your name and ID **on every page**.
- Total time is 1 hour 50 minutes. READ THE EXAM FIRST and organize your time; don't spend too long on any one problem.
- Please **write clearly** and **show all your work**.
- Please write your final answer **in the provided box** where possible.
- You may use **one** sheet containing handwritten notes for reference, and a **basic** calculator; no other electronics allowed.
- If you need clarification on a problem, please raise your hand and wait for the instructor or TA to come over.

Problems

Total, *(74 points.)*

This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

Name: ________ ID#: ________

Problem 1 Decision Trees, *(10 points.)*

Consider the table of measured data given at right. We will use a decision tree to predict the outcome y using the three features, $x_1, \ldots, x_3$. In the case of ties, we prefer to use the feature with the smaller index (x_1 over x_2, etc.) and prefer to predict class 1 over class 0. You may find the following values useful:

$\log_2(1) = 0$ $\log_2(2) = 1$ $\log_2(3) = 1.59$ $\log_2(4) = 2$
$\log_2(5) = 2.32$ $\log_2(6) = 2.59$ $\log_2(7) = 2.81$ $\log_2(8) = 3$
$\log_2(9) = 3.17$ $\log_2(10) = 3.32$
Note also that $\log(a/b) = \log(a) - \log(b)$.

x_1	x_2	x_3	y
1	1	0	1
0	1	0	0
1	0	1	1
0	1	0	0
1	1	0	1
1	1	0	1
1	1	1	0
1	0	1	0
0	1	1	1

(1) What is the entropy of y? *(2 points.)*

(2) Which variable would you split first? Justify your answer. *(2 points.)*

(3) What is the information gain of the variable you selected in part (2)? *(2 points.)*

(4) Draw the full decision tree learned on these data. *(4 points.)*

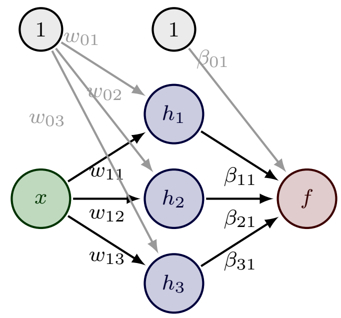
This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

Name: ID#:

Problem 2 Neural Networks, *(9 points.)*

Consider a small neural network designed to classify a scalar feature x as one of $y \in \{-1, +1\}$. We have three hidden nodes h_1, h_2, h_3 and a single output node f_1.



You are given the weights W of the hidden layer,

$$W = \begin{bmatrix} w_{01} & w_{11} \\ w_{02} & w_{12} \\ w_{03} & w_{13} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 3 & -2 \\ -3 & 3 \end{bmatrix}$$

and the weights B of the output layer,

$$B = \begin{bmatrix} \beta_{01} & \beta_{11} & \beta_{21} & \beta_{31} \end{bmatrix} = \begin{bmatrix} -1 & 3 & -1 & -1 \end{bmatrix}.$$

(For example, w_{12} is the weight connecting x_1 to h_2; w_{02} is the constant (bias) term for h_2, etc.)

The network uses a ReLU activation function, $a(z) = \max(0, z)$, for the hidden layer, and a logistic sigmoid activation function,

$$\sigma(z) = \frac{1}{1 + \exp(-z)} = \frac{1}{1 + \frac{1}{\exp(z)}} = \frac{\exp(z)}{\exp(z) + 1},$$

for the output layer (the value of which, as usual, corresponds to the model's probability that the class is +1). You may find the following values useful:
$\exp(0) = 1$ $\exp(1) = 2.72$ $\exp(2) = 7.39$ $\exp(3) = 20.09$ $\exp(4) = 54.60$ $\exp(5) = 148.40$

---

(1) What class is predicted by the model given the input $x_1 = 2$? *(3 points.)*

(2) What is the model's estimated *probability* $p(y = +1 \mid x_1 = 2)$? *(3 points.)*

(3) Suppose our input is $x_1 = 0$: what is the probability $p(y = +1 \mid x_1 = 0)$? *(3 points.)*

This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

Problem 3 True/False, *(10 points.)*

For each of the scenarios below, circle one of "true" or "false" to indicate whether you agree with the statement.

True or **false**: Changing our classifier from a nearest centroid classifier to a Gaussian Bayes classifier is likely to reduce our model's bias.

True or **false**: When training a 1-nearest neighbor classifier, if we decide to double the amount of data available to the learner, we would expect the bias to decrease.

True or **false**: When learning a decision tree model, if we restrict our learner to only use odd-numbered features, we will reduce our variance.

True or **false**: In a random forest model, we use the best-fitting tree in our ensemble for prediction.

True or **false**: In a random forest model, we usually learn using a random subset of features at each node in order to speed up the learning process (reduce computation).

True or **false**: If, when training a random forest model, we decide to use the same subset of features across all levels of a given tree (i.e., the entire tree uses only those features), we will increase our model's bias.

True or **false**: When clustering using k-means, we can select the number of clusters k using a hold-out (validation) data set.

True or **false**: Computing the expected discounted sum of rewards in a Markov Reward Process can be solved using either matrix algebra or dynamic programming.

True or **false**: The larger the value of gamma, the less far-sighted (less interested in future rewards) the RL agent becomes.

True or **false**: Once the optimal state-action value function is computed, the reinforcement learning problem can be considered solved.

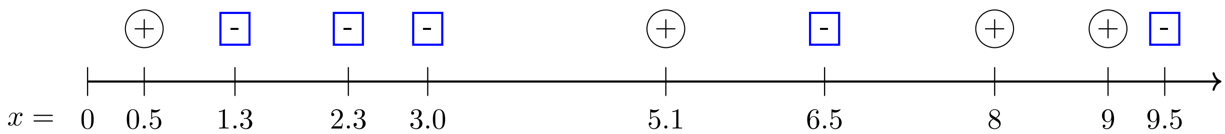
This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

Name: ID#:

Problem 4 K-Nearest Neighbors, *(12 points.)*

Consider the following dataset with *nine* points shown below, for a binary classification task ($y = +, -$) with a scalar feature x. In case of ties, **prefer the negative class**. Put final answers in the box.



(1) Compute the **training** error of a 1-nearest neighbor classifier. *(3 points.)*

(2) Compute the **leave-one-out** cross-validation error of 1-nearest neighbor classifier. *(3 points.)*

(3) Compute the **training** error of 2-nearest neighbor classifier. *(3 points.)*

(4) Compute the **leave-one-out** cross-validation error of 2-nearest neighbor classifier. *(3 points.)*

This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

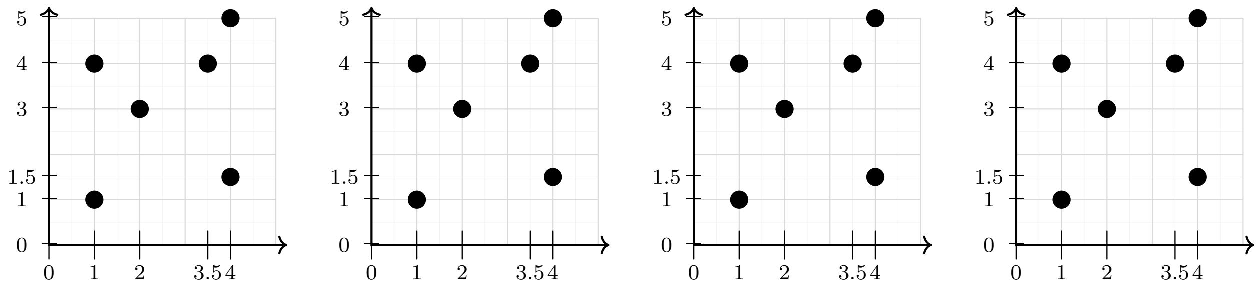
Name: ______ ID#: ______

Problem 5 Hierarchical Clustering, *(12 points.)*

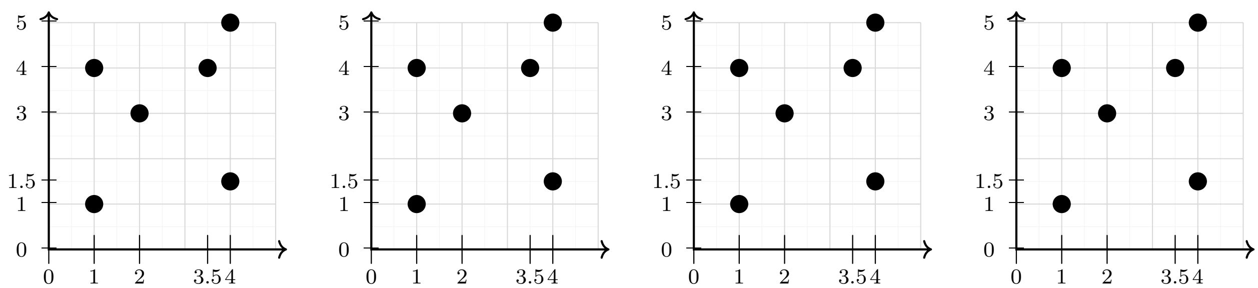
Consider the two-dimensional data points plotted in each panel. In this problem, we will cluster these data.

Linkage

(a) Execute the hierarchical agglomerative clustering (linkage) algorithm on these data points, using "single linkage" (minimum distance) for the cluster scores. Stop when the algorithm would terminate, or after 4 steps, whichever is first. Show each step separately in a panel. *(6 points.)*



(b) Now repeat your agglomerative clustering algorithm, this time using "complete linkage" (maximum distance) for the cluster scores. Stop when the algorithm would terminate, or after 4 steps, whichever is first. Show each step separately in a panel. *(6 points.)*



This page is intentionally blank.

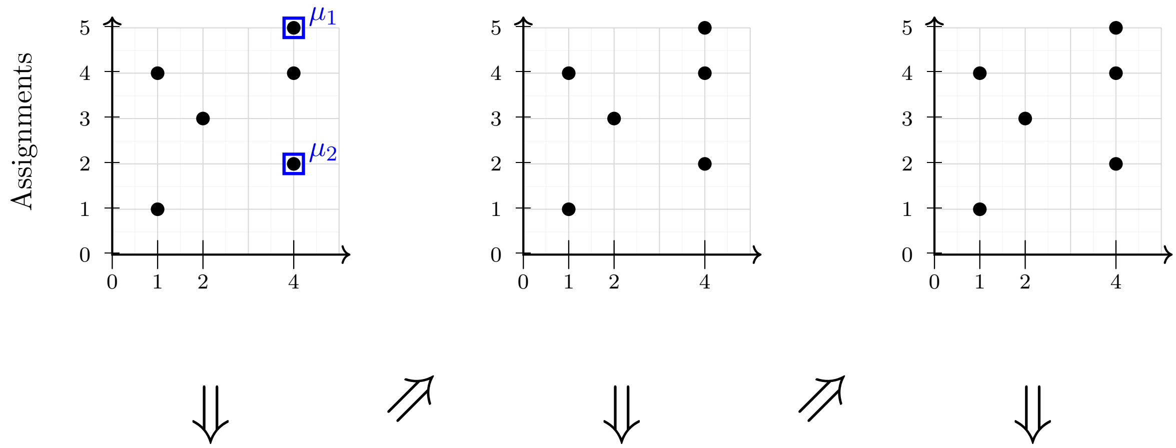
While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

Name: ID#:

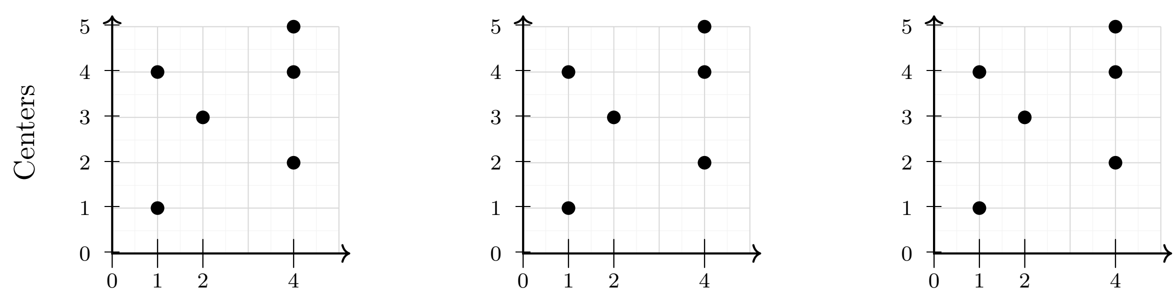
Problem 6 K-Means Clustering, *(9 points.)*

Consider the 2-D data points plotted in each panel. In this problem, we will cluster these data using the k-means algorithm, where each panel is used to show a single step of the algorithm.

Starting from the two cluster centers indicated by squares, perform k-means clustering on the data. In the top panels, indicate the assignment of the data, and then in the panel below give the values of the updated cluster centers and sketch their location in the plot, so each **column** of panels shows an iteration of k-means. Stop when converged, or after 6 steps (3 iterations), whichever is first. In the case of any ties, we will prefer to assign to cluster 1. It may be helpful to recall from our nearest neighbor classifier that the set of points nearer to A than B is separated by a line.



$\mu_1 = (\quad , \quad)$	$\mu_1 = (\quad , \quad)$	$\mu_1 = (\quad , \quad)$
$\mu_2 = (\quad , \quad)$	$\mu_2 = (\quad , \quad)$	$\mu_2 = (\quad , \quad)$



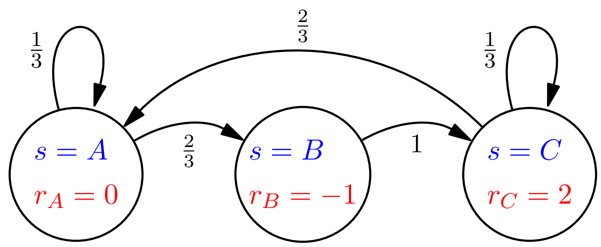
This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

Name: ____ ID#: ____

Problem 7 Markov Processes, *(12 points.)*

Consider the Markov reward process model shown here:



$\Pr[A \to A] = 0.33$
$\Pr[A \to B] = 0.67$
$\Pr[B \to C] = 1.0$
$\Pr[C \to A] = 0.67$
$\Pr[C \to C] = 0.33$

where the transition probabilities are shown next to each arc and at right, and the rewards r_s associated with each state s are shown inside the circles. We will use dynamic programming to (start) computing the expected discounted sum of rewards. Assume a future discounting factor of $\gamma = \frac{1}{2}$.

(1) Compute $J^1(s)$, the expected discounted sum of rewards for state sequences of length 1 (e.g., $[A]$) starting in each state s. *(4 points.)*

$J^1(A) =$	$J^1(B) =$	$J^1(C) =$

(2) Compute $J^2(s)$, the expected discounted sum of rewards for state sequences of length 2 (e.g., $[C \to A]$) starting in each state s. *(4 points.)*

$J^2(A) =$	$J^2(B) =$	$J^2(C) =$

(3) Compute $J^3(s)$, the expected discounted sum of rewards for state sequences of length 3 (e.g., $[C \to A \to A]$) starting in each state s. *(4 points.)*

$J^3(A) =$	$J^3(B) =$	$J^3(C) =$

This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

This page is intentionally blank.
It may be used for scratch work or to provide additional space for solutions if necessary.

This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.

This page is intentionally blank.
It may be used for scratch work or to provide additional space for solutions if necessary.

This page is intentionally blank.

While you may use it as scratch paper, it may not be scanned for grading, so please do not include final solutions.