

CS273A Final Exam  
Introduction to Machine Learning: Winter 2020  
Tuesday March 17th, 2020

Your name:

Your ID #(e.g., 123456789)

UCINetID (e.g. ucinetid@uci.edu)

- Due to the ongoing health emergency, this exam is **take home** and may be submitted in person or on Canvas (scanned).
- Total time is 2 hours 15 minutes for either in-person delivery (to the classroom) or submission to Canvas.
- READ THE EXAM FIRST and organize your time; don't spend too long on any one problem.
- Please put your name and ID **on every page**.
- Please **write clearly** and **show all your work**.
- If you need clarification on a problem, please post **privately** on Piazza and we will try to answer it.

## Problems

1	Bayes Classifiers, <i>(12 points.)</i>	3
2	Cross-Validation, <i>(9 points.)</i>	5
3	Decision Trees, <i>(10 points.)</i>	7
4	Dimensionality Reduction, <i>(10 points.)</i>	9
5	Hierarchical Clustering, <i>(11 points.)</i>	11
6	K-Means Clustering, <i>(10 points.)</i>	13
7	Markov Processes, <i>(12 points.)</i>	15

Total, *(74 points.)*

---

*This page is intentionally blank, use as you wish.*

Name:

ID#:

**Bayes Classifiers, (12 points.)**

In this problem you will use Bayes Rule:  $p(y|x) = p(x|y)p(y)/p(x)$  to perform classification. Suppose we observe some training data with two binary features  $x_1, x_2$  and a binary class  $y$ . After learning the model, you are also given some validation data.

Table 1: Training Data

$x_1$	$x_2$	$y$
0	0	0
0	1	0
0	1	1
0	1	1
1	0	1
1	0	1
1	1	0
1	1	0

Table 2: Validation Data

$x_1$	$x_2$	$y$
0	0	1
0	1	0
1	0	0
1	1	0

In the case of any ties, we will prefer to predict class 0.

- (1) Give the predictions of a joint Bayes classifier on the validation data. What is the validation error rate? (Put final answers in boxes.) (4 points.)

$x_1$	$x_2$	$y$
0	0	
0	1	
1	0	
1	1	

Error Rate:

- (2) Give the required probabilities to define a **naïve** Bayes classifier. (4 points.)

- (3) Give the predictions of a naïve Bayes classifier on the validation data. What is the validation error rate? (Put final answers in boxes.) (4 points.)

$x_1$	$x_2$	$y$
0	0	
0	1	
1	0	
1	1	

Error Rate:

---

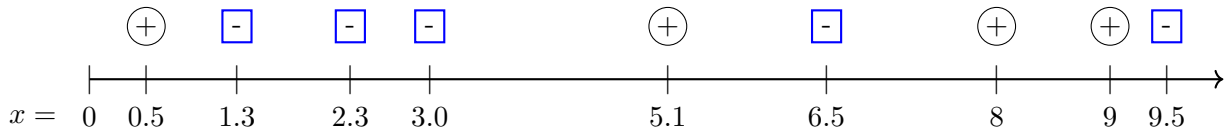
*This page is intentionally blank, use as you wish.*

Name:

ID#:

**Cross-Validation, (9 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 **leave-one-out** cross-validation error rate of a 1-nearest neighbor classifier. (3 points.)

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

- (3) Compute the **leave-one-out** cross-validation error rate of an 8-nearest neighbor classifier. (3 points.)

---

*This page is intentionally blank, use as you wish.*

Name:

ID#:

**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,  $x_1, \dots, x_3$ . 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 (although you may also leave logs unexpanded):

$$\begin{aligned} \log_2(1) &= 0 & \log_2(2) &= 1 & \log_2(3) &= 1.59 & \log_2(4) &= 2 \\ \log_2(5) &= 2.32 & \log_2(6) &= 2.59 & \log_2(7) &= 2.81 & \log_2(8) &= 3 \end{aligned}$$

$x_1$	$x_2$	$x_3$	$y$
0	0	0	0
1	1	0	1
0	1	0	0
0	1	0	0
0	1	0	0
1	0	0	0
1	1	0	1
1	0	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)? (3 points.)

(4) Draw the rest of the decision tree learned on these data. (3 points.)

---

*This page is intentionally blank, use as you wish.*

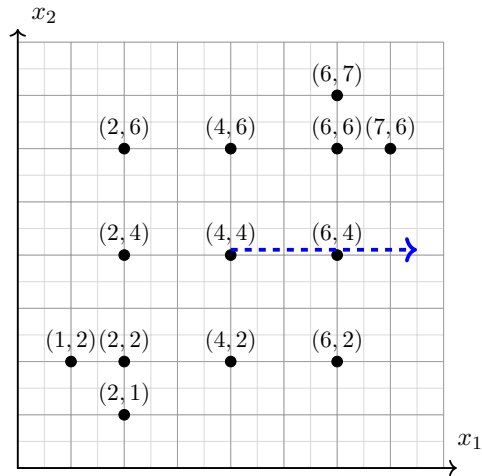


Name:

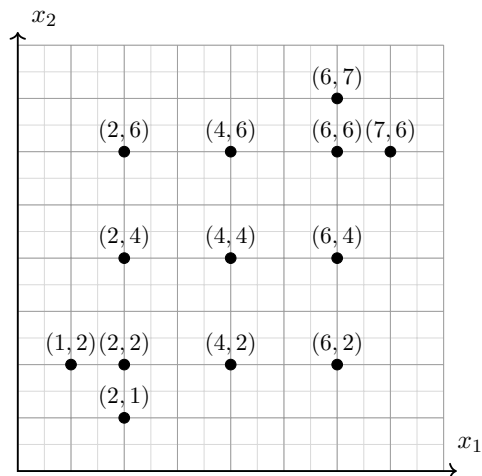
ID#:

**Dimensionality Reduction, (10 points.)**

- (1) For the following points in two dimensions, consider performing linear dimensionality reduction along the given vector (dashed line). What is the reconstruction error, in MSE, when using this vector? (4 points.)



- (2) On the figure below, draw the directions of the first two principal components. (2 points.)
- (3) What is the reconstruction error (MSE) of these points when only the first principal component is used to reconstruct each point? (4 points.)



---

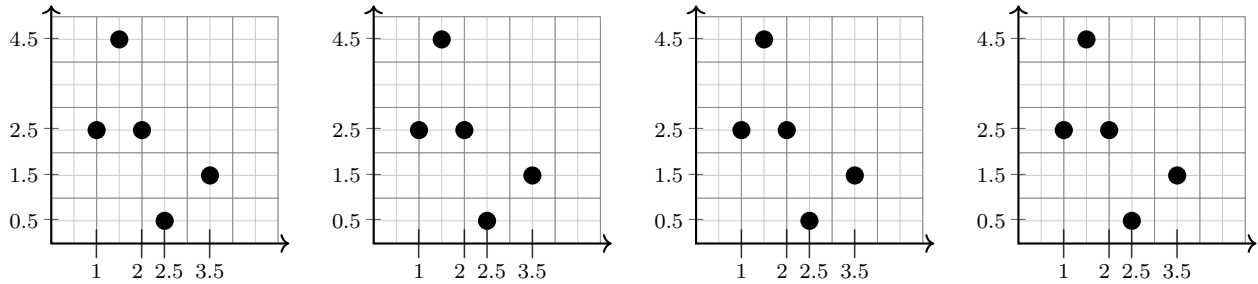
*This page is intentionally blank, use as you wish.*

## Hierarchical Clustering, (11 points.)

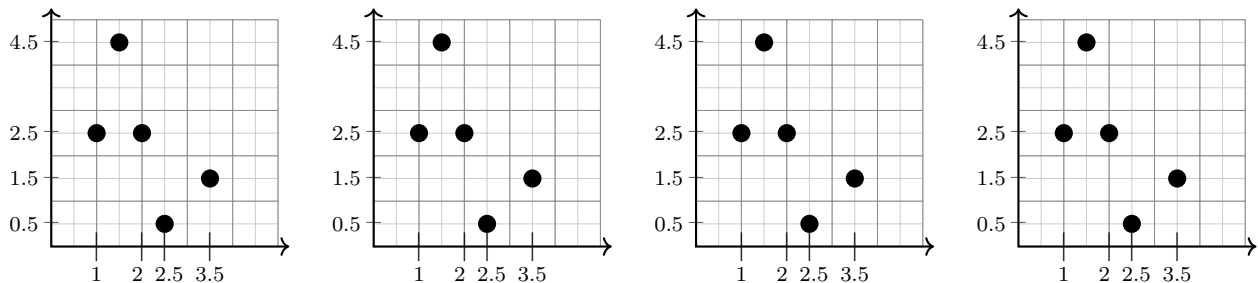
Consider the two-dimensional data points plotted in each panel. In this problem, we will cluster these data.

### 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. (4 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. (4 points.)



(c) What is the (big-O) computational complexity of the hierarchical agglomerative clustering algorithm? Justify your answer **briefly** in 1-2 sentences. (3 points.)

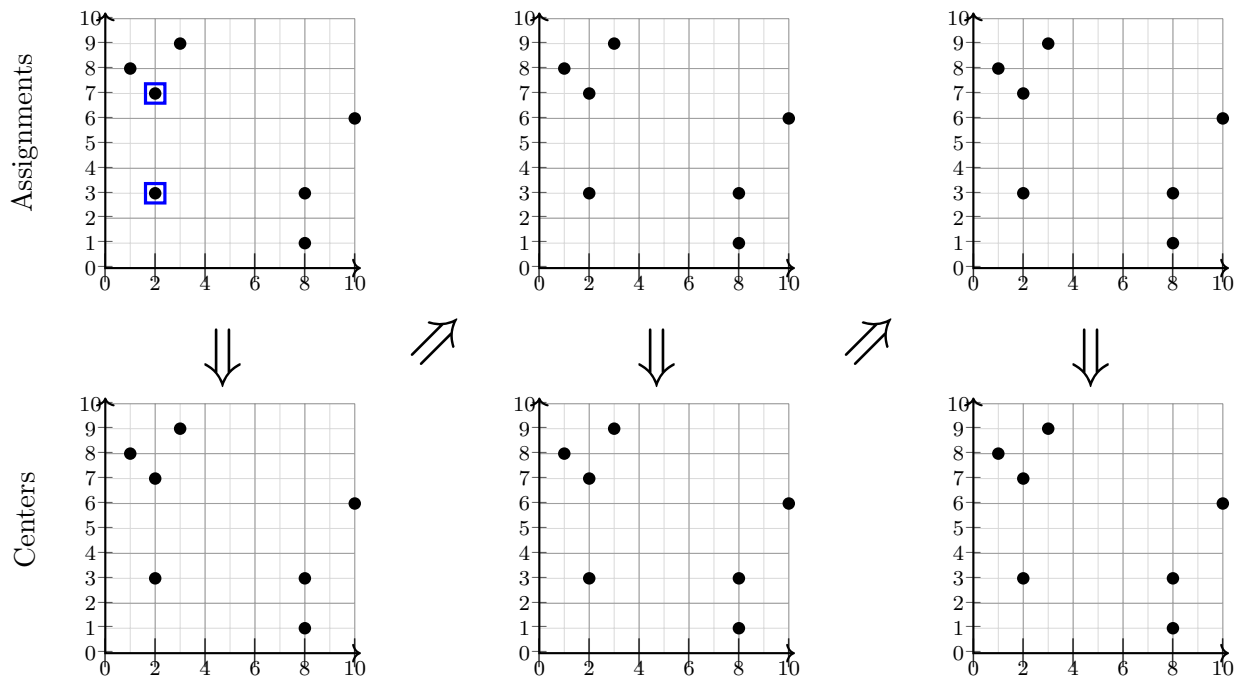
---

*This page is intentionally blank, use as you wish.*

**K-Means Clustering, (10 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.

(a) 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 show the new cluster centers, so each **pair** of panels shows an iteration of  $k$ -means. Stop when converged, or after 6 steps (3 iterations), whichever is first. 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. (6 points.)



(b) Write down the cost function optimized by the  $k$ -means algorithm, in terms of the data locations  $x^{(i)}$ , cluster centers  $\mu_c$ , and cluster assignments  $z^{(i)}$ . (2 points.)

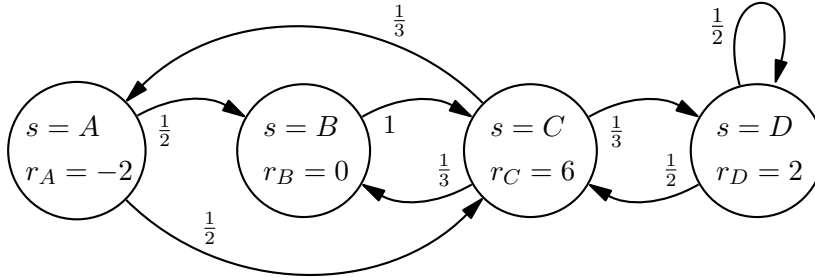
(c) What is the (big-O) computational complexity of each iteration of  $k$ -means (naïve computation), in terms of the data size  $m$  and number of clusters  $k$ ? (2 points.)

---

*This page is intentionally blank, use as you wish.*

Name: ID#: **Markov Processes, (12 points.)**

Consider the Markov reward process model shown here:



$\Pr[A \rightarrow B] = 0.5$

$\Pr[A \rightarrow C] = 0.5$

$\Pr[B \rightarrow C] = 1.0$

$\Pr[C \rightarrow A] = 0.33$

$\Pr[C \rightarrow B] = 0.33$

$\Pr[C \rightarrow D] = 0.33$

$\Pr[D \rightarrow C] = 0.5$

$\Pr[D \rightarrow D] = 0.5$

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.)

$J^1(A) =$

$J^1(B) =$

$J^1(C) =$

$J^1(D) =$

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

$J^2(A) =$

$J^2(B) =$

$J^2(C) =$

$J^2(D) =$

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

$J^3(A) =$

$J^3(B) =$

$J^3(C) =$

$J^3(D) =$

---

*This page is intentionally blank, use as you wish.*



Name:

ID#:

---

*This page is intentionally blank, use as you wish.*

---

*This page is intentionally blank, use as you wish.*