Test 2 Group **A** Data Science

UF ID:

Name:

– State any simple, reasonable assumption used to arrive at your answer.
– An 'yes' or 'no' answer without reasoning is worth 0 points.
– Zero points if the writing is hard to decipher. Use a black pen if in doubt.

1 Regression [3+3+1]

a Compute the best linear polynomial fit to the data

$$(x_i, y_i)_{i=0,1,2} = (-1,-1), (0,0), (1,-1). \tag{1}$$

Show the linear equation to be solved.

b Show the least squares reformulation and solve for linear approximation.

c Draw the points and the solution.

2 K-means [2+2+2+2]

a In the Initialization shown below, • are the initial centers and ○ the 1 dimensional data points:

○ ○ • • ○ ○. (2)

Show and explain the steps of one k-means iteration by minimally adding to the three diagrams below (you can add stages or need not use all, but display only one step of the iteration per drawing!)

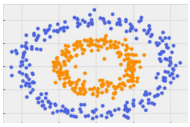
○ ○ ○ ○ (3)

○ ○ ○ ○ (4)

b In one or two sentences, what is a Voronoi Diagram? Show the Voronoi Diagram in the Initialization (2).

c There are three clusters of points that each fit inside a circle and no circles overlap. If you choose k=3 and run the generic k-means algorithm sufficiently long, will k-means find an optimal clustering? Explain.

d Will k-means with Euclidean distance be able to separate the two concentric, orange and blue groups below? Explain why or why not.



3 Dimension reduction [2 + 2 + 2 + 2]

Use your knowledge of the relationship between the singular value decomposition of a matrix A and the eigendecomposition of $M := AA^T$ to answer the following.

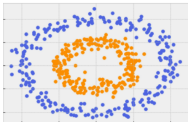
a Applying the principal component analysis (PCA) using SVD to a matrix of data where all entries of one column are zero, one singular value is zero. Explain.

b Compute the singular values of

$$A := \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \tag{5}$$

c How can one decide in practice how many principle components to retain for a learning application?

d Will PCA dimension reduction from $\mathbb{R}^2$ to $\mathbb{R}^1$ keep the orange and blue data clusters below separated? Explain why or why not.



4 Learning [2 + 1 + 1 + 1 + 1]

A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Note: To obtain points you must give a short explanation (one sentence or less).

a To predict weather based on historical weather data what is a good choice for E and T ?

b To predict whether or not it will be raining at 5pm tomorrow would you choose a learning algorithm that performs classification or one that performs regression?

c Suppose you are working on stock market prediction and would like to predict the price of a particular stock tomorrow (measured in dollars). Would you treat this as a classification or a regression problem?

d Would you use supervised or unsupervised learning to predict the gender of a new manuscript's author, based on 50 articles written by male authors, and 50 articles written by female authors?

e Would you use supervised or unsupervised learning to identify schools of thought in a collection of 1000 essays written on the US Economy?

5 Large data [1]

a What problem does MinHash solve?

6 Random Walk [1 + 1 + 2]

a Draw the directed graph of a random walk with equal probablilities that connects node

A to B, C
B to C
C to A

b Write the matrix of the corresponding Markov process.

c What is the steady state of Markov process?

7 Support Vector Machine (SVM) [2 + 1]

a Specify a simple function for the kernel trick so that a SVM can separate the set of • from the set of ○:

• ○ • ○ • ○ •

b Draw the function on the data set.

8 Perceptron [2 + 1 + 2]

A multi-layer perceptron neural net consists of 2 input nodes x_1 and x_2, a hidden layer of 2 nodes y_1 and y_2 and an output layer of 1 node z. The initial weights are $w(x_1, y_1) = 1$, $w(x_1, y_2) = 0$, $w(x_2, y_1) = .5$, $w(x_2, y_2) = .5$, $w(y_1, z) = 1$, $w(y_2, z) = 1$. The target (actual) value is 1. The input values are $x_1 = 1$ and $x_2 = 0$.

a Draw the perceptron (nodes and directed edges). For $i, j \in \{1, 2\}$ label the nodes x_i, y_i, z and add the weights.

b Perform one forward pass of the neural net to compute the predicted value of the output node.

c Perform one step of backpropagation with learning rate 0.05 to reduce the error.