

AIDI 1000: AI Algorithms and Mathematics – Final Exam Fall 2022

Due Date : December 16, 2022, 6:00 PM - 8:30 PM

- Part I: True/False and Reasoning Questions (2 points each)

1. TRUE or FALSE? K-means clustering algorithms can find clusters of arbitrary shape
2. TRUE or FALSE? A binary classifier having accuracy 0.8 is considered to be more useful than a binary classifier having accuracy 0.1.
3. TRUE or FALSE? Both LDA and PCA are linear transformation techniques. The main difference is LDA is supervised whereas PCA is unsupervised.
4. TRUE or FALSE? One can copy a training set 10 times to form a larger training set in order to learn a better classifier.
5. TRUE or FALSE? PCA maximize the variance of the data, whereas LDA maximize the separation between different classes.

- Part II: Short Answer Questions (10 point each) (**Attempt any 4 Questions**)

1. (Dimensionality Reduction)

Consider PCA over 3D dataset (in R^3), which produces the first two principal components as $u_1 = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0)$ and $u_2 = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0)$. If one of the original data points was $x = (1, 2, 3)$, what will be its representation in the projected space?

2. (Hierarchical Clustering)

Based on Euclidean distance in \mathbb{R} , we obtained following three clusters: $A = (0, 5, 6), B = (3, 9), C = (12)$. In the next iteration of the clustering, which two clusters will be merged by complete linkage and single linkage approaches, respectively?

3. (Regression)

Consider the covariance matrix $\Sigma = Cov \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 9 & -2 \\ -2 & 1 \end{bmatrix}$

3.1 What are the standard deviations for X_1 and X_2 ?

3.2 Assume that means of X_1 and X_2 are zero. Find the regression equation to predict X_1 as a function of X_2 .

4. (Logistic Regression)

Consider a Logistic Regression model with ReLU activation function, which has input $x \in R$ and a bias term, and the output value of $y = \max(0, w_0 + w_1 x)$. What would be the input value x that produces output $y > 0$?

5. (K-Means Clustering)

Perform one iteration of k -means on the 1D dataset $X = \{2, 4, 7, 9\}$ with $k = 2$ and initial centers at $c_1 = 0$ and $c_2 = 10$ using Manhattan distance.

5.1 Show initial cluster assignments

5.2 Show the new resulting cluster centers.

• Part III: Long Answer Questions

1. (Decision Tree and KNN) (25 points) Consider a training set provided below which contains two boolean features and a continuous feature.

	A	B	C	Class
row 1	F	T	115	-
row 2	T	F	890	-
row 3	T	T	257	+
row 4	F	F	509	+
row 5	T	T	733	+

1.1 How much information about the class would be gained by knowing whether or not the value of feature *C* is less than 400?

1.2 What is the information gain for feature *A* and *B*?

1.3 Is the above data set preprocessed well for kNN algorithms?

2. (KNN Classification) (25 points)

Given a dataset with binary labels $(x,y) = \{(2,+), (3,+), (5,-), (7,+), (11,-)\}$, compute

- find the kNN training set error with $k = 5$, using 0-1 loss. If there is a tie, always favor the positive class.
- Similar to above but with $k = 4$
- Similar to above but with $k = 3$
- Similar to above but with $k = 2$
- Similar to above but with $k = 1$
- Should we choose the k with the smallest training set error? Why?

The 0-1 Loss formula is as follows:

$$\mathcal{L}_{0/1}(h) = \frac{1}{n} \sum_{i=1}^n \delta_{h(\mathbf{x}_i) \neq y_i}, \text{ where } \delta_{h(\mathbf{x}_i) \neq y_i} = \begin{cases} 1, & \text{if } h(\mathbf{x}_i) \neq y_i \\ 0, & \text{o.w.} \end{cases}$$