CSCI 57800 ML Fall 2023 Homework 1

August 31, 2023

Instructions

We will be using Canvas to collect your assignments. Please read the following instructions to prepare your submission.

1. Submit your solution in a pdf file and a zip file (<yourLastName_FirstName>.pdf/zip). Your write-up must be in pdf. Your code must be in the zip file.
2. In your pdf file, the solution to each problem should start on a new page.
3. Latex is strongly encouraged to write your solutions, e.g., using Overleaf (https://www.overleaf.com/). Neither scanned handwritten copies nor hard copies are acceptable.
4. You need to add screen captures of your code and the output in your write-up.
5. You may discuss the problems and potential directions for solving them with another student. However, you need to write your own solutions and code separately, and not as a group activity. Please list the students you collaborated with on your submission.

Problem 1 (12 points)

Generally, if $\mathbf{A} \in \mathbb{R}^{M\times N}$ and $\mathbf{B} \in \mathbb{R}^{N\times P}$, then $\mathbf{AB} \in \mathbb{R}^{M\times P}$ and $(AB)_{ij} = \sum_k A_{ik}B_{kj}$.

Given $\mathbf{A} = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 3 & -2 & 2 \\ 1 & 1 & -1 \\ 4 & -1 & 3 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$,

Answer the following questions, (b) to (f). Show your work.

(a) (2 pts) What is $\mathbf{AB}$? Does $\mathbf{AB} = \mathbf{BA}$? What is $\mathbf{Bu}$?

(b) (2 pts) What is rank of $\mathbf{A}$?

(c) (2 pts) What is $\mathbf{A}^T$?

(d) (2 pts) Calculate $\mathbf{uv}^T$?

(e) (2 pts) What are the eigenvalues of $\mathbf{A}$?

(f) (2 pts) What is the inner product of $\mathbf{u}$ and $\mathbf{v}$?

Problem 2 (6 points)

Given a line $x + y = 3$ in the 2-D plane,

(a) (2 pts) If a given point (α, β) satisfies $\alpha + \beta > 3$, where does it lie relative to the line?

(b) (2 pts) What is the relationship of vector $\mathbf{v} = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$ to this line?

(c) (2 pts) What is the distance from origin to this line?

Problem 3 (14 points)

(a) (2 pts) What is the expectation of X where X is a single roll of a fair 6-sided dice ($S = 1, 2, 3, 4, 5, 6$)? What is the variance of X?

(b) (2 pts) We got a new dice where the sides are ($S = 5, 6, 7, 8, 9, 10$). What is the expectation of X where X is a single roll of this new fair 6-sided dice? What is the variance?

(c) There are two random variables, X and Y, and each can have two values, x_1, x_2 and y_1, y_2, respectively.

(1) (2 pts) Let's assume that x_1 and y_2 are mutually exclusive. $P(X = x_1) = 0.5, P(Y = y_2) = 0.5$. What is $P(X = x_1, Y = y_2)$?

(2) (2 pts) Let's assume that x_1 and y_2 are not mutually exclusive, but X and Y are independent. $P(X = x_1) = 0.5, P(Y = y_2) = 0.5$. What is $P(X = x_1, Y = y_2)$?

(d) A student is recalling her past study activities, and notices that she seems to get better quiz scores when she reviews previous class materials. She creates a table to track her study habits and quiz scores, with each row representing a day.:

Review	Good_quiz_score	Probability
yes	yes	0.4
yes	no	0.2
no	yes	0.1
no	no	0.3

(1) (2 pts) What is $P(Good\_quiz\_score = yes|Review = yes)$?

(2) (2 pts) Why doesn't $P(Good\_quiz\_score = yes, Review = yes) = P(Good\_quiz\_score = yes) \cdot P(Review = yes)$?

(3) (2 pts) The student merges her sleeping pattern data (whether she got a good sleep last night) with the current data, and finds that the $P(Good_sleep = yes|Review = yes, Good_quiz_score = yes)$ is 0.7. What is the probability of all three happening?

Problem 4 (28 points)

This is a dataset about what a reader of a newsgroup is going to do given a new message, depending on features of the message and the user location. The Action column shows the label, and other columns show the features.

We want to build a decision stump based on this dataset using all 6 samples for training. All the features are categorical. We will rely on the natural split of the categories for rules. To decide the predicted labels, we find the majority class of each child.

Answer the following questions. Show your work.

	Action	Author	Thread	Length	Where
e1	skips	known	new	long	home
e2	reads	unknown	new	short	work
e3	skips	unknown	old	long	work
e4	skips	known	old	long	home
e5	reads	known	new	short	home
e6	skips	known	old	long	work

(a) (2 pts) What would be the accuracy when you make a decision stump with the Thread feature?

(b) (2 pts) What would be the accuracy when you make a decision stump with the Author feature?

(c) (2 pts) What would be the accuracy when you make a decision stump with the Length feature?

(d) (4 pts) What would be the accuracy when you make a decision stump with the Where feature? Do you see any problem with the accuracy here? Discuss it.

(e) (2 pts) Which feature is the decision stump that gives the best accuracy based on?

Now we will use information gain as a score metric.

(f) (2 pts) Calculate the entropy of tossing a fair coin.

(g) (2 pts) What would be the information gain when you make a decision stump with the Thread feature?

(h) (2 pts) What would be the information gain when you make a decision stump with the Author feature?

(i) (2 pts) What would be the information gain when you make a decision stump with the Length feature?

(j) (2 pts) What would be the information gain when you make a decision stump with the Where feature?

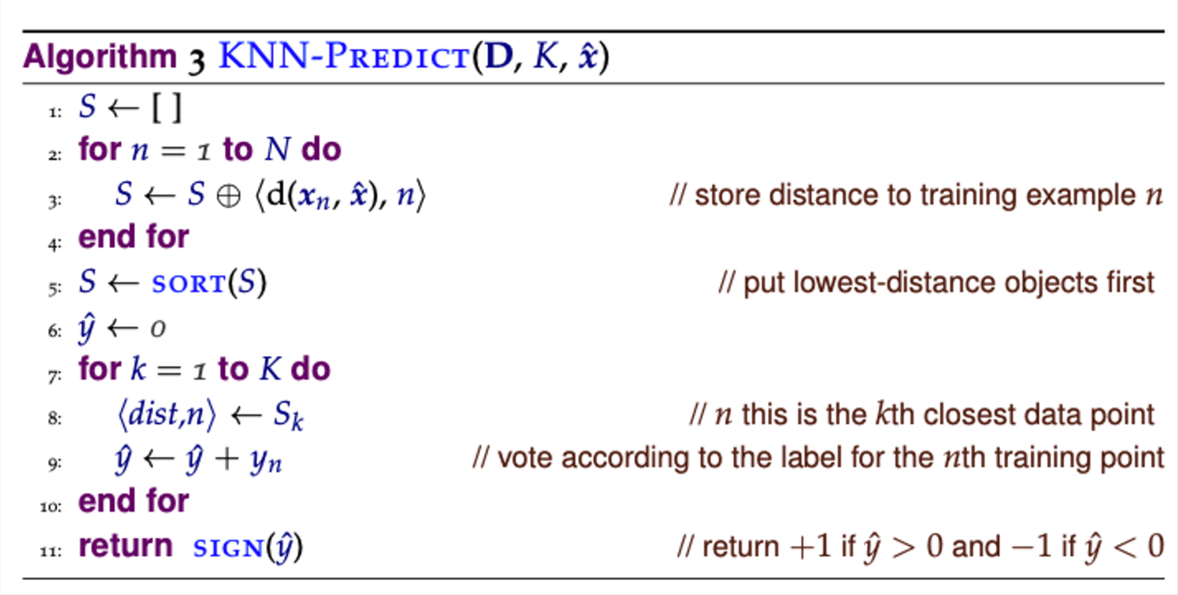
(k) (2 pts) Which feature is the decision stump that gives the best information gain based on?

(l) (2 pts) Compare (e) and (k). Are they the same? different? If different, which one do you think is better and why?

(m) (2 pts) Draw the decision stump resulting from (k). You can draw it in any tool (e.g., PowerPoint) and add the image in your write-up.

Problem 5 (40 points, Programming involved)

We are going to implement the kNN algorithm from scratch (not using a package such as `scikit-learn` or `pandas`) here. You are allowed to use the package `numpy`.We will use python 3. We will follow the pseudo code in the figure below. The function name will be `knn_predict` like in the pseudo code.



```
Algorithm 3 KNN-Predict(D, $K$, $\hat{x}$)
$S \leftarrow [\,]$
for $n = 1$ to $N$ do
    $S \leftarrow S \oplus \langle \mathrm{d}(x_n, \hat{x}), n \rangle$        // store distance to training example $n$
end for
$S \leftarrow$ SORT($S$)        // put lowest-distance objects first
$\hat{y} \leftarrow 0$
for $k = 1$ to $K$ do
    $\langle dist, n \rangle \leftarrow S_k$        // $n$ this is the $k$th closest data point
    $\hat{y} \leftarrow \hat{y} + y_n$        // vote according to the label for the $n$th training point
end for
return SIGN($\hat{y}$)        // return $+1$ if $\hat{y} > 0$ and $-1$ if $\hat{y} < 0$
```

(a) (5 pts) We will use the dataset from Problem 4 as our training data. Convert each feature value as binary (0 or 1) (e.g., skips = 1, reads = 0), and make `data.txt`. Then, write code for a function called `read_data` that reads the input data.

(b) (5 pts) Write code for a function called `distance` that computes the Hamming distance. The input to this function will be two data samples (two vectors), and the output will be the distance between them.

(c) (10 pts) Write code for a function `knn_predict` following the flow of the pseudo code above.

(d) (10 pts) Test your code with the test sample below. You need to write the code for reading this test sample and calling the function `knn_predict` as well.

	Action	Author	Thread	Length	Where
e7	??	unknown	old	long	home

(e) (5 pts) When $k = 1$, what is the prediction from your function `knn_predict`?

(f) (5 pts) When $k = 3$, what is the prediction from your function `knn_predict`?

3 bonus points will be given if your homework is easy to review.