NAME \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Pujitha kallu\_\_\_\_\_\_\_\_\_\_ CSEN 240 MIDTERM EXAM

* Open books/notes/slides. 25 points. Show your work as appropriate.
* Sign the honor code:

***I completed this exam entirely on my own with no collaboration or consultation with anyone.***

# \_\_\_\_\_\_\_\_\_\_\_\_\_\_Pujitha kallu\_\_\_\_\_\_\_\_\_\_\_\_

Signature

PROBLEM 1 (5 points)

Consider the following dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| **Length** | **Width** | **Weight** | **Class** |
| 6.4 | 3.3 | 0.6 | red |
| 7.1 | 3.1 | 1.2 | green |
| 6.9 | 3.2 | 0.9 | blue |
| 7.2 | 4.1 | 0.8 | blue |
| 8.2 | 2.5 | 1.7 | green |
| 7.5 | 3.2 | 1.4 | green |
| 4.2 | 3.9 | 0.8 | red |

(a). What is the dimensionality of the dataset?

(b). What is 𝐱𝐱(5) ?

(c). What is 𝐱𝐱1(3) ?

(d). What is 𝐱𝐱3(1) ?

(e). What is y(4)?

# PROBLEM 2 (5 points)

This dataset is used to train the Bayesian Decision Theoretic classifier with equal prior probabilities.

|  |  |  |  |
| --- | --- | --- | --- |
| Weight | Height | Value | Class Label |
| 5.2 | 56 | 50 | 1 |
| 1.6 | 57 | 51 | 2 |
| 2.8 | 56 | 48 | 1 |
| 8.3 | 55 | 56 | 2 |
| 2.3 | 57 | 62 | 2 |

Recall the discriminant function: 𝑔𝑔𝑖𝑖(𝑥𝑥) = − (𝑥𝑥 − 𝜇𝜇𝑖𝑖)𝑇𝑇Σ𝑖𝑖−1(𝑥𝑥 − 𝜇𝜇𝑖𝑖) − 𝑑𝑑2 ln 2𝜋𝜋 −  ln|Σ𝑖𝑖| + ln 𝑃𝑃(𝜔𝜔𝑖𝑖)

(a). What is/are the value(s) for 𝑖𝑖?

(b). What is the value for 𝑑𝑑?

(c). What is the dimension of 𝜇𝜇𝑖𝑖?

(d). What is the dimension of Σ𝑖𝑖?

(e). What would be a reasonable value for 𝑃𝑃(𝜔𝜔𝑖𝑖)?

PROBLEM 3 (8 points) -- Choose one best answer.

A). Which of the following is NOT true for unsupervised learning?

1. Dimensionality reduction
2. Unlabeled dataset
3. Anomaly detection (d) Logistic regression

(e) Clustering

B). Which regularization do you want use if the dataset has more features than observations?

1. Linear regression
2. Lasso regression
3. Ridge regression
4. Logistic regression
5. Elastic regression

C). Insufficient quantity of training data may cause what problem?

1. Inaccurate feature extraction
2. Inaccurate feature selection
3. High generalization error
4. Low generalization error
5. None of the above

D). Which classifier would you use OVA and AVA for?

1. Linear regression
2. Softmax regression
3. Logistic regression
4. Ridge regression
5. Lasso regression

E). Which one of the following statement accurately describes Early Stopping?

1. Stop training when the validation error reaches minimum
2. Stop training after the training error reaches minimum
3. Stop training when the validation error drops below the training error (d) Stop training when the training error drops below the validation error

(e) Stop training when the validation error is equal to the training error

F). Let 𝑝𝑝̂𝑘𝑘 be the estimated probability that instance **x** is a member of class *k* in Softmax Regression. If 𝑝𝑝̂0, 𝑝𝑝̂1, 𝑝𝑝̂2 and 𝑝𝑝̂3 generated by the test sample **x** are 0.19, 0.35, 0.36, 0.1, respectively, what is the class label that the model might assign to **x**?

1. 0
2. 1
3. 2
4. 3
5. None of the above

G). An appropriate learning model to classify a data set with 4 classes is

1. Linear Regression
2. Logistic Regression
3. Softmax Regression
4. Ridge Regression
5. Lasso Regression

H). Unlike a logistic function, a linear function is not suitable for classification because

1. It does not work with regularization
2. It causes the algorithm to diverge
3. It is susceptible to outliers
4. All of the above
5. None of the above

PROBLEM 4 (2 point)

A classifier trained by the MNIST dataset is classifying the digit “3”. Calculate the precision and the recall if the threshold is placed where “x” is on the decision line below.

3 7 9 2 3 9 3 3 8 x 6 3 3 1 3 3 0 3

Negative predictions   Positive predictions

PROBLEM 5 (2 points)

The optimal weight vector of a 4D linear regression model without the bias term (𝑤𝑤0) is 𝐰𝐰 = [1, −2, 3, −4]. Calculate the following. Show your work.

a). The predicted value 𝑦𝑦 for a test instance if the feature vector is 𝐱𝐱 = [10, 20, 30, 40].

b). The penalty term of the Ridge regression if the regularization parameter is set to 0.6.

PROBLEM 6 (3 points)

Use the CART algorithm (equation 6.2) to train the model on this dataset:

𝑋𝑋 = [50, 60, 70, 80], 𝑌𝑌 = [𝑐𝑐, 𝑎𝑎, 𝑐𝑐, 𝑏𝑏]

Assume maximum depth = 1 (root node and its two children).

a). Determine the minimum CART cost function 𝐽𝐽(𝑘𝑘, 𝑡𝑡𝑘𝑘).

• You may calculate it manually or write code to find it, but you must show your work by showing the value of the cost function for each iteration.

b). Draw the decision tree. In each node, show the GINI score, the number of samples and the value.