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? 3

(b). What is $\mathbf{x}^{(5)}$? [8.2 2.5 1.7]

(c). What is $\mathbf{x}_1^{(3)}$? 6.9

(d). What is $\mathbf{x}_3^{(1)}$? 0.6

(e). What is y_4? blue

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: $g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$

(a). What is/are the value(s) for i? 1, 2

(b). What is the value for d? 3

(c). What is the dimension of μ_i? 3x1 or 1x3

(d). What is the dimension of Σ_i? 3x3

(e). What would be a reasonable value for $P(\omega_i)$? ½

PROBLEM 3 (8 points) -- Choose only one answer per question.

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

(a) Dimensionality reduction
(b) Unlabeled dataset
(c) Anomaly detection
(d) Logistic regression
(e) Clustering

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

(a) Linear regression
(b) Lasso regression
(c) Ridge regression
(d) Logistic regression
(e) Elastic regression

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

(a) Inaccurate feature extraction
(b) Inaccurate feature selection
(c) High generalization error
(d) Low generalization error
(e) None of the above

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

(a) Linear regression
(b) Softmax regression
(c) Logistic regression
(d) Ridge regression
(e) Lasso regression

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

(a) Stop training when the validation error reaches minimum
(b) Stop training after the training error reaches minimum
(c) 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 $\hat{p}_k$ be the estimated probability that instance **x** is a member of class k in Softmax Regression. If $\hat{p}_0, \hat{p}_1, \hat{p}_2$ and $\hat{p}_3$ 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**?

(a) 0
(b) 1
(c) 2
(d) 3
(e) None of the above

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

(a) Linear Regression
(b) Logistic Regression
(c) Softmax Regression
(d) Ridge Regression
(e) Lasso Regression

H). What are the sufficient statistics for the Gaussian distribution?

(a) Mean and standard deviation
(b) Mean and kurtosis
(c) Variance and standard deviation
(d) Standard deviation and kurtosis
(e) All of the above

PROBLEM 4 (3 points)

Consider the "Country Favorite Sports" example discussed in class (see Decision Tree slides). Find the GINI impurity score computed at the root.

At the root:

Sport	**Class probability**
Soccer	5/12
Baseball	2/12 = 1/6
Hockey	2/12 = 1/6
Cricket	3/12 = 1/4

GINI Impurity Score: $G_i = 1 - \sum_{k=1}^{4} p_{i,k}^2$

$$G_{root} = 1 - (5/12)^2 - (1/6)^2 - (1/6)^2 - (1/4)^2 = \mathbf{0.7083}$$

PROBLEM 5 (1 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

TP = 5; FP = 3; FN = 4

Precision = TP / (TP+FP) = 5 / (5+3) = 5/8 = 62.5%

Recall = TP / (TP+FN) = 5 / (5+4) = 5/9 = 55.6%

PROBLEM 6 (3 points)

Polynomial Regression uses a polynomial to fit the data. For the prediction function defined by the polynomial:

$$f\left(\mathbf{x}^{(i)}\right) = \omega_0 + \omega_1 \mathbf{x}^{(i)} + \omega_2 \mathbf{x}^{(i)^2} + \omega_3 \mathbf{x}^{(i)^3}$$

Derive the update equation for ω_3 if the cost function is MSE.

Cost function:

$$J(\boldsymbol{\omega}) = \frac{1}{2m} \sum_{i=1}^{m} \left(f\left(\mathbf{x}^{(i)}\right) - y_i\right)^2$$

Update equation:

$$\omega_3 = \omega_3 - \eta \frac{\partial J(\boldsymbol{\omega})}{\partial \omega_3}$$

$$\frac{\partial J(\boldsymbol{\omega})}{\partial \omega_3} = \frac{1}{m} \sum_{i=1}^{m} \left(f\left(\mathbf{x}^{(i)}\right) - y_i\right) . \mathbf{x}^{(i)^3} = \frac{1}{m} \sum_{i=1}^{m} \left(\omega_0 + \omega_1 \mathbf{x}^{(i)} + \omega_2 \mathbf{x}^{(i)^2} + \omega_3 \mathbf{x}^{(i)^3} - y_i\right) . \mathbf{x}^{(i)^3}$$