PROBLEM 1 (2 points)

A fully connected neural network is used to classify a dataset that contains 10,000 color images of uppercase and lowercase letters (A-Z, a-z), where each image has 30x30x3 pixels. What are the sizes of the input layer and the output layer?

Input layer = 30 x 30 x 3 = **2700**

Output layer = 26 + 26 = **52**

PROBLEM 2 (4 points)

A labeled dataset has 5000 observations, 50 features and 20 classes. The feature matrix $X$ is first processed by PCA. Three principal components are used to reconstruct the dataset $\hat{X}$. Combined with the class label, it is then used to train a Softmax Regression classifier.

What is the total number of *unused* principal components? 47

What is the dimension of feature matrix $X$? 5000x50

What is the dimension of reconstructed matrix $\hat{X}$ 5000x50

What is the dimension of the matrix that contains *all* of the labels for training the Softmax Regression?

Possible answers depending on how you interpret the question:

- 5000x20 (20 one hot encoded labels)
- 5000x70 (50 features + 20 one-hot encoded labels)

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

A). Which of the following is NOT true about Mercer's Theorem?

(a) It can be used to find the mapping function $\phi$
(b) Computing the kernel is sufficient
(c) It is not necessary to find the dimensionality of $\phi$
(d) Both (b) and (c)
(e) None of the above

B). Which of the following is NOT true for SVM?

(a) Training examples that are support vectors can be discarded
(b) Training examples that are support vectors cannot be discarded
(c) The decision boundary is placed in the middle of the "street"
(d) The decision boundary is placed at the optimal location
(e) None of the above

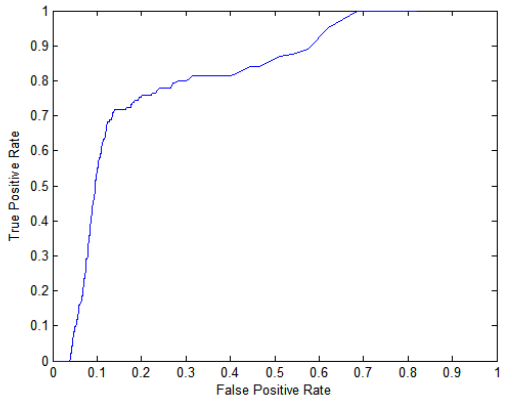
C). Your friend is training a classifier using a dataset whose underlying probability distribution is unknown. What should they use?

(a) Linear Regression
(b) Logistic Regression
(c) Bayes Decision Theoretic
(d) K-means
(e) None of the above

D). How many eigenvalues and eigenvectors does matrix $X = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}$ have?

(a) 3 eigenvalues, 4 eigenvectors
(b) 4 eigenvalues, 3 eigenvectors
(c) 3 eigenvalues, 3 eigenvectors
(d) 4 eigenvalues, 4 eigenvectors
(e) None of the above

E). What is the most likely AUC for the following ROC curve?




(a) 0.95
(b) 0.70
(c) 0.30
(d) 0.15
(e) 0.05

F). Deep neural networks can have a problem where the gradients drop dramatically during back-propagation. What is this problem called?

(a) Unknown gradients
(b) Exploding gradients
(c) Vanishing gradients
(d) Descending gradients
(e) None of the above

G). Which of the following is true about GANs?

(a) During training the weights of the discriminator and the generator are optimized at the same time
(b) During training the weights of the discriminator and the generator are "frozen" at the same time
(c) The generator takes fake images as input and produces real images
(d) GAN-generated images come from the output of the discriminator
(e) None of the above

H). Considerable degradation in statistical significance caused by data sparsity in the hyperspace is not uncommon in high-dimensional data. What is this phenomenon known as?

(a) Hebb's phenomenon
(b) Curse of the pharaohs
(c) Curse of sparsity
(d) Curse of dimensionality
(e) None of the above

PROBLEM 4 (2 points)

List all the training examples generated from the sentence ***Mary had two little lambs*** by Word2Vec if the center word is ***little*** and the window size is 3 words.

{little, Mary}

{little, had}

{little, two}

{little, lambs}

PROBLEM 5 (3 points)

SVD is a powerful decomposition technique. It decomposes matrix $A$ into the product of three matrices: $A = U.S.V^T$ where $U$ is the left singular vectors of $A$ (the eigenvectors of $AA^T$), $S$ is the singular values of $A$ (the square root of the eigenvalues of $AA^T$) and $V$ is the right singular vectors of $A$ (the eigenvectors of $A^TA$). If $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$, determine its singular values.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} . \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix} \rightarrow \begin{vmatrix} 14-\lambda & 10 \\ 10 & 14-\lambda \end{vmatrix} = 0$$

$$\rightarrow (14-\lambda)(14-\lambda) - 100 = 0 \rightarrow (14-\lambda) = \pm 10$$

$$\rightarrow \lambda_1 = 24, \lambda_2 = 4$$

Singular values: $\sqrt{24} = 4.8990$ and $\sqrt{4} = 2$

PROBLEM 6 (3 points)

Suppose the principal components and the eigenvectors in PCA are given by matrix P and matrix V, respectively. Assume the mean of the dataset is 0.

$$P = \begin{bmatrix} 3 & 2 & 3 \\ 4 & 3 & 5 \\ 1 & 1 & 7 \end{bmatrix} \qquad V = \begin{bmatrix} 2 & 0 & 3 \\ 5 & 1 & 2 \\ 4 & 4 & 0 \end{bmatrix}$$

What is the reconstructed dataset $\hat{X}$ if only the first principal component is used?

$$\hat{X} = P.V^T$$

Where $P$ is 3x1 and $V$ is 3x1.

$$\hat{X} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \times [2 \quad 5 \quad 4] = \begin{bmatrix} 6 & 15 & 12 \\ 8 & 20 & 16 \\ 2 & 5 & 4 \end{bmatrix}$$

PROBLEM 7 (3 points)

A two-dimensional dataset is used to train an SVM classifier with a polynomial kernel $(\mathbf{x}^T\mathbf{z} + 1)^2$.

| $i$ | $\mathbf{x}^T$ | $y$ | Langrage Multiplier | Support Vector |
|---|---|---|---|---|
| 1 | 1, 0 | -1 | 0.1 | Yes |
| 2 | 2, 4 | -1 | 0.2 | Yes |
| 3 | 1, 1 | -1 | 0.3 | No |
| 4 | 3, 1 | +1 | 0.4 | Yes |
| 5 | 1, 3 | +1 | 0.5 | No |
| 6 | 4, 4 | +1 | 0.6 | No |

Determine the class label for the instance (-1,1). Assume the bias is 0.

Use only the support vectors: M = 3

$$h = \sum_{i=1}^{M} \alpha_i y_i K(\mathbf{x}_i^T \mathbf{z}) = \sum_{i=1}^{M} \alpha_i y_i (\mathbf{x}_i^T \mathbf{z} + 1)^2$$

For $\boldsymbol{z}^T = -1,1$: $h = -(0.1)(-1+1)^2 - (0.2)(2+1)^2 + (0.4)(-2+1)^2 = -1.4 < 0$

→ **negative class**