CSCI4390/6390 – Data Mining, Fall 2020, Exam II
CSCI4390:100pts, CSCI6390:110pts, Bonus: 10pts

Name: _____ RIN: _____ RCS ID:_____

You must show all work for full credit; it is not enough to show only final answers. You can use a calculator, but not any other device or software. Use at least two decimals for precision (more when necessary).

	X_1	X_2	Y
$\mathbf{x}_1$	2	2	1
$\mathbf{x}_2$	1	2	0
$\mathbf{x}_3$	3	1	1
$\mathbf{x}_4$	2	3	0

Table 1: Regression data

1. (CSCI4390:30 points, CSCI6390:40 points) Consider the points in Table 1. Ignore the response variable Y. Define the homogeneous quadratic kernel between any two point as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^2$$

   (a) (10 points) Compute the kernel matrix $\mathbf{K}$.

   (b) (10 points) Compute the total variance in feature space using only $\mathbf{K}$.

   (c) (10 points) Compute the angle between $\mathbf{x}_1$ and $\mathbf{x}_3$ using the Gaussian kernel, given the gaussian kernel matrix:

$$\mathbf{K} = \begin{pmatrix} 1 & 0.37 & 0.14 & 0.37 \\ 0.37 & 1 & 0.007 & 0.14 \\ 0.14 & 0.007 & 1 & 0.007 \\ 0.37 & 0.14 & 0.007 & 1 \end{pmatrix}$$

   (d) (CSCI6390: 10 points) Let the dominant eigenvector of the kernel matrix $\mathbf{K}$ be

$$\mathbf{c}_1 = (0.00325, 0.085, -0.75, 0.66)^T$$

   Let the centered homogeneous quadratic kernel be given as:

$$\overline{\mathbf{K}} = \begin{pmatrix} 0.625 & 1.125 & -0.875 & -0.875 \\ 1.125 & 18.625 & -11.375 & -8.375 \\ -0.875 & -11.375 & 33.625 & -21.375 \\ -0.875 & -8.375 & -21.375 & 30.625 \end{pmatrix}$$

   What is the variance along the first kernel PC, and what fraction of the total variance is captured by the first kernel PC?

2. (30 points) Consider the data shown in Table 1. Let Y be the response variable, and X_1 and X_2 the predictor variables. The task is to compute the predicted response $\hat{Y}$ for a linear regression model using the geometric approach. Answer the following questions:

(a) (15 points) Compute an orthogonal basis for independent variables via the QR factorization approach for the linear regression model.

(b) (15 points) Using the orthogonal basis, find the predicted response vector $\hat{Y}$.

3. (30 points) Given the data in Table 1, assume that the initial augmented weight vector for logistic regression is $\tilde{\mathbf{w}} = (1, 0.5, 0.5)^T$. Answer the following questions:

(a) (15 points) What is the cross-entropy error for this dataset?

(b) (15 points) Update the weight vector $\tilde{\mathbf{w}}$ by computing the gradient at $\mathbf{x}_1$; assume the step size for gradient ascent is $\eta = 1$ for maximum likelihood estimation.
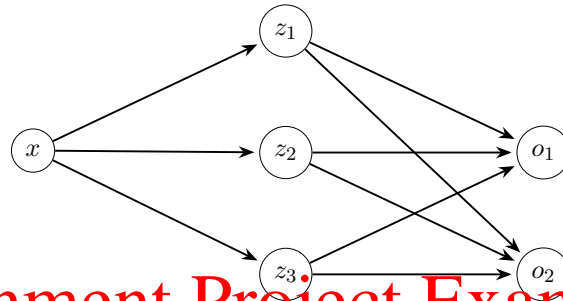
Figure 1: Neural network

4. (10 points) Given the neural network in Fig. 1, let bias values be fixed at 0, and let the weight matrices between input and hidden, and hidden and output layers, be given as:

$$\mathbf{W}_h = (1 \quad 1 \quad -2) \qquad \mathbf{W}_o = \begin{pmatrix} 0.5 & 1 \\ 1 & 2 \\ 2 & 0.5 \end{pmatrix}$$

Assume that hidden layer uses ReLU, whereas output layer uses linear activation. Assume that the loss is given via the squared error. Compute the loss when the input is $x = 2$ and the true response is $\mathbf{y} = (5, 9)^T$:

X_1	Y
1	1
2	3
4	4
6	3

Table 2: Kernel Regression Data

5. (**Bonus:** 10 points) Assume that a kernel ridge regression model is applied on the dataset in Table 2. Find the bias term and regression coefficients assuming we use an inhomogeneous quadratic kernel (with the kernel constant $c = 4$), and with mixture coefficient vector $\mathbf{c} = (1, 0.5, -0.25, 0)^T$, and regression constant $\alpha = 10$.