First Mid Term Examination, ME-781, September 14, 2018

Name: **Roll No:**

Total Time 2 hours; Total Marks 100

Open notes (self hand-written) examination.

20 1. A data set contains the life of a pressure cylinder (y) as a function of the number of moles of gas (n) in the cylinder, the pressure (P) of the gas and the temperature (T) of the gas. If a linear regression model is used to predict the life of the pressure cylinder as a function of the predictors then comment whether it is possible to use fewer than the three predictors for the regression. Please provide explanation for your answer. On the other hand, if nearest neighbour regression is used to predict the life of the pressure cylinder, then will you be able to use fewer than three predictors, please explain. Assume that it is an ideal gas which follows the relation $PV = nRT$ and all the cylinders are of the same volume.

40 2. Provide brief reasoning to show whether the kth nearest neighbor regression would perform better, similar or worst when compared to a quadratic regression model for the following cases:

a.) The underlying true model being linear and the test data is very sparse with small random error

b.) The underlying true model being linear and extensive test data is available with small random error

c.) The underlying true model being linear and extensive test data is available with large random error

d.) The underlying true model is cubic and extensive test data is available with small random error

Assume the best possible value of k is chosen in the kth nearest neighbor regression and the random error has zero mean and is not a function of predictor. (PS: small or large error implies small or large variance in error)

20 3. To a data set (of X and Y) a polynomial model with increasing degrees is fitted. The test MSE is a function of variance and bias of the fitting model $\hat{f}(x_0)$ and the variance of the error in data. Please state as to how would MSE of training compare with MSE of test with comments of the origin of variance in test MSE model (<u>with a brief schematic pictorial reason</u>) when:

a.) A polynomial model with increasing degrees is fitted to an underlining linear model with large error in data.

b.) A polynomial model with increasing degrees is fitted to an underlining cubic model with small error in data.

10 4. Multiple linear regression model has the form

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \qquad \text{with} \qquad RSS(\beta) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2$$

This can be written in matrix form as: $$Y = X\beta$$

Where, Y, X and β are matrix of the size n x 1, n x (p+1) and (p+1) x 1, respectively. (Note that n is the number of training data points, and p is the number of predictors)

Where, $RSS(\beta) = (Y - X\beta)^T (Y - X\beta)$ and $\frac{\partial RSS(\beta)}{\partial \beta} = -2X^T (Y - X\beta)$

Show that the choice of β which minimizes the RSS leads to residual vector $(Y - X\beta)$ becoming orthogonal to column space of X.

20 5. Show that in PCA the variance of the transformed space variables are the eigenvalues and that the covariance of the transformed space variables is a diagonal matrix.

30 6. For a data set (with two predictors (X1, X2)) the covariance matrix is given by $C = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$.

This corresponds to eigenvalues and eigenvectors as follows:

Eigenvalues $\quad$ *λ1 = 0.8727* $\quad$ and $\quad$ *λ2 = 0.1273*

Eigenvectors $\quad a1 = \begin{bmatrix} -0.85065 \\ -0.52571 \end{bmatrix}$ and $\quad a2 = \begin{bmatrix} 0.52571 \\ -0.85065 \end{bmatrix}$

What would be the eigenvalues and eigenvectors for data sets which have covariance Matrix as:

a. $\quad C1 = \frac{1}{3}\begin{bmatrix} 2+0.6 & 1 \\ 1 & 1+0.6 \end{bmatrix}$ $\qquad$ b. $\quad C2 = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \frac{1}{9}\begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$

20 7. In a classification problem (with two predictors X1 and X2) multiple Logistic regression was found to give excellent results. The logit for the multiple Logistic regression is given by $\beta_0 + \beta_1 X_1 + \beta_2 X_2$
It was found that same classification can be achieved after dimensionality reduction to one variable by principal component analysis.
Given the above information find the rotation matrix used for the principal component analysis.

40 8. A data set of circles and squares in two dimensional space is provided below. A classification model needs to be developed which can identify circles from squares for an unknown point in the 2D space. Schematically draw the first nearest neighbour decision boundary and the linear SVM decision boundary for the given data set.