In this assignment, we will use the mathematical notations from the following reference books:

- **Reference books:**

  1. *An Introduction to Statistical Learning with Applications in R* **(ISLR)**. Authors: James, G., Witten, D., Hastie, T., Tibshirani, R.

  2. *The Elements of Statistical Learning Data Mining, Inference, and Prediction* **(ESL)**. Authors: Hastie, T., Tibshirani, R., Friedman, J.

1 Support Vector Machines

We have seen that in p = 2 dimensions, a linear decision boundary takes the form $\beta_0+\beta_1 X_1+\beta_2 X_2 = 0$. We now investigate a non-linear decision boundary.

(a) **(3 points)** Sketch the curve $(1 + X_1)^2 + (2 - X_2)^2 = 5$.

(b) **(3 points)** On your sketch, indicate the set of points for which

$(1 + X_1)^2 + (2 - X_2)^2 > 5$

as well as the set of points for which

$(1 + X_1)^2 + (2 - X_2)^2 \leq 5$

(c) **(4 points**) Suppose that a classifier assigns an observation to the blue class if

$(1 + X_1)^2 + (2 - X_2)^2 > 5$

and to the red class otherwise. To what class is the observation $(0, 0)$ classified? $(-1, 1)$?$(2, 2)$?$(3, 8)$?

(d) **(5 points)** Argue that while the decision boundary in (c) is not linear in terms of X_1 and X_2, it is linear in terms of X_1, X_1^2, X_2, X_2^2.

2 Model Selection

(a) **(4 points)**Why does Ridge Regression improve over least squares? Explain.

(b) **(4 points)** What are the disadvantages of Ridge Regression? Explain.

(c) **(7 points)** Why is it that the LASSO, unlike Ridge Regression, results in coefficient estimates that are exactly equal to zero? Explain.

3 Unsupervised learning

In this problem, you will perform K-means clustering manually, with K = 2, on a small example with n = 6 observations and p = 2 features. The observations are as follows.

(a) **(2 points)** Plot the observations

(b) **(2 points)** Randomly assign a cluster label to each observation.

Observations	X1	X2
1	1	4
2	1	3
3	0	4
4	5	1
5	6	2
6	4	0

(c) **(3 points)** Compute the centroid for each cluster.

(d) **(5 points)** Assign each observation to the centroid to which it is closest, in terms of Euclidean distance. Report the cluster labels for each observation.

(e) **(6 points)** Repeat (c) and (d) until the answers obtained stop changing.

(f) **(2 points)** In your plot from (a), color the observations according to the cluster labels obtained.

4 Non-linearity (Extra Credit: 10 points)

Consider two curves, $\hat{g}_1$ and $\hat{g}_2$, defined by

$$\hat{g}_1 = \arg\min_g \Big(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int [g^{(3)}(x)]^2 dx \Big),$$

$$\hat{g}_2 = \arg\min_g \Big(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int [g^{(4)}(x)]^2 dx \Big),$$

where $g^{(m)}$ denotes the m^{th} derivative of g.

(a) **(3 points)** As $\lambda \to \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training RSS?

(b) **(3 points)** As $\lambda \to \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller test RSS?

(c) **(4 points)** As $\lambda = 0$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training and test RSS?

Submission Instruction: You need to provide the followings:

- Provide your answers to the problems in a PDF file, named as `IS777_hw#_fa19_LastName.pdf`. You need to submit the homework in electronic version as pdf file on Blackboard. If you choose handwriting instead of typing all the answers, you will get 40% points deducted.

Collaboration: You may collaborate. However, collaboration has to be limited to discussion only and you need to write your own solution and submit separately. You also need to list with whom you have discussed.