USC Viterbi
School of Engineering
Daniel J. Epstein Department of Industrial and Systems Engineering

ISE529 Predictive Analytics
Instructor: Dr. Tao Ma
2025 Fall

Homework 5
Due by: **April 21, 2025,** 11:59 PM

Instructions:

1. Print your First and Last name and NetID on your answer sheets
2. Submit all your answers including Python scripts and report in a single Jupyter Lab file (.ipynb) or along with a single PDF to Brightspace by due date. No other file formats will be graded. No late submission will be accepted.
3. Total 4 problems. Total points: 100

---

1. (20 points)
Predict per capita crime rate in the *Boston.csv* data set. Split the data set into 70% for a training set and 30% for a test set. Fit a lasso model, ridge regression model, and PCR model respectively. Use cross-validation method to determine λ and M (the number of PCs). Present the test error and discuss results for the approaches that you consider.

2. (20 points)
Predict the number of applications received using the other variables in the *College.csv* data set. Split the data set into 60% for a training set and 40% for a test set.

(a) Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.

(b) Fit a lasso model on the training set, with λ chosen by cross-validation. Report the test error obtained, along with the number of non-zero coefficient estimates.

(c) Fit a PLS model on the training set, with M chosen by cross-validation. Report the test error obtained, along with the value of M selected by cross-validation.

3. (30 points)

Use the following code to generate a data set with $n = 500$ and $p = 2$, such that the observations belong to two classes with a quadratic decision boundary between them.

```
rng = np.random.default_rng(5)
x1 = rng.uniform(size=500) - 0.5
x2 = rng.uniform(size=500) - 0.5
y = x1**2 - x2**2 > 0
```

(a) Plot the observations, colored according to their class labels. Your plot should display X_1 on the x-axis, and X_2 on the y-axis.

(b) Fit a logistic regression model to the data using X_1, X_2, X_1^2, X_2^2, and $X_1 \times X_2$ as predictors. Obtain a class prediction for each training observation (using full data set). Plot the observations, colored according to the predicted class labels.

(c) Fit a SVM using a non-linear kernel (polynomial with d>1 or RBF kernel) to the data. Obtain a class prediction for each training observation (using full data set). Plot the observations, colored according to the predicted class labels.

(d) Comment on your results.

4. (30 points)

Fit a neural network to the *Default.csv* data set. Default is a response variable. Split the data into 70% training set and 30% test set. Use tensorflow.keras to build your neural network with the following specification for the hyperparameters:

- a single hidden layer with 10 neurons
- dropout regularization rate 40%
- activation function: relu or sigmoid
- epochs 30
- batchsize 128
- validation split 0.2
- loss = 'mean_squared_error'
- optimizer = 'rmsprop'
- metrics = ['mae']

Tasks:

- Perform prediction with test data set.
- Plot training and test error curves.
- Show test set accuracy.