# Problem Set 2

Name: Rupali Roy
UT EID: rur73

1. **Supervised Learning: Generalization [5 points]:**

   (a) Describe the motivation for splitting data into a training dataset, validation dataset, and testing dataset and the purpose of each subset for developing machine learning algorithms.

   Ans. The main motivation for splitting the data into a training, validation and testing dataset is to ensure that our machine learning models performs well on previously unseen set of data and it generalizes well to new cases with good prediction accuracy. Using testing and validation set helps us to prevent the machine learning model from overfitting and reducing error on new dataset.

   Training Dataset:

   We train our model using training dataset and fit the model on this data. The model learns to predict the target variable using this data.

   Validation Dataset: For selecting best performing model from a set of machine learning models, we first need to estimate the prediction error for each model. In order to get the best model, we use the training dataset and train multiple models on this dataset using different set of hyperparameters. The models are then tested using the validation dataset, then the model and the hyperparameters that has the least error on the validation set is used for further prediction.

   Test Dataset:

   After testing the model performance on validation dataset, we then check the performance of our final model on new set of data using the test dataset. We can evaluate our model on the test data by calculating the generalization error.

   (b) If your model performs poorly on the training data and poorly on new instances, is the model overfitting or underfitting?

   Ans. Underfitting

   (c) If your model performs well on the training data and generalizes poorly to new instances, is the model overfitting or underfitting?

   Ans. Overfitting

2. **Model Parameters versus Hyperparameters [5 points]:**

(a) In your own words, describe what is a model parameter and hyperparameter and how they differ.

Ans. A model parameter is the parameter that is learned by the machine, a model learns this parameter on its own using processes like Gradient Descent. For eg. In case of Linear Regression, a model learns the weights of independent variables and biases. When we train the model using training dataset, the learning algorithm tries to find the best values for the model parameters so that the generalization error is less.

A hyperparameter is a parameter that is set manually, and it is the parameter of the learning algorithm that is used for the prediction analysis. For example – Learning Rate and Activation Function are example of hyperparameter in case of neural network.

The main difference between the two is that we estimate the model parameters while training the model whereas for hyperparameters we can set the values manually before-hand. Since the model parameters are obtained by training, the value is dependent on the training dataset. But in case of hyperparameter it is independent of the training dataset.

(b) What is/are the model parameter(s) for both the ridge regression and lasso regression models?

Ans. The model parameter(s) for both ridge and lasso regression models are the weight coefficients for different features and the slope of the regression line (constant).

(c) What is/are the hyperparameter(s) for both the ridge regression and lasso regression models?

Ans. Regularization Parameter (alpha) is the hyperparameter used in both ridge and lasso regression. In Ridge regression we use alpha to add constraint to penalize the squared weight values and in Lasso regression we use alpha to add a constraint to penalize absolute weight values.

3. **Machine Learning Notation [5 points]:**

For this task, show that you can convert data into machine learning notation. The goal of the toy prediction task is to anticipate what would be the overall score Gertrude would get in this class if she took it next year (i.e., between 0% and 100%). We will design our task to predict her performance based on the time she spends per week on her assignments, the number of hours she sleeps per week, and the number of caffeinated drinks she consumes each week. We know for her calculus class she got a score of 70%, spent 8 hours per week on assignments, slept 56 hours per week, and drank 5 caffeinated drinks per week. We know for her ethics class she got a score of 91%, spent 14 hours per week on assignments, slept 50 hours per week, and drank 7 caffeinated drinks per week. We know for her writing class she got a score of 85%, spent 7 hours per week on assignments, slept 65 hours per week, and drank 2 caffeinated drinks per week. Finally, we know for her statistics class she got a score of 92%, spent 10 hours per week on assignments, slept 60 hours per week, and drank 4 caffeinated drinks per week.

(a) Show the matrix that represents the features for the aforementioned dataset. In your own words, also describe what each row and column represents

Features:

a. X1 = time she spends per week on her assignments

b. X2 = the number of hours she sleeps per week

c. X3 = number of caffeinated drinks she consumes each week

Ans. The first row has data for three features X1, X2 and X3 for Calculus. The second row has data for Ethics class, it shows that 14 hours per week on her Ethics class assignment, 50 hours of sleep and had 7 caffeinated drinks that week. The third and fourth row has data for X1, X2,X3 and X4 for Writing and Statistics class. Each column has data for different features mentioned above.

Q3.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2d} \\ \vdots & \vdots & & & & \\ x_{n1} & x_{n2} & & x_{nj} & & x_{nd} \end{bmatrix} \qquad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

feature column $x_{*j}$ $\qquad$ $y$

Feature Matrix

a)

| | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| Calculus → | 8 | 56 | 5 |
| Ethics → | 14 | 50 | 7 |
| Writing → | 7 | 65 | 2 |
| Stats → | 10 | 60 | 4 |

b) Target Variable Vector

$$y = \begin{bmatrix} 0.7 \\ 0.91 \\ 0.85 \\ 0.92 \end{bmatrix}$$

X1 X2 X3

$$\begin{bmatrix} 8 & 56 & 5 \\ 14 & 50 & 7 \\ 7 & 65 & 2 \\ 10 & 60 & 4 \end{bmatrix}$$

(b) Show the vector that represents the target variables.

$$Y = \begin{bmatrix} 0.7 \\ 0.91 \\ 0.85 \\ 0.92 \end{bmatrix}$$

4. **Regression: Evaluation Metrics [5 points]:** Table 1 shows the predicted and actual salaries offered to Gertrude by different companies for machine learning positions.

   (a) Report the mean absolute error for the predictions.

   (b)Report the mean square error for the predictions.

   (c)Which metric is more affected by outliers andwhy?

| Company | Predicted Value | Actual Value |
|---|---|---|
| XYZ Machine Learning | $60,000 | $55,000 |
| LifePredictions | $65,000 | $75,000 |
| GettingRich Investments | $115,000 | $85,000 |
| Vision Forever Incorporated | $90,000 | $85,000 |
| Know It All Firm | $88,000 | $88,000 |

Table 1: Data.

Ans.

| Company | Predicted Value(Ypred) | Actual Value(Y) | Error (abs(predicted - actual)) | Squared Error Square(Predicted-actual) |
|---|---|---|---|---|
| **XYZ Machine Learning** | 60000 | 55000 | 5000 | 25000000 |
| **LifePredictions** | 65000 | 75000 | 10000 | 100000000 |
| **GettingRich Investments** | 115000 | 85000 | 30000 | 900000000 |
| **Vision Forever Incorpora** | 90000 | 85000 | 5000 | 25000000 |
| **Know It All Firm** | 88000 | 88000 | 0 | 0 |
| | | | **Mean Absolute Error** | **Mean Square Error** |
| | | | **10000** | **210000000** |

a. Mean Absolute Error is **10000**

b. Mean Square Error is **210000000**

c. I believe that Mean Square Error is more affected by outlier since we are squaring the difference between the predicted and actual value. In case of MAE the residual for each term contributes in a similar proportion to the final error but in case if MSE the proportion is different. So if we have huge difference between predicted and actual value then they will be affecting the error more in case of MSE than in MAE since we are squaring the error. In our case “GettingRich Investments is the outlier” as the diff is around 30,000.