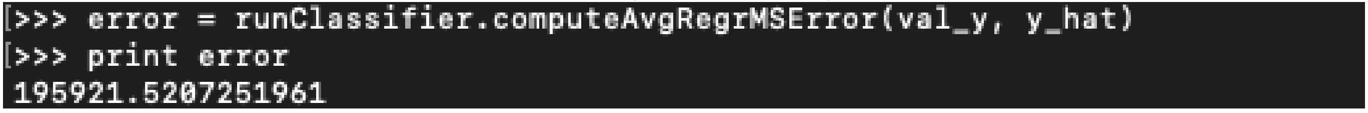
EC4213/ET5402/CT5303: Machine learning and deep learning **Fall 2019**

Report for CA2

Instructor: Jonghyun Choi *Jisu Shin (20175095)*

REPORT1.

linear regression error.png



```
[>>> error = runClassifier.computeAvgRegrMSError(val_y, y_hat)
[>>> print error
195921.5207251961
```

FIGURE 1 – error printed by vanila linear regression

Above figure shows the result error of the vanila linear regression I implemented. The error seems quite large but the given data size was also large so I think it is reasonable. To reduce the error, kernel can be used as it is mentioned, and also there are many kinds of input data, so finding the relationship on those data and handle those data differently will be helpful. Also to reduce error, we should prevent overfitting by set the limit of parameter w, or using the ridge regression, etc.

REPORT2.

of the model as a function of the regularization.png

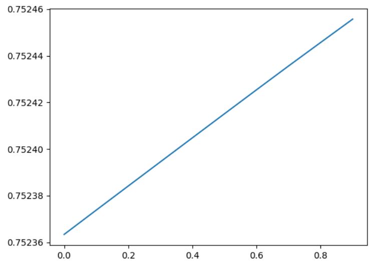


FIGURE 2 – Graph with x-axis : gamma, y-axis : accuracy

Figure 2 is the graph with x-axis is gamma which sweeps from 0.0 to 1.0 and y-axis is the accuracy of the model. By this figure, we can see that the accuracy gets higher as the gamma increases to 1.0 and also it is proportional. Also on figure 3, we can find out that when the gamma is set to 0.0, which works the same with the vanila linear regression model, the error(195921.xxx..) is same with the figure 1. From this, gamma works as the regularization to reduce the error and it can be verified because the ridge regression error is 195920.xxx... when gamma is set to 1.0.

REPORT3. Your writing goes here.

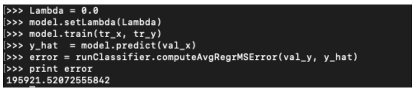
REPORT4. I failed to get the error because of the error related to tensorflow. To improvement its effect, we can add regularizers or change the value of how we quantize $Y(tr_y)$.

REPORT5. We use quantizeY to solve the regression problem by converting it classification problem.

REPORT6.

As you can see on this figure 4, the error from OLS model in scikit-learn(195894.xxx..) is almost the same with the error from the model I implemented on report 1(195921.xxx). The very slight difference can be interpreted by the existence of the bias term. I did not consider the bias term but the scikit-learn model might have considered this.

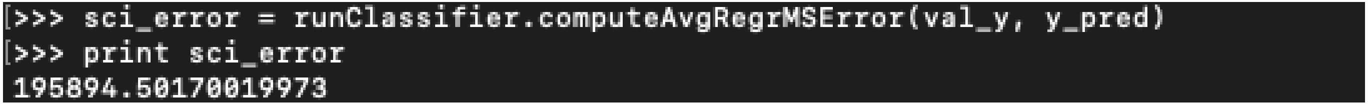
regression error when gamma is 0.png



```
[>>> Lambda = 0.0
[>>> model.setLambda(Lambda)
[>>> model.train(tr_x, tr_y)
[>>> y_hat  = model.predict(val_x)
[>>> error = runClassifier.computeAvgRegrMSError(val_y, y_hat)
[>>> print error
195921.52072555842
```

FIGURE 3 – Ridge regression error (gamma = 0.0)

model in scikit-learn.png

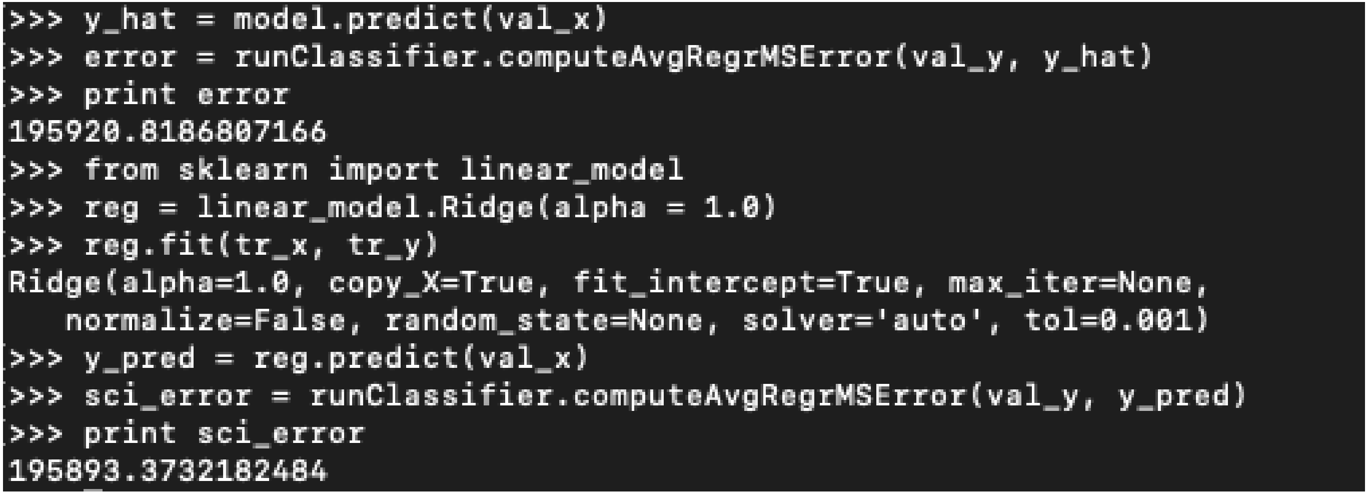


```
[>>> sci_error = runClassifier.computeAvgRegrMSError(val_y, y_pred)
[>>> print sci_error
 195894.50170019973
```

FIGURE 4 – error printed by OLS model in scikit-learn

REPORT7.

regression model by myself and by scikit-learn.png



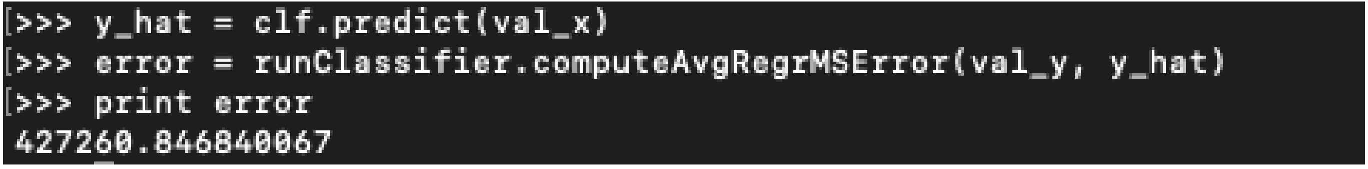
```
>>> y_hat = model.predict(val_x)
>>> error = runClassifier.computeAvgRegrMSError(val_y, y_hat)
>>> print error
195920.8186807166
>>> from sklearn import linear_model
>>> reg = linear_model.Ridge(alpha = 1.0)
>>> reg.fit(tr_x, tr_y)
Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=None,
   normalize=False, random_state=None, solver='auto', tol=0.001)
>>> y_pred = reg.predict(val_x)
>>> sci_error = runClassifier.computeAvgRegrMSError(val_y, y_pred)
>>> print sci_error
195893.3732182484
```

FIGURE 5 – error printed by ridge regression I implemented and scikit-learn both

As you can see on this figure 5, the error from ridge regression model in scikit-learn(195893.xxx..) is almost the same with the error from the model I implemented on report 1(195920.xxx). The very slight difference can be interpreted by the existence of the bias term. I did not consider the bias term but the scikit-learn model might have considered this. It looks like the same situation with the report 6.

REPORT8.

regression model in scikit-learn model.png



```
[>>> y_hat = clf.predict(val_x)
[>>> error = runClassifier.computeAvgRegrMSError(val_y, y_hat)
[>>> print error
 427260.846840067
```

FIGURE 6 – error printed by logistic regression model in scikit-learn model

On figure 6, the error printed by logistic regression model in scikit-learn model is 427260.xxx... . I could not get the error of the logistic regression model I implemented, but I think when I fix the slight error on my setting it will print the similar error like this figure.

REPORT9. I used ridge regression that I implemented on kaggle uploaded file as result.csv. My private score is 205714.19400 and the public score is 223480.70964.