Giovanni Mueco

March 28, 2021

CS489\_1001 - Machine Learning

HW #3

For this assignment, we were to implement Linear Regression in order to classify the given MNIST image data as either a 5 or a 6, and use K-Fold Cross Validation as our method of testing accuracy in order to find the best Linear Regression Model.

I begin with normalizing the data, by dividing all the data by 255, giving us norm\_data. I set my k-value, which for this assignment is 10. Using this k, I get the size of each “fold” array by dividing the size of the norm\_data by k, giving us f\_rows which will be used to increment through each fold during the cross-validation process.

I then set the arrays for TPR, FPR, and average accuracies which I call “scores”. These arrays will be used to hold this kind of data that we obtain each time we use a new fold to validate.

Next comes the outer loop which will iterate k times. Each iteration, it uses a new fold of the training data as validation data in order to test our model. As well as this, with each iteration we obtain the Average Accuracy (score) of the model, the TPR, and the FPR. Using this data, we can determine which model works best for making predictions.

The inner loop will iterate through the other folds or “training folds” that are not currently being used as validation data. Each time we do, we will create a new model and make calculations which help us determine whichever one is best.

We make our final calculations at the end, specifically for overall accuracy. According to this program, the most accurate fold for predicting this data is Fold 7. Below, I’ve included the Accuracy, TPR, and FPR for each fold as well as the Average Overall Accuracy.

Fold #0:

Accuracy: 0.8039215686274509

TPR: 1.7058823529411764

FPR: 0.17647058823529413

Fold #1:

Accuracy: 0.8464052287581699

TPR: 1.5238095238095237

FPR: 0.0

Fold #2:

Accuracy: 0.9150326797385622

TPR: 2.5833333333333335

FPR: 0.09090909090909091

Fold #3:

Accuracy: 0.934640522875817

TPR: 2.1333333333333333

FPR: 0.05263157894736842

Fold #4:

Accuracy: 0.8954248366013073

TPR: 1.4090909090909092

FPR: 0.16666666666666666

Fold #5:

Accuracy: 0.84640522875817

TPR: 1.4736842105263157

FPR: 0.2

Fold #6:

Accuracy: 0.8954248366013071

TPR: 1.4761904761904763

FPR: 0.07692307692307693

Fold #7:

Accuracy: 0.9411764705882353

TPR: 2.357142857142857

FPR: 0.0

Fold #8:

Accuracy: 0.9052287581699348

TPR: 1.7647058823529411

FPR: 0.17647058823529413

Fold #9:

Accuracy: 0.8199233716475096

TPR: 1.5625

FPR: 0.07692307692307693

Average Accuracy Overall: 0.8803583502366464