# Lab 1 Report

EE 385T Intro to Machine Learning

## Code

https://github.com/bennycooly/INF_385T_Intro_To_ML/blob/master/labs/lab1

## Questions

1. Quadratic Dataset
    d. Performance Metrics

| Regression Type | Mean Absolute Error (MAE) | Correlation Coefficient (r) |
|---|---|---|
| **Linear** | 1.3120 | 0.9757 |
| **Ridge** | 1.3124 | 0.9757 |
| **Lasso** | 1.5054 | 0.9757 |
| **Polynomial** | 0.7543 | 0.9917 |

    e. Discussion

In my case I used a slight variation of the standard quadratic expression ($ax^2 + bx + c$) by setting a=1, b=2, and c=1 just to make it interesting and hopefully create a larger difference between the models.

The results show that the standard linear regression model performs slightly better in terms of the Mean Absolute Error (MAE) than the ridge and lasso counterparts, which I first thought was surprising but it makes sense; regularization aims to solve the overfitting problem, and in this case there is not an overfitting problem but an underfitting problem because we are trying to model a quadratic relationship with a linear one. Thus, any attempts to further adjust the model to avoid overfitting actually causes the underfitting issue to be worse. This is also why the lasso model has a much higher MAE than the ridge model, since penalizing the weights using the absolute error may actually result in some coefficients being set to 0, which makes the model even less accurate. However, surprisingly the correlation coefficient stayed the same between all three linear regression models. I am guessing that the dataset did not generate enough noise to make enough of a difference between these models for the correlation coefficient.

Finally, it was obvious that the polynomial regression with a degree of 2 will model our quadratic dataset almost perfectly, and the results show a much lower MAE and higher correlation coefficient, which was expected.

2. Custom Dataset: Red Wine Quality

d. Performance Metrics

| Regression Type | Mean Absolute Error (MAE) | Correlation Coefficient (r) |
|---|---|---|
| **Linear** | 0.5035 | 0.6389 |
| **Ridge** | 0.5058 | 0.6358 |
| **Lasso** | 0.6610 | 0.1238 |
| **Polynomial** | 0.4948 | 0.6460 |

e. Discussion

I was looking online for some interesting datasets and found one that related the subject quality of wine with many features such as acidity, sugar level, chlorides, pH, alcohol, and many others. The actual quality of the wine is a subjective measurement from a scale of 0 to 10 given by "wine experts". There are a total of 11 different features, which makes this problem interesting because not all of them may be correlated to the quality of the wine.

The results show that the standard linear regression model performs better than the regularized counterparts because it yielded a lower MAE and higher correlation coefficient. This is again not surprising since we likely do not have an overfitting problem with a purely linear model. Thus, regularization attempts to cancel out weights may actually make the model even more inaccurate. The most surprising find was the sudden increase in MAE and huge decrease in the correlation coefficient in the lasso model. My guess here is that since the lasso model may cause many of the weights to be zero, the solution plane becomes too flat and the model is basically a flat line in 2D, which results in almost no correlation.

The polynomial regression showed a slightly lower MAE and higher correlation coefficient but not enough to be significant. It would be interesting to test higher degree polynomials and see if they make the model better or worse.

3. Hyper-Parameters
   a. Ridge regression: when alpha is 0, there is no weight penalization in the error equation and the model is the exact same as a standard linear regression.
   b. Ridge regression: when alpha is 1, the maximum weight penalization is applied, so higher-degree polynomial models will be reduced to simpler ones and the solution curve is smoother.
   c. Lasso regression: when alpha is 0, there is no weight penalization in the error equation and the model is the exact same as a standard linear regression.
   d. Lasso regression: when alpha is 1, the maximum weight penalization is applied, so higher-degree polynomial models will be reduced to simpler ones and the solution curve is smoother. The big difference between this and ridge regression is that with lasso, weights can be set to 0, which will cancel out features that have no effect on the solution.