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

Report for CA2

Instructor: Jonghyun Choi *Jwa Younkyung (20165174)*

REPORT1. Report the error. Discuss any ideas to reduce the errors
Using my implementation of vanilla linear regression, the MSE(mean squared error) is 38285489834.36369. By standard scaling with normal distribution, error is 0.28161601241260686. For reducing the error, we need to delete some feature (sqft living) because there is related with other features (sqft living = sqft above + sqft basement). If we don't delete that feature, error is 0.607168576563699.

REPORT2. Sweep gamma from 0.0 to 1.0 (or some other reasonable values), plot a graph (x-axis : gamma, y-axis : accuracy) and discuss the effect of the gamma (especially comparing with vanilla linear when gamma=0.)

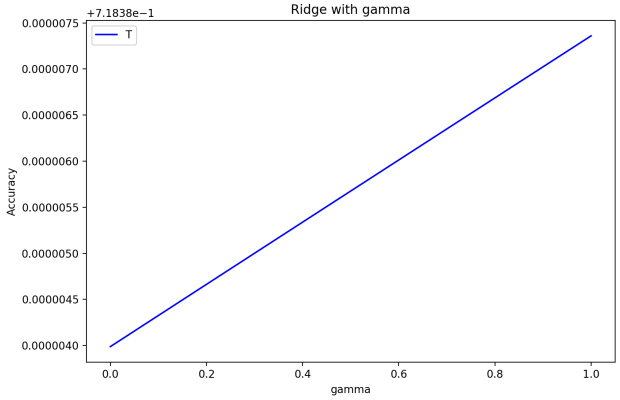

FIGURE 1 – DT on AI.

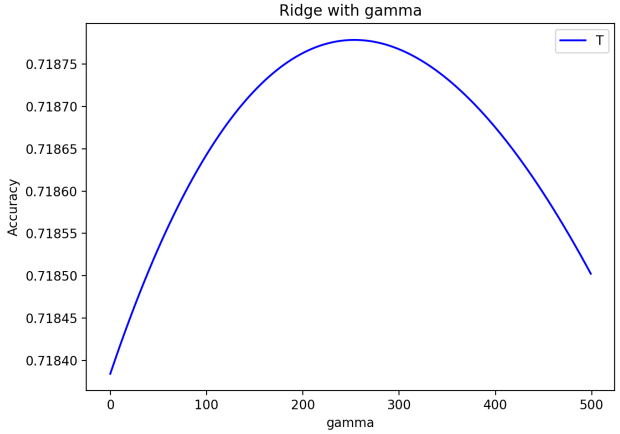

FIGURE 2 – DT on AI.

In gamma = 0, the accuracy is about 0.71840. As gamma increases up to about 250, accuracy also increases (about 0.71875). But when gamma is greater than 250, the accuracy decreases. Using gamma can reduce overfitting and penalize.

REPORT4. Report the error. Discuss any improvemental ideas (eg., add regularizers.).

REPORT5. Discuss any idea to solve the regression problem by converting it classification problem.

REPORT6. Compare the error by your implementations of vanilla linear regression and OLS model in scikit-learn and discuss the reason for the difference. If they are identical, report and claim you're awesome!
My result is very similar with scikit.

error	non scaling	scaling
my	38285489834	0.281616
scikit	38093303528	0.281616

REPORT7. Compare the error by your implementations of ridge regression and ridge regression model in scikit-learn and discuss the reason for the difference. If they are identical, report and claim you're awesome!
My result is very similar with scikit.

error	non scaling	scaling
my	38284833768	0.281612
scikit	38092446800	0.281613

REPORT8. Compare the error by your implementations of logistic regression and logistic regression model in scikit-learn and discuss the reason for the difference. If they are identical, report and claim you're awesome!

REPORT9 : Discuss all trials you've done with either with your implementations or scikit-learn library's functions.