**hw6 linear model (II)**

**郑苏育 经62 2016012602**

1. Given a Gaussian linear regression model, Maximum likelihood estimation of 𝒘 under Gaussian noise assumption is equivalent to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Please prove it.

A: Result:

Proof: In the set D=, we want to use MLE to maximize .Then we assume (discriminative model), we can get = , and when we want to maximize it, it is the same as equation . For simpler calculation, we use the log-likelihood, so the equation we want to maximize become .

For Gaussian distribution, we have , and . So = = . So when we want to do the maximization, it equals .

1. Given a Laplacian linear regression model, Maximum likelihood estimation of 𝒘 under Laplacian noise assumption is equivalent to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Please prove it.

A: Result:

Proof: In the set D=, we want to use MLE to maximize .Then we assume (discriminative model), we can get = , and when we want to maximize it, it is the same as equation . For simpler calculation, we use the log-likelihood, so the equation we want to maximize become .

For Laplacian distribution, we have , and . So = = . So when we want to do the maximization, it equals .

1. Given a linear regression model, please write down the Tikhonov Form and Ivanov Form of Ridge Regression, and these two forms of Lasso Regression as well.  
   A: Ridge Regression:

Tikhonov:

Ivanov:

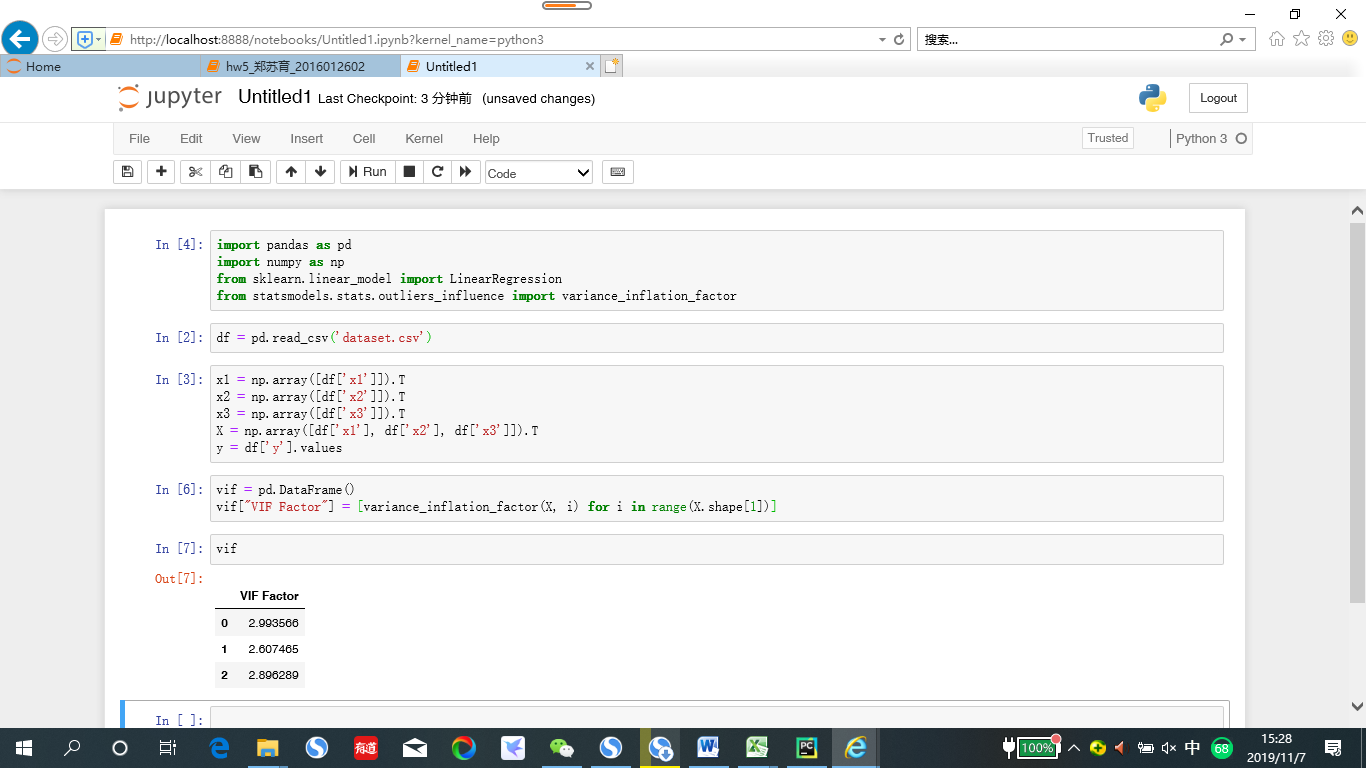
Lasso Regression:

Tikhonov:

Ivanov:

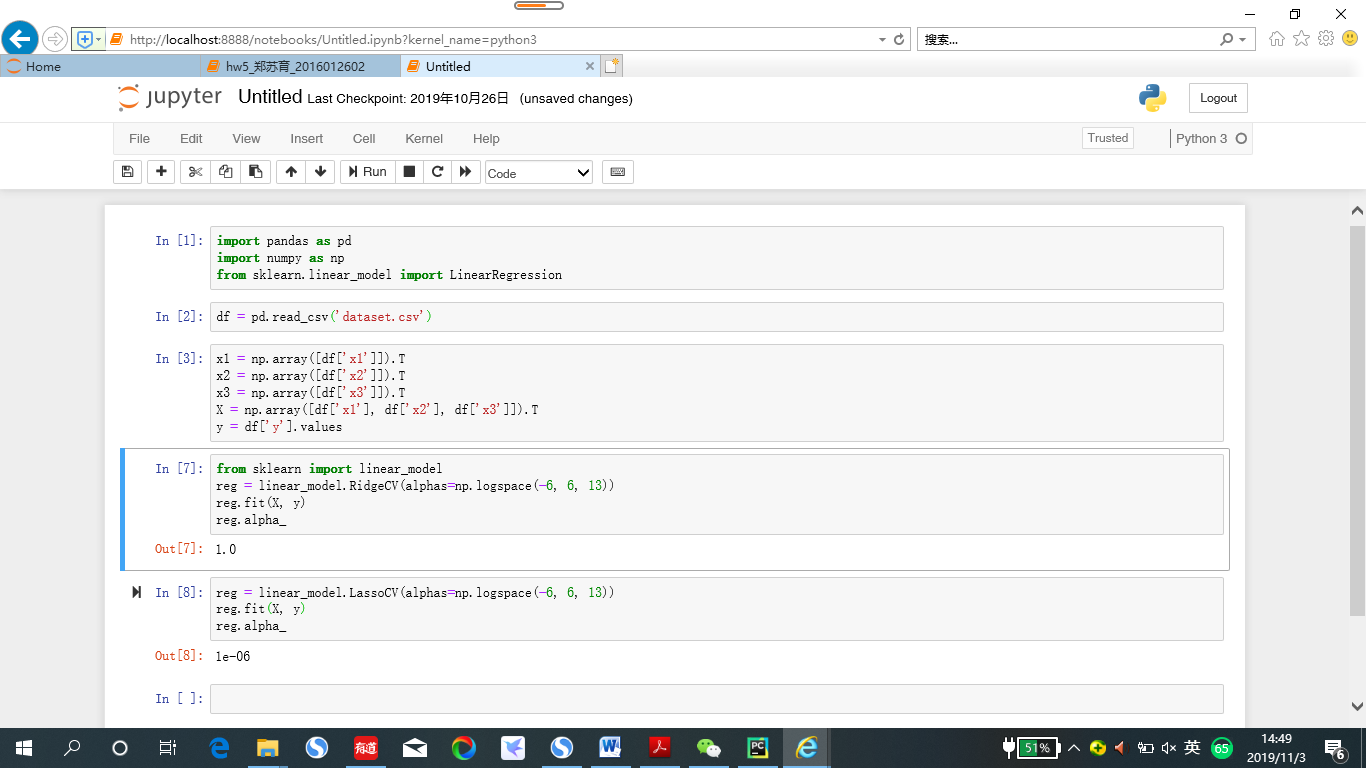
1. By adding a Ridge Regression in the linear regression model of *Question 4 in hw5-linear-model,* can we get a lower generalization error? If yes, use cross validation to attain the best regularization parameter 𝜆, whose possible values are [1.e-06, 1.e-05, 1.e-04, 1.e-03, 1.e-02, 1.e-01, 1.e+00, 1.e+01, 1.e+02, 1.e+03, 1.e+04, 1.e+05, 1.e+06]. If no, please explain why. See the tutorial of linear model in sklearn: <https://scikit-learn.org/stable/modules/linear_model.html> if you need some help.

A：我认为 Ridge regression 是用来解决多重共线性导致的过拟合问题的，所以我使用了计算VIF的方法来判断x1,x2,x3之间的相关性。



可以看到三个变量的VIF都不大（一般大于10才算大），所以使用Ridge或Lasso效果可能都不会有太好的效果，但我仍然使用这两种模型进行了回归。

Best 𝜆 is 1.0.



1. By adding a Lasso Regression in the linear regression model of *Question 4 in hw5-linear-model,* can we get a lower generalization error? If yes, use cross validation to attain the best regularization parameter 𝜆, whose possible values are [1.e-06, 1.e-05, 1.e-04, 1.e-03, 1.e-02, 1.e-01, 1.e+00, 1.e+01, 1.e+02, 1.e+03, 1.e+04, 1.e+05, 1.e+06]. If no, please explain why. See the tutorial of linear model in sklearn: <https://scikit-learn.org/stable/modules/linear_model.html> if you need some help.

A：Best 𝜆 is 1e-06.

