

Stat406 Spring 2010: homework 2

1 Ridge regression using SVD

Exercise 11.5 from book (p345).

2 Ridge regression with diagonal prior

Exercise 11.6 from book (p346).

3 Linear and ridge regression on prostate cancer data (Matlab)

Consider the prostate cancer dataset discussed in [HTF01]. There are 8 continuous inputs and 1 continuous response, namely *lpsa*, which stands for log of prostate-specific antigen. The (standardized) data is in the file `prostate.mat` which contains the following variables (amongst others)

Listing 1: :

| Name | Size | Bytes | Class | Attributes |
|--------|------|-------|--------|------------|
| Xtest | 30x8 | 1920 | double | |
| Xtrain | 67x8 | 4288 | double | |
| names | 1x9 | 624 | cell | |
| ytest | 30x1 | 240 | double | |
| ytrain | 67x1 | 536 | double | |

1. Fit a simple linear model $\hat{y}(x) = w_0 + w_1x_1 + \dots + w_8x_8$ by maximum likelihood on the training set. What coefficients w do you get? What is the mean squared error and its standard error on the test set? Turn in your numbers and code. (You should get the same results as Table 1.)

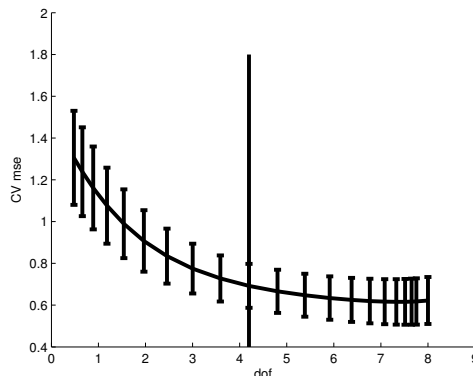


Figure 1: Cross-validation error vs dof for ridge regression on the prostate cancer data. [Based on Figure 3.6 of [HTF01].] Produced by `ridgeProstateDemo`, which is part of Exercise 3.

| Term | LS | Ridge |
|------------|--------|--------|
| intercept | 2.480 | 2.472 |
| lcavol | 0.676 | 0.366 |
| lweight | 0.303 | 0.228 |
| age | -0.141 | -0.021 |
| lbph | 0.209 | 0.151 |
| svi | 0.304 | 0.207 |
| lcp | -0.287 | 0.039 |
| gleason | -0.021 | 0.044 |
| pgg45 | 0.266 | 0.117 |
| Test MSE | 0.586 | 0.541 |
| Std. error | 0.184 | 0.170 |

Table 1: Coefficients and accuracy of least squares and ridge regression on the prostate cancer data. [Based on Table 3.3 of [HTF01].] Produced by `ridgeProstateDemo`, which is part of Exercise 3.

2. Fit the same model using ridge regression. Use 5-fold CV to select λ from the range `[logspace(3, 0, 20) 0]`. Use the `fitCv` function to compute the CV error and to pick the best model. Plot the CV error vs $df(\lambda)$ and indicate the best value of λ chosen, as in Figure 1. (You can use `dofRidge` to compute `df`.) What coefficients \mathbf{w} do you get? What is the mean squared error and its standard error on the test set? Turn in your numbers, plot and code. (You should get similar results to Table 1.)

References

[HTF01] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.