Statistical Learning, Midterm 1

Bring calculator. Make sure how to use the tables of the cumulative distribution function of the normal distribution.

a) Given the inverse of the matrix X'X and the standard deviation of the error term
- calculate the standard deviation of the least square estimator of a given coordinate of the beta vector
- given the value of this estimator - construct a confidence interval for the respective coordinate of beta, perform a z-test for the significance of this coordinate
- given the true value of beta, calculate the power of the respective z-test.

b) Consider the situation where $p<n$ and the elements of X are iid random variables from N(0,1). Find the expected value of the variance of the least squares estimates of individual components of beta.

c) Given a set of p-values perform the multiple testing procedure using Bonferroni or Benjamini-Hochberg multiple testing procedure.

d) Find the expected value of the number of false discoveries when individual tests are performed at the significance level alpha.

e) Given the result of the multiple testing and the indicators of true hypothesis, calculate the False Discovery proportion.

f) Given RSS and the standard estimation of the error term, use SURE to estimate the prediction error.

g) Given RSS for different regression models use AIC, BIC, RIC etc to identify the 'best' model.

h) Given residuals and the elements on the diagonal of the projection matrix H calculate RSS and the cross-validation estimator of the prediction error.

i) Assuming the X'X=I calculate the expected number of false discoveries for any of the considered model selection criteria.

j) When would you use AIC ? BIC ? RIC ? mBIC ? mBIC2 ?

k) Given RSS and the eigenvalues of the inverse of X'X use SURE to estimate the prediction error of ridge regression.

l) Assuming the X'X=I and given values of the least squares estimator provide values of ridge and LASSO estimators.

m) Given X'X = I calculate MSE of the ridge estimator.

n) Given X'X=I calculate the expected value of false discoveries and the power of LASSO

estimator.

o) Be able to verify the irrepresentability condition for LASSO and understand what it implies.

p) Given RSS, the number of variables selected by elastic net (LASSO) [and eigenvalues of the inverse of X_A'X_A], estimate the prediction error of elastic net (LASSO).

q) When would you prefer elastic net over LASSO ? What are the main differences between these methods ?

r) Why do LASSO and elastic net perform variable selection and ridge regression does not ?