Statistical Learning, Midterm 1

Midterm 1 covers Lectures 1-5, Assignments 1, 2 and Problems 1-5 from Assignment 3.

Below you can find some possible midterm problems, but the midterm does not need to be restricted to
these topics.

a) Understand the spectral decomposition of the symmetric matrix. Find the eigenvalues and the
eigenvectors. Apply the spectral decomposition to find the powers, the square root and the inverse
of the symmetric matrix.

b) Multivariate normal distribution and the respective marginal and conditional distributions.
Predict the value of some coordinates given the remaining ones. Be able to construct the
corresponding predictive intervals. Find the distribution of the linear combination of multivariate
normal vectors.

c) Understand the relationship between the following distributions: normal, chi-square, Student, F
and their limits when the number of degrees of freedom diverges to infinity.

d) Perform the Hotelling T-test for the hypothesis about the vector of means for the multivariate
normal distributions.

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

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

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

h) Given the maximum likelihood estimator of the vector of means, the standard deviation of the error term and the norm of the observation vector or its sample variance, calculate the corresponding version of the James Stein estimator.

i) What are the properties of the James-Stein estimator ? Is it unbiased ? How does its variance and the mean squared error compare to the respective characteristics of the maximum likelihood estimator ?

j) What is the value of the mean squared error of the maximum likelihood estimator

of the mean vector for the multivariate normal distribution ?

k) What is the value of the prediction error for the multiple regression model fitted by the least squares method ?

l) Given RSS, the number of explanatory variables and the standard estimation of the error term,
estimate the prediction error.
m) Given RSS for different regression models use AIC, BIC, RIC etc to identify the ‘best’ model.
n) Given residuals and the elements on the diagonal of the projection matrix H calculate RSS and
the cross-validation estimator of the prediction error.
o) Assuming the X’X=I calculate the expected number of false discoveries for any of the considered
model selection criteria.
p) When would you use AIC ? BIC ? RIC ? mBIC ? mBIC2