Problem set 5
Stat 154, SP 21; N. El Karoui
Due date: 4/22/21, 11:59pm

Problem 1

HTF 5.7. Add the following question as question d)
d): Replace the squared error loss in question c) by logistic loss and solve the corresponding problem.

Problem 2

HTF 5.15: Questions c) and d) **You can take a and b for granted**

Problem 3

HTF Problem 5.16

Problem 4

1. Prove the statements made in Table 12.1 of HTF (p.427)

2. **optional**Is using the hinge loss somewhat less informative than using other losses?

3. **[optional]** In light of our discussion in class, characterize the set of functions $L(Y, f(x))$ which make sense from the point of you of classification. Be as general as you can. (Suppose we are working with two classes, coded ± 1. For simplicity, you can limit yourselves to functions of the type $L(y, f(x)) = \phi(yf(x))$.)

Problem 5

optional

1. Use Support vector machines to analyze the south african heart-disease data. Describe your analysis and results.

2. Use a random cosine basis that mimicks the kernel you used in the SVMs analysis you did in the first part. Discuss briefly the impact of the number of basis on predictive performance. If possible look into in the question of speed gain obtained by using random cosine basis.