Homework # 1: Concepts of Statistical Learning (Chap. 1-2)

Due January 30 by 12:00 noon, uploaded on Blackboard $\Rightarrow$ Contents $\Rightarrow$ Homework Assignments.
Quiz 1 is on January 31

1. (#2-1, p.52) For each of parts (a) through (d), indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.

   (a) The sample size n is extremely large, and the number of predictors p is small.
   (b) The number of predictors p is extremely large, and the number of observations n is small.
   (c) The relationship between the predictors and response is highly non-linear.
   (d) The variance of the error terms, i.e. $\sigma^2 = \text{Var}(\varepsilon)$, is extremely high.

2. (#2, p.52) Explain whether each scenario is a classification or regression problem, and indicate whether we are most interested in inference or prediction. Finally, provide n and p.

   (a) We collect a set of data on the top 500 firms in the US. For each firm we record profit, number of employees, industry and the CEO salary. We are interested in understanding which factors affect CEO salary.

   (b) We are considering launching a new product and wish to know whether it will be a success or a failure. We collect data on 20 similar products that were previously launched. For each product we have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.

   (c) We are interesting in predicting the % change in the US dollar in relation to the weekly changes in the world stock markets. Hence we collect weekly data for all of 2012. For each week we record the % change in the dollar, the the % change in the British market, and the German market.

3. (#2-4, p. 53) You will now think of some real-life applications for statistical learning. In each example, describe the response and the predictors and state the goal - inference or prediction.

   (a) Describe two real-life applications in which *classification* might be useful.
   (b) Describe two real-life applications in which *regression* might be useful.
   (c) Describe two real-life applications in which *cluster analysis* might be useful.