Statistical Learning - Milan, Fall2019

Homework problem 7

Short intuition problems

Answer and explain *very* briefly. If you need additional assumptions to reach your conclusion, specify them.

1. If I have a very large amount of data in a reasonably low dimension (very large n, small p), which regression method would be most likely to minimize my model's EPE?
   (a) Linear regression
   (b) 1-NN (i.e., nearest neighbor with 1 neighbor)
   (c) k-NN with $k = \log(n)$
   (d) k-NN with $k = \sqrt{n}$
2. Same question for two-class classification (explain the difference, if any)
   (a) Logistic regression
   (b) LDA
   (c) 1-NN (i.e., nearest neighbor with 1 neighbor)
   (d) k-NN with $k = \log(n)$
   (e) k-NN with $k = \sqrt{n}$
3. If I believe that only a small number of my variables are important, which one (or more) of these four regularization approaches should I use?
   (a) Ridge
   (b) Lasso
   (c) Variable selection
   (d) PCA regression
4. Same question, except that now I believe that only a low-dimensional linear subspace of the span of my variables is important.
5. We are given a problem (like say our Netflix example) with training set of size 8000 and test set of size 2000. Ruth and Naomi each build a regression tree on this training data, but each uses a different half of the variables/columns (say, Ruth uses the even ones and Naomi the odd ones). They both use all 8000 observations/rows. Obviously the trees look completely different, as the two trees never use the same variables. When applying their models to the test set, they are surprised to find out that the predictions are very similar, say with correlation 0.98 between them. Ruth says "Uh-oh, this makes no sense we must have a bug", and Naomi replies:"Actually, it means that two different models reach the same conclusions, so our models must be very accurate!". Which one of them is correct and why? Note they may both be correct, or both wrong.
6. We mentioned in class that in bootstrap sampling, used in Bagging and Random Forest, the probability of each observation to get selected zero times is about $1/e$. What is the corresponding probability of each observation to get selected exactly once?