Dear students,

I have some information about the exam in Machine learning that we had recently. First, I want you to know that me and Jose correct it differently: I put my comments and highlight your mistakes in the exam report without sending correct answers while Jose only reports the points for each assignment but provides you with a correct solution instead. We are also open for questions about your corrected report if you have such questions.

I have already corrected a large portion of the exam reports and these are the typical mistakes that I detected for “my assignments”:

TDDE01

Assignment 1:

- ID column used as feature
- Family=”binomial” is not specified in glm – serious mistake
- Regression tree is used instead of classification, i.e. as.factor(Class) forgotten
- Combination of classifiers is wrongly constructed. We discussed this example in the lectures, i.e. if one classifier says P(Y)=0.45 and another classifier says P(Y)=0.9 -> you can get a combined decision by averaging the probabilities, i.e. (0.45+0.9)/2 (unless you know apriori that one classifier is better than another) . This should of course be done per prediction point

Assignment 2:

Just a comment: we solved exactly same kind of problem at the lectures on whiteboard.

732A99

Assignment 1:

- ID column used as feature
- Family=”Binomial” is not specified in glm – serious mistake

Assignment 2:

- Wrong basis functions used in lm() model. In Hasties book, there is a clear description of what the basis fuctions should be for an order-M spline.

Assignment 3:

- Set.seed forgotten before pamr.cv(). Note that according to the exam instruction you were supposed to use set.seed before every piece of code containing randomness.
- Inability to interpret the centroid plot correctly, i.e. misspecifying what positive and negative numbers mean (again, Hasties book describes it)