STA 314, Review problems 7

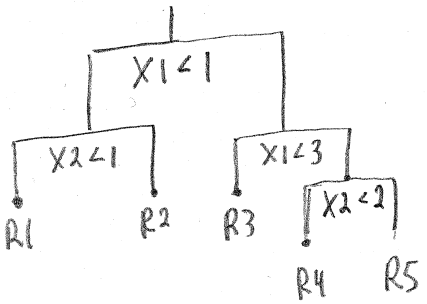
1. Which of the following will lead to less flexible models?
   (a) Increasing the number of leafs (aka terminal nodes) in a regression tree.
   (b) Pruning large trees.

2. Which are true for bagging?
   (a) Bagging can help to improve the variance of predictions.
   (b) Bagging will not reduce the variance of predictions but can help to improve the bias of predictions.
   (c) Bagging will never lead to better prediction performance compared with just growing a single large tree.

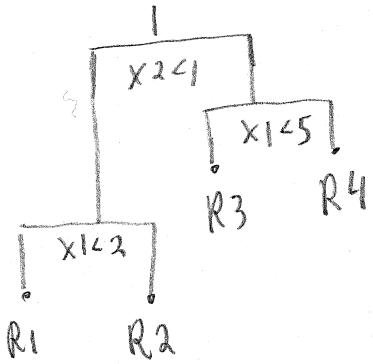
3. Which of the following statements are true for boosting?
   (a) A slower learning rate means that less trees are required to obtain a good predictive model.
   (b) Using too many trees can lead to models that overfit the data.
   (c) Using interaction depth 1 can will lead to additive models, i.e. models of the form
   $$f(X) = f_1(x_1) + f_2(x_2) + ... + f_d(x_d).$$
   (d) Choosing a learning rate close to 1 will usually lead to models with good predictive performance.
   (e) Interaction depth should always be chosen as large as possible.
   (f) Choosing a smaller learning rate will require less computation to obtain a good model.

4. For each of the trees on the next page, draw the corresponding partition of the predictor space and mark the corresponding regions. For each partition of the predictor space, decide if it corresponds to a tree and draw the corresponding tree (if possible) annotating the leafs with corresponding regions.

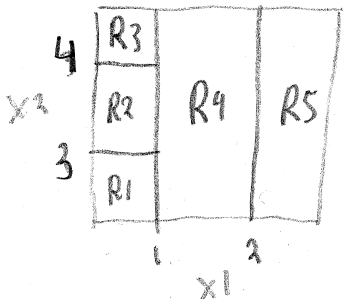
①


X1<1
X2<1
X1<3
X2<2
R1
R2
R3
R4
R5

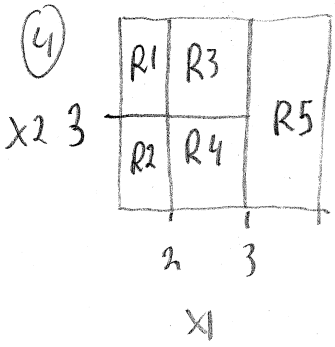
②


X2<1
X1<5
X1<2
R3
R4
R1
R2

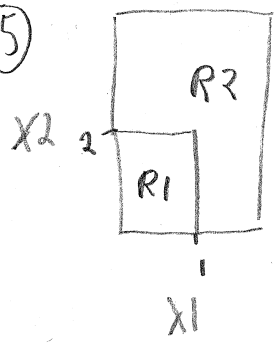
③


4
3
R3
R2
R1
R4
R5
1
2
X1

④


X2 3
R1
R3
R2
R4
R5
2
3
X1

⑤


X2
2
R2
R1
1
X1