STATS 784 Data Mining: Possible questions for test

NEURAL NETWORKS
1. What is the basic function being fitted in a neural network?
2. How are starting values chosen?
3. What problems might you encounter when fitting a neural network?
4. What does LINOUT do?
5. What is the skip layer?
6. If you run the same model with the same data, will you always get the same answer?

PPR/GAMS
7. What is the basic idea of PPR?
8. How does R choose the number of projections?
9. What is the difference between gams and PPR?
10. Can PPR handle factors? How?
11. What is backfitting?

MARS
12. What are basis functions?
13. What basis functions are used in MARS?
14. How does the software decide which basis functions to fit?

Trees
15. What is recursive partitioning?
16. What function is being fitted by a tree?
17. How do we decide how to split?
18. How does pruning work?
19. What are the disadvantages of trees?
20. How does using a tree for prediction differ from using a tree for classification?

Regularisation, Boosting and Bagging

1. What is the basic purpose of regularisation?
2. What is the difference between ridge regression and the lasso?
3. What packages would you use?
4. Are ridge estimates biased? If so why would we use them?
5. Why does the lasso zero coefficients but not ridge?
6. What is the difference between boosting and bagging?
7. What extra wrinkles (over basic bagging) do random forests incorporate?
8. What is the significance of the argument mtry in the R function randomForest?
9. How do we measure variable importance in random forests?