

Oral Questions 2015: STAT 760

Name: _____

Plotting

1. What is the difference between the `q` function in `ggplot2` and the `ggplot` function?
2. What are aesthetics?
3. How is the trellis idea implemented in `ggplot2`?
4. What is the main difference between `ggplot2` and standard R graphics?

Linear models

1. What is a contrast matrix in R?
2. Suppose we have a simple model $y \sim A$ where A is a factor. How does R build up the model matrix when we are using sum contrasts?
3. What numerical method does R use to solve the normal equations?
4. Explain interaction between a continuous variable and a factor in terms of lines (or planes).
5. How does R decide if a variable is a factor when creating a data frame?
6. If we have two factors and one continuous variable in a model, what does the interaction between the two factors refer to?

Data mining

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 is the connection between AdaBoost and forward stagewise modelling?
8. What extra wrinkles (over basic bagging) do random forests incorporate?
9. What is the significance of the argument mtry in the R function randomForest?
10. How do we measure variable importance in random forests?

Survival Analysis

1. What is the survival function?
2. What is the hazard function?
3. What is the difference between the accelerated failure time model and the Cox proportional hazards model?
4. How does the Kaplan-Meier estimate work? What does it estimate?
5. In the output from a Cox model, the estimated beta coefficient for a variable x is positive. Is an increase in x associated with an increase or a decrease in survival?
6. In the output from an accelerated failure time model, the estimated beta coefficient for a variable x is positive. Is an increase in x associated with an increase or a decrease in survival?

Spatial statistics

1. What is a trend surface?
2. What is the downside of using OLS to fit trend surfaces?
3. How do we estimate covariance functions?
4. How do we fit a model using GLS?

5. Can you give an example where correlation might not be a function of distance alone?
6. What is kriging?
7. How do you define a Poisson process?
8. What is the difference between a homogeneous and inhomogeneous Poisson process?
9. Do Poisson processes allow for attraction between points?

Time series

1. What does stationarity mean?
2. When is an AR(1) process stationary?
3. What is the difference between an AR process and an ARMA process?
4. What does the "I" in ARIMA mean?
5. What feature of a time series is captured in a GARCH model?
6. How is this incorporated in the model?
7. What is the periodogram?
8. What does it mean when the periodogram is flat?
9. Suppose a time series is essentially a sine wave (perhaps with some random noise added.) What will the spectrum look like?