Midterm, CS 595, Summer I 2015

Study Guide

1. Before the creation of computing technology statistical methods were limited to simple models. Describe two models for statistical learning that are practically impossible without computing. Why are linear models desired ?
2. Fill in blanks with numbers from list in the box at right. The formalism used in the test conceptualizes a problem as one where we are trying to learn a relationship
3. Closed form
4. Dependent variable
5. Error
6. Features
7. Independent variable
8. Irreducible
9. Prediction
10. Reducible

Y = f (X) + ε

where X = (x1, x2, … xn). Y is \_\_\_ , X is \_\_\_\_ , ε is termed

\_\_\_\_ and the xi  are ­­­\_\_\_\_\_ . We wish to estimate the

function f as f’ with two purposes in mind:

\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_ .

We may wish to describe f with some \_\_\_\_\_\_ or simply make \_\_\_\_\_\_\_

with some accuracy which is controlled by \_\_\_\_ and \_\_\_\_ error.

1. Why is some error irreducible ?
2. The irreducible error of the expected value of squared error of our estimate can be formalized as … ?
3. What are some of the *inference* questions we might ask about our estimate of f ?
4. Distinguish between parametric and non-parametric methods to estimate f.
5. Why does model interpretability increase as accuracy decreases ? Why would we choose a less accurate model ?
6. Give an example of a ‘qualitative’ (vs quantitative) dependent variables. Why is linear regression inappropriate for qualitative dependent variables ?
7. What is the typical measure of ‘fit’ for a supervised model that has been trained ?
8. The ‘flexibility’ of a model increases as the fit of test data decreases. Explain.
9. What is the ‘bias’ of a model ? How do bias and variance relate to reducible and irreducible error ?
10. True of False ?
    1. A model has variance
    2. Data has variance
    3. A model has bias
    4. Data has bias
11. What is the error measure for supervised learning based on a classification problem ?
12. Describe the K-nearest neighbor algorithm. What are we optimizing ?
13. Describe the K-means algorithm for clustering.
14. What is the objection that Bayesians have with ‘frequentists’. Why have Bayesian statistics come into such pre-eminence ?
15. Suppose there are 3 Classes, C1, C2, and C3 with frequencies as shown. Assume further that each class has frequencies for words w1,w2,w3, w4 as shown. Build a Naïve Bayes Classifier from these data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Freq | w1 | w2 | w3 | w4 |
| C1 | 8 | 3 | 5 | 2 | 7 |
| C2 | 6 | 5 | 1 | 2 | 3 |
| C3 | 9 | 4 | 6 | 3 | 5 |

1. a) What is a joint probability ?

b) What is a conditional probability ?

c) State Bayes law.

d) What is the ‘naïve’ assumption of a Bayes Classifier ?

19) Lab Possibilities for Exam

Be able to take some simple data and create a plot in R that will (for example)

a) show the features in a linear regression model that are most important

b) show that clustering can significantly vary if features discovered as above are used for the clustering

Generally, understand the R assignments you have been asked to do and be able to imitate that same work.

20) Why is it not overwhelming evidence for conviction based solely on a positive DNA test ?