Gabriel West (notes)

Chapter 3: Linear Regression

Lecture 1

- Simple is better
- Single dimension linear regression is finding the line that best represents the underlying relationship between predictor variables and response variables.
- Least squares minimization is used to find the unique ‘best’ beta parameters.
- Standard Error is used to describe how well a line fits the data, it involves the standard deviation of the data, as well as the spread of the x values (good spread means the like has less slop to pivot upon new data arrival)
- Confidence Intervals, closely related to hypothesis testing, give the interval for the slope where we are 95% sure the true slope lies.

Lecture 2

- Hypothesis Testing
    - Confidence

- t statistic
- p value
- assessing overall accuracy
- R^2 is how much we reduced TSS with our estimation
- Always remember the domain when evaluating how good an R^2 is

Lecture 3 Multiple linear regression

- Line becomes a hyperplane
- Looking at the effect of multiple predictors
- Trouble of interpreting beta estimations
    - If predictors are correlated, how can we make a statement of causality?
    - You have to do a controlled perturbation to truly make a statement of causality

- Predictors can be heavily correlated, resulting in the effects of one being ‘soaked up’ by the other in the model. This doesn’t necessarily mean that one is more

important than the other.

##Lecture 4 Model Selection and Qualitative Predictors - Are any predictors useful? Which ones? How well do they predict? - F-Statistic - Tells you how well the model predicts - Deciding on important variables is done with an automated search approach, because the number of models is exponential on the number of parameters - Forward Selection - Start with the null model - Add the single (simple) linear model that results in the lowest overall RSS - Continue until stopping condition - Backward Selection - start with all variables - remove variable that does the 'least damage'/least significance according to the t statistic - repeat until stop condition - Qualitative predictors - aka categorical or factor predictors - not ordered - Use dummy varialels to define a base, and a deviation from the base - RSS (fit) is the same no matter which is baseline, though the contrasts change

Lecture 5 Interactions and non linearity

- Interaction/synergy
- Make the product of two variables a new variable
- Hierarchy principle: Leave the primary features in when you do interactions, even if they have a large p value
    - why? interpretability
    - can be done with qualitative as well

- Nonlinearity
    - To model nonlinear predictors, make a variable squared, and fit a linear model with a coefficient for that new variable. This can be done to arbitrary degree, though you run the risk of overfitting.

- Didn't cover:
    - Outliers, Non-constant variance of error terms, high leverage points, Collinearity

- Discussion of future topics ##Lecture 6