Midterm study guide

MATH 456

Navbar: [Schedule][Week 7 Overview]

Linear Regression Model Building

1. What is the purpose of the model?
    a. Predictive
    b. Interpretation
2. How do we build the model?
    a. Variable selection
    b. Splines and non-linear terms
3. Assess model fit
    a. Assumptions
    b. R^2, RMSE / CV error
    c. Effect of outliers and influential points
4. Report
    a. Final model
    b. Estimates (if interpreting) in context
    c. RMSE & R^2

Things to look out for with Regression

1. Violation of assumptions
2. Nonsencical/blind/uninformative variable selection
3. Multicolinearity
4. Outliers
5. Poor model fit

Model assumptions

1. Linearity between Y and X.
2. Residuals: $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

Variable selection procedures

1. Pro's and cons for each method
    a. Manual selectio
    b. Forward/Backward/Stepwise selections
    c. Stepwise selection
2. Measures of fit

    a. Information criteria: AIC/BIC
    b. R-squared
    c. Cross-validation (prediction/RMSE) error

Outliers

1. Measures of influence
    a. DFFITS
    b. Leverage
    c. Cooks Distance
    d. Studentized residuals
2. Identifying outliers
    a. Residual plots
    b. Determining outliers in Y
    c. Determining outliers in X
    d. Determining influential points
3. Deciding what to do with them
    a. Removal? Needs justification.
    b. Assess impact of removal.