

ISYE6740 Midterm
Kinds of Questions to Expect

1. Simple Linear Regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

where ϵ_i is a random variable.

- (a) What are the usual assumptions about ϵ_i ?
- (b) Derive the least squares estimators of β_0 and β_1 .
- (c) How would you test the hypothesis that $\beta_1 = 0$?
(Tell what quantities you would use, but you do not need to show the exact formulas.)
- (d) If r_0 is the residual at the point x_0 , what is the variance of r_0 ?
- (e) Suppose in the application of this model, there are two different groups of observations, say male and female.
 - i. Write a model for the situation in which we think the x affects the y the same in both groups.
 - ii. Write a model for the situation in which we think the x affects the y differently in the two groups.

2. Logistic Models

- (a) What is the response in a logistic model?
- (b) Why do we use an expression of the form $\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$?

3. KNN Methods

- (a) Use of a logistic model is a global method; use of KNN is a *local method*. In this context, what is the difference in a global method and a local method?
- (b) Use a sketch of a plot of 10 pairs of points (x, y) , and show how you would do KNN regression of y on x .

4. LDA and QDA

- (a) What is the underlying rationale of LDA and QDA?
- (b) Why do we need to assume a particular distribution in LDA and QDA? (That is, the the development of the method where and why do we make an assumption about some distribution?)
- (c) How do the assumptions about an underlying distribution differ in LDA and QDA.
- (d) Derive the QDA decision boundary for a classification problem with one response y and one predictor x .