MA 751
M. Kon

Problem Set 3
Due Thurs. 2/17/22

Lectures 6, 7

Reading: Sections 4.1-4.3,

Scheduling: Note there will be no class next Tuesday Feb. 15 (Monday Schedule). We will not need to make adjustments to the discussion sections - the Wednesday discussion section will not be affected, and we will keep the Monday discussion section (now on Tuesday).

Notes: We will start with classification using linear discriminant analysis. Using some natural assumptions this method divides the feature space into regions, each of which corresponds to a unique predicted class. We will then compare it to another method with a similar purpose but very different thinking, logistic regression.

Problems:

1. Hastie, problem 4.2

2. In the discussion of section 4.2, consider a dataset $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{24}$ with 4 categories ($k = 1, ..., 4$), in which the category $\mathbf{y}_i$ of data point $\mathbf{x}_i$ is i mod $4 + 1$ (i.e., the category alternates successively to the next one in the group of 4).
(a) What does the data matrix $\mathbf{Y}$ look like?
(b) Let $\mathbf{y}$ be the first column of $\mathbf{Y}$. If we replace the data matrix $\mathbf{Y}$ by $\mathbf{y}$, show that the vector $\widehat{\beta}$ for this standard regression is just the first column of the matrix $\widehat{\mathbf{B}}$.
(c) Generalize this statement to all columns of $\widehat{\mathbf{B}}$.
(d) What does this tell you that the regression using $\mathbf{Y}$ is doing? Can it be interpreted precisely in terms of the individual regressions arising from replacing $\mathbf{Y}$ by one of its columns?

3. Derive equation (4.9)