Machine Learning - Assignment II
Linear regression

Souhaib BEN TAIEB

23 March 2020

Note: It is strongly recommended to read Sections 3.1 to 3.4 in ISLR. Also, please provide justifications for every step you take in your derivations.

Question 1

Do Exercise 7 in Chapter 3.7 in ISLR.

Question 2

Read Section 3.6.2 in ISLR.

Do Exercise 8 in Chapter 3.7 in ISLR. For 8(c), look only at the plot which gives the residuals vs fitted values.

Question 3

Do Exercise 9 (a), (b), (c), (e) and (f) in Chapter 3.7 in ISLR.

Question 4

Read Section 3.5 in ISLR.

Let us assume we observe a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ where the x_i are fixed (not random) and $y_i = f(x_i) + \varepsilon_i$ with $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = \sigma^2$.

If we let the kth nearest neighbor of x_0 be x_k for $k = 1, 2, \ldots, K$ and $K \geq 1$, the prediction of the KNN regression algorithm for a point x_0 can be written as

$$\hat{f}(x_0) = \frac{1}{K} \sum_{i=1}^{K} y_i.$$

(a) What is the squared bias?

(b) What is the variance?

(c) What is the mean squared error?

TURN IN

- Your `.Rmd` file (which should knit without errors and without assuming any packages have been pre-loaded)
- Your pdf file that results from knitting the Rmd.
- DUE: April 19, 11:55pm (late submissions not allowed), loaded into Moodle