UNIVERSITY OF TEXAS AT AUSTIN

HW Assignment 6

Prep for In-Term One.

Please, provide your **complete solutions** to the following problems. Final answers only, even if correct will earn zero points for those problems.

Problem 6.1. ($4 \times 3 = 12$ points) Solve Problem **2.4.2** from the textbook (pp.52-53).

Solution:

(a) $n = 500, p = 3$, regression, inference
(b) $n = 20, p = 13$, classification, prediction
(c) $n = 52, p = 3$, regression, prediction

Problem 6.2. ($4 + 4 = 8$ points) Solve Problem **2.4.4 (a) and (b)** from the textbook (p.53).

Solution: Solutions will vary!

Problem 6.3. ($3 + 2 + 2 + 5 = 12$ points) Solve Problem **2.4.7** from the textbook (p.54).

Solution:

(a)

$$1: \sqrt{3^2} = 3$$
$$2: \sqrt{2^2} = 2$$
$$3: \sqrt{1^2 + 3^2} = \sqrt{10} = 3.162278$$
$$4: \sqrt{1^2 + 2^2} = \sqrt{5} = 2.236068$$
$$5: \sqrt{1^2 + 1^2} = \sqrt{2} = 1.414214$$
$$6: \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3} = 1.732051$$

(b) *Green*, because observation #5 is the closest
(c) *Red*, because observations #2 and #6 join observation #5 in the neighborhood
(d) **Small** is better because its more local.

Problem 6.4. ($4 \times 2 = 8$ points) Solve Problem **3.7.4** from the textbook (p.122).

Solution:

(a) The RSS for the cubic should be lower for the training set.
(b) The RSS for the cubic will potentially be higher for the test set due to overfitting.
(c) Again, the RSS for the cubic should be lower for the training set.
(d) We cannot conclude anything here since the nature of nonlinearity is not provided.

Problem 6.5. (10 points) Solve Problem **4.8.1** from the textbook (p.189).

Solution:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 x} \quad \Leftrightarrow \quad p(X) = (1-p(X))e^{\beta_0+\beta_1 x}$$
$$\Leftrightarrow \quad p(X) = e^{\beta_0+\beta_1 x} - p(X)e^{\beta_0+\beta_1 x}$$
$$\Leftrightarrow \quad p(X)(1+e^{\beta_0+\beta_1 x}) = e^{\beta_0+\beta_1 x} \quad \Leftrightarrow \quad p(X) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$