Stat 627/427 (Statistical Machine Learning)

Homework # 9: Polynomials and Splines (Chap. 7, 7.1–7.5)

Due April 10 by noon on Blackboard. Quiz #7 is on April 11.

1. (**Chap. 7, # 3, p. 298**) Suppose we fit a curve with basis functions $b_1(X) = X$ and $b_2(X) = (X-1)^2 I\{X \geq 1\}$. We fit the regression model

$$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \varepsilon$$

and obtain the estimated slopes $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = -2$. Sketch the estimated curve between $X = -2$ and $X = 2$. Report the intercepts, slopes, and other relevant information.

2. (**Chap. 7, # 4, p. 298**) Repeat the previous exercise with basis functions $b_1(X) = I\{0 \leq X \leq 2\} - (X-1)I\{1 \leq X \leq 2\}$ and $b_2(X) = (X-3)I\{3 \leq X \leq 4\} + I\{4 < X \leq 5\}$ and estimated slopes $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = 3$.

3. (**Chap. 7, ≈# 8, p. 299**) Apply some of the non-linear models discussed in this chapter to the **Auto** data set to predict the vehicle's **acceleration** time based on the **horsepower** of its engine.

   (a) Use cross-validation to select the optimal degree for the polynomial regression.

   (b) Looking at the scatterplot of acceleration vs horsepower, choose some knots and fit a regression spline.

   (c) Fit a smoothing spline, selecting the smoothing parameter by cross-validation.

   For each method, make a plot of the resulting fitted line, and estimate its prediction mean-squared error by some cross-validation technique. Which approach resulted in the best prediction accuracy?

4. (**For Stat-627 only... Chap. 7, # 2, p. 298**) Suppose that a curve g is computed to smoothly fit a set of n points using the following formula

$$\hat{g} = \arg\min_g \left\{ \sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int \left[g^{(m)}(x) \right]^2 dx \right\},$$

where $g^{(m)}$ is the m-th derivative of g (and $g^{(0)} = g$). Provide example sketches of $\hat{g}$ in each of the following scenarios.

(a) $\lambda = \infty$, $m = 0$. (b) $\lambda = \infty$, $m = 1$. (c) $\lambda = \infty$, $m = 2$.
(d) $\lambda = \infty$, $m = 3$. (e) $\lambda = 0$, $m = 3$.

This problem does not require you to take evaluate derivatives or integrals. Recall, however, that $g' = g'' = 0$ for a constant, $g' = const$ and $g'' = 0$ for a linear function $g(x)$, and $g'' = const$ for a quadratic function $g(x)$.