Homework: Ch 07

STAT 4510/7510

Due Tuesday, April 12, 11:59 pm

Problem 1

Let's consider the following data set.

y	x
3.5	0.2
1.2	0.7
6.9	1.2
2.6	3.6
4.4	3.8
3.0	4.0
8.5	4.2

Suppose we try to fit the step function model with two cutpoints $c_1 = 1$ and $c_2 = 3.7$.

$$y_i = \beta_0 + \beta_1 I(1 \leq x_i < 3.7) + \beta_2 I(x_i \geq 3.7) + \epsilon_i$$

where $I(\cdot)$ is an indicator function that returns a 1 if the condition is true, and returns a 0 otherwise.

Find the estimates $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$. You can do this manually. Note that the predicted values of the response should be the mean of y_i's that belong to the corresponding range of x.

Problem 2

Let's consider the following function.

$$f(x) = 3 + 5x + 3x^2 + 1.5x^3 + (x-1)^3_+$$

where $(x-1)^3_+ = (x-1)^3$ when $x > 1$ and $(x-1)^3_+ = 0$ when $x \leq 1$. This function consists of the basis functions $b_1(x) = x, b_2(x) = x^2, b_3(x) = x^3, b_4(x) = (x-1)^3_+$. We will validate this function is the piecewise cubic polynomial that is continuous at the knot point $x = 1$ up to the second derivative.

(a) Find the cubic polynomial function $f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3$ when $x \leq 1$.

(b) Find the cubic polynomial function $f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3$ when $x > 1$.

(c) Validate if $f(x)$ is continuous at $x = 1$, that is $f_1(1) = f_2(1)$.

(d) Validate if $f'(x)$ is continuous at $x = 1$, that is $f_1'(1) = f_2'(1)$.

(d) Validate if $f''(x)$ is continuous at $x = 1$, that is $f_1''(1) = f_2''(1)$.

Problem 3

In this problem, we will use the `Auto` data set. Load the data set onto the global environment by running the following code.

```r
library(ISLR)
Auto <- Auto
attach(Auto)
```

(a) Suppose we are interested in the relationship between the `horsepower` and `acceleration`. Fit the model to predict `horsepower` with polynomial functions of `acceleration` with different degrees from 1 to 3. Compare the models using `anova()` function. Comment on the outcome.

(b) Fit a natural cubic spline model with 4 degrees of freedom. Make predictions on the range of `acceleration`. Plot the data points and predicted curve with 95% confidence interval.

(c) Fit the smoothing spline model with the tuning parameter λ found by cross-validation. Find the corresponding effective degrees of freedom. Plot the data points and predicted curve.

(d) Fit the local regression model with span 0.2 and 0.5. Plot the data points and both curves with the appropriate legend. Comment on two models regarding variance-bias trade-offs.

(e) Let's now consider to use `weight` in addition to `acceleration` to predict `horsepower`. We will use GAM to fit the model. Fit 3 different models as follows and compare them with `anova()` function. Comment on the outcome.

- gam1 - without `weight` and smoothe spline of `acceleration` with 5 dof.
- gam2 - with linear term of `weight` and smoothe spline of `acceleration` with 5 dof.
- gam3 - with smooth spline of `weight` with 5 dof and smoothe spline of `acceleration` with 5 dof.

(f) Using the best model you found from (e), plot the fitted curves for each predictor.