STA 35C: Homework 8

Instructor: Akira Horiguchi

Student name: ABCDE FGHIJ; Student ID: 123456789

Nov 26, 2025 (Wednesday), 22:59 PST

The assignment must be done in an R Markdown or Quarto document. The assignment must be submitted by the due date above by uploading:

- a .pdf file in GRADESCOPE (if you can knit/compile your .rmd to a .html file only, please save the created .html file as a .pdf file (by opening the .html file -> print -> save to .pdf)).

Email submissions will not be accepted.

Each answer has to be based on `R` code that shows how the result was obtained. The code has to answer the question or solve the task. For example, if you are asked to find the largest entry of a vector, the code has to return the largest element of the vector. If the code just prints all values of the vector, and you determine the largest element by hand, this will not be accepted as an answer. No points will be given for answers that are not based on `R`. This homework already contains chunks for your solution (you can also create additional chunks for each solution if needed, but it must be clear to which tasks your chunks belong).

There are many possible ways to write `R` code that is needed to answer the questions or do the tasks, but for some of the questions or tasks you might have to use something that has not been discussed during the lectures or the discussion sessions. You will have to come up with a solution on your own. Try to understand what you need to do to complete the task or to answer the question, feel free to search the Internet for possible solutions, and discuss possible solutions with other students. It is perfectly fine to ask what kind of an approach or a function other students use. However, you are not allowed to share your code or your answers with other students. Everyone has to write the code, do the tasks and answer the questions on their own.

During the discussion sessions, you may be asked to present and share your solutions.

1. Basis functions

(a) Suppose we fit a curve with basis functions $b_1(X) = X$ and $b_2(X) = (X-1)^2 1_{[1,\infty)}(X)$. We fit the linear regression model

$$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \varepsilon, \tag{1}$$

and obtain the coefficient estimates $\hat{\beta}_0 = 1, \hat{\beta}_1 = 1, \hat{\beta}_2 = -2$. Sketch the estimated curve for all $X \in [-1, 2]$ by evaluating it at $X = -1, X = -0.75, X = -0.5, \dots, X = 1.75, X = 2$, and connect the values in a meaningful way.

```
### Your Solution (Code)
```

(b) Suppose we fit a curve with basis functions $b_1(X) = 1_{[0,2)}(X) - X1_{[1,2)}(X), b_2(X) = (X-3)1_{[2,3)}(X), b_3(X) = X^2 1_{[1,3)}(X)$. We fit the linear regression model

$$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \beta_3 b_3(X) + \varepsilon, \tag{2}$$

and obtain the coefficient estimates $\hat{\beta}_0 = 5, \hat{\beta}_1 = 1, \hat{\beta}_2 = 3, \hat{\beta}_3 = 1$. Sketch the estimated curve for all $X \in [-0.5, 3.5]$ by evaluating it at $X = -0.5, X = -0.25, X = 0, \dots, X = 3.25, X = 3.5$ and connect the values in a meaningful way.

```
### Your Solution (Code)
```

2. Applied

ISLR Chapter 7, exercise 6. In this exercise, you will further analyze the `Wage` data set considered throughout this chapter.

(a) Perform polynomial regression to predict `wage` using `age`. Use cross-validation to select the optimal degree d for the polynomial. What degree was chosen? Make a plot of the resulting polynomial fit to the data.

```
### Your Solution (Code)
```

(b) Fit a step function to predict `wage` using `age`, and perform cross-validation to choose the optimal number of cuts. Make a plot of the fit obtained.

```
### Your Solution (Code)
```

3. K-means clustering

Recall the K-means clustering algorithm. Consider the following dataset where $n = 6$ and $p = 2$:

init	x1	x2
1	1	4
1	0	3
2	0	4
1	5	2
2	6	2
2	6	0

Consider the clustering induced by using `init` as the cluster labels. Using this clustering as step 1 of the algorithm, perform each iteration of step 2 of the algorithm until the induced clusters stop changing.