# MCE5901 Homework Set 3

October 22, 2023

**NOTICE:** The homework is due on Nov. 3 (Friday) 11:59pm. **Please draft your solution with steps and provide the codes that you use for the applied problems.** You will need to submit two files to the blackboard: a PDF file including the solutions and results for all problems, and a file containing the codes for the applied problems. You are allowed, and even encouraged, to discuss the homeworks with your classmates. However, you must **write up the solutions on your own**. Plagiarism and other anti-scholarly behavior will be dealt with severely.

Problem 3(g) and Problem 4 are optional. You will receive 20 points bonus if you solve it and the maximum score of the assignment is 100 points.

**Problem 1. Logistic regression and LDA**

An experiment is conducted among students to see the relationship between $X$ = study hours per day and $Y$ = receive a grade A. In the experiment, 120 students are divided into 6 groups and each group has 20 students. We name the groups by *Group A*, *Group B*, *Group C*, *Group D*, *Group E* and *Group F*. The study hours for *Group A* to *Group F* are 0, 1, 2, 3, 4 and 5, respectively. The number of students in each group that receive an A are recoded and given below

| Group | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ |
|---|---|---|---|---|---|---|
| study hours per day | 0 | 1 | 2 | 3 | 4 | 5 |
| receive a grade A | 1 | 3 | 9 | 12 | 17 | 20 |

(a) What is the observed proportion of students who receive a grade A (i.e., $Y = 1$), given that the study hours per day is 3? What is the observed odds of $Y = 1$ at $X = 3$?

(b) We fit a **logistic regression** and produce estimated coefficient $\hat{\beta}_0 = -2.9161$ and $\hat{\beta}_1 = 1.2176$. Give the logistic regression expression for $\hat{Pr}(Y = 1|X = x)$. Predict the posterior probability and log-odds for $X = 5$.

(c) Will $\hat{Pr}(Y = 1|X = x)$ increase or decrease if $X$ increases? Why?

(d) If we increase the $X$ from 4 to 5, by what multiple will the predicted log-odds change? How about the change in the predicted posterior probability? What would be the case if we increase $X$ from 3 to 4?

(e) Now we apply the **linear discriminant analysis** to fit the data. What are the discriminant functions $\hat{\delta}_1(x)$ and $\hat{\delta}_2(x)$? (Give steps how you derive them).

(f) What is the decision boundary? Predict the class when $X = 0, 1, \ldots, 5$. Write down the confusion matrix. What is the average training error rate?

**Problem 2. Cubic spline**

It was mentioned in class that a cubic regression spline with one knot at $\xi$ can be obtained using a basis of the form $X$, $X^2$, $X^3$, $(X-\xi)^3_+$, where$(X-\xi)^3_+ = (X-\xi)^3$ if $X > \xi$ and equals 0 otherwise. We will now show that a function of the form

$$f(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 (X-\xi)^3_+$$

is indeed a cubic regression spline, regardless of the values of $\beta_0, \beta_1, \beta_2, \beta_3$ and $\beta_4$.

(a) Find a cubic polynomial

$$f_1(X) = a_1 + b_1 X + c_1 X^2 + d_1 X^3$$

such that $f(X) = f_1(X)$ for all $x \leq \xi$. Express $a_1, b_1, c_1, d_1$ in terms of $\beta_0, \beta_1, \beta_2, \beta_3$ and $\beta_4$.

(b) Find a cubic polynomial

$$f_2(X) = a_2 + b_2 X + c_2 X^2 + d_2 X^3$$

such that $f(X) = f_2(X)$ for all $x > \xi$. Express $a_2, b_2, c_2, d_2$ in terms of $\beta_0, \beta_1, \beta_2, \beta_3$ and $\beta_4$. We have now established that $f(x)$ is a piecewise polynomial.

(c) Show that $f_1(\xi) = f_2(\xi)$. That is, $f(x)$ is continuous at $\xi$.

(d) Show that $f_1'(\xi) = f_2'(\xi)$. That is, $f'(x)$ is continuous at $\xi$.

(e) Show that $f_1''(\xi) = f_2''(\xi)$. That is, $f''(x)$ is continuous at $\xi$.

**Problem 3. Applied problem: non-linear modeling**

In this exercise, you will try some non-linear models discussed in class. We use the Wage data set in "Wage.csv". We only use the feature age to predict wage.

(a) Divide the data into a train part of 70% and a test part of 30%. Use "train_test_split" from "sklearn.model_selection". Set random_state = 42.

(b) Visualize the relationship between age and wage by producing a scatterplot of the training data.

(c) Fit polynomial regression of degree from 1 to 10. Report the mean square error (MSE) computed on the training data and test data, respectively, in a table. Plot the data and the polynomial fits. Explain your results. (Hint: you can generate a new feature matrix consisting of all polynomial combinations of the features by "PolynomialFeatures" from "sklearn.preprocessing")

(d) Fit a cubic spline for a range of degrees of freedom from 4 to 6. Plot the resulting fits and report the resulting train and test MSE in the same table. (Hint: you can use "dmatrix" from "pasty". The "bs()" function generates the entire matrix of basis functions for splines with the specified set of knots or specified degree of freedom.)

(e) Fit a nature cubic spline for a range of degrees of freedom from 4 to 6. Plot the resulting fits and report the resulting train and test MSE in the same table. (Hint: The "cr()" function generates the entire matrix of basis functions for splines.)

(f) Compare all the models in the table and comment on the results.

(g) (optional) Use the features year, education and age to predict wage using Generalized Additive Model (GAM).

Reference link: http://www.science.smith.edu/~jcrouser/SDS293/labs/lab13-py.html

**Problem 4. Linear discriminant analysis** (optional)

Suppose that we have a training data set $\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ consisting of $n$ independent examples, where $x_i \in \mathbb{R}$ is a scalar and $y_i \in \{0, 1\}$. We will model the joint distribution of $(x, y)$ according to: where

$$p(y) = \begin{cases} \pi & \text{if } y = 1 \\ 1 - \pi & \text{if } y = 0 \end{cases}$$

$$p(x|y = 1) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(x - \mu_1)^2}{2\sigma^2})$$

$$p(x|y = 0) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(x - \mu_0)^2}{2\sigma^2})$$

Here, the parameters of our model are $\pi, \mu_0, \mu_1$ and $\sigma$. The log-likelihood of the data is

$$\begin{aligned} \ell(\pi, \mu_0, \mu_1, \sigma) &= \log \prod_{i=1}^{n} p(x_i, y_i; \pi, \mu_0, \mu_1, \sigma) \\ &= \log \prod_{i=1}^{n} p(x_i|y_i; \mu_0, \mu_1, \sigma) p(y_i; \pi) \end{aligned}$$

(a) Show that the maximum likelihood estimates of $\pi, \mu_0, \mu_1$ and $\sigma$ are given in the following formulas

$$\pi = \frac{1}{n} \sum_{i=1}^{n} I(y_i = 1)$$

$$\mu_1 = \frac{\sum_{i=1}^{n} I(y_i = 1) x_i}{\sum_{i=1}^{n} I(y_i = 1)}$$

$$\mu_0 = \frac{\sum_{i=1}^{n} I(y_i = 0) x_i}{\sum_{i=1}^{n} I(y_i = 0)}$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{y_i})(x_i - \mu_{y_i})^T$$

(b) Show that the postierior probability takes the form of a logsitic function, and can be written as

$$p(y = 1|x; \pi, \mu_0, \mu_1, \sigma) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x)}$$

where $\beta_0$ and $\beta_1$ are some appropriate functions of $\pi, \mu_0, \mu_1$ and $\sigma$.