Theodore Kim
CS-UY 4563-A: Introduction to Machine Learning
Homework #2

Question 1: Data – ((0,0), 1), ((0, 1), 4), ((1, 0), 3), ((1, 1), 7)

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 7 \end{bmatrix}$$

$$w = (X^T X)^{-1} X^T y$$

$$w = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 4 \\ 3 \\ 7 \end{bmatrix}$$

$$w = \left(\begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 15 \\ 10 \\ 11 \end{bmatrix}$$

$$w = \left(\begin{bmatrix} 0.75 & -0.5 & -0.5 \\ -0.5 & 1 & 0 \\ -0.5 & 0 & 1 \end{bmatrix} \right) \cdot \begin{bmatrix} 15 \\ 10 \\ 11 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 2.5 \\ 3.5 \end{bmatrix}$$

$$w_0 = 0.75, w_1 = 2.5, w_2 = 3.5$$

$$\hat{y} = Xw = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.75 \\ 2.5 \\ 3.5 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 4.25 \\ 3.25 \\ 6.75 \end{bmatrix}$$

$$RSS = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum \left(\begin{bmatrix} 1 \\ 4 \\ 3 \\ 7 \end{bmatrix} - \begin{bmatrix} 0.75 \\ 4.25 \\ 3.25 \\ 6.75 \end{bmatrix} \right)^2 = \sum \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}^2 = \sum \begin{bmatrix} 0.0625 \\ 0.0625 \\ 0.0625 \\ 0.0625 \end{bmatrix} = 0.25$$

$$TSS = \sum_{i=1}^{N} (y_i - \bar{y})^2 = (1 - 3.75)^2 + (4 - 3.75)^2 + (3 - 3.75)^2 + (7 - 3.75)^2$$

$$= 7.5625 + 0.0625 + 0.5625 + 10.5625 = 18.75$$

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{0.25}{18.75} = 0.987$$

$$\hat{y} = Xw = \begin{bmatrix} 1 & 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 0.75 \\ 2.5 \\ 3.5 \end{bmatrix} = 3.75$$

98.7% of the variance in *y* is explained by *x*.

Question 2: Compute *RSS*, *TSS*, and R^2

$$RSS = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = 14517.5517$$

$$TSS = \sum_{i=1}^{N} (y_i - \bar{y})^2 = 42716.2954$$

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{14517.5517}{42716.2954} = 0.6601$$

66.01% of the variance in *y* is explained by *x*

Question 3: Suppose you were interested in crop yields and you had collected data on the amount of rainfall, the amount of fertilizer, the average temperature, and the number of sunny days. How could you formalize this a as regression problem?

After collecting the features of the sample data ($d = 4$, amount of rainfall, amount of fertilizer, average temperature, and number of sunny days) for a sample set (some N number of fields treated to the recorded conditions), collect data on the crop yields in some quantifiable way (i.e. pounds of crops harvested), then organize the sample data into X and y.

$$X = \begin{bmatrix} 1 & x_{1,rain} & x_{1,fertilizer} & x_{1,temp} & x_{1,sun} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & x_{N,sun} & x_{N,sun} & x_{N,sun} & x_{N,sun} \end{bmatrix} y = \begin{bmatrix} y_{1,yield} \\ \vdots \\ y_{N,yield} \end{bmatrix}$$

Then solve for w and evaluate the fit of the resulting model.

Question 4: Running gradient descent with the functions:

$$f_1(b_1, b_2) = (b_1 - 4)^2 + (b_2 + 3)^2$$

$$f_2(b_1, b_2) = (4 - b_1)^2 + 34 \cdot \left((b_1 + 4) - (b_2 - 4)^2\right)^2$$

Given, α =0.5, `num_iters=4,` $(b_1, b_2) = (0, 0)$

$$f_1(b_1, b_2) = \left({b_1}^2 - 8b_1 + 16\right) + \left({b_2}^2 + 6b_2 + 9\right)$$

$$\frac{\partial f_1(b_1, b_2)}{\partial b_1} = 2b_1 - 8$$

$$\frac{\partial f_1(b_1, b_2)}{\partial b_2} = 2b_2 + 6$$

$$temp1 = b_1 - \alpha \frac{\partial f_1(b_1, b_2)}{\partial b_1} = b_1 - 0.5(2b_1 - 8)$$

$$temp2 = b_2 - \alpha \frac{\partial f_1(b_1, b_2)}{\partial b_2} = b_2 - 0.5(2b_2 + 3)$$

After Iteration 1: $(b_1, b_2) = (4, 3)$
After Iteration 2: $(b_1, b_2) = (4, -3)$
After Iteration 3: $(b_1, b_2) = (4, -3)$
After Iteration 4: $(b_1, b_2) = (4, -3)$

$$f_2(b_1, b_2) = \left({b_1}^2 - 8b_1 + 16\right) + 34 \cdot \left((b_1 + 4) - ({b_2}^2 - 8b_2 + 16)\right)^2$$
$$\frac{\partial f_1(b_1, b_2)}{\partial b_1} = -68{b_2}^2 + 544b_2 + 70b_1 - 824$$
$$\frac{\partial f_1(b_1, b_2)}{\partial b_2} = -136(b_2 - 4)(-{b_2}^2 + 8b_2 + b_1 - 12)$$
$$temp1 = b_1 - \alpha \frac{\partial f_1(b_1, b_2)}{\partial b_1} = b_1 - 0.5(68{b_2}^2 + 544b_2 + 70b_1 - 824)$$
$$temp2 = b_2 - \alpha \frac{\partial f_1(b_1, b_2)}{\partial b_2} = b_2 - 0.5(-68(2b_2 - 8)(-{b_2}^2 + 8b_2 + b_1 - 12))$$

After Iteration 1: $(b_1, b_2) = (412, 3264)$
After Iteration 2: $(b_1, b_2) = (3.61e8, 2.37e12)$
After Iteration 3: $(b_1, b_2) = (1.9059E+26, 9.02e38)$
After Iteration 4: $(b_1, b_2) = (2.7691E+79, 4.99e118)$