ECE1513-Introduction to Machine Learning

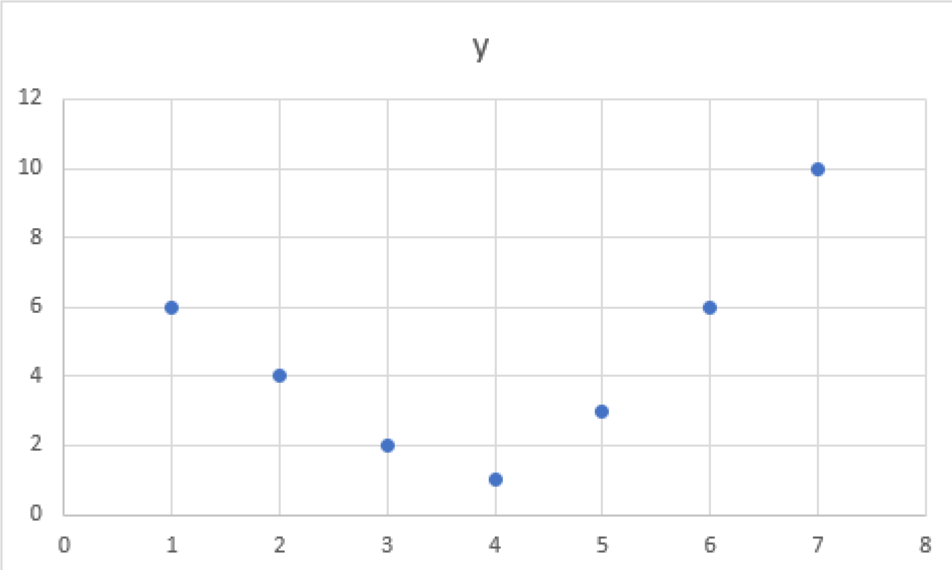
Assignment 1

Zihao Jiao (1002213428)

Date: 2020.1.22

Problem 1 Assume we collected a dataset D $= \{(Xi, Yi)\}_{i\in 1..7}$ of N = 7 points (i.e., observations) with inputs X = (1, 2, 3, 4, 5, 6, 7) and outputs Y = (6, 4, 2, 1, 3, 6, 10) for a regression problem.

1. Draw a scatter plot of the dataset on a spreadsheet software (e.g., Excel).



2. Let us use a linear regression model $gw, b(x) = wx + b$ to model this data. Write down the analytical expression of the mean squared loss of this model on dataset D. Your loss should take the form of

$$\frac{1}{2N}\sum_{i\in 1.N} A\,w^2 + Bb^2 + Cwb + Dw + Eb + F$$

where A, B, C, D, E, and F are expressed only as a function of Xi and Yi and constants. Do not fill-in any numerical values yet.

Solution:

Known $g_{w,b}(x) = wx + b$ and loss function is $L_{(y,t)} = \frac{1}{2}(y - t)^2$.

Thus, $\varepsilon(w, b) = \frac{1}{2N}\sum_{i\in 1.N}(y^{(i)} - Y^{(i)})^2 = \frac{1}{2N}\sum_{i=1}^{N}[(wx^i + b) - Y^i]^2$

$=\frac{1}{2N}\sum_{i=1}^{N}[w^2(x^i)^2 + 2wx^ib + b^2 - 2wx^iY^i - 2Y^ib + (Y^i)^2]$

In this case, A = $(x^i)^2$; B = 1; C = $2x^i$; D = $-2x^iY^i$; E = $-2Y^i$; F = $(Y^i)^2$

3. Derive the analytical expressions of w and b by minimizing the mean squared loss from the previous question. Your expressions for parameters w and b should only depend on A, B, C, D and E. Do not fill-in any numerical values yet.

Solution: Known the mean squared error is:

$$\varepsilon(w, b) = \frac{1}{2N}\sum_{i\in 1.N} A\,w^2 + Bb^2 + Cwb + Dw + Eb + F$$

in order to minimize it, we need to find $\frac{\delta\varepsilon}{\delta w} = 0$ *and* $\frac{\delta\varepsilon}{\delta b} = 0.$

Thus, $\frac{\delta\varepsilon}{\delta b} = \frac{1}{2N}\sum_{i=1}^{N}(2Bb + Cw + E) = 0$

$$\frac{\delta\varepsilon}{\delta w} = \frac{1}{2N}\sum_{i=1}^{N}(2Aw + Cb + D) = 0$$

4. Give approximate numerical values for w and b by plugging in numerical values from the dataset D.
   Solution:

   substitute A, B, C, D, E, F into $\frac{\delta\varepsilon}{\delta b}, \frac{\delta\varepsilon}{\delta w}$ from question 3.

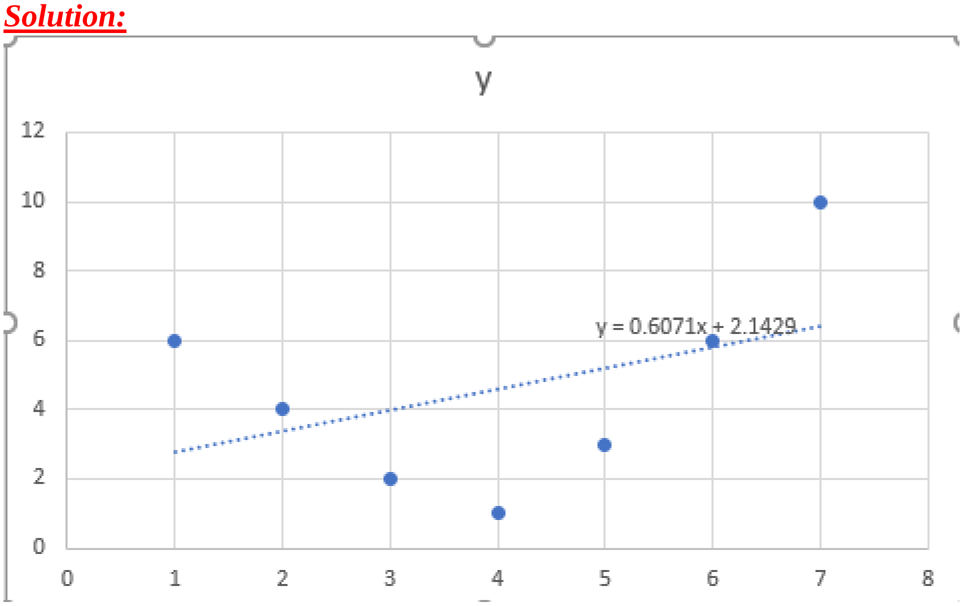
   $$\frac{\delta\varepsilon}{\delta b} = \frac{1}{2N}\sum_{i=1}^{N}(2b + 2x^i w - 2Y^i) = \bar{x}w - \bar{Y} + b = 0$$

   $$\frac{\delta\varepsilon}{\delta w} = \frac{1}{2N}\sum_{i=1}^{N}[2(x^i)^2 w + 2x^i b - 2x^i Y^i] = \bar{x}^2\, w + \bar{x}b - \overline{xY} = 0$$

   $\bar{x}\ from\ dataset = \frac{1+2+3+4+5+6+7}{7} = 4$; $\bar{Y}\ from\ dataset = \frac{6+4+2+1+3+6+10}{7} = 4.571$; and $\overline{xY} = 20.714$.

   Thus, $\begin{cases} \bar{x}w - \bar{Y} + b = 0 \\ \bar{x}^2\, w + \bar{x}b - \overline{xY} = 0 \end{cases}$, we can get $\begin{cases} w = 0.607 \\ b = 2.14 \end{cases}$.

5. Double-check your solution with the scatter plot from the question earlier: e.g., you can use Excel to find numerical values of w and b.
   Solution:



The answers from excel and analytical solutions are identical.

Problem 2 Let us now assume that D is a dataset with d features per input and N > 0 inputs. We have D = $\{((X_{ij})_{j\in 1..d}, Y_i)\}_{i\in 1..N}$. In other words, each Xi is a column vector with d components indexed by j such that X_{ij} is the jth component of X_i . The output Y_i remains a scalar (real value).

Let us assume for simplicity that b = 0 so we have a simplified linear regression model:

$$g_{\vec{w}}(X) = \vec{w}X$$

where $\vec{w}$ is now a vector of dimensionality d. Each component of ~w multiplies the corresponding feature of X, which gives the following: $g_{\vec{w}}(X_i) = \sum_{j\in 1..d} w_j X_{ij}$

We would like to train a regularized linear regression model, where the mean squared loss is augmented with an L2 regularization penalty $\frac{1}{2}\|\vec{w}\|_2^2$ on the weight parameter $\vec{w}$:

$$L(\vec{w}, D) = \frac{1}{2N}\sum_{i\in 1..N}(Y_i - g_{\vec{w}}(X_i))^2 + \frac{\lambda}{2}\|\vec{w}\|_2^2$$

where λ > 0 is a hyperparameter that controls how much importance is given to the penalty.

1. Let $A = \sum_{i\in 1..N} X_i X_i^T$. Give a simple analytical expression for the components of A.
   Solution:
   Known X∈ R^{Nxd}

$$A = \sum_{i\in 1..N} X_i X_i^T = \begin{pmatrix} X_1^2 & \cdots & X_1 X_d \\ \vdots & \ddots & \vdots \\ X_d X_1 & \cdots & X_d^2 \end{pmatrix}$$

$$A = \sum_{i=1}^{N}\sum_{j=1}^{d} X_{ij}X_{ji} = XX^T$$

2. Let us write $B = \sum_{i\in 1..N} Y_i X_i$, prove that the following holds:

$$\nabla L(\vec{w}, D) = \frac{1}{N}(A\vec{w} - B) + \lambda\vec{w}$$

   Solution:

$L(\vec{w}, D) = \frac{1}{2N}\sum_{i\in 1..N}(Y_i - g_{\vec{w}}(X_i))^2 + \frac{\lambda}{2}\|\vec{w}\|_2^2$

$=\frac{1}{2N}\sum_{i=1}^{N}(Y_i^2 - 2\sum_{j=1}^{d} Y_i\vec{w}_j X_{ij} + (\sum_{j=1}^{d}\vec{w}_j X_{ij})^2) + \frac{\lambda}{2}\|\vec{w}\|_2^2$

Thus, $\nabla L(\vec{w}, D) = \frac{\delta L}{\delta\vec{w}} = \frac{1}{N}(\vec{w}\sum_{i=1}^{N}(\sum_{j=1}^{d} X_{ij}^2)^2 - \sum_{i=1}^{N} Y_i X_i) + \lambda\vec{w}$

So, the yellow part represents A and B respectively, our solution match it with:

$$\nabla L(\vec{w}, D) = \frac{1}{N}(A\vec{w} - B) + \lambda\vec{w}$$

3. Write down the matrix equation $\vec{w}^*$ that should satisfy, where:

$$\vec{w}^* = \arg min\ L(\vec{w}, D)$$

Your equation should only involve A, B, λ, N, and $\vec{w}^*$.

Solution:

$$\vec{w}^* = \arg min\ L(\vec{w}, D) = \frac{1}{N}(A\vec{w}\ -\ B) +\ \lambda\vec{w} = 0$$

$$\frac{A}{N}\vec{w}\ -\ \frac{B}{N} + \lambda\vec{w} = 0$$

$$\left(\frac{A}{N} +\ \lambda I\right)\vec{w} = \frac{B}{N}$$

$$\text{Thus, } \vec{w}^* = \frac{\frac{B}{N}}{\frac{A}{N}+\lambda I} = \frac{B}{A+\lambda \mathrm{N}I}$$

4. Prove that all eigenvalues of A are positive.

Solution:

We know that matrix A is symmetric matrix, assume x is <u>full rank non-zero</u> vector, that we can get:

Note that: A= $[a_{ij}]$ and X = $[x_i]$

$$X^T AX = \lambda X^T X = \lambda||X||^2 > 0$$

$$X^T AX = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i\neq j} a_{ij}x_i x_j = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_i x_j$$

Thus, A is positive definite matrix, which leads that all eigenvalues of A are positive

5. Demonstrate that matrix $A + \lambda \mathrm{N}I$ is invertible by proving that none of its eigenvalues are zero. Here, I is the identity matrix of dimension d.

Solution:

Knowing A is positive definite matrix,

$$X^T(A + \lambda \mathrm{N}I\ )X = X^T AX + X^T\lambda \mathrm{N}IX$$

From Q4 we know that $X^T AX > 0$, $X^T\lambda \mathrm{N}IX = X_1^2\lambda\mathrm{N} + \cdots + X_d^2\lambda\mathrm{N}$

Since X is full rank nonzero vector, and N>0, thus, $X^T\lambda \mathrm{N}IX > 0$

Which leads $X^T(A + \lambda \mathrm{N}I\ )X > 0$, so, $A + \lambda \mathrm{N}I$ is positive definite matrix, all eigenvalues are not equal to zero. It is invertible.

6. Using the invertibility of matrix $A + \lambda \mathrm{N}I$, solve the equation stated in question 3 and deduce an analytical solution for $\vec{w}^*$. You've obtained a linear regression model regularized with an L2 penalty.

Solution:

From question 3, $\vec{w}^* = \dfrac{B}{A+\lambda \mathrm{N}I}$, and $\begin{cases} A = XX^T; \\ B = XY^T, \end{cases}$

By substituting A, B, and using the invertibility of matrix $A + \lambda \mathrm{N}I$,

$$\vec{w}^* = \frac{XY^T}{XX^T + \lambda \mathrm{N}I} = (XX^T + \lambda \mathrm{N}I)^{-1}XY^T$$