Machine Learning para mineria de datos
Homework 2

Salazar Martinez Miguel Angel

22 de enero de 2025

1. Given the following data set

   - https://dev.mysql.com/doc/employee/en/employees-installation.html

   (a) Given a of the people and its salary, extract information into a pandas data frame using by pulling

      i. End Data

      ii. Salary

   (b) Generate a simple Linear Regression (Which needs to be programmed using Jax) to estimate given a title, anyone you want, to say the salary of a person in 2025.

      i. Using the Canonical Version

2. Implement the regularized version of the following equation

$$L(\boldsymbol{w}) = \sum_{i=1}^{N} \left(y_i - \boldsymbol{x}_i^T \boldsymbol{w}\right)^2 + \lambda \sum_{i=1}^{d+1} w_i^2$$

   (a) Give me the $\Delta L(\boldsymbol{w})$ of gradient descent by deriving the function $L(\boldsymbol{w})$. Here is the function:

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \alpha \Delta L(\boldsymbol{w}_t)$$

   (b) Implement the gradient descent step size without using the operation grad at Jax using the previous information.

   (c) Answer the problem using this new model.

   (d) Use a search grid on λ to obtain the best lambda for it.

1. Argue that in the case of simple linear regression, the least squares line always passes through the point $(\overline{\boldsymbol{x}}, \overline{y})$. Here,

   (a) $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$

   (b) And $\widehat{y}_i = g(x_i) + \epsilon$. Hint argue that $\sum_{i=1}^{N} \epsilon = 0$

2. Prove that he R^2 statistic is equal to the square of the correlation between X and Y. For simplicity, you may assume that $\overline{\boldsymbol{x}} = \overline{\boldsymbol{y}} = 0$.

3. Carefully explain the differences between the KNN classifier and KNN regression methods.

4. Approximating a vector as a multiple of another one. In the special case n = 1, the general least squares problem reduces to finding a scalar x that minimizes $\|ax - b\|^2$, where a and b are d-vectors. We write the matrix A here in lower case, since it is an d-vector.

   (a) Assuming a and b are nonzero, show that $\|a\widehat{x} - b\|^2 = \|b\| \sin\theta$. θ is the angle between a and b.

5. Suppose the $m \times n$ matrix A has linearly independent columns, and b is an d-vector. Let $\widehat{\boldsymbol{x}} = \left(A^T A\right)^{-1} A^T \boldsymbol{x}$ denote the least squares approximate solution of $A\boldsymbol{x} = \boldsymbol{b}$.

   (a) Show that for any d-vector $\boldsymbol{x}$, $(A\boldsymbol{x})^T \boldsymbol{b} = (A\boldsymbol{x})^T (A\widehat{\boldsymbol{x}})$.

   (b) Least angle property of least squares. The choice $\boldsymbol{x} = \widehat{\boldsymbol{x}}$ minimizes the distance between $A\boldsymbol{x}$ and $\boldsymbol{b}$. Show that $\boldsymbol{x} = \widehat{\boldsymbol{x}}$ also minimizes the angle between $A\boldsymbol{x}$ and $\boldsymbol{b}$. (You can assume that $A\boldsymbol{x}$ and $\boldsymbol{b}$ are nonzero.) For any positive scalar α,