Homework 1

Che-Hsun Liu

August 12, 2020

1 整理

此次作業是要利用某個時間點之前的數小時觀測數據來預測該時間點的 PM2.5 數值 y。假設該模型為 linear model ，數學式可表示為

$$y = f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$$

其中 $\mathbf{x} = \{1, x_1, \ldots, x_k\}$ 為給定測試資料所取出的 $k+1$ 維特徵值，$\mathbf{w} = \{w_0, w_1, \ldots, w_k\}$ 則是該模型參數。

用來估計模型參數的 n 筆訓練資料，每一筆一樣有著 $k+1$ 維特徵以及單一標記，以數學式表示如下：

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^n \end{pmatrix}, \ \mathbf{y} = \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix} .$$

這組訓練資料帶入模型所得 PM2.5 值和實際值的誤差 loss function 則可表示為

$$L(\mathbf{w}) = \sum_{i=1}^{n} \left(y^i - \mathbf{w} \cdot \mathbf{x}^i\right)^2,$$

使該函式最小化之參數即為所求

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} L(\mathbf{w}) = \arg\min_{\mathbf{w}} \sum_{i=1}^{n} \left(y^i - \mathbf{w} \cdot \mathbf{x}^i\right)^2 .$$

最佳解或次佳解可用 gradient descent 演算法來逼近。首先隨機抓取一組參數

$$\mathbf{w}^0 = \{w_0^0, w_1^0, \ldots, w_k^0\}$$

然後進行以下運算數次

$$\left(\mathbf{w}^{t+1}\right)^\intercal = \begin{pmatrix} w_0^{t+1} \\ w_1^{t+1} \\ \vdots \\ w_k^{t+1} \end{pmatrix} = \begin{pmatrix} w_0^{t} \\ w_1^{t} \\ \vdots \\ w_k^{t} \end{pmatrix} - \eta \begin{pmatrix} \frac{\partial}{\partial w_0^t} L(\mathbf{w}^t) \\ \frac{\partial}{\partial w_1^t} L(\mathbf{w}^t) \\ \vdots \\ \frac{\partial}{\partial w_k^t} L(\mathbf{w}^t) \end{pmatrix}$$

$$\frac{\partial}{\partial w_h^t} L(\mathbf{w}^t) = \sum_{i=1}^{n} 2\left(y^i - \mathbf{w}^t \cdot \mathbf{x}^i\right)\left(-x_h^i\right) .$$