深度学习个人笔记

Amamiyaren

my own deep learning notes

更新：2024 年 5 月 17 日

1 多元线性回归

1.1 基本形式

线性模型试图学得一个通过属性的线性组合来预测的函数，基本形式如下：

$$f(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b \tag{1.1}$$

向量形式如下:

$$f(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\mathbf{x} + \mathbf{b} \tag{1.2}$$

1.2 一元情形

一般采用最小二乘法确定模型参数,一元情形推导如下：

$$\begin{aligned}(w^*, b^*) &= \arg\min_{(w,b)} \sum_{i=1}^{m} (f(x_i) - yi)^2 \\ &= \arg\min_{(w,b)} \sum_{i=1}^{m} (y_i - wx_i - b)^2 \end{aligned} \tag{1.3}$$

我们令$E_{(w,b)} = \underset{(w,b)}{\arg\min} \sum_{i=1}^{m} (y_i - wx_i - b)^2$，等式两边同时对$w$和$b$求导有：

$$\frac{\partial E_{(w,b)}}{\partial b} = 2\Big(mb - \sum_{i=1}^{m}(y_i - wx_i)\Big) \tag{1.4}$$

$$\begin{aligned}\frac{\partial E_{(w,b)}}{\partial w} &= \sum_{i=1}^{m} 2(y_i - wx_i - b)(-x_i) \\ &= 2\Big(w\sum_{i_1}^{m} x^2 - \sum_{i=1}^{m}(y_i - b)x_i\Big)\end{aligned} \tag{1.5}$$

然后令式(1.4)和(1.5)为零得到：

$$\begin{aligned}b &= \frac{1}{m}\sum_{i=1}^{m} y_i - wx_i \\ &= \bar{y} - w\bar{x}\end{aligned} \tag{1.6}$$

$$w\sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} (y_i - b)x_i \tag{1.7}$$

将(1.6)带入(1.7)得到：

$$\begin{aligned}w\sum_{i=1}^{m} x_i^2 &= \sum_{i=1}^{m} (y_i x_i) - \sum_{i=1}^{m} (\bar{y} - w\bar{x})x_i \\ &= \sum_{i=1}^{m} y_i x_i - \bar{y}\sum_{i=1}^{m} x_i + w\bar{x}\sum_{i=1}^{m} x_i\end{aligned} \tag{1.8}$$

对(1.8)式进行化简，最终可得:

$$w = \frac{\sum_{i=1}^{m} y_i x_i - \bar{y}\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} \bar{x} x_i} \tag{1.9}$$

又由于$\bar{y}\sum_{i=1}^{m}x_i = \frac{1}{m}\sum_{i=1}^{m}y_i\cdot\sum_{i=1}^{m}x_i = \bar{x}\sum_{i=1}^{m}y_i$和$\bar{x}\sum_{i=1}^{m}y_i = \frac{1}{m}\sum_{i=1}^{m}x_i\sum_{i=1}^{m}x_i = \frac{1}{m}\left(\sum_{i=1}^{m}x_i\right)^2$将这两个式子带入(1.9)式进行进一步化简得到：

$$w = \frac{\sum_{i=1}^{m}y_i(x_i-\bar{x})}{\sum_{i=1}^{m}x_i^2-\left(\sum_{i=1}^{m}x_i\right)^2} \tag{1.10}$$

1.2.1 一元情形的矩阵形式

由于 python 中累加形式只能使用`for`循环进行迭代，运行效率比较低，如果将(1.20)式转化为矩阵形式可以使用`numpy`等第三方库对其矩阵运算进行加速，公式推导如下：

$$\bar{y}\sum_{i=1}^{m}x_i\bar{x}\sum_{i=1}^{m}y_i=\sum_{i=1}^{m}\bar{y}x_i=\sum_{i=1}^{m}\bar{x}y_i=m\bar{x}\bar{y}=\sum_{i=1}^{m}\bar{x}\bar{y} \tag{1.11}$$

$$\sum_{i=1}^{m}x_i\bar{x}=\bar{x}\sum_{i=1}^{m}x_i=\bar{x}\cdot m\cdot\frac{1}{m}\sum_{i=1}^{m}x_i=m\bar{x}^2=\sum_{i=1}^{m}\bar{x}^2 \tag{1.12}$$

将式(1.21)和式(1.22)带入式(1.9)可以得到：

$$\begin{aligned} w &= \frac{\sum_{i=1}^{m}(y_ix_i-y_i\bar{x}-x_i\bar{y}+\bar{x}\bar{y})}{\sum_{i=1}^{m}(x_i^2-\bar{x}x_i-\bar{x}x_i+\bar{x}^2)} \\ &= \frac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{m}(x_i-\bar{x})^2} \end{aligned} \tag{1.13}$$

记$x=(x_1;x_2...x_m)$,$x_d=(x_i-\bar{x};x_2-\bar{x}...x_m-\bar{x})$,$y_d$同$x_d$,再由向量内积的定义可以得到:

$$w=\frac{x_d^Ty_d}{x_d^Tx_d} \tag{1.14}$$

1.2.2 多元

多元的情形可由一元的情形举一反三推导出，如下：

$$\begin{aligned} w^* &= \mathop{\arg\min}_{(\hat{w})} \sum_{i=1}^{m} (y_i - \hat{w}^T x_i)^2 \\ &= \mathop{\arg\min}_{(\hat{w})} \sum_{i=1}^{m} (y_i - x_i^T \hat{w})^2 \\ &= \mathop{\arg\min}_{(\hat{w})} \left[y_1 - x_1^T \hat{w} \ldots y_m - x_m^T \hat{w} \right] \cdot \begin{bmatrix} y_1 - x_1^T \hat{w} \\ \ldots \\ y_m - x_m^T \hat{w} \end{bmatrix} \end{aligned} \tag{1.15}$$

其中：

$$\begin{aligned} \begin{bmatrix} y_1 - x_1^T \hat{w} \\ y_2 - x_2^T \hat{w} \\ \ldots \\ y_m - x_m^T \hat{w} \end{bmatrix} &= \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_m \end{bmatrix} - \begin{bmatrix} x_1^T \hat{w} \\ x_2^T \hat{w} \\ \ldots \\ x_m^T \hat{w} \end{bmatrix} \\ &= y - \begin{bmatrix} x_1^T \\ x_2^T \\ \ldots \\ x_m^T \end{bmatrix} \cdot \hat{w} \\ &= y - X\hat{w} \end{aligned} \tag{1.16}$$

所以:

$$\hat{w}^* = \mathop{\arg\min}_{(\hat{w})} (y - X\hat{w})^T (y - X\hat{w}) \tag{1.17}$$

现引入两个定理对$\hat{w}$的最优解做进一步推导。

1.2.3 w预测值的求解

$Def_{1.28}$：设$D \subset R^n$为非空开凸集$f(x)$是定义在D上的实值函数，且$f(x)$在D上二阶连续可微，若$f(x)$的$Hession$矩阵$\nabla^2 f(x)$在D上半正定则$f(x)$是D上

的凸函数，若$\nabla^2 f(x)$在D上正定则$f(x)$为D上的严格凸函数。 $Def_{1.29}$：若$f(x)$为凸函数且$f(x)$一阶连续可微则x^*是全局最优解的充要条件为其梯度等于零向量即:

$$\nabla f(x^*) = \mathbf{0}$$

$$\begin{aligned} E_{\hat{w}} &= (y - X\hat{w})^T(y - X\hat{w}) \\ &= y^T y - y^T X\hat{w} - \hat{w}^T X^T y + \hat{w}^T X^T X\hat{w} \end{aligned} \tag{1.18}$$

$$\begin{aligned} \frac{\partial E\hat{w}}{\partial \hat{w}} &= 0 - X^T y - X^T y + (X^T X + X^T X)\hat{w} \\ &= 2X^T(X\hat{w} - y) \end{aligned} \tag{1.19}$$

由式(1.31)再对$\hat{w}^T$求偏导可以得到$E_{\hat{w}}Hession$矩阵的表达式，如下：

$$\begin{aligned} \nabla^2 E_{\hat{w}} &= \frac{\partial}{\partial w^T}\Big(\frac{\partial E\hat{w}}{\partial w}\Big) \\ &= 2X^T X \end{aligned} \tag{1.20}$$

由$Def_{1.28}, Def_{1.29}$当$X^T X$为正定矩阵时，令其梯度为零向量可解得方程的全局最优解即：

$$\hat{w}^* = (X^T X)^{-1} X^T y \tag{1.21}$$