Trivial Big Data

wu

2022 年 11 月 7 日

目录

1 一元线性回归

方差

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

标准差

$$S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

协方差

$$Cov = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

相关系数

$$\rho = \frac{Cov}{S_x S_y}$$

一元线性回归：找一条直线，拟合数据点

$$y = \beta_0 + \beta_1 x$$

最小二乘法

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} ((\beta_0 + \beta_1 x_i) - y_i)^2$$

$$f(\beta_0,\beta_1) = \sum_{i=1}^{n} ((\beta_0 + \beta_1 x_i) - y_i)^2$$

$$\begin{aligned} &\frac{\partial f}{\partial \beta_0} = 2\sum_{i=1}^{n} (\beta_0 + \beta_1 x_i - y_i) = 0 \\ &\Rightarrow n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0 \\ &\Rightarrow \beta_0 + \beta_1 \bar{x} - \bar{y} = 0 \end{aligned}$$

$$\begin{gathered} \frac{\partial f}{\partial \beta_1} = 2\sum_{i=1}^{n} x_i(\beta_0 + \beta_1 x_i - y_i) = 0 \\ \sum_{i=1}^{n} x_i(\beta_0 + \beta_1 x_i - y_i) = 0 \end{gathered}$$

注意到

$$\sum_{i=1}^{n} \bar{x}(\beta_0 + \beta_1 x_i - y_i) = \bar{x}(n\beta_0 + \beta_1 \sum x_i - \sum y_i) = 0$$

所以

$$\sum_{i=1}^{n} (x_i - \bar{x})(\beta_0 + \beta_1 x_i - y_i) = 0$$

而

$$\sum_{i=1}^{n} (x_i - \bar{x})(\beta_0 + \beta_1 \bar{x} - \bar{y}) = (\beta_0 + \beta_1 \bar{x} - \bar{y}) \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

所以

$$\sum_{i=1}^{n} (x_i - \bar{x})(\beta_1(x_i - \bar{x}) - (y_i - \bar{y})) = 0$$

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{Cov}{S_x^2}$$
$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}} \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \rho\frac{S_y}{S_x}$$

因此

$$y = \bar{y} - \beta_1\bar{x} + \beta_1 x \Rightarrow y - \bar{y} = \beta_1(x-\bar{x}) = \rho\frac{S_y}{S_x}(x-\bar{x})$$

2 多元线性回归

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

最小二乘法

$$\min_{\beta_0,\ldots,\beta_m} \sum_{i=1}^{n}((\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}) - y_i)^2$$

令

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1m} \\ 1 & x_{21} & x_{22} & \ldots & x_{2m} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nm} \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

最小二乘形式

$$\min_{\beta} \|X\beta - y\|^2$$

$$g(\beta) = \langle w, \beta \rangle = w^T\beta = \sum_{i=0}^{m} w_i\beta_i$$

$$\nabla g = \begin{pmatrix} \frac{\partial g}{\partial \beta_0} \\ \frac{\partial g}{\partial \beta_1} \\ \vdots \\ \frac{\partial g}{\partial \beta_m} \end{pmatrix} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_m \end{pmatrix} = w$$

假设 $A = A^T$

$$h(\beta) = \langle A\beta, \beta \rangle = \beta^T A \beta = \sum_{i,j} a_{ij}\beta_i\beta_j$$

定义 $p(u,v) = \langle Au, v\rangle = \langle Av, u\rangle$ 令

$$u(\beta) = \beta, v(\beta) = \beta \Rightarrow h(\beta) = p(u(\beta), v(\beta))$$

$$\nabla h = \frac{\partial p}{\partial u}\frac{\partial u}{\partial \beta} + \frac{\partial p}{\partial v}\frac{\partial v}{\partial \beta} = Av(\beta) + Au(\beta) = 2A\beta$$

$$\begin{aligned} f(\beta) &= (X\beta - y)^T(X\beta - y) \\ &= (\beta^T X^T - y^T)(X\beta - y) \\ &= \beta^T X^T X\beta - \beta^T X^T y - y^T X\beta + y^T y \end{aligned}$$

$$\nabla_\beta f = 2X^T X\beta - X^T y - X^T y = 2(X^T X\beta - X^T y) = 0$$

因此

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

为了增加鲁棒性，通常会最小化如下目标函数

$$\|X\beta - y\|^2 + \lambda \|\beta\|^2 \, (\lambda > 0)$$

此时

$$\beta = (X^T X + \lambda I)^{-1} X^T y$$

3 逻辑回归

分类问题建模:

$$f : \mathbb{R}^m \to \{0, 1\}$$
$$f(x) = \sigma(\beta^T x)$$

σ 的一种取法:

$$\sigma(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases}$$

问题:σ 在 0 点处间断

光滑化:逻辑函数(Logistic Function)

$$\sigma(z) = \frac{1}{1+e^{-z}}$$
$$f(x) = \sigma(\beta^T x) = \frac{1}{1+e^{-z}}$$

求解如下优化问题:

$$\min_{\beta} \sum_{i=1}^{n} \left(\frac{1}{1+e^{-\beta^T x_i}} - y_i \right)^2$$

对于输入 x,当 $f(x) \geq 0.5$ 时预测 1,当 $f(x) < 0.5$ 时预测 $y = 0$。

如何求 β: **梯度下降法**

$$\min_{\beta} C(\beta)$$
$$\beta_{m+1} = \beta_m - \lambda \nabla C(\beta_m)$$
$$C(x) \approx C(x') + \nabla C(x')(x - x')$$

$$\begin{aligned} C(\beta_{m+1}) &= C(\beta_m - \lambda \nabla C(\beta_m)) \\ &\approx C(\beta_m) - \lambda \left\| \nabla C(\beta_m) \right\|^2 \\ &\leq C(\beta_m) \end{aligned}$$

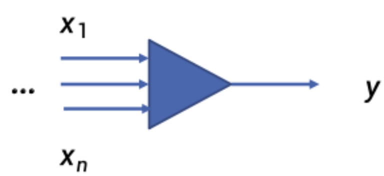
混淆矩阵

		Actual class	
		positive class	negative class
Predicted class	positive class	True Posotive(TP)	False Positive(FP)
	negative class	False Negative(FN)	True Negative(TN)

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
$$\text{precision} = \frac{TP}{TP + FP}$$
$$\text{recall} = \frac{TP}{TP + FP}$$
$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$

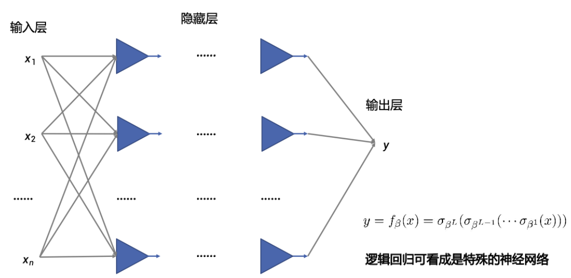
4 深度学习

单个神经元



$$y = \sigma_\beta(x_1, \dots, x_n) = \sigma(\beta_1 x_1 + \dots + \beta_n x_n) = \sigma(\beta^T x)$$

激活函数：$\sigma(z) = \frac{1}{1+e^{-z}}$ 或者 $\sigma(z) = \max\{z, 0\}$ 或 ...



神经网络参数 $\beta = (\beta^L, \beta^{L-1}, \dots, \beta^1)$ 的计算方法：

$$y = f_\beta(x) = \sigma_{\beta^L}(\sigma_{\beta^{L-1}}(\cdots\sigma_{\beta^1}(x)))$$

$$C(\beta) = \sum_{i=1}^{n}(f_\beta(x_i) - y_i)$$

$$\min_\beta(C(\beta)) = \min_\beta \sum_{i=1}^{n}(f_\beta(x_i) - y_i)^2$$

梯度下降法：

$$\min_\beta C(\beta)$$

$$\beta^{k,m+1} = \beta^{k,m} - \lambda\nabla C(\beta^{k,m}), \quad k = 1, 2, \dots, L$$

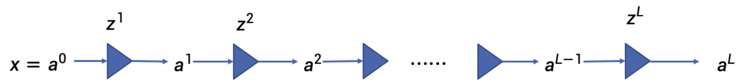
记

$$\beta^{*\ ,m} = (\beta^{1,m}, \beta^{2,m}, \dots, \beta^{L,m})$$

$$\begin{aligned} C(x) &\approx C(x') + \nabla C(x')(x - x') \\ C(\beta^{*\ ,m+1}) &= C(\beta^{*,m} - \lambda\nabla C(\beta^{*,m})) \\ &\approx C(\beta^{*,m}) - \lambda\left\|\nabla(\beta^{*,m})\right\|^2 \\ &\leq C(\beta^{*,m}) \end{aligned}$$

关键在于计算 C 关于 β 的导数，以 1 维为例，并省略上标 m

$$x = a^0 \to z^1 \to a^1 \to \dots \to a^{L-1} \to z^L \to z^L$$



其中

$$
\begin{aligned}
& z^{i+1} = \beta^{i+1} a^i, i = 0, 1, \dots, L-1 \\
& a^i = \sigma(z^i), i = 1, \dots, L \\
& C(\beta) = (a^L - y)^2
\end{aligned}
$$

需要计算

$$
\frac{\partial C}{\partial \beta^L}, \frac{\partial C}{\partial \beta^{L-1}}, \dots, \frac{\partial C}{\partial \beta^1}
$$

$$
\begin{aligned}
& \frac{\partial C}{\partial a^L} = 2(a^L - y) \\
& \frac{\partial C}{\partial \beta^L} = \frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L} \frac{\partial z^L}{\partial \beta^L} = 2(a^L - y) \cdot \sigma'(z^L) \cdot a^{L-1}
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\partial C}{\partial a^{L-1}} = \frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L} \frac{\partial z^L}{\partial a^{L-1}} = 2(a^L - y) \cdot \sigma'(z^L) \cdot \beta^L \\
& \frac{\partial C}{\partial \beta^{L-1}} = \frac{\partial C}{\partial a^{L-1}} \frac{\partial a^{L-1}}{\partial z^{L-1}} \frac{\partial z^{L-1}}{\partial \beta^{L-1}} = 2(a^L - y)\sigma'(z^L)\beta^L \cdot \sigma'(z^{L-1}) \cdot a^{L-2}
\end{aligned}
$$

因此

$$
\begin{aligned}
& \frac{\partial C}{\partial a^i} = \frac{\partial C}{\partial a^{i+1}} \frac{\partial a^{i+1}}{\partial z^{i+1}} \frac{\partial z^{i+1}}{\partial a^i} = \frac{\partial C}{\partial a^{i+1}} \sigma'(z^{i+1}) \beta^{i+1}, \quad i = 0, 1, \dots, L-1 \\
& \frac{\partial C}{\partial \beta^i} = \frac{\partial C}{\partial a^i} \frac{\partial a^i}{\partial z^i} \frac{\partial z^i}{\partial \beta^i} = \frac{\partial C}{\partial a^i} \sigma'(z^i) a^{i-1}, \quad i = 1, 2, \dots, L
\end{aligned}
$$

其中，σ 为激活函数，当 $\sigma(z) = \frac{1}{1+e^{-z}}$ 时，$\sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \sigma(z)(1 - \sigma(z))$

反向传播：

1. 根据 $\frac{\partial C}{\partial a^L} = 2(a^L - y)$ 和 $\frac{\partial C}{\partial a^i} = \frac{\partial C}{\partial a^{i+1}} \sigma'(z^{i+1}) \beta^{i+1}, i = 0, 1, 2, \dots, L-1$，反向计算出 $\frac{\partial C}{\partial a^L}, \frac{\partial C}{\partial a^{L-1}}, \dots, \frac{\partial C}{\partial a^0}$

2. 根据 $\frac{\partial C}{\partial \beta^i} = \frac{\partial C}{\partial a^i}\sigma'(z^i)a^{i-1}, i = 1, 2, ..., L$,依次计算出 $\frac{\partial C}{\partial \beta^L}, \frac{\partial C}{\partial \beta^{L-1}}, ..., \frac{\partial C}{\partial \beta^1}$
3. 根据 $\beta^i \leftarrow \beta^i - \lambda \frac{\partial C}{\partial \beta^i}, i = 1, 2, ..., L$ 对参数进行更新，并重复上述步骤直至收敛。