logistic回归

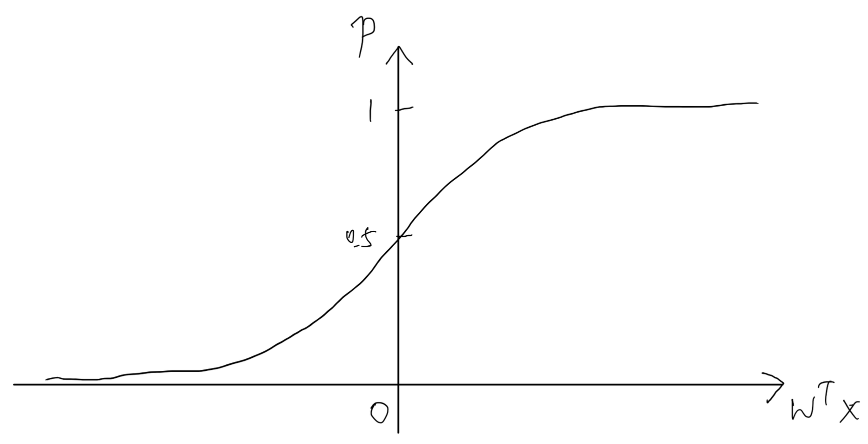
2020年8月13日 星期四 下午8:55

sigmoid曲线（S型曲线）

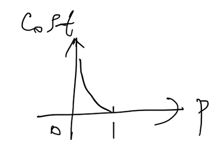
$\sigma(x) = \frac{1}{1+e^{-x}}$ ，$\sigma'(x) = \sigma(x)(1-\sigma(x))$

增广权重向量 $W = [w_1, \dots, w_k, b]$，增广特征向量 $x = [x_1, \dots, x_k, 1]$，预测函数 $= w_1x_1 + w_2x_2 + \dots + w_kx_k + b = W^Tx$

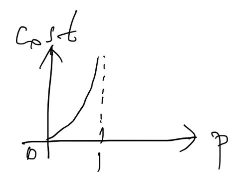
$P(Y=1|x) = \sigma(W^Tx) = \frac{1}{1+e^{-W^Tx}}$ = x是1类的概率 $= P_w(x) = \sigma(z)$，$z = W^TX$



误分类的损失 $Cost(p(x), Y) = \begin{cases} -\log(p), & Y=1 \\ -\log(1-p), & Y=0 \end{cases}$



p越接近真实分类1，损失越小

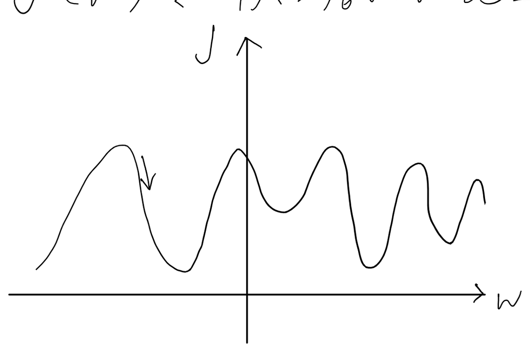


$\therefore Cost(p(x), Y) = Y \cdot [-\log(p)] + (1-Y) \cdot [-\log(1-p)]$ （交叉熵损失函数）

总损失 $$J(W) = \frac{1}{N}\sum_{i=1}^{N} Cost(p(x_i), Y_i)$$
$$= -\frac{1}{N}\sum_{i=1}^{N} Y_i \log(p(x_i)) + (1-Y_i)\log(1-p(x_i))$$
$$= -\frac{1}{N}\sum_{i=1}^{N} Y_i \log(\sigma(z_i)) + (1-Y_i)\log(1-\sigma(z_i)), \quad z_i = W^Tx_i$$

$$\frac{\partial J(W)}{\partial W} = -\frac{1}{N}\sum_{i=1}^{N} Y_i \frac{\sigma'(z_i)}{\sigma(z_i)} + (Y_i - 1)\frac{\sigma'(z_i)}{1-\sigma(z_i)}$$
$$= -\frac{1}{N}\sum_{i=1}^{N} Y_i \frac{\sigma(z_i)(1-\sigma(z_i)) \cdot z_i'}{\sigma(z_i)} + (Y_i - 1)\frac{\sigma(z_i)(1-\sigma(z_i)) \cdot z_i'}{1-\sigma(z_i)}$$
$$= -\frac{1}{N}\sum_{i=1}^{N} Y_i(1-\sigma(z_i))z_i' + (Y_i - 1)\sigma(z_i) \cdot z_i'$$
$$= -\frac{1}{N}\sum_{i=1}^{N} z_i'(Y_i - \sigma(z_i))$$
$$= -\frac{1}{N}\sum_{i=1}^{N} (W^Tx_i)'(Y_i - \sigma(z_i))$$
$$= -\frac{1}{N}\sum_{i=1}^{N} x_i(Y_i - \sigma(z_i))$$
$$= -\frac{1}{N}\sum_{i=1}^{N} x_i(\underbrace{Y_i}_{\text{真实值}} - \underbrace{P_W(x_i)}_{\text{关于W下的预测值}})$$

要令J(w)尽快减小到最小值.



梯度下降法：$\theta_{t+1} = \theta_t - \underbrace{\alpha}_{\text{步长，学习率}}\frac{\partial J(\theta)}{\partial \theta}$ ，随机梯度下降：每次只选一个样本，
（批量）

α随迭代次数增多而减小

$$\therefore W_{t+1} = W_t - \alpha\frac{\partial J(W)}{\partial W}$$
$$= W_t + \alpha\frac{1}{N}\sum_{i=1}^{N} x_i(Y_i - P_{W_t}(x_i))$$

$$P_{W_t}(x_i) \to \frac{1}{1+e^{-W^Tx_i}}$$

$x_1k_1 + x_2k_2 + \dots + x_nk_n$ （令 $k_i = Y_i - P_i$）

$$[k_1 \; k_2 \cdots k_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$