

# 回归(Regression)

本质是利用已有的数据，对未来的数据进行预测

## 1. 定义损失函数

$$L(w, b) = \sum_{i=1}^n (\hat{y}_i - (b + wx_i))^2$$

## 2. 利用共轭梯度法更新参数w,b

$$\frac{\partial L}{\partial w} = \sum 2(\hat{y}_i - (b + wx_i))(-x_i)$$

$$\frac{\partial L}{\partial b} = \sum 2(\hat{y}_i - (b + wx_i))$$

## 3. 正则化

$$y = b + \sum_{i=1}^n w_i x_i$$

$$L = \sum_i^n (\hat{y}_i - (b + \sum w_i x_i))^2 + \lambda \sum (w_i)^2$$

## 4. 补充

- 若数据存在很大偏差：
  - . 如果模型对训练数据都无法进行很好的拟合：模型欠拟合(Underfitting)
  - . 如果模型对训练数据拟合的很好，但是在测试集上表现得很差：模型过拟合(Overfitting)
- 对以上问题的解决方法：
  - . 针对输入数据增加更多的特征
  - . 设计更加复杂的模型
- 局部加权线性回归(LWLR)
  - . 针对线性回归中遇到的欠拟合问题，因为求的是具有最小均方误差的无偏估计，所以如果模型不能得到较好的预测结果，允许引入一些偏差，从而降低预测的均方误差。

$$\hat{w} = (X^T W X)^{-1} X^T W y$$

其中 $X$ 是一个系数阵， $W$ 是一个对角阵，对系数赋予不同的权重。

## 共轭梯度法(Gradient Descent)补充

也是通过计算损失函数的梯度，更新参数

- Adagrad

$$w_1 = w_0 - \frac{\eta_0}{\sigma_0} g_0 \quad \sigma_0 = \sqrt{(g_0)^2}$$

$$w_2 = w_1 - \frac{\eta_1}{\sigma_1} g_1 \quad \sigma_1 = \sqrt{\frac{1}{2}(g_0^2 + g_1^2)}$$

...

$$w_{t+1} = w_t - \frac{\eta_t}{\sigma_t} g_t \quad \sigma_t = \sqrt{\frac{1}{t+1} \sum_{i=0}^t (g_i)^2}$$

## 分类(Classification)

- 可以结合不同的概率模型运算
- 定义损失函数(最大似然估计)

$$L(\mu, \Sigma) = \int_{u, \Sigma} (x_1) \int_{u, \Sigma} (x_2) \dots \int_{u, \Sigma} (x_n)$$

$$\mu^*, \Sigma^* = \arg \underbrace{\max}_{\mu, \Sigma} L(\mu, \Sigma)$$

$$\mu^* = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\Sigma^* = \frac{1}{n} \sum_{i=1}^n (x_i - \mu^*)(x_i - \mu^*)^T$$

$$P(c_1|x) = \frac{P(x|c_1)P(c_1)}{P(x|c_1)P(c_1) + P(x|c_2)P(c_2)}$$

其中，针对 $P(c_1|x)$ 的计算，可以采用不同的概率模型，比如朴素贝叶斯。

- 逻辑回归(Logistic Regression)

*Step 1.* 定义分类的目标函数

$$f_{w,b}(x) = \sigma\left(\sum_i^n w_i x_i + b\right)$$

$\sigma$ 函数输出0或1。

*Step 2.* 对数据集 $(x_n, y_n)$ 进行训练，定义损失函数：

$$L(f) = \sum_n C(f(x_n, \hat{y}_n))$$

定义交叉熵(Cross entropy):

$$C(f(x_n, \hat{y}_n)) = -[\hat{y}_n \ln f(x_n) + (1 - \hat{y}_n) \ln(1 - f(x_n))]$$

*Step 3.* 寻找到最好的函数

令 $z = \sum_i w_i x_i$ ，对 $\ln L(w, b)$ 求偏导，最后可以得到：

$$\frac{\partial \ln(1 - \sigma(z))}{\partial z} = \sigma z$$

$$\frac{\partial \ln L(w, b)}{\partial w_i} = \sum_n (\hat{y}_n - f_{w,b}(x_n)) x_i^n$$

所以可以得到：

$$w_i \leftarrow w_i - \eta \sum_n [-(\hat{y}_n - f_{w,b}(x))x_i^n]$$

- 关于逻辑回归损失函数补充

如果和回归模型一样，不采用交叉熵，会出现以下问题：

假设损失函数模型为：

$$L(f) = \frac{1}{2} \sum_n (f_{w,b}(x_n) - \hat{y}_n)^2$$

对 $L(f)$ 进行求导：

$$\frac{\partial (f_{w,b}(x) - \hat{y})^2}{\partial w_i} = 2(f_{w,b}(x) - \hat{y})f_{w,b}(x)(1 - f_{w,b}(x))x_i$$

假设 $\hat{y}_n = 1$ ，若 $f_{w,b}(x_n) = 1$ ， $\frac{\partial L}{\partial w_i} = 0$ ，模型的预测结果符合真实结果；若 $f_{w,b}(x_n) = 0$ ，同样可以得到 $\frac{\partial L}{\partial w_i} = 0$ 。

综上，不可以用最小二乘法作为该二分类问题的损失函数。