线性模型

- 基本形式
  给定由 d 个属性描述的示例 $x=(x_1;x_2;\cdots;x_d)$，其中x_i是x在第i个属性上的取值，线性模型试图学得一个通过属性的线性组合来进行预测的函数，即

$$f(x)=w_1x_1+w_2x_2+\ldots+w_dx_d+b$$

- 向量形式

$$f(x)=w^Tx+b$$

单元线性回归

$$f(x)=wx_i+b\text{，使得}f(x_i)\approx y_i$$

性能度量：

$$(w^*,b^*)=\arg\min\nolimits_{(w,b)}\sum_{i=1}^{m}(f(x_i)-y_i)^2=\underset{(w,b)}{\arg\min}\sum_{i=1}^{m}(y_i-wx^i-b)^2$$

对w和b分别求偏导即可得到最优解

多元线性回归

$$f(x_i)=w^Tx_i+b\text{使得}f(x_i)\approx y_i\text{，其中}x_i=(x_{i1};x_{i2};\cdots;x_{id})$$

把w和b吸收入向量形式&\hat{w}=(w;b)&，把数据集D表示为一个$m\times(d+1)$大小的矩阵X，有

$$X=\begin{bmatrix} x_{11} & a_{12} & \cdots & a_{1d} & 1 \\ a_{21} & a_{22} & \cdots & a_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{md} & 1 \end{bmatrix}=\begin{bmatrix} x_1^T & 1 \\ x_2^T & 1 \\ \vdots & \vdots \\ x_m^T & 1 \end{bmatrix}$$

标记写成向量形式$y=(y_1;y_2;\cdots;y_m)$，有

$$\hat{w}^*=\underset{\hat{w}}{\arg\min}(y-X\hat{w})^T(y-X\hat{w})$$

令$E^{\hat{w}}=(y-X\hat{w})^T(y-X\hat{w})$，对$\hat{w}$求导得到

$$\frac{\partial E_{\hat{w}}}{\partial \hat{w}}=2X^T(X\hat{w}-y)$$

当上式为零可得$\hat{w}$最优解的闭式解。

最终得到多元线性回归模型为

$$f(\hat{x}_i)=\hat{x}_i^T(X^TX)^{-1}X^Ty$$

广义线性模型

现实中不可能每次都能用线性模型进行拟合，需要对输出做空间的非线性映射，便可得到广义的线性模型，从线性到非线性

$$y=g^{-1}(w^Tx+b)$$

线性判别分析（LDA）

给定训练样例集，设法将样例投影到一条直线上，使得同类样例的投影点尽可能接近、异类样例的投影点尽可能远离；在对新样本进行分类时，将其投影到同样的这条直线上，再根据投影点的位置来确定新样本的类别。（监督降维技术）

$$\max J=\frac{\left\|w^T\mu_0-w^T\mu_1\right\|_2^2}{w^T\sum_0 w+w^T\sum_1 w}=\frac{w^T(\mu_0-\mu_1)(\mu_0-\mu_1)^Tw}{w^T(\sum_0+\sum_1)w}$$

逻辑回归

用线性回归模型的预测结果去逼近真实标记的对数几率。实际上是一种分类学习办法。

对数几率函数

$$y=\frac{1}{1+e^{-(w^Tx+b)}}$$

变换为

$$\ln\frac{y}{1-y}=w^Tx+b$$

多分类学习

- 多分类学习可在二分类基础上进行。将原问题拆分为多个二分类任务，然后每个二分类训练一个分类器，然后再进行集成获得最终的多分类结果。
- 拆分策略

  1. One vs One
  2. One vs Rest
  3. Many vs Many

类别不平衡问题

- 再放缩

$$\frac{y'}{1-y'} = \frac{y}{1-y} \times \frac{m^-}{m^+}$$

- 解决策略

  1. 欠采样
  2. 过采样
  3. 阈值移动