线性模型

线性回归

有d个属性的示例 $x = (x_1; x_2; ...; x_d)$，线性模型通过属性的线性组合来求预测函数：

$$f(x) = w_1x_1 + w_2x_2 + ... + w_dx_d + b,$$

向量形式：

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

其中$w = (w_1; w_2; ...; w_d)$，w和b确定模型。

有数据集

$$D = (x_1, y_1), (x_2, y_2), ..., (x_m, y_m)$$

其中 $x_i = (x_{i1}; x_{i2}; ...; x_{id})$

先考虑属性数目为1的情况。 线性回归试图学得$f(x_i) = wx_i + b$，使得$f(x_i) \simeq y_i$

如何衡量f(x)和y之间的差别？ **均方误差。**

$$\begin{aligned}（w^*, b^*） &= argmin_{(w,b)} \Sigma_{i=1}^m (f(x_i) - y_i)^2) \\ &= argmin_{(w,b)} \Sigma_{i=1}^m (y_i - wx_i - b)^2\end{aligned}$$

求解w,b变成求
$E_{(w,b)} = \Sigma_{i=1}^m (y_i - w_i - b)^2$最小化的过程，该过程称为线性回归模型的 最小二乘“参数估计”。

由于$E_{(w,b)}$是关于w和b的凸函数。所以讲$E_{(w,b)}$分别对w和b求偏导，令偏导为0即可获得解析解（或叫闭式解）。

- 为什么用最小二乘法?
  - 非常好的几何意义：对应欧氏距离
  - 由最大似然估计可以推导出来

- 什么是凸函数?
  - 碗型
  - $f(\frac{x_1+x_2}{x}) \leq \frac{f(x_1)+f(x_2)}{2}$

多元线性回归

上面是假设属性数目为1，若有d个属性，则：

$$f(x_i) = w^T x_i + b \text{ , 使得} f(x_i) \backsimeq y_i,$$

把w和b吸收入向量形式$\hat{w} = (w; b)$，把数据集D表示成一个$m \times (d+1)$矩阵X。 标记 $y = (y_1; y_2; \ldots; y_m)$，则有：

$$\hat{w}^* = argmin_{\hat{w}} (y - X\hat{w})^T (y - X\hat{w})$$

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

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

上式为零可得$\hat{w}$的最优解的解析解。解出线性回归模型为：

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

但是涉及到矩阵逆运算。 矩阵

- $X^T X$可逆（满秩矩阵或正定矩阵）
- 若不可逆，可解出多个$\hat{w}$，选择哪一个作为输出，需要引入正则化项。

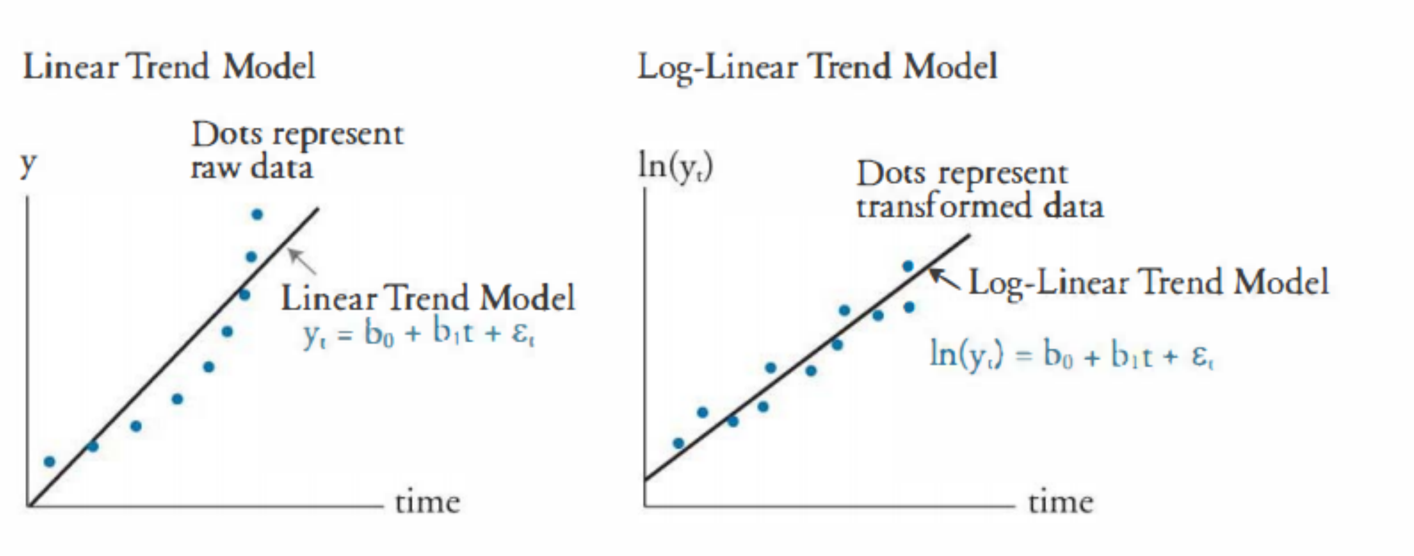
广义线性回归

线性模型的扩展很广。我们把线性模型简写为：

$$y = w^T x + b$$

若令模型预测值逼近y的衍生物，比如lny则就是“对数线性回归”

Figure 1: Linear vs. Log-Linear Trend Models



实际上就是用e^{w^Tx+b}逼近y。实现的效果是输入空间到输出空间的非线性函数映射。

更一般的，若有单调可微函数$g(.)，令：

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

为广义线性模型

LR回归

在线性模型的基础上加一个对率函数：

$$y = \frac{1}{1 + e^{-z}}$$

将z值转换为一个接近0或1的y值，在z=0处变化很陡峭。

对率函数：

- Sigmoid函数，性质，作用。
- 模拟单位跃阶函数

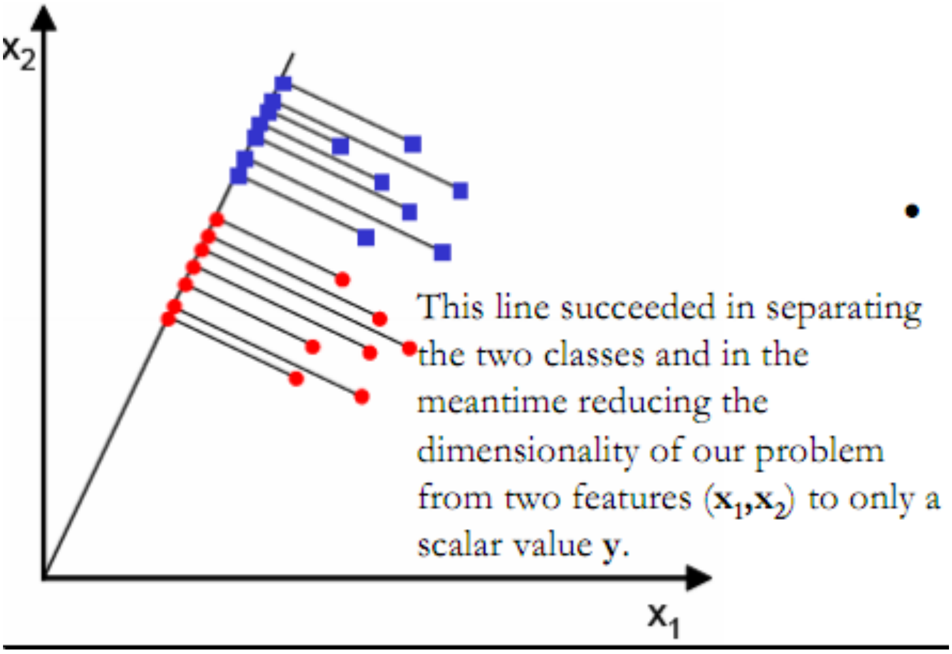
对数几率：

$$ln \frac{y}{1 - y}$$

线性判别分析（LDA）

用于二分类问题。

将样例投影到一条直线上，让同类投影点的尽可能近，异类投影点尽可能远。



- 同类投影点协方差尽量小
- 异类中心距离尽量大

类内散度矩阵。

类间散度矩阵。