回归

线性回归方程 $h(x) = X\cdot w$

即找到回归系数 w，使预测值 $h(x)$ 和真实值 y 之间的误差最小，通常使用平方误差。

$$\therefore Err(w) = \sum_{i=1}^{n}[y_i - h(x_i)]^2 = \sum_{i=1}^{n}[y_i - x_i w]^2$$

$$= [Y - X\cdot w]^T[Y - X\cdot w] = Y^TY - Y^TXw - w^TX^TY + w^TX^TXw$$

对其求导，并令为0，得

$$\frac{d}{dw}Err(w) = \frac{d}{dw}Y^TY - \frac{d}{dw}Y^TXw - \frac{d}{dw}w^TX^TY + \frac{d}{dw}w^TX^TXw$$

$$= 0 - X^TY - \frac{d}{dw}(Y^TXw)^T + 2X^TXw$$

$$= -X^TY - X^TY + 2X^TXw = -2X^TY + 2X^TXw = 0$$

$$\therefore w = (X^TX)^{-1}X^TY$$

上式即 w 的最优解，最后即可得到一条预测直线 $h(x) = X\cdot w$

可利用相关系数来求预测值与真实值的匹配程度

线性回归的一个问题是，会出现欠拟合现象，因为求的是最小均方误差的无偏估计。

基于此，提出局部加权线性回归，给待预测点附近的每个点赋予一定的权重。

$$\therefore \hat{w} = (X^TWX)^{-1}X^TWY$$

其中 W 常用的是高斯核。

$$W(i,i) = \exp\left[\frac{|x^{(i)} - x|^2}{-2k^2}\right]$$

可见，W 是只含对角元素的权重矩阵，并且点 x 与 $x^{(i)}$ 越近，$W(i,i)$ 会越大。

给出测试点 x 与每行 $x^{(i)}$ 作运算，将结果放置在对角线上。

k 是决定了对附近点赋予多大权重。

另外还介绍了两种缩减系数的方法。

① 岭回归。

当 X 的特征比样本多时 $(n>m)$，会导致 X^TX 为奇异矩阵，所以引入 λI 来使其变成非奇异，从而对 $X^TX+\lambda I$ 求逆。

$$\hat{w} = (X^TX + \lambda I)X^TY$$

其中 I 为单位矩阵，λ 为惩罚项，能够缩减不重要的系数。

为了使用岭回归和缩减技术，首先需要对特征做标准化处理

② 前向逐步回归。

只是比岭回归更加高效。

是一个贪心算法，每一步所做的决策是对某个权重增加或减少一个很小的值，通过比较其最小误差来记录每次变动的情况。

当应用缩减技术时，模型会增加了偏差，与此同时却减少了方差

偏差是预测值与真实值的偏离程度，反映了拟合程度。

方差是同等大小的不同测试集对学习器的干扰，即数据干扰，也可以理解为模型复杂度。