线性回归

Luke

2020 年 7 月 1 日

1 最小二乘

对于$y = b + w \cdot x_{cp}$，定义损失函数(Loss function)来度量函数的好坏：

$$L(f) = L(w, b) = \sum_{n=1}^{10} (\hat{y}^n - (b + w \cdot x_{cp}^n))^2 \tag{1}$$

角标n表示第n个数据点。可以预见，当L取最小值时的函数f就是最好的函数，因此对L求导：

$$\frac{\partial}{\partial w} L = \sum_{n=1}^{10} 2(\hat{y}^n - (b + w \cdot x_{cp}^n)) x_{cp}^n = 0 \tag{2}$$

$$\frac{\partial}{\partial b} L = \sum_{n=1}^{10} 2(\hat{y}^n - (b + w \cdot x_{cp}^n)) = 0 \tag{3}$$

代入数据即可求出一组w和b的值。

2 梯度下降

任意取一$L(w_0, b_0)$，计算

$$\frac{\partial L}{\partial b} \mid_{w=w_0, b=b_0}, \frac{\partial L}{\partial w} \mid_{w=w_0, b=b_0}$$

而

$$(-\eta \frac{\partial L}{\partial b}, -\eta \frac{\partial L}{\partial w})$$

的方向就是L下降最快的方向。取

$$b_1 = b_0 - \eta \frac{\partial L}{\partial b}$$

$$w_1 = w_0 - \eta \frac{\partial L}{\partial w}$$

在$L(w_1, b_1)$重复此过程。（η是选定的步长）

3 欠拟合和过拟合

对于同样的100组数据，选择多项式拟合的阶数较少会造成欠拟合，阶数较多会造成过拟合。欠拟合表现为拟合的模型偏差较大 (large bias) 但是方差较小 (small variance)，过拟合表现为拟合的模型偏差较小 (small bias) 但是方差较大 (large variance) 。此处的偏差与方差描述不同训练数据得出的模型的集合与理想模型(图中的靶心)的对比。

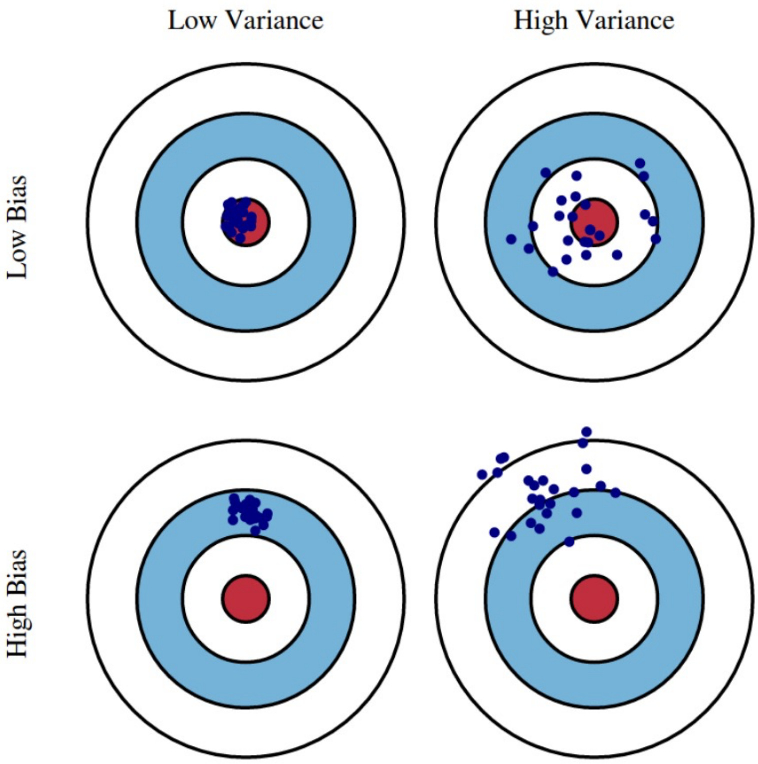

图 1: bias和variance的示意图

通过增加数据的方法可以解决variance过大的问题，而在没有更多数据的时候，可以用regularzation(正则化)的方法减小variance，但是可能会同时增大bias.