一些统计学习的基本概念——概率不等式、泛化边界

张树恒

2018-06-20

模型的假设空间为 $f \in \mathcal{F}$，损失函数为 $\ell(f, X, Y)$，输入输出的联合概率分布为 $Pr(X, Y)$，可以是条件概率分布或者决策函数，则损失函数在整个输入输出空间上的期望为

$$\begin{aligned} R_{\text{exp}}(f) &= \mathrm{E}_{P_{XY}}[\ell(f, X, Y)] \\ &= \int_{X \times Y} \ell(f, x, y) Pr(x, y) \mathrm{d}x \mathrm{d}y \end{aligned} \tag{1}$$

R_{exp} 叫做期望风险，但由于联合概率分布 $Pr(X, Y)$ 是未知的是需要进行学习估计的。现有训练数据集 $\mathcal{D}$，根据该训练数据集会得到一个数据集的平均损失叫做经验风险

$$R_{\text{emp}}(f) = \frac{1}{N} \sum_{i=1}^{N} \ell(f, X, Y) \tag{2}$$

根据大数定理，当样本容量 N 趋于无穷时，经验风险 R_{exp} 趋于期望风险 R_{emp}

1

Eq. (1) 可以看成是关于 f 的能量泛函，我们要做的就是最小化该能量泛函，得到最优的 f，故把期望误差风险转换成条件概率的形式

$$\begin{aligned} R_{\text{exp}}(f) &= \int_{X \times Y} \ell(f, x, y) Pr(x, y) \mathrm{d}x \mathrm{d}y \\ &= \iint \ell(f, x, y) p(y \mid x) \mathrm{d}y \cdot p(x) \mathrm{d}x \\ &= \int E_{Y|x}[\ell(f, x, Y)] \cdot p(x) \mathrm{d}x \\ &= E_X[E_{Y|X}[\ell(f, X, Y)]] \end{aligned} \tag{3}$$

对其求偏导

$$\begin{aligned} \frac{\delta R_{\text{exp}}}{\delta f} &= \int \frac{\delta \ell(f, x, y)}{\delta f} p(y \mid x) \mathrm{d}y \cdot p(x) \\ &= E_{Y|x}[\frac{\delta \ell(f, x, Y)}{\delta f}] p(x) \end{aligned} \tag{4}$$

损失函数取方差 $\ell(f, X, Y) = (Y - f(X))^2$，

$$\frac{\delta R_{\text{exp}}}{\delta f} = \int 2(f(x) - y) p(y \mid x) \mathrm{d}y \cdot p(x) = 0 \tag{5}$$

整理得

$$\begin{aligned}\int f(x)p(y\mid x)\mathrm{d}y &= \int yp(y\mid x)\mathrm{d}y \\ f(x) &= E_{Y|x}[Y] \\ &= E[Y\mid X=x]\end{aligned} \tag{6}$$

因此，Y 的最优的预测是点 x 处的条件均值。则单纯基于统计的或者近邻算法均是根据该原理，计算每一点处的 y 对应的条件均值。

2 泛化边界