课程作业

杨佳乐

2021 年 9 月 30 日

题目 1. $min_k \|t_k - \hat{y}\|$

解答.

$$argmin_k \|t_k - \hat{y}\| = argmin_k (t_k - \hat{y})^2$$

$$= argmin_k (t_k^2 - 2t_k\hat{y} + \hat{y}^2)$$

$$= argmin_k (t_k^2 - 2t_k\hat{y})$$

$$= argmin_k \sum_{i=1}^{K} (t_k)_i^2 - 2(t_k)_i\hat{y}_i$$

因为 $\sum_{i=1}^{K} (t_k)_i^2 = 1$, $\sum_{i=1}^{K} 2(t_k)_i y_i = 2y_k$，所以，

$$argmin_k \|t_k - \hat{y}\| = argmin_k (-2y_k)$$

$$= argmax_k (y_k)$$

题目 2. 推导 The median distance from the origin to the closest data point 公式

解答. 将 $d(p,N)$ 记为 r，用 D_{min} 表示到原点距离最近的数据点, 则:

$$P(D_{min} >= r) = \frac{1}{2}$$

用 $\{X_1, X_2, ...X_N\}$ 表示所有点到原点的距离，可以推出:

$$P(X_1 >= r \cap X_2 >= r \cap ... \cap X_N >= r) = \frac{1}{2}$$

因为 $\{X_1, X_2, ...X_N\}$ 是独立随机变量，所以:

$$\prod_{i=1}^{N} P(X_i >= r) = \frac{1}{2}$$

半径为 r 的高维球的体积和高维单位球体积的比值为 $\frac{r^p}{1}$，可以推导出:

$$P(X_i >= r) = 1 - r^p, i \in 1, 2, ..., N$$

$$(1 - r^p)^N = \frac{1}{2}$$

所以

$$r = d(p,N) = (1 - \frac{1}{2}^{\frac{1}{N}})^{\frac{1}{p}}$$

题目 3. Show that z_i is distributed $N(0,1)$ and show that the squared distance from any sample points to the origin has a χ_p^2 distribution with mean p.

解答. 已知 x_0，关联向量 $a = \frac{x_0}{\|x_0\|}$，$z = \sum_{i=1}^{p} a_i x_i = a^T x$。

因为 x $N(0,1)$, 所以 $E(z) = E(a^T x) = a^T E(x) = a^T 0 = 0$

$Var(z) = Var(a^T x) = a^T Var(x) a = a^T I_p a = a^T a = \frac{x_0^T x_0}{\|x_0\|^2} = 1$ 独立正态分布随机变量的线性组合服从正态分布，以上证明了 z 的均值为 0，方差为 1，所以 z $N(0,1)$

点 x 到原点的距离 $x^t x = \sum_{i=1}^{p}$。x 向量的每一分量 x_i 独立且服从标准正态分布，所以距离服从卡方分布 χ_p^2

题目 4. 2.7

解答. (a)

k 近邻：

$Nearest_k(x_0)$ 表示离 x_0 最近的 k 个点的集合，定义：

$$l_i(x_0;\chi) = \begin{cases} \frac{1}{k} & if x_i \in Nearest(x_0) \\ 0 & otherwise \end{cases}$$

可以得到函数在 x_0 的估计：

$$f(x) = \sum_{i=1}^{N} l_i(x_0,\chi)y_i = \frac{1}{k}\sum_{if x_i \in Nearest_k(x_0)} y_i$$

线性回归：

可以假设 $\hat{f}(x_0) = x_0^T\hat{\beta}$，其中 $\hat{\beta} = (X^TX)^{-1}X^Ty$。所以，$\hat{f}(x_0) = x_0^T(X^TX)^{-1}X^Ty$。

将 $x_0^T(X^TX)^{-1}X^T$ 看作 $l(x_0,\chi)$，它是一个 $1\times N$ 的向量，将它的每一行看作 $l_i(x_0,\chi)$，即可得到题目所给形式。

(b)

$$E_{y|x}[(f(x_0) - \hat{f}(x_0))^2] = E_{y|x}[(f(x_0) - E_{y|x}[\hat{f}(x_0)])^2]$$

$$+E_{y|x}[(E_{y|x}[\hat{f}(x_0)] - \hat{f}(x_0))^2]$$

$$+E_{y|x}[2(f(x_0) - E_{y|x}[\hat{f}(x_0)])(E_{y|x}[\hat{f}(x_0)] - \hat{f}(x_0))]$$

$$Bias^2(\hat{f}(x_0)) = E_{y|x}[(f(x_0) - E_{y|x}[\hat{f}(x_0)])^2] = (f(x_0) - E_{y|x}[\hat{f}(x_0)])^2$$

$$Var(\hat{f}(x_0)) = E_{y|x}[(E_{y|x}[\hat{f}(x_0)] - \hat{f}(x_0))^2]$$

第三项等于 $2(f(x_0) - E_{y|x}[\hat{f}(x_0)]) \times E_{y|x}[(E_{y|x}[\hat{f}(x_0)] - \hat{f}(x_0))] = 2(f(x_0) - E_{y|x}[\hat{f}(x_0)]) \times (E_{y|x}[\hat{f}(x_0)] - E_{y|x}[\hat{f}(x_0)]) = 0$，完成偏差方差分解。

(c)

$$E_{X,Y}[(f(x_0)-\hat{f}(x_0))^2] = E_X[E_{Y|X}[(f(x_0)-\hat{f}(x_0))^2]]$$
$$= E_X[Bias^2(\hat{f}(x_0)) + Var(\hat{f}(x_0))]$$
$$= \int Bias^2(\hat{f}(x_0))(\prod_{i=1}^{N} h(x_i))dx_1...dx_N + \int Var(\hat{f}(x_0))(\prod_{i=1}^{N} h(x_i))dx_1...dx_N$$

(d)

上一小题结果里面 $\hat{f}(x_0)$ 是在训练数据集 χ 下的预测值，偏差平方和方差（在训练数据集为 χ 下）之和就等于期望平方损失（Expected Squared Error）。