ZHAOYANG XU + CHM564

December 10, 2019

Contents

1 The Role of Independence

For the first time, we can get

$$P(X_1 = 1) = \frac{1}{2}$$
$$P(X_1 = 0) = \frac{1}{2}$$

If the first time we get 1, then all the other $X_i = 1$;if we get 0, all the other $X_i = 0$.

$$P(X_i = 1|X_1 = 1) = 1 \qquad i \in 2, 3, 4, ..., n$$
$$P(X_i = 0|X_1 = 0) = 1 \qquad i \in 2, 3, 4, ..., n$$

We can calculate μ:

$$\mu = E[X_i] = 1 * \frac{1}{2} + 0 * \frac{1}{2}$$
$$P(|\mu - \frac{1}{n}\sum_{i=1}^{n} X_i|) = P(|\frac{1}{2} - 1 or 0| \geq \frac{1}{2}) = 1$$

so,we get

$$P(|\mu - \frac{1}{n}\sum_{i=1}^{n} X_i| \geq \frac{1}{2}) = 1$$

2 How to Split a Sample into Training and Test Set

1. We can derive

$$P(L(\hat{h}^*_{S_train}) \leq \hat{L}(\hat{h}^*_{S_train}, S_{test}) + \sqrt{\frac{ln\frac{1}{\delta}}{2n_{test}}}) \geq 1 - \delta$$

2. for $i \in \{1, 2, ..., m\}$:

$$P(L(\hat{h}^*_i) \leq \hat{L}(\hat{h}^*_i, S_{test}) + \sqrt{\frac{ln\frac{1}{\delta}}{2n_{test}}}) \geq 1 - \delta$$

for $\forall i \in \{1, 2, ..., m\}$

$$P(\forall i \in \{1,2,...,m\} : L(\hat{h}_i^*) \leq \hat{L}(\hat{h}_i^*, S_{test}) + \sqrt{\frac{ln\frac{1}{\delta}}{2n_{test}}}) = P(\bigcap_i^m (L(\hat{h}_i^*) \leq \hat{L}(\hat{h}_i^*, S_{test}) + \sqrt{\frac{ln\frac{1}{\delta}}{2n_{test}}}))$$
$$= P(L(\hat{h}_i^*) \leq \hat{L}(\hat{h}_i^*, S_{test}) + \sqrt{\frac{ln\frac{1}{\delta}}{2n_{test}^*}})$$
$$(when\ n_{test}^*\ is\ the\ biggest\ n_{test}(i))$$

3. We assume $\pi(h) \geq 0$

$$P(L(\hat{h}_{S_train}^*) \leq \hat{L}(\hat{h}_{S_train}^*, S_{test}) + \sqrt{\frac{ln\frac{1}{\delta\pi(h)}}{2n_{test}}}) \geq 1 - \delta$$

The modelss trained on more data, performence became better. So,$\sqrt{\frac{ln\frac{1}{\delta\pi(h)}}{n_{test}}}$ will become smaller.$\pi(h) \geq 0$ should be decreased with n_{test}
Let $\pi(h) = \frac{1}{2^{n_{test}}}$:

$$P(L(\hat{h}_{S_train}^*) \leq \hat{L}(\hat{h}_{S_train}^*, S_{test}) + \sqrt{\frac{ln\frac{2^{n_{test}}}{\delta}}{n_{test}}}) \geq 1 - \delta$$

3 Occam's Razor

1. For $h \in H_d$

$$P(\forall h \in H_d : L(h) \leq \hat{L}(h) + \sqrt{\frac{ln\frac{M}{\pi(h)\delta}}{2n}}) \geq 1 - \delta$$
$$P(\forall h \in H_d : L(h) \leq \hat{L}(h) + \sqrt{\frac{\frac{ln(2^{2^{d(h)}})}{\delta}}{2n}}) \geq 1 - \delta$$

$|\sum| = 27$

$$P(\forall h \in H_d : L(h) \leq \hat{L}(h) + \sqrt{\frac{\frac{ln(2^{2^27})}{\delta}}{2n}}) \geq 1 - \delta$$

2. For $h \in H$

$$P(\forall h \in H : L(h) \leq \hat{L}(h) + \sqrt{\frac{\frac{ln(2^{2^{d(h)}} \cdot 2^{d(h)+1})}{\delta}}{2n}}) \geq 1 - \delta$$
$$P(\forall h \in H : L(h) \leq \hat{L}(h) + \sqrt{\frac{\frac{ln(2^{2^{27}} \cdot 2^{28})}{\delta}}{2n}}) \geq 1 - \delta$$

3. From the inequality

$$P(\forall h \in H_d : L(h) \leq \hat{L}(h) + \sqrt{\frac{ln\frac{M}{\pi(h)\delta}}{2n}}) \geq 1-\delta$$

with d increased, $\pi(h)$decreased, $ln\frac{M}{\pi(h)\delta}$increased, $\sqrt{\frac{ln\frac{M}{\pi(h)\delta}}{2n}}$increased.

4. Let $M = |H| = 2^{2n}$

$$\begin{aligned}\sqrt{\frac{ln\frac{M}{\pi(h)\delta}}{2n}} &= \sqrt{\frac{ln\frac{2^{2n}}{\pi(h)\delta}}{2n}} \\ &\geq \sqrt{\frac{ln2^{2n}}{2n}} \\ &= \sqrt{ln2} \\ &\geq 0.8 \geq 0.25\end{aligned}$$

so,there is no contradiction.

4 Kernels

1. We assume $\phi(x) = (x_{ij}) \quad \phi(z) = (z_{ij})$
so

$$\begin{aligned}\|\phi(x) - \phi(z)\| &= \sqrt{\sum_i \sum_j (x_{ij} - z_{ij})^2} \\ &= \sqrt{\sum_i \sum_j ({x_{ij}}^2 - 2x_{ij}z_{ij} + {z_{ij}}^2)} \\ &= \sqrt{\sum_i \sum_j {x_{ij}}^2 - 2\sum_i \sum_j x_{ij}z_{ij} + \sum_i \sum_j {z_{ij}}^2}\end{aligned}$$

Because its defining on RKHS, so the second term is symmetrical.

$$\begin{aligned}&= \sqrt{\langle\phi(x),\phi(x)\rangle - 2\langle\phi(x),\phi(z)\rangle + \langle\phi(z),\phi(z)\rangle} \\ &= \sqrt{k(x,x) - 2k(x,z) + k(z,z)}\end{aligned}$$

2. We know k_1, k_2 are positive-definite kernels. So,

$$\begin{aligned}K_1 &= k_1(x,z) \qquad && \forall c_1, c_2 ... \in R : \sum c_x c_z K_1 \geq 0 \\ K_2 &= k_2(x,z) \qquad && \forall c_1, c_2 ... \in R : \sum c_x c_z K_2 \geq 0\end{aligned}$$

From these, we can get

$$\begin{aligned} \forall c_1, c_2 ... \in R : \sum c_x c_z K_1 + \sum c_x c_z K_2 \geq 0 \\ \sum c_x c_z (K_1 + K_2) \geq 0 \end{aligned}$$

So $K_1 + K_2 = K = k(x,z), \quad k(x,z)$ is positive-definite.

3. The Gram matrix from m input patterns is $m * m$. So the maximum rank of Gram matrix is m.