ML Assignment 3

Philip Lassen vgh804

December 2018

1 The Role of Independence

2 To Split or not to Split

1. Applying the inequality (3.4) from the lecture notes we can bound $L(\hat{h^*})$ as follows.

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

2. Since we are only using a single Hypothesis for each of the divisions, we can use the alternative formulation of equation (3.1) from the lecture notes.

$$P(L(\hat{h^*}) \leq \hat{L}(\hat{h^*}, S_{Val}^i) + \sqrt{\frac{ln\frac{1}{\delta}}{\frac{2n}{M}}}) \geq 1 - \delta$$

3. (a) Using the process described in 3 gives us a result of

$$P(L(\hat{h^*}) \leq \hat{L}(\hat{h^*}, S_{Val}) + \sqrt{\frac{ln\frac{1}{\delta}}{n}}) \geq 1 - \delta$$

   (b) We need to compare the values $\sqrt{\frac{ln\frac{1}{\delta}}{n}}$ and $\sqrt{\frac{ln\frac{M}{\delta}}{2n}}$. Doing this comparison we get

$$\sqrt{\frac{ln\frac{1}{\delta}}{n}} \geq \sqrt{\frac{ln\frac{M}{\delta}}{2n}}$$
$$\frac{ln\frac{1}{\delta}}{n} \geq \frac{ln\frac{M}{\delta}}{2n}$$
$$2ln\frac{1}{\delta} \geq ln\frac{M}{\delta}$$
$$(\frac{1}{\delta})^2 \geq \frac{M}{\delta}$$
$$\frac{1}{\delta} \geq M$$

Thus we can say that (1) gives a tighter a bound when M is less than $\frac{1}{\delta}$ and (3) gives a tighter bound otherwise.

(c) One of the draw back could be depending on on how the data is split there may be a possibility that the data looks different the first half of the data it was trained on. This could happen if we make an unlucky split.

4. One a small value of α is chosen we increase the probabiliy that we choose a hypothesis that doesn't best fit the data. And if α is too large then we increase Variance of the expected loss.

3 Occam's Razor

4 Kernels

1.

$$\begin{aligned}||\phi(x) - \phi(z)||^2 &= \phi(x)\phi(x) - 2\phi(x)\phi(z) + \phi(z)\phi(z) \\ &= k(x,x) + k(z,z) - 2k(x,z)\end{aligned}$$

Taking the square root of both sides gives us

$$||\phi(x) - \phi(z)|| = \sqrt{k(x,x) + k(z,z) - 2k(x,z)}$$

2. We know from the definition of a positive definite Kernel that

$$\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} c_x c_z k_1(x,z) \geq 0 \tag{1}$$

and

$$\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} c_x c_z k_2(x,z) \geq 0 \tag{2}$$

It then follows that

$$\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} c_x c_z k(x,z) = \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} c_x c_z (k_1(x,z) + k_2(x,z)) \tag{3}$$

$$= \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} c_x c_z k_1(x,z) + \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} c_x c_z k_2(x,z) \tag{4}$$

Substituting Equation (1) and (2) into equation (4) we get

$$\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} c_x c_z k(x,z) \geq 0$$

Thus we have shown k is positive definite.

3. Let X be the matrix where the columns are the m input vector. Then we can write the Gram matrix as follows

$$G = X^T X$$

I will start by showing the null space of $X^T X$ and X are the same

$$\begin{aligned} X^T X z = 0 &\implies z X^T X z = 0 &(5)\\ &\implies ||Xz||^2 = 0 &(6)\\ &\implies Xz = 0 &(7)\end{aligned}$$

The other direction proceeds as follows

$$Xz = 0 \implies X^T X z = 0$$

Thus we have shown their nullspaces are the same. Since the dimension of their nullspaces are the same and the dimension of X and $(X^T X)$ is the same, from the rank nullity theorem it follows that the rank$(X^T X) =$ rank(X). We can bound the Rank of X with min(m, d) thus a bound of rank for $(X^T X)$ is min(m, d).