作业4

2017年4月17日 8:10

1. 注意 Deterministic Noise 的定义：
   difference between best $h^* \in \mathcal{H}$ and f
   当于假设空间变小，更可能找不到合适的 h，Deterministic Noise 会变大

2. $\sum_{q=0}^{2} w_q L_q \cap \sum_{q=0}^{3} w_q L_q = \sum_{q=0}^{2} w_q L_q = \mathcal{H}_2$

3. $E_{aug}(w) = E_{in}(w) + \frac{\lambda}{N} w^T w$

   对 w 求导 $\nabla E_{aug}(w) = \nabla E_{in}(w) + \frac{2\lambda}{N} w$
   按梯度下降法更新

$$w(t+1) \leftarrow w(t) - \eta (\nabla E_{aug}(w))$$
$$= w(t) - \eta (\nabla E_{in}(w(t)) + \frac{2\lambda}{N} w(t))$$
$$= (1 - \frac{2\lambda\eta}{N}) w(t) - \eta \nabla E_{in} w(t)$$

$\alpha = 1 - \frac{2\lambda\eta}{N}$ $\beta = -\eta$

4. 正则化等价于 $\min E_{in}(w) = \frac{1}{N}(Zw - y)^T(Zw - y)$ s.t. $w^T w \le C$
   若 W_{lin} 满足 $w^T w \le C$，则 $W_{reg} = W_{lin}$ $\|W_{reg}\|$ 不随 λ 变化，
   若 W_{lin} 不满足 $w^T w \le C$，此时 $\lambda \uparrow$，$C \downarrow$ $\|W_{reg}\| \downarrow$
   (此说明只是示意性的)

5. $E_{loocv}(\mathcal{H}_0, \mathcal{A}) = \frac{1}{N} \sum_{n=1}^{N} err(g_n^-(x_n), y_n)$

   $g_1^-(x) = \frac{1}{2}$ $g_2^-(x) = 0$ $g_3^- x = \frac{1}{2}$

$$E_{loocv} = \frac{(\frac{1}{2} - 0)^2 + (0 - 1)^2 + (\frac{1}{2} - 0)^2}{3} = \frac{1}{2}$$

对 $\mathcal{H}_1$ $x_1^- = \begin{pmatrix} 1 & \rho \\ 1 & 1 \end{pmatrix}$ $y_1^- = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $w_1^- = (x_1^{-T} x_1^-)^{-1} x_1^{-T} y_1^-$

$x_2^- = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ $y_2^- = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $w_2^- = (x_2^{-T} x_2^-)^{-1} x_2^{-T} y_2^-$

$x_3^- = \begin{pmatrix} 1 & -1 \\ 1 & \rho \end{pmatrix}$ $y_3^- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $w_3^- = (x_3^{-T} x_3^-)^{-1} x_3^{-T} y_3^-$

$$E_{loocv} = \frac{(x_1^T w_1^- - y_1)^2 + (x_2^T w_2^- - y_2)^2 + (x_3^T w_3^- - y_3)^2}{3}$$

令二者相等，求得 $\rho = \sqrt{9 + 4\sqrt{6}}$

6. ① 给32个人发第一场比赛的预测，一半预测赢，一半预测输
② 给①中已输的16人发第二场比赛的预测，一半预测赢，一半预测输
③ 依此类推，每次减半
共需邮件 $32+16+8+4+2$

7. 共发邮件 $32+16+8+4+2+1=63$，赚 $1000-63\times 10=370$

8. 算法是数学推导的确定性公式，$M=1$

9. $P_D(BAD\ D) \le 2M\exp(-2\epsilon^2 N) = 0.271$

10. 训练数据来自 $a(x)$，是有偏的，要保证实际测试时与训练时偏向的一致，故可以先用 $a(x)$ 筛选，再用 $g(x)$ 筛选

11. $E_{in}(w) = \frac{1}{N+K} = \frac{1}{N+K}\left(wX^TXw - 2w^TX^Ty + y^Ty + w\tilde{X}^T\tilde{X}w - 2w^T\tilde{X}^T\tilde{y} + \tilde{y}^T\tilde{y}\right)$

$\nabla E_{in}(w) = \frac{1}{N+K}\left(X^TXw - X^Ty + \tilde{X}^T\tilde{X}w - \tilde{X}^T\tilde{y}\right)$

$\nabla E_{in}(w) = 0 \Rightarrow w = (X^TX + \tilde{X}^T\tilde{X})^{-1}(X^Ty + \tilde{X}^T\tilde{y})$

12. $w_{reg} = (X^TX + \lambda I)^{-1}X^Ty$

与11中式子相等，则 $\begin{cases}\tilde{X}^T\tilde{X} = \lambda I \\ \tilde{X}^T\tilde{y} = 0\end{cases}$ 比如可令 $\begin{cases}\tilde{X} = \sqrt{\lambda} I \\ \tilde{y} = 0\end{cases}$ 满足条件

13～20 参见代码

13. $\lambda=11.26$ 时 $E_{in}=0.055$ $E_{out}=0.052$
14. $index=-8$ $E_{in}=0.015$ $E_{out}=0.02$
15. $index=-7$ $E_{in}=0.03$ $E_{out}=0.015$
16. $index=-8$ $E_{train}=0.0$ $E_{val}=0.05$ $E_{out}=0.025$
17. $index=0$ $E_{train}=0.0333$ $E_{val}=0.0375$ $E_{out}=0.0280$
18. $E_{in}=0.035$ $E_{out}=0.02$
19. $index=-8$ $E_{cv}=0.03$

201 $Z_{in} = 0.015$ $Z_{out} = 0.02$