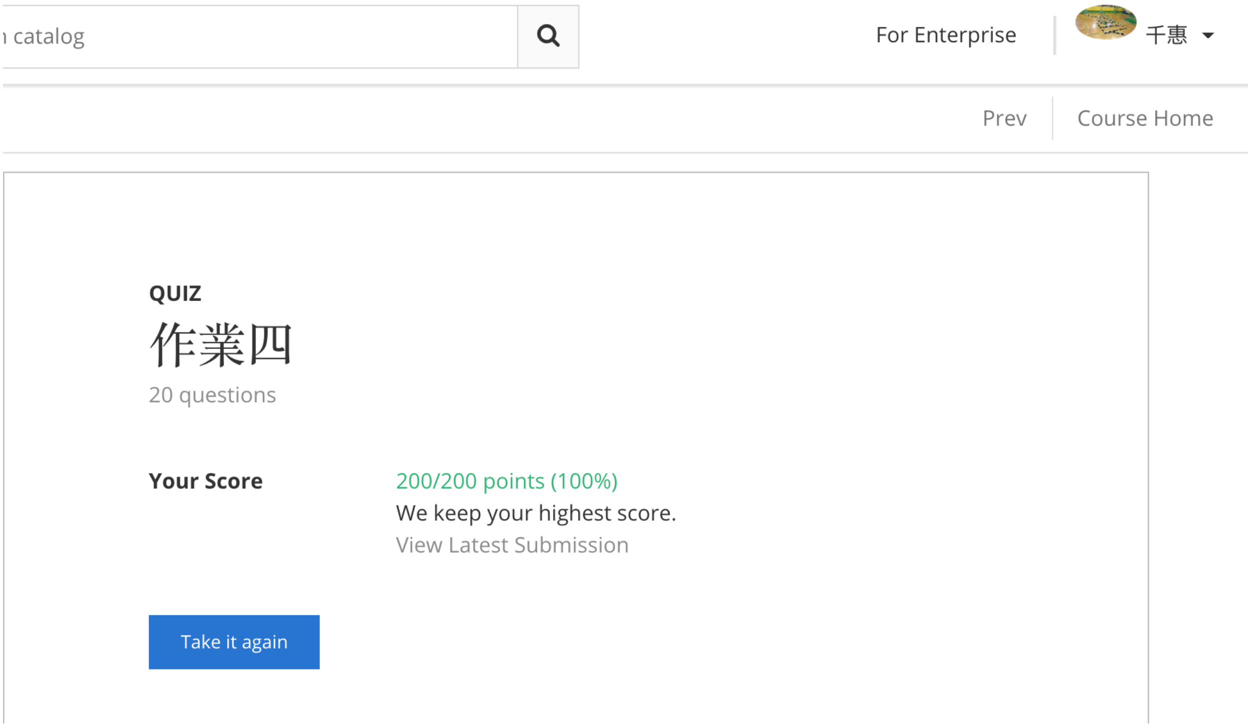
Homework #4

資工三 B04902009 蕭千惠

17th January, 2018

Problem 1

Score: 200 / 200



Problem 2

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

$$\Rightarrow \nabla E_{aug}(\mathbf{w}) = \nabla E_{in}(\mathbf{w}) + \frac{2\lambda}{N}\mathbf{w}$$

$$\mathbf{w}(t+1) \leftarrow \mathbf{w}(t) - \eta \nabla E_{aug}(w)$$

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

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

Problem 3

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

$\|\mathbf{w}_{reg}(\lambda)\|^2 = \mathbf{w}^T\mathbf{w} = C$

When $\lambda = 0, \mathbf{w}_{reg}(\lambda) = \mathbf{w}_{lin} \Rightarrow \sqrt{C} = \|\mathbf{w}_{reg}(\lambda)\| = \|\mathbf{w}_{lin}\|$

Larger $\lambda \Leftrightarrow$ prefer shorter $\mathbf{w} \Leftrightarrow$ effectively smaller C

$\Rightarrow$ For any $\lambda > 0, \sqrt{C} = \|\mathbf{w}_{reg}(\lambda)\| \leq \|\mathbf{w}_{lin}\|$

Problem 4

$E_{loocv} = \frac{1}{3}(e_1 + e_2 + e_3)$

1. Model: constant $h_0(x) = b_0$

   $g_1^- = \frac{1}{2} \qquad g_2^- = 0 \qquad g_3^- = \frac{1}{2}$

   $e_1 = err(g_1^-(-1), 0) = err(\frac{1}{2}, 0) = (\frac{1}{2})^2 = \frac{1}{4}$

   $e_2 = err(g_2^-(-1), 1) = err(0, 1) = 1$

   $e_3 = err(g_3^-(-1), 0) = err(\frac{1}{2}, 0) = (\frac{1}{2})^2 = \frac{1}{4}$

   $\Rightarrow E_{loocv}(h_0) = \frac{1}{3}(\frac{1}{4} + 1 + \frac{1}{4}) = \frac{1}{2}$

2. Model: linear $h_1(x) = a_1 x + b_1$

   $g_1^- = \frac{1}{\rho - 1}x - \frac{1}{\rho - 1} \qquad g_2^- = 0 \qquad g_3^- = \frac{1}{\rho + 1}x + \frac{1}{\rho + 1}$

   $e_1 = err(g_1^-(-1), 0) = err(\frac{2}{\rho - 1}, 0) = (\frac{2}{\rho - 1})^2$

   $e_2 = err(g_2^-(\rho), 1) = err(0, 1) = 1$

   $e_3 = err(g_3^-(1), 0) = err(\frac{2}{\rho + 1}, 0) = (\frac{2}{\rho + 1})^2$

   $\Rightarrow E_{loocv}(h_1) = \frac{1}{3}((\frac{2}{\rho - 1})^2 + 1 + (\frac{2}{\rho + 1})^2)$

$E_{loocv}(h_0) = E_{loocv}(h_1)$

$\Rightarrow \frac{1}{2} = \frac{1}{3}((\frac{2}{\rho - 1})^2 + 1 + (\frac{2}{\rho + 1})^2)$

$\Rightarrow \rho = \sqrt{9 + 4\sqrt{6}}$

Problem 5

$E_{in}(\mathbf{w}_{lin}) = \min_{\mathbf{w}} \frac{1}{N+K}\Big(\sum_{n=1}^{N}(y_n - \mathbf{w}^T\mathbf{x}_n) + \sum_{k=1}^{K}(\tilde{y}_k - \mathbf{w}^T\tilde{\mathbf{x}}_k)\Big)$

$E_{in}(\mathbf{w}_{lin}) = \min_{\mathbf{w}} \frac{1}{N+K}[(\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} + 2\mathbf{w}^T\mathbf{X}^T\mathbf{y} + \mathbf{y}^T\mathbf{y}) + (\mathbf{w}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\mathbf{w} + 2\mathbf{w}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{y}} + \tilde{\mathbf{y}}^T\tilde{\mathbf{y}})]$

$\Rightarrow \nabla E_{in}(\mathbf{w}_{lin}) = \frac{2}{N+K}(\mathbf{X}^T\mathbf{X}\mathbf{w}_{lin} - \mathbf{X}^T\mathbf{y} + \tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\mathbf{w}_{lin} - \tilde{\mathbf{X}}^T\tilde{\mathbf{y}}) = 0$

$\Rightarrow (\mathbf{X}^T\mathbf{X} + \tilde{\mathbf{X}}^T\tilde{\mathbf{X}})\mathbf{w}_{lin} = \mathbf{X}^T\mathbf{y} + \tilde{\mathbf{X}}^T\tilde{\mathbf{y}}$

$\Rightarrow \mathbf{w}_{lin} = (\mathbf{X}^T\mathbf{X} + \tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}(\mathbf{X}^T\mathbf{y} + \tilde{\mathbf{X}}^T\tilde{\mathbf{y}})$

Problem 6

$\mathbf{w}_{lin} = (\mathbf{X}^T\mathbf{X} + \tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}(\mathbf{X}^T\mathbf{y} + \tilde{\mathbf{X}}^T\tilde{\mathbf{y}})$

$\mathbf{w}_{reg} = \operatorname*{argmin}_{\mathbf{w}} \frac{\lambda}{N}\|\mathbf{w}\|^2 + \frac{1}{N}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2$

$\Rightarrow \frac{2\mathbf{w}_{reg}\lambda}{N} + \frac{2}{N}(\mathbf{X}^T\mathbf{X}\mathbf{w}_{reg} - \mathbf{X}^T\mathbf{y}) = 0$

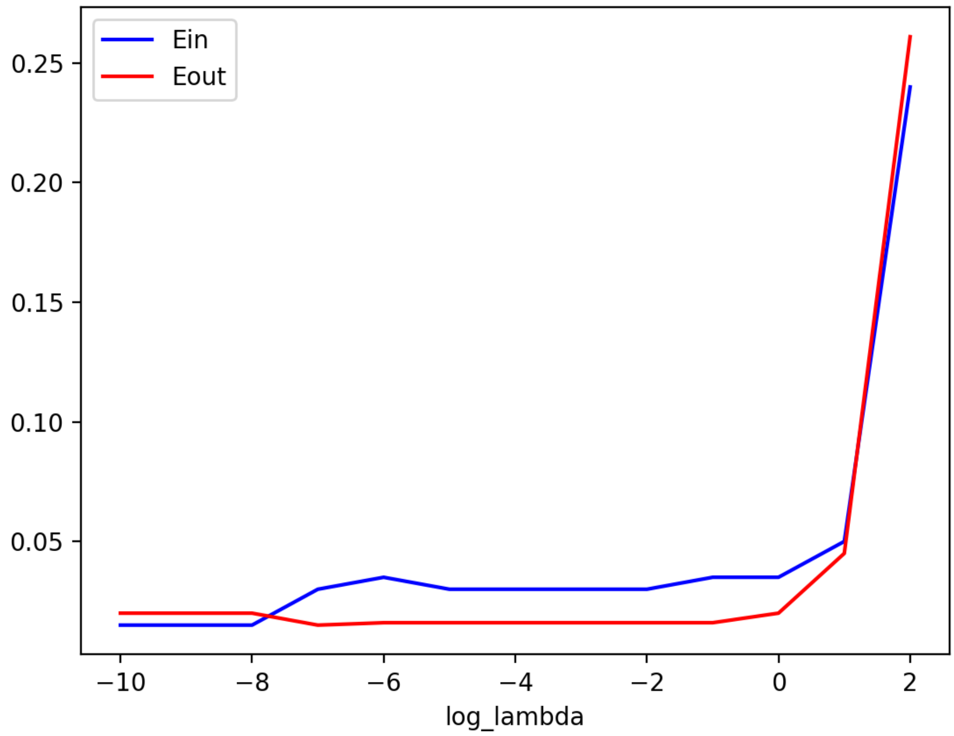
$\Rightarrow (\lambda\mathbf{I} + \mathbf{X}^T\mathbf{X})\mathbf{w}_{reg} - \mathbf{X}^T\mathbf{y} = 0$

$\Rightarrow \mathbf{w}_{reg} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$

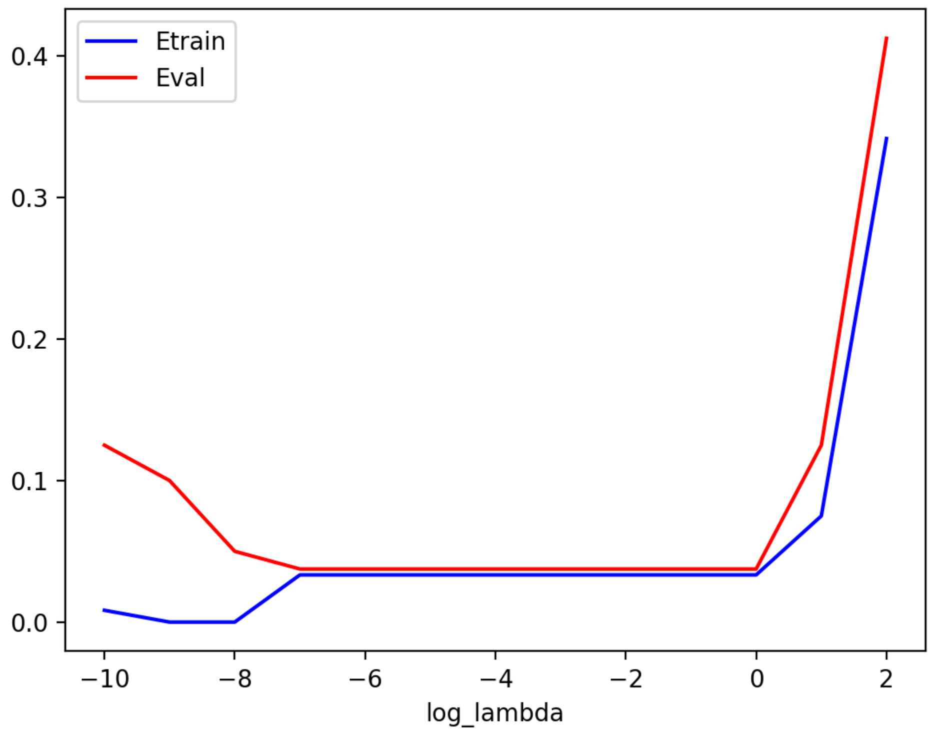
$\mathbf{w}_{lin} = \mathbf{w}_{reg}$

$\Rightarrow \tilde{\mathbf{X}} = \sqrt{\lambda}\mathbf{I}, \tilde{\mathbf{y}} = 0$

Problem 7



Problem 8



Problem 9(Bonus)

(a)

1. Algorithm $\mathcal{A}_{majority}$

   Let $g \in \mathcal{A}_{majority}$, $\begin{cases} g = +1 \text{ when the majority class is positive} \\ g = -1 \text{ when the majority class is negative} \end{cases}$

   $$E_{loocv}(\mathcal{A}_{majority}) = \frac{1}{2252}\sum_{i=1}^{2252} e_i = \frac{1}{2252}\sum_{i=1}^{2252} err(g_i^-, y_i)$$

   $$\begin{cases} x_i > 0 \Rightarrow g_i^- = -1, \;\; y_i = +1 \Rightarrow e_i = 1 \\ x_i < 0 \Rightarrow g_i^- = +1, \;\; y_i = -1 \Rightarrow e_i = 1 \end{cases}$$

   $$\Rightarrow E_{loocv}(\mathcal{A}_{majority}) = \frac{1}{2252}\sum_{i=1}^{2252} 1 = 1$$

2. Algorithm $\mathcal{A}_{minority}$

   Let $g \in \mathcal{A}_{minority}$, $\begin{cases} g = +1 \text{ when the minority class is positive} \\ g = -1 \text{ when the minority class is negative} \end{cases}$

   $$E_{loocv}(\mathcal{A}_{minority}) = \frac{1}{2252}\sum_{i=1}^{2252} e_i = \frac{1}{2252}\sum_{i=1}^{2252} err(g_i^-, y_i)$$

   $$\begin{cases} x_i > 0 \Rightarrow g_i^- = +1, \;\; y_i = +1 \Rightarrow e_i = 0 \\ x_i < 0 \Rightarrow g_i^- = -1, \;\; y_i = -1 \Rightarrow e_i = 0 \end{cases}$$

   $$\Rightarrow E_{loocv}(\mathcal{A}_{minority}) = \frac{1}{2252}\sum_{i=1}^{2252} 0 = 0$$

$\Rightarrow \mathcal{A}_{minority}$ algorithm should be chosen if we use E_{loocv} for algorithm selection.

(b)

Let $g \in \mathcal{A}_{average}$ predicts the average value within the data set that it sees.

$$E_{loocv}(\mathcal{A}_{average}) = \frac{1}{N}\sum_{i=1}^{N} e_i = \frac{1}{N}\sum_{i=1}^{N} err(g_i^-, y_i)$$

$$\forall x_i, \begin{cases} y_i = x_i \\ g_i^- = \dfrac{N\bar{x} - x_i}{N-1} \end{cases}$$

$$\Rightarrow err(g_i^-, y_i) = (x_i - \frac{N\bar{x} - x_i}{N-1})^2 = (\frac{N}{N-1}(x_i - \bar{x}))^2$$

$$\Rightarrow E_{loocv}(\mathcal{A}_{average}) = \frac{N^2}{(N-1)^2}\sum_{i=1}^{N}\frac{(x_i - \bar{x})}{N} = \frac{N^2}{(N-1)^2}\sum_{i=1}^{N}\frac{(y_i - \bar{y})}{N} = \frac{N^2}{(N-1)^2}\mathrm{Var}\left[\{y_n\}_{n=1}^{N}\right]$$

$\Rightarrow E_{loocv}(\mathcal{A}_{average})$ is a scaled version of the variance of $\{y_n\}_{n=1}^{N}$