Machine Learning Techniques Homework #2

tags: `mlt` , `ntu` , `course` , `homework`

Cheng-Shih Wong, R04945028, mob5566@gmail.com (mailto:mob5566@gmail.com)

1.

For

$$F(A,B) = \frac{1}{N}\sum_{n=1}^{N} \ln\Big(1+\exp\Big(-y_n\left(A\cdot\left(\mathbf{w}_{\mathsf{SVM}}^T\phi(\mathbf{x}_n)+b_{\mathsf{SVM}}\right)+B\right)\Big)\Big)$$

Let $z_n = \mathbf{w}_{\mathsf{SVM}}^T\phi\left(\mathbf{x}_n\right)+b_{\mathsf{SVM}}, s_n = Az_n + B,\ \text{and}\ p_n = \theta\left(-y_n s_n\right)$

where $\theta\left(s\right) = \frac{\exp(s)}{1+\exp(s)}$

$$\begin{aligned}
\frac{\partial F(A,B)}{\partial A} &= \frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial A}\ln\left(1+e^{-y_n s_n}\right)\\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\left(1+e^{-y_n s_n}\right)}\frac{\partial}{\partial A}\left(1+e^{-y_n s_n}\right)\\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\left(1+e^{-y_n s_n}\right)}\frac{\partial}{\partial A}e^{-y_n s_n}\\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{e^{-y_n s_n}}{\left(1+e^{-y_n s_n}\right)}\frac{\partial}{\partial A}(-y_n(Az_n+B))\\
&= \frac{1}{N}\sum_{n=1}^{N}\theta\left(-y_n s_n\right)\left(-y_n z_n\right)\\
&= -\frac{1}{N}\sum_{n=1}^{N} y_n z_n p_n
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial F(A,B)}{\partial B} &= \frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial B}\ln\left(1+e^{-y_n s_n}\right) \\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{(1+e^{-y_n s_n})}\frac{\partial}{\partial B}\left(1+e^{-y_n s_n}\right) \\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{(1+e^{-y_n s_n})}\frac{\partial}{\partial B}e^{-y_n s_n} \\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{e^{-y_n s_n}}{(1+e^{-y_n s_n})}\frac{\partial}{\partial B}(-y_n(Az_n+B)) \\
&= \frac{1}{N}\sum_{n=1}^{N}\theta\left(-y_n s_n\right)\left(-y_n\right) \\
&= -\frac{1}{N}\sum_{n=1}^{N}y_n p_n
\end{aligned}
$$

$$
\nabla F(A,B) = \begin{bmatrix} \frac{\partial F(A,B)}{\partial A} \\ \frac{\partial F(A,B)}{\partial B} \end{bmatrix} = \begin{bmatrix} -\frac{1}{N}\sum_{n=1}^{N} y_n z_n p_n \\ -\frac{1}{N}\sum_{n=1}^{N} y_n p_n \end{bmatrix}
$$

2.

$$
\begin{aligned}
\frac{\partial^2 F(A,B)}{\partial A^2} &= -\frac{1}{N}\sum_{n=1}^{N} y_n z_n \frac{\partial p_n}{\partial A} \\
&\because \frac{\mathrm{d}\theta\left(s\right)}{\mathrm{d}s} = \theta\left(s\right)\left(1-\theta\left(s\right)\right) \\
&= -\frac{1}{N}\sum_{n=1}^{N} y_n z_n p_n(1-p_n)\frac{\partial}{\partial A}(-y_n(Az_n+B)) \\
&= -\frac{1}{N}\sum_{n=1}^{N} y_n z_n p_n(1-p_n)(-y_n z_n) \\
&= \frac{1}{N}\sum_{n=1}^{N} y_n^2 z_n^2 p_n(1-p_n)
\end{aligned} \tag{2}
$$

Similar to (2)

$$
\frac{\partial^2 F(A,B)}{\partial A \partial B} = \frac{\partial^2 F(A,B)}{\partial B \partial A} = \frac{1}{N}\sum_{n=1}^{N} y_n^2 z_n p_n(1-p_n)
$$

$$
\frac{\partial^2 F(A,B)}{\partial B^2} = \frac{1}{N}\sum_{n=1}^{N} y_n^2 p_n(1-p_n)
$$

$$
\begin{aligned}
H(F) &= \begin{bmatrix} \frac{\partial^2 F(A,B)}{\partial A^2} & \frac{\partial^2 F(A,B)}{\partial A \partial B} \\ \frac{\partial^2 F(A,B)}{\partial B \partial A} & \frac{\partial^2 F(A,B)}{\partial B^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{N}\sum_{n=1}^{N} y_n^2 z_n^2 p_n(1-p_n) & \frac{1}{N}\sum_{n=1}^{N} y_n^2 z_n p_n(1-p_n) \\ \frac{1}{N}\sum_{n=1}^{N} y_n^2 z_n p_n(1-p_n) & \frac{1}{N}\sum_{n=1}^{N} y_n^2 p_n(1-p_n) \end{bmatrix}
\end{aligned}
$$

3.

$$\because \text{ all } \mathbf{x}_n \text{ are different}$$

$$\therefore \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \begin{cases} 0, & \text{if } i = j \\ \|\mathbf{x}_i - \mathbf{x}_j\|^2 > 0, & \text{else} \end{cases}$$

$$\therefore \text{ Gaussian kernel } K(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \exp(-\gamma * 0) = 1, & \text{if } i = j \\ \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2), & \text{else} \end{cases}$$

$$\therefore \lim_{\gamma \to \infty} \exp(-\gamma\|\mathbf{x}_i, \mathbf{x}_j\|^2) = \begin{cases} \exp(-\gamma * 0) = 1, & \text{if } i = j \\ \exp(-\infty) = \frac{1}{\infty} = 0, & \text{else} \end{cases}$$

$$\text{The kernel matrix K} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix} \text{ is an identity matrix}$$

From course slide **Support Vector Regression** page 4
The analytic solution of kernel ridge regresssion

$$\begin{aligned} \beta &= (\lambda \mathbf{I} + \mathsf{K})^{-1}\mathbf{y} \\ &= ((\lambda + 1)\mathbf{I})^{-1}\mathbf{y} \\ &= (\frac{1}{\lambda + 1}\mathbf{I})\mathbf{y} \\ &= \frac{\mathbf{y}}{\lambda + 1} \end{aligned}$$

4.

$$\because \lim_{\gamma \to 0} \exp(-\gamma\|\mathbf{x}_i, \mathbf{x}_j\|^2) = \exp(0) = 1$$

$$\therefore \text{ The kernel matrix K} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

As in **Q.3**
The analytic solution of kernel ridge regresssion

$$\begin{aligned} \beta &= (\lambda \mathbf{I} + \mathsf{K})^{-1}\mathbf{y} \\ &= \left(\begin{bmatrix} \lambda + 1 & 1 & \cdots & 1 \\ 1 & \lambda + 1 & \ddots & 1 \\ 1 & 1 & \ddots & \vdots \\ 1 & 1 & \cdots & \lambda + 1 \end{bmatrix} \right)^{-1} \mathbf{y} \end{aligned}$$

😄
😄
😄
😄
😄

5.

$$(P_2) \quad \min_{b,\mathbf{w},\xi^\vee,\xi^\wedge} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N}\left((\xi_n^\vee)^2 + (\xi_n^\wedge)^2\right)$$
$$\text{s.t.} \quad -\epsilon - \xi_n^\vee \le y_n - \mathbf{w}^T\phi(\mathbf{x}_n) - b \le \epsilon + \xi_n^\wedge$$

The constraint can be write as the absolute value form

$$\min_{b,\mathbf{w},\boldsymbol{\xi}} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N}(\xi_n)^2$$
$$\text{s.t.} \quad |y_n - \mathbf{w}^T\phi(\mathbf{x}_n) - b| \le \epsilon + \xi_n$$

where the ξ_n is the **tube violation**

$$\xi_n = \max\left(0, |\mathbf{w}^T\phi(\mathbf{x}_n) + b - y_n| - \epsilon\right)$$

That is the equivalent unconstrained form of (P_2) is

$$\min_{b,\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N}\left(\max\left(0, |\mathbf{w}^T\phi(\mathbf{x}_n) + b - y_n| - \epsilon\right)\right)^2 \tag{5}$$

6.

Substitute optimal $\mathbf{w}_*$ into (5)

$$\min_{b,\boldsymbol{\beta}} F(b,\boldsymbol{\beta}) = \frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N}\beta_n\beta_m K(\mathbf{x}_n,\mathbf{x}_m) + C\sum_{n=1}^{N}\left(\max\left(0, \left|\sum_{m=1}^{N}\beta_m K(\mathbf{x}_n,\mathbf{x}_m) + b - y_n\right| - \epsilon\right)\right)^2$$

Let $s_n = \sum_{m=1}^{N}\beta_m K(\mathbf{x}_n,\mathbf{x}_m) + b$

$$\min_{b,\boldsymbol{\beta}} F(b,\boldsymbol{\beta}) = \frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N}\beta_n\beta_m K(\mathbf{x}_n,\mathbf{x}_m) + C\sum_{n=1}^{N}(\max(0, |s_n - y_n| - \epsilon))^2$$

To obtain $\frac{\partial F(b,\boldsymbol{\beta})}{\partial \beta_m}$

The partial derivative of **regularizer term** of $F(b,\boldsymbol{\beta})$ is

$$\frac{\partial\left(\frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N}\beta_n\beta_m K(\mathbf{x}_n,\mathbf{x}_m)\right)}{\partial\beta_m} = \sum_{n=1}^{N}\beta_n K(\mathbf{x}_m,\mathbf{x}_n)$$

For the **regression error term** of $F(b,\boldsymbol{\beta})$

Let

$$D_n = \frac{\partial(\max(0, |s_n - y_n| - \epsilon))^2}{\partial\beta_m}$$

and

$$\frac{\partial s_n}{\beta_m} = K(\mathbf{x}_n,\mathbf{x}_m)$$

We should discuss three conditions $s_n - y_n < -\epsilon, -\epsilon \le s_n - y_n \le \epsilon, \epsilon < s_n - y_n$

If $-\epsilon \leq s_n - y_n \leq \epsilon$

$$\begin{aligned} D_n &= \frac{\partial (0)^2}{\partial \beta_m} \\ &= 0 \end{aligned}$$

If $s_n - y_n < -\epsilon$

$$\begin{aligned} D_n &= \frac{\partial (y_n - s_n - \epsilon)^2}{\partial \beta_m} \\ &= 2(y_n - s_n - \epsilon)\frac{\partial (y_n - s_n - \epsilon)}{\partial \beta_m} \\ &= 2(y_n - s_n - \epsilon)\frac{\partial - s_n}{\partial \beta_m} \\ &= -2(y_n - s_n - \epsilon)K(\mathbf{x}_n, \mathbf{x}_m) \end{aligned}$$

If $\epsilon < s_n - y_n$

$$\begin{aligned} D_n &= \frac{\partial (s_n - y_n - \epsilon)^2}{\partial \beta_m} \\ &= 2(s_n - y_n - \epsilon)\frac{\partial (s_n - y_n - \epsilon)}{\partial \beta_m} \\ &= 2(s_n - y_n - \epsilon)\frac{\partial s_n}{\partial \beta_m} \\ &= 2(s_n - y_n - \epsilon)K(\mathbf{x}_n, \mathbf{x}_m) \end{aligned}$$

Finally

$$\frac{\partial F(b, \boldsymbol{\beta})}{\partial \beta_m} = \sum_{n=1}^{N} \beta_n K(\mathbf{x}_m, \mathbf{x}_n) + C \sum_{n=1}^{N} D_n$$
$$\text{where } D_n = \begin{cases} -2(y_n - s_n - \epsilon)K(\mathbf{x}_n, \mathbf{x}_m) & \text{, if } s_n - y_n < -\epsilon \\ 0 & \text{, if } -\epsilon \leq s_n - y_n \leq \epsilon \\ 2(s_n - y_n - \epsilon)K(\mathbf{x}_n, \mathbf{x}_m) & \text{, if } \epsilon < s_n - y_n \end{cases}$$

7.

Denote x_1, x_2 are the values we sample from random variable $X \overset{\text{i.i.d}}{\sim} U(0, 1)$

Given any two points on 2D space, we can always obtain a **straight line** passing through both point to minimize the mean squared error by the linear regression hypothesis $h(x) = w_1 x + w_0$.

$$w_1 x_1 + w_0 = x_1^2 \tag{7.1}$$
$$w_1 x_2 + w_0 = x_2^2 \tag{7.2}$$

$$\begin{aligned} (7.1) - (7.2) &\Rightarrow w_1(x_1 - x_2) = (x_1^2 - x_2^2) = (x_1 - x_2)(x_1 + x_2) \\ &\Rightarrow w_1 = x_1 + x_2 \qquad (7.3) \\ \text{Substitute (7.3) into (7.1)} &\Rightarrow (x_1 + x_2)x_1 + w_0 = x_1^2 \\ &\Rightarrow w_0 = -x_1 x_2 \end{aligned}$$

$$\because \text{ For any two independent } x_1, x_2$$
$$\text{We can construct a line } h(x) = (x_1 + x_2)x - x_1 x_2$$
$$\text{and}$$
$$\mathcal{E}[X] = \frac{0+1}{2} = 0.5$$
$$\therefore \bar{g}(x) = (0.5 + 0.5)x - 0.5 * 0.5 = x - 0.25$$

8.

For any $\tilde{y}_n$ in $\{\tilde{y}_n\}_{n=1}^{\tilde{N}}$

Let

$$h_0(\mathbf{x}) = 0, \text{ for any } \mathbf{x}$$
$$\text{and}$$
$$h_n(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} = \tilde{\mathbf{x}}_n \\ 0, & \text{else} \end{cases}$$

We can obtain

$$\begin{aligned} r_z =& \text{RMSE}(h_0) \\ =& \sqrt{\frac{1}{\tilde{N}} \sum_{m=1}^{\tilde{N}} (\tilde{y}_m - h_0(\tilde{\mathbf{x}}_m))^2} \\ =& \sqrt{\frac{1}{\tilde{N}} \sum_{m=1}^{\tilde{N}} \tilde{y}_m^2} \end{aligned} \tag{8.1}$$

$$\begin{aligned} &\text{and} \\ r_n =& \text{RMSE}(h_n) \\ =& \sqrt{\frac{1}{\tilde{N}} \sum_{m=1}^{\tilde{N}} (\tilde{y}_m - h_n(\tilde{\mathbf{x}}_m))^2} \\ =& \sqrt{\frac{1}{\tilde{N}} \sum_{m \neq n} \tilde{y}_m^2 + (\tilde{y}_n - 1)^2} \\ =& \sqrt{\frac{1}{\tilde{N}} \sum_{m=1}^{\tilde{N}} \tilde{y}_m^2 - 2\tilde{y}_n + 1} \end{aligned}$$

Then we can subtract r_z^2 by r_n^2 to obtain

$$\begin{aligned} r_z^2 - r_n^2 =& \frac{1}{\tilde{N}} \left(\sum_{m=1}^{\tilde{N}} \tilde{y}_n^2 - \left(\sum_{m=1}^{\tilde{N}} \tilde{y}_n^2 - 2\tilde{y}_n + 1 \right) \right) \\ =& \frac{1}{\tilde{N}} (2\tilde{y}_n - 1) \end{aligned}$$

After that

$$\tilde{y}_n = \frac{\tilde{N}(r_z^2 - r_n^2) + 1}{2} \tag{8.2}$$

Therefore, if we have r_z and $\{r_n\}_{n=1}^{\tilde{N}}$, we can obtain all $\tilde{y}_n$ by (8.2).

As a consequence, if $T = \tilde{N} + 1$, we can construct a

$$g(\mathbf{x}) = \begin{cases} \tilde{y}_n, & \text{if } \mathbf{x} = \tilde{\mathbf{x}}_n \\ \text{something}, & \text{else} \end{cases}$$

with $\text{RMSE}(g) = 0$

9.

Given any hypothesis g

$$\begin{aligned} \mathbf{g} &= (g(\tilde{\mathbf{x}}_1), g(\tilde{\mathbf{x}}_2), \cdots, g(\tilde{\mathbf{x}}_{\tilde{N}})) \\ \tilde{\mathbf{y}} &= (\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_{\tilde{N}}) \end{aligned}$$

Denote

$$\begin{aligned} r_g &= \text{RMSE}(g) \\ &= \sqrt{\frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} \left(\tilde{y}_n - g(\tilde{\mathbf{x}}_n)\right)^2} \\ &= \sqrt{\frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} \left(\tilde{y}_n^2 - 2\tilde{y}_n g(\tilde{\mathbf{x}}_n) + g(\tilde{\mathbf{x}}_n)^2\right)} \\ &= \sqrt{\frac{1}{\tilde{N}} \left(\sum_{n=1}^{\tilde{N}} \tilde{y}_n^2 - 2\sum_{n=1}^{\tilde{N}} \tilde{y}_n g(\tilde{\mathbf{x}}_n) + \sum_{n=1}^{\tilde{N}} g(\tilde{\mathbf{x}}_n)^2\right)} \\ &= \sqrt{\frac{1}{\tilde{N}} \left(\sum_{n=1}^{\tilde{N}} \tilde{y}_n^2 - 2\mathbf{g}^T\tilde{\mathbf{y}} + \mathbf{g}^T\mathbf{g}\right)} \\ &\quad \text{By (8.2)}, \sum_{n=1}^{\tilde{N}} \tilde{y}_n^2 = \tilde{N} r_z^2 \\ &= \sqrt{\frac{1}{\tilde{N}} \left(\tilde{N} r_z^2 - 2\mathbf{g}^T\tilde{\mathbf{y}} + \mathbf{g}^T\mathbf{g}\right)} \end{aligned}$$

Then, we can obtain

$$\mathbf{g}^T\tilde{\mathbf{y}} = \frac{\tilde{N}(r_z^2 - r_g^2) + \mathbf{g}^T\mathbf{g}}{2}$$

Therefore, we only need $\text{RMSE}(h_0)$ and $\text{RMSE}(g)$ to calculate $\mathbf{g}^T\tilde{\mathbf{y}}$, i.e. we only need 2 queries.

10.

Denote that

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_K \end{bmatrix}^T$$

$$G = \begin{bmatrix} g_1(\tilde{\mathbf{x}}_1) & g_2(\tilde{\mathbf{x}}_1) & \cdots & g_K(\tilde{\mathbf{x}}_1) \\ g_1(\tilde{\mathbf{x}}_2) & g_2(\tilde{\mathbf{x}}_2) & \cdots & g_K(\tilde{\mathbf{x}}_2) \\ \vdots & \vdots & \vdots & \vdots \\ g_1(\tilde{\mathbf{x}}_{\tilde{N}}) & g_2(\tilde{\mathbf{x}}_{\tilde{N}}) & \cdots & g_K(\tilde{\mathbf{x}}_{\tilde{N}}) \end{bmatrix}_{\tilde{N} \times K}$$

The goal is to

$$
\begin{aligned}
\min_{\boldsymbol{\alpha}} \mathrm{RMSE}\left(\sum_{k=1}^{K} \alpha_k g_k\right) &= \min_{\boldsymbol{\alpha}} \sqrt{\frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} \left(\tilde{y}_n - \sum_{k=1}^{K} \alpha_k g_k(\tilde{\mathbf{x}}_n)\right)^2} \\
&= \min_{\boldsymbol{\alpha}} \sum_{n=1}^{\tilde{N}} \left(\tilde{y}_n - \sum_{k=1}^{K} \alpha_k g_k(\tilde{\mathbf{x}}_n)\right)^2 \\
&= \min_{\boldsymbol{\alpha}} \sum_{n=1}^{\tilde{N}} \left[\tilde{y}_n^2 - 2\tilde{y}_n \sum_{k=1}^{K} \alpha_k g_k(\tilde{\mathbf{x}}_n) + \left(\sum_{k=1}^{K} \alpha_k g_k(\tilde{\mathbf{x}}_n)\right)^2\right] \\
&= \min_{\boldsymbol{\alpha}} \sum_{n=1}^{\tilde{N}} \tilde{y}_n^2 - 2\sum_{n=1}^{\tilde{N}} \tilde{y}_n \sum_{k=1}^{K} \alpha_k g_k(\tilde{\mathbf{x}}_n) + \sum_{n=1}^{\tilde{N}} \left(\sum_{k=1}^{K} \alpha_k g_k(\tilde{\mathbf{x}}_n)\right)^2 \\
&= \min_{\boldsymbol{\alpha}} \tilde{N} r_z^2 - 2\tilde{\mathbf{y}}^T G \boldsymbol{\alpha} + \boldsymbol{\alpha}^T G^T G \boldsymbol{\alpha}
\end{aligned}
$$

Let

$$
\begin{aligned}
E(\boldsymbol{\alpha}) &= \tilde{N} r_z^2 - 2\tilde{\mathbf{y}}^T G \boldsymbol{\alpha} + \boldsymbol{\alpha}^T G^T G \boldsymbol{\alpha} \\
\nabla E(\boldsymbol{\alpha}) &= 2\left(G^T G \boldsymbol{\alpha} - \tilde{\mathbf{y}}^T G\right)
\end{aligned}
$$

We want $\nabla E(\boldsymbol{\alpha}) = 0$

$$
\alpha = \left(G^T G\right)^{-1} \tilde{\mathbf{y}}^T G
$$

that minimize $E(\boldsymbol{\alpha})$

and

$$
\tilde{\mathbf{y}}^T G = \begin{bmatrix} \mathbf{g}_1^T \tilde{\mathbf{y}} \\ \mathbf{g}_2^T \tilde{\mathbf{y}} \\ \vdots \\ \mathbf{g}_K^T \tilde{\mathbf{y}} \end{bmatrix}
$$

For each $\left\{\mathbf{g}_k^T \tilde{\mathbf{y}}\right\}_{k=1}^{K}$, we need RMSE$(h_0)$ and RMSE$(\mathbf{g}_k)$ as **Q.9**.

Therefore, we need $K+1$ queries to solve $\boldsymbol{\alpha}$.

11.

E_{in}

$\gamma \setminus \lambda$	0.001	1	1000
32	**0**	**0**	**0**
2	**0**	**0**	**0**
0.125	**0**	0.03	0.2425

The minimum $E_{\text{in}}(g)$ is 0, and occurred at $\gamma = 32$, $\gamma = 2$, and $\gamma = 0.125 \wedge \lambda = 0.001$.

😄

12.

E_{out}

$\gamma \setminus \lambda$	0.001	1	1000
32	0.45	0.45	0.45
2	0.44	0.44	0.44
0.125	0.46	0.45	**0.39**

The minimum $E_{\text{out}}(g)$ is 0.39, and occurred at $\gamma = 0.125 \wedge \lambda = 1000$.

13.

In Q.13 and Q.14, I use `sklearn.svm.SVR` to train support vector regression model.

E_{in}

$\gamma \setminus C$	0.001	1	1000
32	0.4	**0**	**0**
2	0.4	**0**	**0**
0.125	0.4	0.035	**0**

The minimum $E_{\text{in}}(g)$ is 0, and occurred at $C = 1000$, and $C = 1 \wedge (\gamma = 32 \vee \gamma = 2)$.

14.

E_{out}

$\gamma \setminus C$	0.001	1	1000
32	0.48	0.48	0.48
2	0.48	0.48	0.48
0.125	0.48	**0.42**	0.47

The minimum $E_{\text{out}}(g)$ is 0.42, and occurred at $\gamma = 0.125 \wedge C = 1$.

15.

I iterate 300 times for bagging.

E_{in}

λ	0.01	0.1	1	10	100
E_{in}	0.32	0.3175	0.3175	0.315	**0.31**

The minimum $E_{\text{in}}(g)$ is 0.31, and occurred at $\lambda = 100$.

😄

16.

E_{out}

λ	0.01	0.1	1	10	100
E_{out}	$\mathbf{0.36}$	$\mathbf{0.36}$	$\mathbf{0.36}$	$\mathbf{0.36}$	0.4

The minimum $E_{\text{out}}(g)$ is 0.36, and occurred at $\lambda = 0.01 \vee 0.1 \vee 1 \vee 10$.