Introduction to Machine Learing: Homework III

Due on Nov 16th, 2022 at 11:59pm

Professor Ziping Zhao

Bingnan Li
2020533092

1. [SVM]

   (a) Hard-margin SVM

      (i) Lagrangian function and dual representation:
      Solution:

      By introducing Lagrangian multipliers $\{\alpha_i\}$, we have the Lagrangian function $\mathcal{L}$

$$\begin{aligned}\mathcal{L}(\boldsymbol{w}, w_0, \{\alpha_i\}) &= \frac{1}{2}||\boldsymbol{w}||_2^2 - \sum_{i=1}^{N} \alpha_i \left[r^i(\boldsymbol{w}^T\boldsymbol{x}^i + w_0) - 1\right] \\ &= \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \boldsymbol{w}^T\sum_{i=1}^{N}\alpha_i r^i \boldsymbol{x}^i - w_0\sum_{i=1}^{N}\alpha_i r^i + \sum_{i=1}^{N}\alpha_i\end{aligned}$$

      Then by zero gradient theorem, we have

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{N}\alpha_i r^i\boldsymbol{x}^i = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{N}\alpha_i r^i\boldsymbol{x}^i \tag{1}$$

$$\frac{\partial\mathcal{L}}{\partial w_0} = -\sum_{i=1}^{N}\alpha_i r^i = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i r^i = 0 \tag{2}$$

      By substituting $\boldsymbol{w}$ and $\sum_{i=1}^{N}\alpha_i r^i$ with equation (1) and (2), we have

$$G(\{\alpha_i\}) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j r^i r^j (\boldsymbol{x^i})^T\boldsymbol{x^j}$$

      Thus, the dual representation is as following:

$$\begin{aligned}&maximize_{\{\alpha_i\}} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j r^i r^j(\boldsymbol{x^i})^T\boldsymbol{x^j} \\ &subject\ to \sum_{i=1}^{N}\alpha_i r^i = 0 \\ &\alpha_i \geq 0, \forall i\end{aligned}$$

      (ii) Show that $\frac{1}{\gamma_{max}^2} = \sum_{i=1}^{N}\alpha_i$.
      Proof:

      Define

$$\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1\\ \vdots \\ \alpha_N\end{bmatrix} \quad \boldsymbol{r} = \begin{bmatrix}r^1\\ \vdots \\ r^N\end{bmatrix}$$

      and symmetric matrix $\boldsymbol{H}$ with $h_{ij} = r^i r^j (x^i)^T x^j$. Then the dual representation of maximum margin problem can be rewritten as following

$$\begin{aligned}&maximize_{\boldsymbol{\alpha}}\ \boldsymbol{\alpha^T 1} - \frac{1}{2}\boldsymbol{\alpha^T H\alpha} \\ &subject\ to\ \boldsymbol{\alpha^T r} = 0 \\ &\boldsymbol{\alpha} \geq \boldsymbol{0}\end{aligned}$$

      Then the Lagrangian function is

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\lambda}, v) = \frac{1}{2}\boldsymbol{\alpha^T H\alpha} - \boldsymbol{\alpha^T 1} + \boldsymbol{\lambda^T\alpha} + v\boldsymbol{\alpha^T r}$$

and the corresponding KKT conditions are

$$\begin{cases} \dfrac{\partial\lambda}{\partial\boldsymbol{\alpha}} = \boldsymbol{H\alpha} - \boldsymbol{1} - \boldsymbol{\lambda} + v\boldsymbol{r} = 0 \\ \boldsymbol{\alpha^T r} = 0, \boldsymbol{\alpha} \geq \boldsymbol{0} \\ \boldsymbol{\lambda} \geq 0 \\ \boldsymbol{\alpha^T \lambda} = 0 \end{cases}$$

Thus, we have

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}\boldsymbol{\alpha^T}(\boldsymbol{1} + \boldsymbol{\lambda} - v\boldsymbol{r}) - \boldsymbol{\alpha^T 1} \\ &= \frac{1}{2}\boldsymbol{\alpha^T 1} - \boldsymbol{\alpha^T 1} \\ &= -\frac{1}{2}\boldsymbol{\alpha^T 1} = -\frac{1}{2}\sum_{i=1}^{N}\alpha_i \end{aligned}$$

Which means the constraint minimum value of $\frac{1}{2}\boldsymbol{\alpha^T H\alpha} - \boldsymbol{\alpha^T 1}$ is $-\frac{1}{2}\sum_{i=1}^{N}\alpha_i$, then the constraint maximum value of $-\frac{1}{2}\boldsymbol{\alpha^T H\alpha} + \boldsymbol{\alpha^T 1}$ is $\frac{1}{2}\sum_{i=1}^{N}\alpha_i$. Moreover, the constraint minimum value of $\frac{1}{2}||\boldsymbol{w}||_2^2$ is also $\frac{1}{2}\sum_{i=1}^{N}\alpha_i$.

Thus, we have

$$||\boldsymbol{w}||_{min}^2 = \frac{1}{\gamma_{max}^2} = \sum_{i=1}^{N}\alpha_i$$

(b) Soft-margin SVM

Proof:

The primal problem of soft margin SVM problem is as following:

$$\begin{aligned} minimize_{\boldsymbol{w},w_0,\{\xi_t\}} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N}\xi_i \\ subject\ to \quad & r^i f(\boldsymbol{x}) \geq 1 - \xi_i,\ \forall i \\ & \xi_i \geq 0,\ \forall i \end{aligned}$$

and KKT conditions are

$$\begin{cases} \boldsymbol{w} - \sum_{i=1}^{N}\alpha_i(r^i x^i) = 0 \\ \sum_{i=1}^{N}\alpha^i r^i = 0 \\ \mu_i = C - \alpha_i \\ \alpha^i \geq 0, \mu_i \geq 0 \\ \alpha_i(r^i f(\boldsymbol{x}) - 1 + \xi_i) = 0, \mu_i\xi_i = 0 \\ r^i f(x) - 1 + \xi_i \geq 0, \xi_i \geq 0 \end{cases}$$

If $\alpha_i = 0$, we have

$$\begin{aligned} \mu_i = C - \alpha_i &= C \\ \mu_i\xi_i = 0 &\Rightarrow \xi_i = 0 \\ \alpha_i(r^i f(\boldsymbol{x}) - 1 + \xi_i) = 0, r^i f(\boldsymbol{x}) - 1 + \xi_i \geq 0 &\Rightarrow r^i f(\boldsymbol{x}) - 1 + \xi_i \geq 0 \\ &\Rightarrow r^i f(\boldsymbol{x}) \geq 1 - \xi_i = 1 \end{aligned}$$

If $0 < \alpha_i < C$, we have

$$\begin{aligned}\mu_i &= C - \alpha_i \neq 0 \\ \mu_i \xi_i &= 0 \Rightarrow \xi_i = 0 \\ \alpha_i(r^i f(\boldsymbol{x}) - 1 + \xi_i) &= 0 \Rightarrow r^i f(\boldsymbol{x}) - 1 = 0 \\ &\Rightarrow r^i f(\boldsymbol{x}) = 1\end{aligned}$$

If $\alpha = C$, we have

$$\begin{aligned}\mu_i &= C - \alpha_i = 0 \\ \mu_i \xi_i &= 0, \xi_i \geq 0 \Rightarrow \xi_i \geq 0 \\ \alpha_i(r^i f(\boldsymbol{x}) - 1 + \xi_i) &= 0 \Rightarrow r^i f(\boldsymbol{x}) = 1 - \xi_i \leq 1\end{aligned}$$

2. [SVM]

   (a) Dual problem of SVR
   Solution:
   In order to convert SVR into dual QP problem, we introduce slack variables ξ_i^+ and ξ_i^- and the primal problem is as following

$$\begin{aligned}minimize_{\boldsymbol{w},\{\xi_i^+\},\{\xi_i^-\}} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N}(\xi_i^+ + \xi_i^-) \\ subject\ to \quad & r^i - \boldsymbol{w}^T\boldsymbol{x}^i \leq \epsilon + \xi_i^+, \quad \forall i \\ & \boldsymbol{w}^T\boldsymbol{x}^i - r^i \leq \epsilon + \xi_i^-, \quad \forall i \\ & \xi_i^+, \xi_i^- \geq 0, \quad \forall i\end{aligned}$$

   and the Lagrangian function is

$$\begin{aligned}&\mathcal{L}(\boldsymbol{w}, \{\xi_i^+\}, \{\xi_i^-\}, \{\alpha_i^+\}, \{\alpha_i^-\}, \{\mu_i^+\}, \{\mu_i^-\}) \\ &= \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N}(\xi_i^+ + \xi_i^-) - \sum_{i=1}^{N}\alpha_i^+[\epsilon + \xi_i^+ - r^i + \boldsymbol{w}^T\boldsymbol{x}^i] - \sum_{i=1}^{N}\alpha_i^-[\epsilon + \xi_i^- + r^i - \boldsymbol{w}^T\boldsymbol{x}^i] \\ &\quad - \sum_{i=1}^{N}(\mu_i^+\xi_i^+ + \mu_i^-\xi_i^-)\end{aligned}$$

   By setting the gradient of $\mathcal{L}$ to 0, we have

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-)\boldsymbol{x}^i = \boldsymbol{0} \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^+} = C - \alpha_i^+ - \mu_i^+ = 0 \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^-} = C - \alpha_i^- - \mu_i^- = 0+ \tag{5}$$

   Then, substitute $\boldsymbol{w}, C$ with equation (3), (4) and (5), we have

$$\begin{aligned}&G(\{\alpha_i^+\}, \{\alpha_i^-\}) \\ &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)(\boldsymbol{x}^i)^T\boldsymbol{x}^j - \epsilon\sum_{i=1}^{N}(\alpha_i^+ + \alpha_i^-) + \sum_{i=1}^{N}r^t(\alpha_i^+ - \alpha_i^-)\end{aligned}$$

Thus, the dual problem is

$$
\begin{aligned}
\mathit{maximize}_{\{\alpha_i^+\},\{\alpha_i^-\}} \quad & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)(\boldsymbol{x}^i)^T\boldsymbol{x}^j \\
& -\epsilon\sum_{i=1}^{N}(\alpha_i^+ + \alpha_i^-) + \sum_{i=1}^{N} r^t(\alpha_i^+ - \alpha_i^-) \\
\mathit{subject\ to} \quad & 0 \le \alpha_i^+ \le C,\ \forall i \\
& 0 \le \alpha_i^- \le C,\ \forall i
\end{aligned}
$$

(b) Kernelized dual problem

Solution:

Define a linear kernel $K(\boldsymbol{x}^i, \boldsymbol{x}^j) = \phi(\boldsymbol{x}^i)^T\phi(\boldsymbol{x}^j) = (\boldsymbol{x}^i)^T\boldsymbol{x}^j$, then the Kernelized dual problem is as following

$$
\begin{aligned}
\mathit{maximize}_{\{\alpha_i^+\},\{\alpha_j^-\}} \quad & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)K(\boldsymbol{x}^i, \boldsymbol{x}^j) \\
& -\epsilon\sum_{i=1}^{N}(\alpha_i^+ + \alpha_i^-) + \sum_{i=1}^{N} r^t(\alpha_i^+ - \alpha_i^-) \\
\mathit{subject\ to} \quad & 0 \le \alpha_i^+ \le C,\ \forall i \\
& 0 \le \alpha_i^- \le C,\ \forall i
\end{aligned}
$$

(c)

3. (a)

   (b)

   (c)

4. (a)

   (b)

   (c)