HW3 : Kernel Methods

(ABISHEK SRIDHAR,
Andrew Id : abisheks)

1. Validity of a Kernel

(a) $K_1(x_1, x_2) = (1+\exp(x_1^T x_2))^{2022}$, $K_1 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$x_1^T x_2$ is a valid kernel (from Assumption 2.)

$\Rightarrow$ $\exp(x_1^T x_2)$ is a valid kernel (Assumption 1 : closure under exponentiation)

$\Rightarrow$ $(1+\exp(x_1^T x_2))^{2022}$ is a valid kernel function (Assumption 1 : closure under combination of polynomial with positive coefficients
↳ the binomial expansion here will have only positive coefficients for the polynomials of valid kernel $\exp(x_1^T x_2)$)

$\Rightarrow$ $K_1(x_1, x_2)$ is a valid kernel function

(b) $K_2(x_1, x_2) = \sigma^2 \exp\left(2022 \cos\left(\frac{2\pi(x_1 - x_2)}{\sigma^2}\right)\right)$, σ : constant , $K_2 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$

$\underbrace{\cos\left(\frac{2\pi x_1}{\sigma^2}\right) \cdot \cos\left(\frac{2\pi x_2}{\sigma^2}\right) \text{ is a valid kernel}}_{(1)}$ (Assumption 3 with $f(x) = \cos\left(\frac{2\pi x}{\sigma^2}\right)$)

$\sin\left(\frac{2\pi x_1}{\sigma^2}\right)\sin\left(\frac{2\pi x_2}{\sigma^2}\right)$ is a valid kernel (Assumption 3 with $f(x) = \sin\left(\frac{2\pi x}{\sigma^2}\right)$)

(2)

From (1) & (2),

$$\Rightarrow \cos\left(\frac{2\pi x_1}{\sigma^2}\right)\cos\left(\frac{2\pi x_2}{\sigma^2}\right) + \sin\left(\frac{2\pi x_1}{\sigma^2}\right)\sin\left(\frac{2\pi x_2}{\sigma^2}\right) = \cos\left(\frac{2\pi(x_1 - x_2)}{\sigma^2}\right)$$

is a valid kernel (Assumption 1: closure under addition)

$\Rightarrow 2022 \cdot \cos\left(\frac{2\pi(x_1 - x_2)}{\sigma^2}\right)$ is a valid kernel (Assumption 1: closure under polynomial with positive coefficients)

$\Rightarrow \exp\left(2022 \cdot \cos\left(\frac{2\pi(x_1 - x_2)}{\sigma^2}\right)\right)$ is a valid kernel (Assumption 1: closure under exponentiation)

$\Rightarrow \sigma^2 \exp\left(2022 \cos\left(\frac{2\pi(x_1 - x_2)}{\sigma^2}\right)\right)$ is a valid kernel (Assumption 1: closure under polynomial with positive coefficients)

$\Rightarrow K_2(x_1, x_2)$ is a valid kernel function

2. Kernelized Soft SVM

(a) $x_i \in \mathbb{R}^2$. $\psi'(x_i) = [x_{i1}^2, x_{i1}x_{i2}, x_{i1}x_{i2}, x_{i2}^2]$

Kernel Function $K'(x_1, x_2) = \langle \psi'(x_1), \psi'(x_2) \rangle$

$$= \langle [x_{11}^2, x_{11}x_{12}, x_{11}x_{12}, x_{12}^2], [x_{21}^2, x_{21}x_{22}, x_{21}x_{22}, x_{22}^2] \rangle$$

$$= x_{11}^2 x_{21}^2 + x_{11}x_{21}x_{12}x_{22} + x_{11}x_{21}x_{12}x_{22} + x_{12}^2 x_{22}^2$$

$$= (x_{11}x_{21})^2 + 2(x_{11}x_{21})(x_{12}x_{22}) + (x_{12}x_{22})^2$$

$$= (x_{11}x_{21} + x_{12}x_{22})^2$$

$$= (\langle [x_{11}, x_{12}], [x_{21}, x_{22}] \rangle)^2 = (x_1^T x_2)^2$$

$$= (\langle x_1, x_2 \rangle)^2$$

, which is a valid kernel function Ofcourse being a polynomial of valid kernel $x_1^T x_2$

Now soft SVM problem is:

$$\min_w \underbrace{\frac{1}{2}\|w\|_2^2}_{\text{non-decreasing function in } \|w\|} + C \sum_{i=1}^{n} \max(0, 1 - y_i(w^T \psi'(x_i)))$$

That is of the form:

$$\min_{w} f\left(\langle w, \Psi(x_1)\rangle, \langle w, \Psi(x_2)\rangle, \dots, \langle w, \Psi(x_n)\rangle\right) + R(\|w\|)$$

Hence, by REPRESENTER THEOREM, $\exists$ optimal $\underset{\in \mathbb{R}^4}{w}$ of the form $\sum_{i=1}^{n} \alpha_i \Psi'(x_i)$

$\alpha = [\alpha_1, \dots, \alpha_n]^T \in \mathbb{R}^n$

Now the soft SVM problem becomes:

$$\min_{\alpha} \frac{1}{2} \left\langle \sum_{i=1}^{n} \alpha_i \Psi'(x_i), \sum_{i=1}^{n} \alpha_i \Psi'(x_i) \right\rangle + C \sum_{i=1}^{n} \max\left(0, 1 - y_i \left(\left(\sum_{j=1}^{n} \alpha_j \Psi'(x_j)\right)^T \Psi'(x_i)\right)\right)$$

$$\Leftrightarrow \min_{\alpha} \frac{1}{2} \sum_{1 \le i,j \le n} \alpha_i \alpha_j \Psi'(x_i)^T \Psi'(x_j) + C \sum_{i=1}^{n} \max\left(0, 1 - y_i \left(\sum_{j=1}^{n} \alpha_j \Psi'(x_j)^T \Psi'(x_i)\right)\right)$$

$$\Leftrightarrow \min_{\alpha} \frac{1}{2} \sum_{1 \le i,j \le n} \alpha_i \alpha_j \langle \Psi'(x_i), \Psi'(x_j) \rangle + C \sum_{i=1}^{n} \max\left(0, 1 - y_i \left(\sum_{j=1}^{n} \alpha_j \langle \Psi'(x_i), \Psi'(x_j) \rangle\right)\right)$$

$$\Leftrightarrow \min_{\alpha} \underbrace{\frac{1}{2} \sum_{1 \le i,j \le n} \alpha_i \alpha_j K'(x_i, x_j) + C \sum_{i=1}^{n} \max\left(0, 1 - y_i \left(\sum_{j=1}^{n} \alpha_j K'(x_i, x_j)\right)\right)}_{\mathcal{L}(\alpha)}$$

Kernelized Soft SVM problem: $\min_{\alpha} \mathcal{L}(\alpha)$

(b) For writing the gradient descent algorithm, main component is $\nabla_\alpha \mathcal{L}(\alpha)$.

$$\mathcal{L}(\alpha) = \frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i^2 K'(x_i, x_i) + 2\sum_{1\le i<j\le n} \alpha_i\alpha_j K'(x_i, x_j)\right) + C\sum_{i=1}^{n}\max(0, f_i(\alpha))$$

$$\text{where } f_i(\alpha) = 1 - y_i\left(\sum_{j=1}^{n} \alpha_j K'(x_i, x_j)\right)$$

Also
$$\frac{\partial(\max(0, f_i(\alpha)))}{\partial\alpha_j} = \frac{\partial(\max(0, f_i(\alpha)))}{\partial(f_i(\alpha))} \cdot \frac{\partial(f_i(\alpha))}{\partial\alpha_j}$$

$$= \mathbb{1}\{f_i(\alpha) > 0\} \cdot (-y_i \cdot K'(x_i, x_j)) \longrightarrow (1)$$

We use sub-gradient here since it is not differentiable at $f_i(\alpha) = 0$.

And by defining, $\frac{\partial(\max(0, y))}{\partial y} = 0$

at $y = 0$, the sub-gradient is valid since:

$$\max(0, y_1) \geq 0$$

$$\Leftrightarrow \max(0, y_1) \geq \max(0, y) + \frac{\partial(\max(0, y))}{\partial y}(y_1 - y)$$

for $y = 0$

$$\frac{\partial \mathcal{L}(\alpha)}{\partial \alpha_j} = \alpha_j K'(x_j, x_j) + \sum_{\substack{1 \le i \le n \\ i \ne j}} \alpha_i \; K'(x_i, x_j) + C \underbrace{\sum_{i=1}^{n} \mathbb{1}\{f_i(\alpha) > 0\} \cdot \left(-y_i \cdot K'(x_i, x_j)\right)}_{(\text{from } (1))}$$

$$= \sum_{i=1}^{n} \alpha_i K'(x_i, x_j) + C \sum_{i=1}^{n} \mathbb{1}\{f_i(\alpha) > 0\} \cdot \left(-y_i \cdot K'(x_i, x_j)\right)$$

Suppose $K' = \begin{bmatrix} K'(x_1,x_1) & K'(x_1,x_2) & \cdots & K'(x_1,x_n) \\ K'(x_2,x_1) & & & \\ \vdots & & \ddots & \\ K'(x_n,x_1) & & & K'(x_n,x_n) \end{bmatrix}$ [$n \times n$ matrix with entries $K'_{ij} = K'(x_i, x_j)$]

$y = [y_1, y_2, \ldots, y_n]^T \in \mathbb{R}^n$

$f(\alpha) = [f_1(\alpha), f_2(\alpha), \ldots, f_n(\alpha)]^T \in \mathbb{R}^n$

In vectorized form,

$$\nabla_\alpha \mathcal{L}(\alpha) = (\alpha^T K')^T - C \cdot \left((y \odot \mathbb{1}\{f(\alpha) > 0\})^T \cdot K' \right)^T$$

, where $f(\alpha) = \mathbf{1} - y^T \cdot K' \alpha$ and $\odot$ represents element-wise multiplication

$$= K'^T \alpha - C \cdot K'^T \left(y \odot \mathbb{1}\{ y^T \cdot K' \alpha < \mathbf{1} \} \right)$$

Here "n" need not be the entire training set's size, but any n-sized set of samples from the training data.

Let $X = \begin{bmatrix} — x_1 — \\ — x_2 — \\ \vdots \\ — x_n — \end{bmatrix}_{n \times d}$

* Mini-batch Stochastic Gradient Descent Algorithm

- Given
  - X : training samples' features, $\in \mathbb{R}^{n \times d}$
  - y : training samples' labels, $\in \mathbb{R}^{n}$

- Initialization Step

  $\alpha^0 \leftarrow 0$

  $\eta \leftarrow$ fixed learning rate

  $\delta \leftarrow$ fixed convergence threshold

  $C \leftarrow$ fixed hyperparameter weighing hinge loss

- Update Step

  For $t = 1, 2, \ldots$:

  $m \leftarrow$ vector of b points sampled uniformly at random from $[n]$

  $X^t, y^t, \alpha \leftarrow X[m], y[m], \alpha^{t-1}[m]$ // selecting the rows indexed by values in set m

  $K^t \leftarrow K[m, m]$ // selecting both rows and columns indexed by values in set m

  $g^t \leftarrow (K^t)^T \alpha - c.(K^t)^T \left(y^t \odot \mathbb{1}\{(y^t)^T K^t \alpha < 1\}\right)$ // $\nabla_\alpha \mathcal{L}(\alpha) \big|_{\alpha = \alpha^{t-1}[m]}$

  $\alpha^t \leftarrow \alpha^{t-1}$

  $\alpha^t[m] \leftarrow \alpha^{t-1}[m] - \eta . g^t$

  If $\|\alpha^t - \alpha^{t-1}\| < \delta$:

  $\alpha^* \leftarrow \alpha^t$

  Exit Loop

- Output Step

  Return α^*

3. Kernel Support Vector Regression

Original Problem: $\min_{w \in \mathbb{R}^m, b \in \mathbb{R}} \frac{1}{2} \|w\|^2$ s.t. $|y_i - w^T\phi(x_i) - b| \leq \varepsilon \quad \forall\ i \in \{1, \ldots, n\}$

1. To allow for some errors in the above problem, we can allow the component $y_i - w^T\phi(x_i) - b$ to vary more than ε in the positive direction, or negative direction, or both for every point. Hence we introduce two slack variables λ_i and λ'_i for every point & formulate the problem as:

$$\min_{\substack{w \in \mathbb{R}^m, \\ b \in \mathbb{R}, \\ \lambda_i \in \mathbb{R}, \\ \lambda'_i \in \mathbb{R}}} \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{n}(\lambda_i + \lambda'_i)$$

$$\text{s.t.} \quad -\varepsilon - \lambda'_i \leq y_i - w^T\phi(x_i) - b \leq \varepsilon + \lambda_i ;$$

$$\lambda_i \geq 0 ;$$

$$\lambda'_i \geq 0 \qquad \forall\ i \in \{1, .., n\} \qquad \longrightarrow (1)$$

2. Re-writing the constraints in (1) to apply KKT conditions directly:

$$y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i \leq 0$$

$$w^T\phi(x_i) + b - y_i - \varepsilon - \lambda'_i \leq 0 \qquad \forall\ i \in \{1, .., n\}$$

$$-\lambda_i \leq 0$$

$$-\lambda'_i \leq 0$$

Lagrangian :

$$\mathcal{L}(w, b, \lambda, \lambda', \bar{p}, \bar{q}, \bar{r}, \bar{s}) = \frac{1}{2}\|w\|^2 + C \cdot \sum_{i=1}^{n} (\lambda_i + \lambda_i') + \Big[\sum_{i=1}^{n} p_i (y_i - w^T \phi(x_i) - b - \varepsilon - \lambda_i)$$

$$+ \sum_{i=1}^{n} q_i (w^T \phi(x_i) + b - y_i - \varepsilon - \lambda_i')$$

$$+ \sum_{i=1}^{n} r_i (-\lambda_i)$$

$$+ \sum_{i=1}^{n} s_i (-\lambda_i') \Big]$$

$\bar{p} = [p_1, \dots, p_n]$

$\bar{q} = [q_1, \dots, q_n]$

$\bar{r} = [r_1, \dots, r_n]$

$\bar{s} = [s_1, \dots, s_n]$

Primal Probleem $= \displaystyle\min_{\substack{w \in \mathbb{R}^m, \\ \lambda, \lambda', b \in \mathbb{R},}} \; \max_{\substack{\bar{p}, \bar{q}, \bar{r}, \bar{s} \\ \in \mathbb{R}^n_{\geq 0}}} \mathcal{L}(w, b, \lambda, \lambda', \bar{p}, \bar{q}, \bar{r}, \bar{s})$

Now, applying KKT conditions,

(Since the functions are differentiable, sub-gradient is replaced by

- (Stationarity)

wrt w:
$$\nabla_w\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\lambda_i+\lambda_i'\right) + \sum_{i=1}^{n} p_i \nabla_w\left(y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i\right) + \sum_{i=1}^{n} q_i \nabla_w\left(w^T\phi(x_i) + b - y_i - \varepsilon - \lambda_i'\right)$$

$$+ \sum_{i=1}^{n} r_i \nabla_w(-\lambda_i) + \sum_{i=1}^{n} s_i \nabla_w(-\lambda_i') = 0$$

$$\Rightarrow \quad w - \sum_{i=1}^{n} p_i \phi(x_i) + \sum_{i=1}^{n} q_i \phi(x_i) = 0$$

wrt b:
$$\nabla_b\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\lambda_i+\lambda_i'\right) + \sum_{i=1}^{n} p_i \nabla_b\left(y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i\right) + \sum_{i=1}^{n} q_i \nabla_b\left(w^T\phi(x_i) + b - y_i - \varepsilon - \lambda_i'\right)$$

$$+ \sum_{i=1}^{n} r_i \nabla_b(-\lambda_i) + \sum_{i=1}^{n} s_i \nabla_b(-\lambda_i') = 0$$

$$\Rightarrow \quad -\sum_{i=1}^{n} p_i + \sum_{i=1}^{n} q_i = 0$$

wrt λ_i:
$$\nabla_{\lambda_i}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\lambda_i+\lambda_i'\right) + \sum_{i=1}^{n} p_i \nabla_{\lambda_i}\left(y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i\right) + \sum_{i=1}^{n} q_i \nabla_{\lambda_i}\left(w^T\phi(x_i) + b - y_i - \varepsilon - \lambda_i'\right)$$

$$+ \sum_{i=1}^{n} r_i \nabla_{\lambda_i}(-\lambda_i) + \sum_{i=1}^{n} s_i \nabla_{\lambda_i}(-\lambda_i') = 0$$

$$\Rightarrow \quad C - p_i - r_i = 0 \qquad \forall \; i \in \{1, .., n\}$$

wrt λ_i':

$$\nabla_{\lambda_i'}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \lambda_i + \lambda_i'\right) + \sum_{i=1}^{n} p_i \nabla_{\lambda_i'}\left(y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i\right) + \sum_{i=1}^{n} q_i \nabla_{\lambda_i'}\left(w^T\phi(x_i) + b - y_i - \varepsilon - \lambda_i'\right)$$

$$+ \sum_{i=1}^{n} r_i \nabla_{\lambda_i'}(-\lambda_i) + \sum_{i=1}^{n} s_i \nabla_{\lambda_i'}(-\lambda_i') = 0$$

$$\Rightarrow \quad C - q_i - s_i = 0 \qquad \forall \; i \in \{1, .., n\}$$

- (complementary slackness)

$$\left.\begin{aligned} p_i\left(y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i\right) &= 0 \\ q_i\left(w^T\phi(x_i) + b - y_i - \varepsilon - \lambda_i'\right) &= 0 \\ r_i(-\lambda_i) &= 0 \\ s_i(-\lambda_i') &= 0 \end{aligned}\right\} \quad \forall \; i \in \{1, .., n\}$$

- (primal feasibility)

$$\left.\begin{aligned} y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i &\leq 0 \\ w^T\phi(x_i) + b - y_i - \varepsilon - \lambda_i' &\leq 0 \\ -\lambda_i &\leq 0 \\ -\lambda_i' &\leq 0 \end{aligned}\right\} \quad \forall \; i \in \{1, .., n\}$$

- (dual feasibility)

$$\left.\begin{array}{l} p_i \geq 0 \\ q_i \geq 0 \\ r_i \geq 0 \\ s_i \geq 0 \end{array}\right\} \forall i \in \{1, \dots, n\}$$

3. Dual Problem $= \max\limits_{\substack{\bar{p}, \bar{q}, \bar{r}, \bar{s} \\ \in \mathbb{R}^n_{\geq 0}}} \min\limits_{\substack{w \in \mathbb{R}^m, \\ b, \lambda, \lambda' \in \mathbb{R}}} \mathcal{L}(w, b, \lambda, \lambda', \bar{p}, \bar{q}, \bar{r}, \bar{s})$

$\bar{p} = [p_1, \dots, p_n]$

$\bar{q} = [q_1, \dots, q_n]$

$\bar{r} = [r_1, \dots, r_n]$

$\bar{s} = [s_1, \dots, s_n]$

$$= \max_{\substack{\bar{p}, \bar{q}, \bar{r}, \bar{s} \\ \in \mathbb{R}^n_{\geq 0}}} \min_{\substack{w \in \mathbb{R}^m, \\ b, \lambda, \lambda' \in \mathbb{R}}} \underbrace{\frac{1}{2}\|w\|^2 + C \cdot \sum_{i=1}^n (\lambda_i + \lambda_i') + \left[\sum_{i=1}^n p_i (y_i - w^T \phi(x_i) - b - \varepsilon - \lambda_i) + \sum_{i=1}^n q_i (w^T \phi(x_i) + b - y_i - \varepsilon - \lambda_i') + \sum_{i=1}^n r_i (-\lambda_i) + \sum_{i=1}^n s_i (-\lambda_i')\right]}_{\text{This is an unconstrained optimization problem in } w,\, b,\, \lambda,\, \lambda'}$$

can be solved by stationarity conditions in KKT $\Longleftarrow$ This is an unconstrained optimization problem in w, b, λ, λ'

Substituting the stationarity conditions,

$$w = \sum_{i=1}^{n} p_i \phi(x_i) - \sum_{i=1}^{n} q_i \phi(x_i) \; ; \quad \sum_{i=1}^{n} (p_i - q_i) = 0 \; ; \quad r_i = C - p_i \; ; \quad s_i = C - q_i$$

$$\Rightarrow \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} (\lambda_i + \lambda_i') + \Big[\sum_{i=1}^{n} p_i (y_i - w^T\phi(x_i) - b - \varepsilon - \lambda_i) + \sum_{i=1}^{n} q_i (w^T\phi(x_i) + b - y_i - \varepsilon - \lambda_i') + \sum_{i=1}^{n} r_i (-\lambda_i) + \sum_{i=1}^{n} s_i (-\lambda_i') \Big]$$

$$= \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \Big(\cancel{p_i \lambda_i} + \cancel{q_i \lambda_i'} + p_i y_i - p_i w^T\phi(x_i) - p_i b - p_i \varepsilon - \cancel{p_i \lambda_i} + q_i w^T \phi(x_i) + q_i b - q_i y_i - q_i \varepsilon - \cancel{q_i \lambda_i'} \Big)$$

$$= \frac{1}{2}\|w\|^2 - w^T \Big(\underbrace{\sum_{i=1}^{n} (p_i \phi(x_i) - q_i \phi(x_i))}_{=w} \Big) + \Big(\cancel{\sum_{i=1}^{n} (p_i - q_i)}^{\,0} \Big)(-b) + \sum_{i=1}^{n} \big(p_i y_i - p_i \varepsilon - q_i \varepsilon - q_i y_i \big)$$

$$= -\frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} (p_i - q_i) y_i - \varepsilon \sum_{i=1}^{n} (p_i + q_i)$$

$$= -\frac{1}{2} \Big(\Big(\sum_{i=1}^{n} (p_i \phi(x_i) - q_i \phi(x_i)) \Big)^T \Big(\sum_{j=1}^{n} (p_j \phi(x_j) - q_j \phi(x_j)) \Big) \Big) + \sum_{i=1}^{n} (p_i - q_i) y_i - \varepsilon \sum_{i=1}^{n} (p_i + q_i)$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} (p_i - q_i)(p_j - q_j) K(x_i, x_j) \right) - \varepsilon \sum_{i=1}^{n} (p_i + q_i) + \sum_{i=1}^{n} (p_i - q_i) y_i \quad //$$

Therefore, the dual problem is:

$$\max_{\bar{p}, \bar{q} \in \mathbb{R}^n} -\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} (p_i - q_i)(p_j - q_j) K(x_i, x_j) \right) - \varepsilon \sum_{i=1}^{n} (p_i + q_i) + \sum_{i=1}^{n} (p_i - q_i) y_i$$

$$\text{s.t.} \quad 0 \le p_i \le C \qquad \forall\, i \in \{1, \dots, n\}$$
$$0 \le q_i \le C$$

[$p_i, q_i \geq 0$ is obvious. Then, $r_i \geq 0 \Rightarrow C - p_i \geq 0 \Rightarrow p_i \leq C$;
$s_i \geq 0 \Rightarrow C - q_i \geq 0 \Rightarrow q_i \leq C$]

4. If a point (x_i, y_i) lies outside the ϵ-tube, there are 2 cases:

Case 1: $y_i - W^T\phi(x_i) - b > \epsilon$ — (1)

$\Rightarrow \lambda_i > 0 \quad (\lambda_i \neq 0)$

so that $y_i - W^T\phi(x_i) - b \leq \epsilon + \lambda_i$

Then by complementary slackness,

$r_i = 0 \Rightarrow p_i = C - r_i = C$

Also (1) $\Rightarrow y_i - W^T\phi(x_i) - b > -\epsilon$ (trivially)

And since we want to minimize λ_i' in the objective, here λ_i' will be 0

$(y_i - W^T\phi(x_i) - b > -\epsilon - \lambda_i'$ satisfied)

$\Downarrow$

$W^T\phi(x_i) + b - y_i - \epsilon - \lambda_i' < 0 \quad (\neq 0)$

$\Rightarrow$ Dual variable $q_i = 0$ (complementary slackness)

$\Rightarrow s_i = C - q_i = C$

$\therefore$, $\boxed{p_i = s_i = C \; ; \; q_i = r_i = 0}$

↳ Dual variables' values

Case 2: $y_i - W^T\phi(x_i) - b < -\epsilon$ — (2)

$\Rightarrow \lambda_i' > 0 \quad (\lambda_i' \neq 0)$

so that $y_i - W^T\phi(x_i) - b \geq -\epsilon - \lambda_i'$

Then by complementary slackness,

$s_i = 0 \Rightarrow q_i = C - s_i = C$

(2) $\Rightarrow y_i - W^T\phi(x_i) - b < \epsilon$ (trivially)

Minimum λ_i that makes above statement true is $\lambda_i = 0$

$(y_i - W^T\phi(x_i) - b < \epsilon + \lambda_i$ satisfied)

$\Downarrow$

$y_i - W^T\phi(x_i) - b - \epsilon - \lambda_i < 0 \quad (\neq 0)$

$\Rightarrow$ Dual variable $p_i = 0$ (complementary slackness)

$\Rightarrow r_i = C - p_i = C$

$\therefore$, $\boxed{p_i = s_i = 0 \; ; \; q_i = r_i = C}$

↳ Dual variables' values