Q2] Problem : $\min_{w,b,\varepsilon} \left(\frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{m} \varepsilon^2 \right)$

st. $y^i(w^T x^i + b) \geq 1 - \varepsilon_i \quad , i = 1, \dots, m$

a) No there is no non-negativity constraint on $\varepsilon \geq 0$ for l2 norm as in the minimization objective there is ε^2 and even for $\varepsilon < 0$ there will be positive loss, hence no need for constraint.

b) For lagrangian we need to find:

minimize $f_0(x)$

subject to $f_i(x) \leq 0 \quad i = 1, \dots m$

$h_i(x) = 0 \quad i = 1, \dots p$

then its solution is

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)$$

For our problem: $f_0(x) = \left(\frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1} \varepsilon_i^2 \right)$

$\forall i = 1, \dots m, \; f_i(x) \Leftrightarrow \left(-y^{(i)}(w^T x^{(i)} + b) + 1 - \varepsilon_i \leq 0 \right)$

$$L(w, b, \varepsilon, \alpha) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{m} \varepsilon_i^2 + \sum_{i=1}^{m} \alpha_i \left(1 - \varepsilon_i - y^{(i)}(w^T x^{(i)} + b) \right)$$

c) $W(\alpha) = \min_{w,b,\varepsilon} L(w, b, \varepsilon, \alpha)$ is objective of dual function.

for dual we need to use,

$\frac{\text{grad}(L)}{\text{grad}(w)} = 0$ or grad L wrt $w = 0$

$\nabla_w L = 0 \Rightarrow \boxed{w = \sum_{i=1}^{m} \alpha_i y^i x^i}$

$\nabla_b L = 0 \Rightarrow \cancel{w} \ \frac{\partial L}{\partial b} = -\sum \alpha_i y^i$

$\therefore \boxed{\sum \alpha_i y_i = 0}$

$\nabla_\varepsilon L = 0 \Rightarrow C\varepsilon = \alpha \Rightarrow \boxed{C\varepsilon_i = \alpha_i} \quad \forall i = 1, \ldots m$

$\therefore$ Dual $W(\alpha) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\alpha_i y^i x^i)^T (\alpha_j y^j x^j)$

$+ \frac{1}{2} \sum \frac{\alpha_i}{\varepsilon_i} \varepsilon_i^2 + \sum_{i=1} \alpha_i \left[-y_i (w^T x_i + b) + 1 - \varepsilon_i \right]$

$W(\alpha) = \frac{1}{2} \sum_{i=1}^{m} \alpha_i \varepsilon_i + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i \varepsilon_i - \left(\sum_{i=1}^{m} \alpha_i y_i \right) b$

$- \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^i y^j (x^i)^T x^j$

$W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^i y^j (x^i)^T x^j$

$- \frac{1}{2} \sum_{i=1}^{m} \frac{\alpha_i^2}{C}$

Hence dual form is $\max_\alpha W(\alpha)$

$\alpha_i \geq 0 \quad \forall i = 1, \ldots m$

$\sum_{i=1}^{m} \alpha_i y_i = 0$

DATE
PAGE

Q3] a) Setting $\alpha_i = 1$ $\forall i = 1, \dots, m$ and $b = 0$
For training sample i, $\{x_i, y_i\}$

$$|f(x^i) - y^i| = \left| \sum_{j=1}^{m} y^j \, k(x^j, x^i) - y^i \right|$$

$$= \left| \sum_{j=1}^{m} y^j e^{(-|x^j - x^i|^2 / \tau^2)} - y^i \right|$$

$$= \left| \sum_{j \neq i}^{m} y^j e^{(-|x^j - x^i|^2 / \tau^2)} \right|$$

$$\leq \sum_{j \neq i} \left| y^j e^{(-|x^j - x^i|^2 / \tau^2)} \right|$$

$$\leq \sum_{j \neq i} |y^{(j)}| . \left(e^{-|x^j - x^i|^2 / \tau^2}\right)$$

$$\leq \sum_{j \neq i} 1 \times e^{(-\varepsilon^2 / \tau^2)} \qquad (|y^j| = 1)$$

Assuming, $|x^j - x^i| \geq \varepsilon$ $\forall i \neq j$

$$= (m-1) e^{(-\varepsilon^2 / \tau^2)}$$

(~~as~~ taking max of all values)

$$\therefore (m-1) e^{-\varepsilon^2 / \tau^2} < 1$$

$$e^{\varepsilon^2 / \tau^2} > m - 1$$

$$\frac{\varepsilon^2}{\tau^2} > \log(m-1) \qquad \therefore m > 1$$

$$\frac{\varepsilon^2}{\log(m-1)} > \tau^2$$

or $$\boxed{\tau^2 < \frac{\varepsilon^2}{\log(m-1)}}$$

Hence found.

b) Yes the classifier will obtain 0 training error.

The SVM without slack variables will return a 0 error if there exists even a single solution and hence, will be 0 slack variables

Let's ~~each~~ consider i^{th} point such that $y^i(w^T x^i + b)$ for $\{x^i, y^i\}$

Now put $b=0$ for simplicity.

We have constraint $= y^i(w^T x^i)$
$= y^i f(x^i)$

Now as $f(x^i)$ and y^i have same sign we get

$y^i f(x^i) > 0$ and

using large α_i we have $(y^i)(w^T x^i + b) > 1$

hence our solution.