Q1. $J_w = l(w) + \lambda/2 \, \|w\|_2$. finding the update direction for stochastic gradient descent. when $l(w)$ is.

a). Quadratic loss $\Rightarrow$ $l(w) = \sum_{i=1}^{N} (w^T x_i - y_i)^2$.

$$\therefore \quad J_w = \sum_{i=1}^{N} (w^T x_i - y_i)^2 + \frac{\lambda}{2} \|w\|_2 .$$

Writing the objective function as an average of N datapoints.

$$(J_{w,\lambda})_i = \left\{ (w^T x_i - y_i)^2 + \frac{\lambda'}{2} \|w\|_2 \right\}$$

$$J_{w,\lambda} = \sum_{i=1}^{N} (J_{w,\lambda})_i$$

where $\lambda' = \lambda/N$

$$\frac{\partial (J_{w,\lambda})_i}{\partial w} = 2 (w^T x_i - y_i)^T x_i + \lambda' w$$

$$\therefore \quad w^{(K+1)} = w^{(K)} - t^{(K)} \left(2 (w^T x_i - y_i)^T x_i + \lambda' w \right)$$

b). Using Logistic loss. $l(w) = \sum_{i=1}^{N} \log(1 + \exp(-y_i (w^T x_i)))$

Writing the objective function as an average of N data points.

$$(J_{w,\lambda})_i = \left\{ \log\left(1 + \exp(-y_i (w^T x_i))\right) + \frac{\lambda'}{2} \|w\|_2 \right\}$$

also here $\lambda' = \lambda/N$.

$$J_{w,\lambda} = \sum_{i=1}^{N} (J_{w,\lambda})_i$$

$$\frac{\partial (J_{w,\lambda})_i}{\partial w} = \frac{1 * \exp(-y_i (w^T x_i))}{1 + \exp(-y_i (w^T x_i))} \cdot -(y_i * x_i) + \lambda' w .$$

$$\therefore \quad w^{(K+1)} = w^{(K)} - t^{(K)} \left(\lambda' w - \frac{y_i \, x_i \, \exp(-y_i (w^T x_i))}{1 + \exp(-y_i (w^T x_i))} \right) .$$

Q2.a). for the dual SVM problem.

$$\max_{\alpha} J(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j).$$

$$\text{s.t} \quad 0 \leq \alpha_i \leq C \quad \forall \ i = 1, \dots, N.$$

Taking the gradient of the objective w.r.t one of the dual variables. α_K.

$$\nabla_{\alpha_K} J(\alpha) = \left\{ 1 - \underbrace{\frac{1}{2} \sum_{j=1}^{N} \alpha_j y_i y_j K(x_i, x_j)}_{i = K} - \underbrace{\frac{1}{2} \sum_{i=1}^{N} \alpha_i y_i y_j K(x_i, x_j)}_{j = K}, \quad 0 \right\}.$$

$$\therefore \nabla_{\alpha_K} J(\alpha) = \begin{cases} 1 - y_i \sum_{j=1}^{N} y_j \alpha_j K(x_i, x_j) & K = i \text{ or } K = j \\ 0 & K \neq i \text{ and } K \neq j \end{cases}$$

$1 - y_i \sum_{j=1}^{N} y_j \alpha_j K(x_i, x_j)$ is a scalar and can be. > 0 or < 0 and. after normalizing. becomes 1. if $1 - y_i \sum_{j=1}^{N} y_j \alpha_j K(x_i, x_j) > 0$ & -1 otherwise.

Now since we are finding the update direction (normalized). w.r.t only one of the directions we disregard the gradient in the other directions or set them as zeros.

$$\therefore \quad \Delta \alpha_j = 0 \quad \forall \ j \neq i$$

$$\Delta \alpha_i = \begin{cases} 1 & \text{if } 1 - y_i \sum_{j=1}^{N} y_j \alpha_j K(x_i, x_j) \\ -1 & \text{otherwise.} \end{cases}$$

for the dual SVM problem. $t = \dfrac{\Delta\alpha^T(1 - H\alpha)}{\Delta\alpha^T H \Delta\alpha}$

Q2 (b)

if only one ~~dimension~~ element in $\Delta\alpha$ is getting updated then $\Delta\alpha$ looks like $[0, 0, \ldots\ldots, \Delta\alpha_i, 0, 0 \ldots, 0]^T$

$$H = N \begin{bmatrix} y_1 y_2 K_{11} & K_{12} & K_{13} \ldots & K_{1N} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ K_{N1} & \ldots & \ldots & K_{NN} \end{bmatrix}_{N} \qquad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \vdots \\ \alpha_N \end{bmatrix}_{N \times 1}$$

$$H\alpha = N \times 1 \quad \left[y_1 \sum_{i=1}^{N} y_i K_{1i} \alpha_i,\ y_2 \sum_{i=1}^{N} y_i K_{2i} \alpha_i, \ldots\ldots, y_N \sum_{i=1}^{N} y_i K_{Ni} \alpha_i \right]$$

$$1 - H\alpha = \left[1 - y_1 \sum_{i=1}^{N} y_i K_{1i} \alpha_i,\ 1 - y_2 \sum_{i=1}^{N} y_i K_{2i} \alpha_i, \ldots, 1 - y_N \sum_{i=1}^{N} y_i K_{Ni} \alpha_i \right]$$

Then. $\Delta\alpha^T(1 - H\alpha)$

$$= \left[\Delta\alpha_1 \left(1 - y_1 \sum_{j=1}^{N} y_j \alpha_j K_{1j}\right), \ldots, \Delta\alpha_N \left(1 - y_N \sum_{j=1}^{N} y_j \alpha_j K_{Nj}\right) \right].$$

$\therefore$ for the i^{th} dual variable $\Delta\alpha^T(1 - H\alpha)$ would be

$$\Delta\alpha_i \left(1 - y_i \sum_{j=1}^{N} y_j \alpha_j K_{ij}\right).$$

Where $K_{ij} = K(x_i, x_j)$.

$$\therefore t = \frac{\Delta\alpha_i \left(1 - y_i \sum_{j=1} y_j \alpha_j K_{ij}\right)}{\cancel{\Delta\alpha_i (1 - y_i \sum_{j=1}^{N} y_j \alpha_j K_{ij})} \quad \Delta\alpha^T H \Delta\alpha}.$$

Simplifying the denominator now,

$$H\Delta\alpha_i = [y_1 y_i K_{1i} \Delta\alpha_i,\ y_2 y_i K_{2i} \Delta\alpha_i, \ldots\ldots, y_N y_i K_{Ni} \Delta\alpha_i]$$

$$\Delta\alpha_i^T H \Delta\alpha_i = y_i y_i K_{ii} \Delta\alpha_i \Delta\alpha_i$$

$$\therefore t = \frac{\Delta\alpha_i \left(1 - y_i \sum_{j=1}^{N} y_j \alpha_j K_{ij}\right)}{\Delta\alpha_i \Delta\alpha_i y_i y_i K_{ii}}$$

Since $\Delta\alpha_i$ is just the direction. for $\Delta\alpha_i \neq 0$ it is either 1 or -1 $\therefore \Delta\alpha_i \Delta\alpha_i = 1$. Similarly for the case when $y_i = \pm 1$ then $y_i y_i = 1$

$$\therefore t = \Delta\alpha_i \left(1 - y_i \sum_{j=1}^{N} y_j \alpha_j K_{ij}\right) / K_{ii}$$

Q 2. c). The duality gap is given as.

$$dg(\alpha) = f_0(w, \xi_{s_i}) - g(\alpha).$$

where $g(\alpha)$ = dual soft-margin SVM for any dual feasible α.

$$\therefore\ g(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j).$$

primal soft-margin SVM

$$= \text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$$

subject to: $1 - y_i(\langle w, \phi(x_i)\rangle) - \xi_{s_i} \leq 0 \Rightarrow 1 - y_i(\langle w, \phi(x_i)\rangle) \leq \xi_{s_i}$

for all $i = 1 \dots N$. $\qquad \xi_{s_i} \geq 0$

$$-\xi_{s_i} \leq 0.$$

or $\xi_{s_i} \geq \max(0, 1 - y_i (\langle \alpha^T \phi(x_i), \phi(x_j)\rangle))$

$$\max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j \phi(x_j)^T \phi(x_i)\right).$$

$$\max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j)\right).$$

$$\boxed{\text{optimal } w = \sum_{i=1}^{N} \alpha_i y_i \phi(x_i).}$$

$$= \text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j)\right)$$

also. $\frac{1}{2} W^T W \Rightarrow \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j).$

$$\therefore\ f_0(w, \xi_{s_i}) \Rightarrow \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

$$+ C \sum_{i=1}^{N} \max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j)\right).$$

$\therefore \quad f_0(\omega, \xi_i) - g(\alpha)$.

$$= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) + C \sum_{i=1}^{N} \max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j)\right)$$

$$- \sum_{i=1}^{N} \alpha_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j).$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) + C \sum_{i=1}^{N} \max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j)\right) - \sum_{i=1}^{N} \alpha_i$$

Duality Gap

re-arranging a few terms.

$$\Rightarrow \boxed{- \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) + C \sum_{i=1}^{N} \max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j K_{ij}\right)}.$$

$$= - \sum_{i=1}^{N} \alpha_i \left(1 - y_i \sum_{j=1}^{N} \alpha_j y_j K_{ij}\right) + C \sum_{i=1}^{N} \max\left(0, 1 - y_i \sum_{j=1}^{N} \alpha_j y_j K_{ij}\right)$$

if $g_i = 1 - y_i \sum_{j=1}^{N} y_j \alpha_j K_{ij}$ Then the above Equation simplifies to.

$$= \boxed{- \sum_{i=1}^{N} \alpha_i g_i + C \sum_{i=1}^{N} \max(0, g_i)}$$

Duality Gap in terms of the gradient g_i , alpha_i and C.