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.

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

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.

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

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

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 <u>one</u> 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 $[0, 0, \dots\dots, \Delta\alpha_i, 0, 0\dots, 0]^T$

$$H = N\ \begin{array}{c} \begin{array}{cccc} y_1y_2K_{11} & K_{12} & K_{13} & \dots K_{1N} \end{array} \\ \begin{bmatrix} \cancel{\dots\dots} & & & \vdots \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ K_{N1} & \dots & \dots & K_{NN} \end{bmatrix} \\ N \end{array} \qquad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \vdots \\ \alpha_N \end{bmatrix}_{N\times 1}$$

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

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

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

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

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

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

Simplifying the denominator now,

$$H\Delta\alpha_i = \left[y_1y_iK_{1i}\Delta\alpha_i,\ y_2y_iK_{2i}\Delta\alpha_i, \dots\dots, y_Ny_iK_{Ni}\Delta\alpha_i\right]$$

$$\Delta\alpha_i^T H\Delta\alpha_i = y_iy_iK_{ii}\Delta\alpha_i\Delta\alpha_i$$

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

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

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 $y_iy_i = 1$

Q2. c). The duality gap is given as.

$$dg(\alpha) = f_0(w, \xi_i) - g(\alpha).$$

where $g(\alpha)$ = dual softmargin 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 softmargin SVM

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

subject to: $1 - y_i(\langle w, \phi(x_i)\rangle) - \xi_i \le 0 \Rightarrow 1 - y_i(\langle w, \phi(x_i)\rangle) \le \xi_i$

for all $i = 1 \dots N$. $\xi_i \geq 0$

$-\xi_i \le 0$.

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

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

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

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

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

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

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

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

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

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

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

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

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.