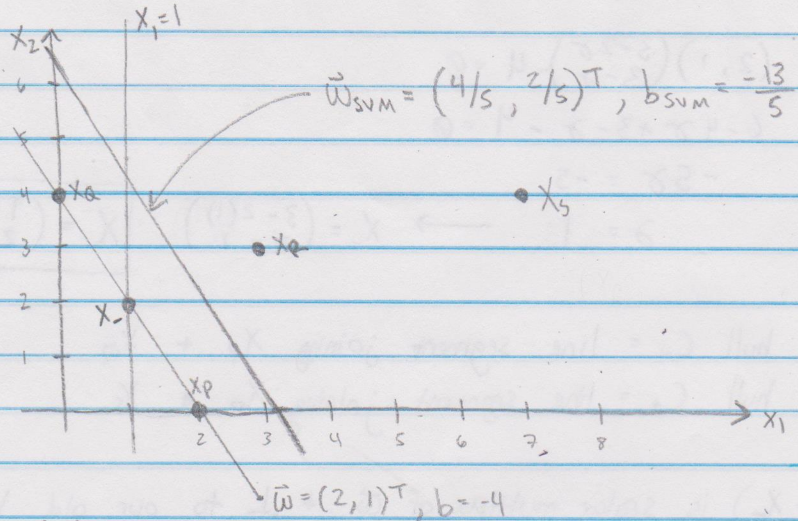
Briana Zhao
U93987617
EC414 HW5

1

(1.1) $\vec{X}_P = (2,0)$ $\vec{X}_Q = (0,4)$ label $y=-1$

$\vec{X}_R = (3,3)$ $\vec{X}_S = (7,4)$ label $y=+1$



(1.2) $(3,0)\begin{pmatrix}x_1\\x_2\end{pmatrix} - 3 = 0$

$3x_1 = 3 \rightarrow x_1 = 1$ (added to graph in 1.1)

$$\min_{x\in\{\vec{x}_P,\vec{x}_Q,\vec{x}_R,\vec{x}_S\}} |w^Tx+b| = \min\left\{\left|(3,0)\begin{pmatrix}2\\0\end{pmatrix}-3\right|, \left|(3\ 0)\begin{pmatrix}0\\4\end{pmatrix}-3\right|, \left|(3,0)\begin{pmatrix}3\\3\end{pmatrix}-3\right|, \left|(3,0)\begin{pmatrix}7\\4\end{pmatrix}-3\right|\right\}$$

$$= \min\{|6-3|, |0-3|, |9-3|, |21-3|\} = \min\{3,3,6,18\} = 3$$

so $\vec{W}_{can} = \frac{\vec{\omega}}{3}$ $b_{can} = \frac{b}{3}$

$$\boxed{\vec{W}_{can} = \begin{pmatrix}1\\0\end{pmatrix},\ b_{can} = -1}$$

(1.3) $\text{slope} = \frac{0-4}{2-0} = -2$ x_2 intercept $= 4$

$x_2 = -2x_1 + 4 \rightarrow 2x_1 + x_2 - 4 = 0 \rightarrow (2\ 1)\begin{pmatrix}x_1\\x_2\end{pmatrix} - 4 = 0$

$$\boxed{\vec{\omega} = (2,1)^T \quad b = -4}$$ (added to graph in 1.1)

(1.4) X_- = the orthogonal projection

$\rightarrow w^TX_- + b = 0$ since X_- is on the hyperplane

$(2,1)\begin{pmatrix}X_{-1}\\X_{-2}\end{pmatrix} - 4 = 0$

$\rightarrow \vec{\omega}$ is $\perp$ to the hyperplane

$X_R - X_-$ is also $\perp$ to the hyperplane

$$\underset{\text{scalar}}{(X_R - X_-) = \gamma\vec{\omega}}$$

$$X_- = X_R - \gamma\vec{\omega}$$

$$X_- = \begin{pmatrix}3\\3\end{pmatrix} - \gamma\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}3-2\gamma\\3-\gamma\end{pmatrix}$$

so $(2,1)\begin{pmatrix}3-2\gamma\\3-\gamma\end{pmatrix} - 4 = 0$

$$6 - 4\gamma + 3 - \gamma - 4 = 0$$

$$-5\gamma = -5$$

$$\gamma = 1 \longrightarrow X_- = \begin{pmatrix}3-2(1)\\3-1\end{pmatrix} \quad \boxed{X_- = \begin{pmatrix}1\\2\end{pmatrix}}$$ (added to graph in 1.1)

(1.5) Convex hull C_- = line segment joining X_P + X_Q
Convex hull C_+ = line segment joining X_R & X_S

$(X_R - X_-)$ is scalar multiple of $\vec{\omega}$ + $\perp$ to our old hyperplane
$(X_S - X_-)$ is scalar multiple of $\vec{\omega}$ & $\perp$ to our old hyperplane but longer than $(X_R - X_-)$

So $\underline{X_R}$ is the closest point to C_- and $\underline{X_-}$ is the point closest to C_+

$$W_{SVM} = \gamma(X_R - X_-) = \gamma\begin{pmatrix}3-1\\3-2\end{pmatrix} = \gamma\begin{pmatrix}2\\1\end{pmatrix}$$

hyperplane will go through the midpoint between X_R and X_-

$$\text{midpoint} = \frac{X_R + X_-}{2} = \frac{1}{2}\begin{pmatrix}3+1\\3+2\end{pmatrix} = \begin{pmatrix}2\\5/2\end{pmatrix}$$

$$W_{SVM}^T\begin{pmatrix}2\\5/2\end{pmatrix} + b_{SVM} = 0$$

$$\rightarrow b_{SVM} = -\gamma(2,1)\begin{pmatrix}2\\5/2\end{pmatrix} = -\gamma(13/2)$$

X_R + X_- are the closest to the hyperplane

Since we want canonical form, we say $W_{SVM}^T X_R + b = 1$

$$\gamma\left[(2,1)\begin{pmatrix}3\\3\end{pmatrix} - 13/2\right] = 1$$

$$\gamma\left[9 - 13/2\right] = 1$$

$$\gamma(5/2) = 1$$

$$\gamma = \frac{2}{5} \rightarrow W_{SVM} = \frac{2}{5}\begin{pmatrix}2\\1\end{pmatrix}, \quad b_{SVM} = -\frac{2}{5}\left(\frac{13}{2}\right)$$

$$\boxed{\vec{W}_{SVM} = (4/5 \;\; 2/5)^T \qquad b_{SVM} = -\frac{13}{5}}$$

(Sketched on graph from 1.1)

1.6

$$w_{svm} = -\alpha_P \vec{x}_P - \alpha_Q \vec{x}_Q + \alpha_R \vec{x}_R + \alpha_S \vec{x}_S$$

$\alpha_S = 0$ because $w^T x_S + b > 1$ so x_S is not a support vector
(x_S is too far from the SVM hyperplane)

$$\begin{pmatrix} 4/5 \\ 2/5 \end{pmatrix} = -\alpha_P \begin{pmatrix} 2 \\ 0 \end{pmatrix} - \alpha_Q \begin{pmatrix} 0 \\ 4 \end{pmatrix} + \alpha_R \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

system of equations

$$\begin{cases} -\alpha_P - \alpha_Q + \alpha_R = 0 & \text{since } \sum \alpha_i y_i = 0 \\ \frac{4}{5} = -2\alpha_P + 3\alpha_R \\ \frac{2}{5} = -4\alpha_Q + 3\alpha_R \end{cases}$$

$$\hookrightarrow \frac{2}{5} + 4\alpha_Q = 3\alpha_R$$

$$\alpha_R = \frac{2}{15} + \frac{4}{3}\alpha_Q$$

$\downarrow$

Plugging into eq. 2

$$\frac{4}{5} = -2\alpha_P + 3\left(\frac{2}{15} + \frac{4}{3}\alpha_Q\right)$$

$$\frac{4}{5} = -2\alpha_P + \frac{6}{15} + 4\alpha_Q$$

$$\frac{2}{5} - 4\alpha_Q = -2\alpha_P \longrightarrow \alpha_P = -\frac{1}{5} + 2\alpha_Q$$

$\downarrow$

Plugging both into eq. 1

$$-\left(-\frac{1}{5} + 2\alpha_Q\right) - \alpha_Q + \left(\frac{2}{15} + \frac{4}{3}\alpha_Q\right) = 0$$

$$\frac{1}{5} - 2\alpha_Q - \alpha_Q + \frac{2}{15} + \frac{4}{3}\alpha_Q = 0$$

$$\frac{1}{3} = \frac{5}{3}\alpha_Q$$

$$\alpha_Q = \frac{1}{5}, \quad \alpha_P = \frac{1}{5}, \quad \alpha_R = \frac{2}{5}$$

so $\boxed{\alpha_P = \frac{1}{5}, \ \alpha_Q = \frac{1}{5}, \ \alpha_R = \frac{2}{5}, \ \alpha_S = 0}$

The α's are unique because there is only 1 way to combine $\vec{x}_P$, $\vec{x}_Q$, and $\vec{x}_R$ to get $\vec{w}_{svm}$. There is only 1 way to get $\vec{x}_R$, and also only 1 way to get $\vec{x}_-$ from $\vec{x}_P$ and $\vec{x}_Q$. And since $(\vec{x}_R - \vec{x}_-)$ is a scalar multiple of $\vec{w}_{svm}$, there is only 1 way to get $\vec{w}_{svm}$. This also makes sense because there are only 1 or 2 points in each convex hull. Thus, α's are unique.

$\boxed{2}$

2.1 $$L(\omega) = -\frac{1}{n}\sum_{i=1}^{n} \underbrace{y_i \log(\sigma(z)) + (1-y_i)\log(1-\sigma(z))} \qquad \text{where } z = \omega^T x_i$$

To derive this part (call it F): $\frac{\partial F}{\partial \omega} = \frac{\partial z}{\partial \omega} \cdot \frac{\partial \sigma(z)}{\partial z} \cdot \frac{\partial F}{\partial \sigma(z)}$

$$\rightarrow \frac{\partial F}{\partial \sigma(z)} = \frac{y_i}{\sigma(z)} - \frac{1-y_i}{1-\sigma(z)}$$

$$\rightarrow \frac{\partial \sigma(z)}{\partial z} = \frac{\partial}{\partial z}(1+e^{-z})^{-1} = \frac{e^{-z}}{(1+e^{-z})^2}$$ which is the same as $\sigma(z)(1-\sigma(z))$

because $$\sigma(z)(1-\sigma(z)) = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) = \frac{1}{1+e^{-z}}\left(\frac{1+e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}}\right)$$

$$= \frac{1}{1+e^{-z}}\left(\frac{e^{-z}}{1+e^{-z}}\right) = \frac{e^{-z}}{(1+e^{-z})^2} \checkmark$$

$$\rightarrow \frac{\partial z}{\partial \omega} = \frac{\partial}{\partial \omega}(\omega^T x_i) = x_i$$

so $$\frac{\partial F}{\partial \omega} = [x_i] \cdot [\sigma(z)(1-\sigma(z))]\left[\frac{y_i}{\sigma(z)} - \frac{1-y_i}{1-\sigma(z)}\right]$$

$$= [x_i]\left[y_i(1-\sigma(z)) - (1-y_i)(\sigma(z))\right]$$

$$= [x_i][y_i - y_i\sigma(z) - \sigma(z) + y_i\sigma(z)]$$

$$= [x_i][y_i - \sigma(z)]$$

Therefore, $$\boxed{\nabla_\omega L(\omega) = -\frac{1}{n}\sum_{i=1}^{n}\left[(x_i)(y_i - \sigma(\overbrace{\omega^T x_i}^{z})\right]}$$