Ryan Gentry
ECE 6254
13 Feb 17

HW #2

(3) Argue that $\rho = \min_i |\langle \theta^*, \tilde{x}_i \rangle| > 0$ = distance from hyperplane θ^* to closest x_i in training data

$$\rho = \min_i |(\theta^*)^T \tilde{x}_i| = \min_i |(w^*)^T x_i + b^*|$$

$$\rho = \min_i \frac{|w^T x_i + b|}{\|w\|_2} = \min_i \frac{|w^T x_i + b|}{\sqrt{w^T w}}$$

a) $|w^T x_i + b| > 0$ & $\sqrt{w^T w} > 0$ by definition

so $\rho = \min_i |\langle \theta^*, \tilde{x}_i \rangle| > 0$

b) $\langle \theta^j, \theta^* \rangle \geq \langle \theta^{j-1}, \theta^* \rangle + \rho$

$$\langle \theta^{j-1} + \frac{y_i - \text{sign}((\theta^{j-1})^T \tilde{x}_i)}{2} \tilde{x}_i, \theta^* \rangle \geq \langle \theta^{j-1}, \theta^* \rangle + \min_i \frac{|w^T x_i + b|}{\|w\|_2}$$

$$(\theta^{j-1})^T \theta^* + \left[\left(\frac{y_i - \text{sign}((\theta^{j-1})^T \tilde{x}_i)}{2}\right) \tilde{x}_i\right]^T \theta^* \geq (\theta^{j-1})^T \theta^* + \min_i \frac{|w^T x_i + b|}{\|w\|_2}$$

$$\frac{1}{2}\left[\left(y_i - \text{sign}((\theta^{j-1})^T \tilde{x}_i\right) \tilde{x}_i\right]^T \theta^* \geq \min_i \frac{|w^T x_i + b|}{\|w\|_2}$$

☆ if $y_i \neq \text{sign}((\theta^{j-1})^T \tilde{x}_i) \rightarrow$ (mislabel requiring update) ☆

$$(y_i \tilde{x}_i)^T \theta^* \geq \min_i \frac{|w^T x_i + b|}{\|w\|_2}$$

$$y_i ((w^*)^T x_i + b^*) \geq \min_i \frac{|w^T x_i + b|}{\|w_2\|}$$

Therefore

$$\langle \theta^j, \theta^* \rangle \geq \langle \theta^{j-1}, \theta^* \rangle + \rho \geq j\rho$$

$$\langle \theta^{j-1}, \theta^* \rangle + \rho \geq j\rho, \quad j \leq \frac{(R^2+1)\|\theta^*\|_2^2}{\rho^2} \text{ b/c mislabel } \rho\,\theta^{j-1}$$

$$\langle \theta^{j-1}, \theta^* \rangle \rho + \rho^2 \geq (R^2+1)\|\theta^*\|_2^2$$

$$\boxed{\rho \langle \theta^j, \theta^* \rangle \geq \rho(\langle \theta^{j-1}, \theta^* \rangle + \rho) \geq (\max_i \|x_i\|_2^2 + 1)\|\theta^*\|_2^2}$$

(3) c) $\|\theta^j\|_2^2 \le \|\theta^{j-1}\|_2^2 + \|\tilde{x}_{ij}\|_2^2$

$$\|\theta^{j-1} + y_i \tilde{x}_{ij}\|_2^2 \le \|\theta^{j-1}\|_2^2 + \|\tilde{x}_{ij}\|_2^2$$

$$\boxed{\begin{array}{l} \text{if } y_i = 1 \rightarrow (\theta^{j-1} + \tilde{x}_{ij}) \le (\theta^{j-1}) + (\tilde{x}_{ij}) \ \checkmark \\ \text{if } y_i = -1 \rightarrow (\theta^{j-1} - \tilde{x}_{ij}) \le (\theta^{j-1}) + (\tilde{x}_{ij}) \ \checkmark \end{array}}$$

d) $\|\theta^j\|_2^2 \le \|\theta^{j-1}\|_2^2 + \|\tilde{x}_{ij}\|_2^2 \le j(1+R^2)$

$$\theta^{j-1} + (1 + x_i) \le \frac{(R^2+1)\|\theta^*\|_2^2}{\rho^2}(1+R^2)$$

$$\rho^2(\theta^{j-1} + (1+x_i)) \le \theta^*(1 + \max_i x_i)^2$$

$$\min_i |\langle \theta^*, \tilde{x}_i \rangle|^2 (\theta^{j-1} + 1 + x_i) \le \theta^*(1 + \max_i x_i)^2$$

$\max_i x_i > x_i$

$1 + \max_i x_i > 1 + x_i$

$(1 + \max_i x_i)^2 = 1 + 2(\max_i x_i) + (\max_i x_i)^2 > 1 + x_i + \theta^{j-1}$

$$\boxed{\theta^*(1 + 2\max_i x_i + (\max_i x_i)^2) > (1 + x_i + \theta^{j-1}) \min_i |\langle \theta^*, \tilde{x}_i \rangle|^2}$$

e) Cauchy-Schwartz: $|\langle u,v\rangle|^2 \le \langle u,u\rangle \cdot \langle v,v\rangle$

$\langle \theta^j, \theta^* \rangle \ge j\rho$ and $\|\theta^j\|_2^2 \le j(1+R^2)$

$j \le \frac{\langle \theta^j, \theta^* \rangle}{\rho}$, need $\langle \theta^j, \theta^* \rangle = \frac{(1+R^2)\|\theta^*\|_2^2}{\rho}$

$|\langle \theta^j, \theta^* \rangle|^2 \le \langle \theta^j, \theta^j \rangle \cdot \langle \theta^*, \theta^* \rangle$

$= \cancel{(\theta^j)^T \theta^j} \cdot \cancel{(\theta^*)}$

$\|\theta^j\|_2^2 \cdot \|\theta^*\|_2^2$

$\le j(1+R^2)\|\theta^*\|_2^2$

$j \le \frac{(j(1+R^2)\|\theta^*\|_2^2)^{\frac{1}{2}}}{\rho}$

$j^{\frac{1}{2}} \le \left[\frac{(1+R^2)\|\theta^*\|_2^2}{\rho}\right]^{\frac{1}{2}}$

$\boxed{j \le \frac{(1+R^2)\|\theta^*\|_2^2}{\rho^2}}$

$j\rho \le \langle \theta^j, \theta^* \rangle \le (j(1+R^2)\|\theta\|_2^2)^{\frac{1}{2}}$