2014 Midterms

① a) $J(\theta) = (X\theta - \vec{y})^T (X\theta - y)$

$$= \theta^T X^T X \theta - 2\vec{y}^T X \theta + \vec{y}^2$$

$$\frac{\partial J}{\partial \theta} = (X^T X + (X^T X)^T)\theta - 2\vec{y} X = 0$$

$$\cancel{2} X^T X \theta = \cancel{2} \vec{y}^T X$$

$$\Theta = (X^T X)^{-1} \vec{y}^T X \quad X^T y$$

$\mathbb{R}^{(n+1)\times(n+1)}$ $\quad \mathbb{R}^{1\times m}$ $\quad \mathbb{R}^{m\times(n+1)}$ $\quad n+1\times m * m\times 1$

$(n+1)\times 1$

$$[X \ \vec{v}] \begin{bmatrix} -X^T \\ -\vec{v}^T \end{bmatrix}$$

$(m\times(n+2))$ $\quad [n+2]\times m$

$$\hat{\theta} = I X^T y$$

$$\Rightarrow J(\hat{\theta}) = (X X^T y - \vec{y})^T (X X^T \vec{y} - \vec{y})$$

b) $\tilde{J}(\hat{\theta}_{new}) = (\tilde{X}\tilde{X}^T\vec{y} - \vec{y})^T(\tilde{X}\tilde{X}^T\vec{y} - \vec{y})$

$$= \left([X \ \vec{v}]\begin{bmatrix} X^T \\ \vec{v}^T \end{bmatrix}\vec{y} - \vec{y}\right)^T \left([X \ \vec{v}]\begin{bmatrix} X^T \\ \vec{v} \end{bmatrix}\vec{y} - \vec{y}\right)$$

$\hat{\Theta}_{new} = (\cancel{\tilde{X}^T \tilde{X}})^{-1} \tilde{X}^T y$

$$([X X^T + \vec{v}\vec{v}^T])\vec{y} - \vec{y})^T(([X X^T + \vec{v}\vec{v}^T])\vec{y} - \vec{y})$$

CS229 2014 Mid-term (cont'd)

(2) a) $c_1, \ldots, c_i, \ldots c_v$ = number of times word i appears in document

(w_j in slot i)

$$p(x|y) = p(c_i = m \mid y = x)$$

$$\cancel{L = (c) \prod_{i=1}^{n} p(w_v^{(i)} = m \mid y = x)^{c_v} \phi_y^{y^{(i)}} (1-\phi_y)^{1-y^{(i)}}}$$

probability 1

~~(c) # of possible combination for~~

$$\cancel{L = (c) \prod_{i=1}^{V} p(w^{(i)} = v}$$

b/c we assume x are ∥, conditional on y.

$$L = (c) \prod_{i=1}^{n} p(w^{(i)} \mid y = x)^{c_i} \phi_y^{(y)} (1-\phi)^{(1-y)}$$

$$p(x^{(i)} \mid y = y)\, \phi^{(y)} (1-\phi)^{(1-y)}$$

$$= \prod_{i=1}^{n} p(x^{(i)} \mid y) + ny \log \phi$$

$p(w_i|y) = p$

$$\ln(L) = \sum_{i=1}^{n} \left(\log(c_i p(x^{(i)}) \mid y) + n(1-y) \log(1-\phi) \right)$$

$$= \sum_{i=1}^{n} (c_i) \log(p(x_i) \mid y)^{c_i} + ny \log \phi + n(1-y)(\log(1-\phi))$$

$p(w_i|y)$:
probability word i is in doc, given y

$\tilde{J}(\tilde{\theta}_{new}) =$

$$\left(\vec{y}^T \begin{bmatrix} x x^T \\ + \\ \vec{v}\vec{v}^T \end{bmatrix} \cancel{y} - \vec{y}^T \right) \left([x x^T + \vec{v}\vec{v}^T] \vec{y} - \vec{y} \right)$$

$$\vec{y}^T \begin{bmatrix} x x^T \\ + \\ \vec{v}\vec{v}^T \end{bmatrix} [x x^T + \vec{v}\vec{v}^T] y - \vec{y}^T \begin{bmatrix} x x^T \\ + \\ \vec{v}\vec{v}^T \end{bmatrix} \vec{y}$$

$$- \vec{y}^T [x x^T + \vec{v}\vec{v}^T] \vec{y} + \vec{y}^T \vec{y})$$

$$\tilde{J}(\tilde{\theta}_{new}) = \vec{y}^T (x x^T)(x x^T) + \cancel{\alpha} 2 x x^T \vec{v}\vec{v}^T + (\vec{v}\vec{v}^T)(\vec{v}\vec{v}^T)] \vec{y} - 2\vec{y}(x x^T + \vec{v}\vec{v}^T)\vec{y} + \vec{y}^T \vec{y}$$

$$= \vec{y}\cancel{x x^T} (x x^T \vec{y} - \vec{y})^T (x x^T \vec{y} - \vec{y}) + 2\vec{y} x x^T (\vec{v}\vec{v}^T \cancel{I}) \vec{y} + \vec{y}^T (\vec{v}\vec{v}^T)(\vec{v}\vec{v}^T) \vec{y}$$

$$- 2\vec{y}^T (\vec{v}\vec{v}^T) \vec{y} + \vec{y}^T (\vec{v}\vec{v}^T)(\vec{v}\vec{v}^T \cancel{I}) \vec{y}$$

$$\tilde{J}(\tilde{\theta}_{new}) = J(\hat{\theta}) + f(x, \vec{v}, \vec{y})$$

$$f(x, \vec{v}, \vec{y}) = 2\vec{y}^T x x^T (\vec{v}\vec{v}^T \cancel{I}) \vec{y} + \vec{y}^T (\vec{v}\vec{v}^T)(\vec{v}\vec{v}^T) \vec{y}$$

$$-2\vec{y}^T (\vec{v}\vec{v}^T) \vec{y} + \vec{y}^T \vec{v}\vec{v}^T y$$

$$= 2\cancel{\vec{y}^T} - 2\vec{y}^T (\vec{v}\vec{v}^T)\vec{y} = -\vec{y}^T (\vec{v}\vec{v}^T)\vec{y} = -(\vec{v}^T y)^2$$

show $f(x, \vec{v}, \vec{y}) \leq 0$

d) no, b/c $\uparrow$ only describes bias reduction for extra features, doesn't account for variance

$p(y=1|x)$

$$= \frac{1}{\left[1 + \underbrace{\prod_{j=1}^{V} \frac{\cancel{D_i}\; p(x_j|y=0)^{c_j}}{\cancel{D_i}\; p(x_j|y=1)} \frac{p(y=0)}{p(y=1)}}_{F}\right]}$$

$\Rightarrow$

$$F \;\cancel{\ldots}\; \leq 1 \Rightarrow \hat{y} = 1$$

$$= \prod_{j=1}^{V} \left[p(x_j|y=0)/p(x_j|y=1)^{c_j} \frac{p(y=0)}{p(y=1)} \right]$$

$$\log(L) = \sum_{i=1}^{V} \left[\log(x_j|y=0)^{c_j} - \log(x_j|y=1)^{c_j} + \log(p(y=0)) - \log p(y=1) \right]$$

$$\leq \log(1) = 0$$

$$\leq 0$$

$$\log(L) = \sum_{i=1}^{V} c_j \overbrace{\left[\log(x_j|y=0) - \log(x_j|y=1)\right]}^{W} + \underbrace{V\left[\log \cancel{\phi}^{(1-\phi)} - \log(\phi)\right]}_{b}$$

$$= \sum_{=1}^{V} \left(W c_j + b\right] \geq 0$$

$$p(y=1|x) \geq p(y=0|x) \Rightarrow$$

$$\Rightarrow \frac{p(x|y=1)\,p(y=1)}{\cancel{p(x)}} \geq \frac{p(x|y=0)\,p(y=0)}{\cancel{p(x)}}$$

$$\Rightarrow N(\mu_0, \Sigma_0)\,\phi \geq N(\mu_1, \Sigma_1)(1-\phi)$$

$$\ln(N(\mu_0, \Sigma_0)) + \ln(\phi) \geq \ln(N(\mu_1, \Sigma_1)) + \ln(1-\phi)$$

$$\cancel{\ln\left(\frac{1}{2n^{n/2}}\right)} + \ln\frac{1}{|\Sigma_0|^{1/2}} + \left(-\frac{1}{2}(x-\mu_0)^T \Sigma_0^{-1}(x-\mu_0)\right)$$

$$+ \ln\phi \geq \cancel{\ln\left(\frac{1}{2n^{n/2}}\right)} + \ln\left(\frac{1}{|\Sigma_1|^{1/2}}\right) +$$

$$+ \left(-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)\right] +$$

$$\ln(1-\phi)$$

$$\sum_{i=0}^{1} -\frac{1}{2}(x-\mu_i)\Sigma_i^{-1}(x-\mu_i)$$

$$\Rightarrow \left[-\frac{1}{2}\ln|\Sigma_0| + \frac{1}{2}\ln|\Sigma_1|\right] \cancel{-\frac{1}{2}(x-\mu_0)^T\left[\Sigma_0^{-1} - \Sigma_1^{-1}\right](x-\mu_0)}$$

$$+ \ln\phi - \ln(1-\phi) \geq 0$$

2b)

$$c \triangleq \left[\frac{1}{2} \ln|\Sigma_0| - \frac{1}{2} \ln|\Sigma_1| \right]$$

$$+ \ln\phi - \ln(1-\phi) + \frac{1}{2}\left[\mu_0^T \Sigma_0^{-1} \mu_0 - \mu_1^T \Sigma_1^{-1} \mu_1\right]$$

$$\cancel{-\frac{1}{2}(x-\mu_0)^T \left[\Sigma_0^{-1} - \Sigma_1^{-1}\right](x-\mu_0)}$$

$$-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)$$

$$= -\frac{1}{2}\left[x^T \Sigma_i^{-1} x \; -2 x^T \Sigma_i^{-1} \mu_i \; \cancel{-\mu_i^T x} \; + \mu_i^T \Sigma_i^{-1} \mu_i \right]$$

$$-2\mu_i \Sigma_i^{-1} x$$

$$\sum_i D = -\frac{1}{2} x^T\left(\Sigma_0^{-1} + \Sigma_1^{-1}\right) x$$

$$\cancel{-2 x^T (\Sigma_0^{-1} + \Sigma_1^{-1}) x + \mu_1}$$

$$\cancel{-2 x^T \Sigma_0^{-1} \mu_0 - 2 x^T \Sigma_1^{-1} \mu_1 +}$$

$$-2 x^T \left(\Sigma_0^{-1} \mu_0 + \Sigma_1^{-1} \mu_1 \right)$$

$$+ \mu_0^T \Sigma_0^{-1} \mu_1 + \mu_1^T \Sigma_1^{-1} \mu_1$$

2b, cont'd)

$$\beta = -2 \left[\Sigma_0^{-1} \mu_0 - \mu_1 \Sigma_1^{-1} \right]$$

$$A = -\frac{1}{2} \left[\Sigma_0^{-1} - \Sigma_1^{-1} \right)$$

---

3 i) $p(y;n) = b(y) \exp(n^T T(y) - a(n))$

$$n^T = \log(1-\phi)$$

$$a(n) = -\log(\phi)$$

$$T(y) = 1$$

$$b(y) = y-1$$

$$e^{n^T} = 1-\phi$$

$$\phi = 1 - e^{n^T}$$

3aii) $\log(\cancel{xy}^{(i)} \mid x^{(i)}; \theta) = (y-1)\log(\cancel{y-1})(1-\phi)\,\phi$

$= \cancel{\Sigma}\,(y-1)\log(1-\phi) + \log(\phi)$

$= y\log(1-\phi) + \log\phi - \log(1-\phi)$

$\phi = 1 - e^{+n} = 1 - e^{+\theta^T x}$

$= y\log(1-(1-e^{+\theta^T x})) + \log(1-e^{\theta^T x})$

$-(\log(1-(1-e^{\theta^T x})))$

$= y(\log(e^{\theta^T x})) + \log(1-e^{\theta^T x})$

$-\log(e^{\theta^T x})$

$\log \prod^n p^{(i)} = \sum_{i=1}^{n} (y-1)\,\theta^T x + \log(1-e^{\pm\theta^T x})$

3 b) (cont'd)

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial \sum_{i=1}^{m} y x^T - \frac{x^T}{(1-e^{\theta^T x})}}{\partial \theta}$$

$$= \sum_{1}^{m} \frac{+x^T(-xe^{\theta^T x})}{(1-e^{\theta^T x})^2} = \sum_{i=1}^{n} \frac{-x^T x e^{\theta^T x}}{(1-e^{\theta^T x})^2} \geq 0$$

$$\theta := \theta - \frac{\nabla \ell(\theta)}{H}$$

$$\tilde{x}^T \quad \sum \left(\frac{-x x^T e^{\theta^T x}}{(1-e^{\theta^T x})^2} \right) \tilde{x}$$

$$= \sum \frac{-(\tilde{x}^T x)(x^T \tilde{x}) e^{\theta^T x}}{(1-e^{\theta^T x})^2}$$

$$= \sum \frac{-(\tilde{x} x)^2 e^{\theta^T x}}{(1-e^{\theta^T x})} \leq 0$$

4) a)

$$\mathcal{L} = \sum_{i=1}^{n} \frac{1}{2}\|w\|^2 + \alpha_i(y^{(i)} - w^T x^{(i)} - b - \epsilon)$$

$$+ \alpha_i^* (w^T x^{(i)} + b - y^{(i)} - \epsilon)$$

$$\max_{\alpha, b}\left(\min_{w,b}(\mathcal{L}) \right) = 0$$

$$\frac{\partial \mathcal{L}}{\partial w} = \sum_{i=1}^{n} \|w\| + -\alpha_i x^{(i)} + \alpha_i^* x^{(i)} = 0$$

$$\|w\| = \sum_{i=1}^{n} \left(x^{(i)} (\alpha_i - \alpha_i^*) \right)$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\alpha_i + \alpha_i^* = 0 \qquad \sum(\alpha_i^* - \alpha_i) = 0$$

$$(\alpha_i^* - \alpha_i) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_i} = y^{(i)} - w^T x^{(i)} - b - \epsilon = 0 = y^{(i)} - x^{(i)T} x^{(i)} (\alpha_i - \alpha_i^*) - b - \epsilon = 0$$

$$b = y - x^T x(\alpha - \alpha^*) - \epsilon$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_i^*} = \left(w^T x^{(i)} + b - y^{(i)} - \epsilon \right) = 0$$

$$x^{(i)T} x (\alpha_i - \alpha_i^*) + b - y^{(i)} - \epsilon = 0$$

$$\cancel{x^{(i)T} x^{(i)} (\alpha_i - \alpha_i^*) + y - x^{(i)T} x^{(i)} (\alpha_i - \alpha^*)} - \epsilon$$

$$\cancel{- y^{(i)}} - \epsilon$$

$$-2\epsilon = 0$$

4a (ii) $\quad l = \log \prod_{i=1}^{m} \left(1 - e^{\theta^T x^{(i)}}\right)$

$\cancel{l = D}$

cdf poisson distribution

$$l = \cancel{\log}\sum_{i=1}^{m} \left[(y-1)\log\left(e^{\theta^T x}\right) + \log\left[\frac{\cancel{-xe^{\theta^T x}}\,(1-e^{\theta^T x})}{\cancel{(1-e^{\theta^T x})}}\right]\right]$$

$$= \sum \left[(y-1)\,\theta^T x - \log \theta^T x \right.$$

$$\left. + \log\left[1 - e^{\theta^T x}\right]\right]$$

4b) $\quad \frac{\partial l}{\partial \theta} = \sum \left[(y-1)\,x^T \;\cancel{-x^T}\; - \frac{+xe^{\theta^T x}}{(1-e^{\theta^T x})} \right]$

$$\cancel{= \sum \left[y x^T + \frac{xe^{\theta^T x}}{(1-e^{\theta^T x})} \right]}$$

$$\cancel{= \sum y \left[\frac{\ldots\, e^{\theta^T x}\, x^T + xe^{\theta^T x}\, x^T}{1-e^{\theta^T x}} \right]}$$

$$\cancel{= \sum \frac{(y-1)\,x\,(1-e^{\theta^T x}) - xe^{\theta^T x}}{(1-e^{\theta^T x})}}$$

$$\cancel{= \sum (y \ldots} \quad = \sum y x^T - \frac{x^T(1-\cancel{e^{\theta^T x}}) + x\cancel{e^{\theta^T x}}}{(1-e^{\theta^T x})}$$

$$= \sum y x^T - \frac{x^T}{(1-e^{\theta^T x})}$$

4c) $L = \frac{1}{2} \sum_{i=1}^{n} x^{(i)T} x^{(i)} (\alpha_i - \alpha_i^*)^2$

$+ \sum_{i=1}^{n} \alpha_i \left[-y^{(i)} + \cancel{\sum_{j=1}^{n} x^{(j)T} (\alpha_i^* - \alpha_i) x^{(i)}} + b + \epsilon \right]$

$+ \sum_{i=1}^{n} \alpha_i^* \left[-y^{(i)} + \cancel{\sum_{j=1}^{n} x^{(j)T} (\alpha_j^* - \alpha_j) x^{(i)}} + b + \epsilon \right]$

$+ \sum_{i,j}^{n} (\alpha_i + \alpha_i^*) \cancel{x^{(j)T}} \cancel{x^{(i)}} \phi(x_j) \phi(x_i)$

$||$

$f(w, x) = \sum_{i=1}^{m} x^{T(i)} (\alpha_i - \alpha_i^*) x + b$

~~n+1 test the, just change n to (n+1), recompute recompute w~~

$f(w, x) \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x^{(i)}, x) + b)$

~~(2015 Practice Midterm)~~

(5) a) $\hat{\varepsilon}(h_0) \leq \varepsilon(h_0) + \gamma$ w/ prob at least $1-\delta$

$\hat{\varepsilon}(h_0) \leq \varepsilon(h^*) + \eta + \gamma$

$\leq \varepsilon(\hat{h}) + \eta + \gamma$

$\leq \hat{\varepsilon}(\hat{h}) + \eta + 2\gamma$

$\Rightarrow \hat{\varepsilon}(h_0) > \hat{\varepsilon}(\hat{h}) + \eta + \gamma$

w/ prob at most δ

b) if $\varepsilon(h_0) > \varepsilon(h^*) + \eta$

show prob of yes is at most δ

$\varepsilon(h_0) ~~- \varepsilon(h^*) > \eta~~$

$\cancel{\varepsilon(h_0) > \varepsilon(\hat{h}) + \eta}$

$>$

$\boxed{5b}$ if $\varepsilon(h_0) > \varepsilon(h^*) + \eta$

show that probability of YES is at most δ

with prob $(1-\delta)$

$\hat{\varepsilon}(h_0) \geq \varepsilon(h_0) - \gamma$

$\varepsilon(h_0) > \varepsilon(h^*) + \eta$ $\leq$ $\varepsilon(h^*) + \eta + \gamma$

$> \varepsilon(h^*) + \eta - 2\gamma$

$> \varepsilon(h_\gamma) - 2\delta$

$> \varepsilon(h^*) + \eta - 2\delta$

$> \hat{\varepsilon}(\hat{h}) + \eta - 2\gamma$

$> \hat{\varepsilon}(h_0) \geq \varepsilon(h_0) - \gamma$

$\geq \varepsilon(h^*) + \eta - \gamma$

$\geq \varepsilon($

$\hat{\varepsilon}(h_0) \geq \varepsilon(h_0) - \gamma$

$\geq \varepsilon(h_0) - 2\gamma$

$\geq \varepsilon(h^*) + \eta - 2\gamma$

5c)

if $m \to \infty$, $\gamma \to 0$, so

$$\hat{\varepsilon}(h_0) = \hat{\varepsilon}(h^*) \le \varepsilon(\hat{h}) + \cancel{x} + \gamma$$

$$\le \varepsilon(h^*) + n + \gamma$$

w/ prob $1-\delta$

if $\gamma \to 0$

$\Rightarrow$

$\hat{\varepsilon} \le \varepsilon(h^*)$

~~we're already shown~~

~~is $\varepsilon(h_0)$ is n-optimal~~

~~$p(\ldots) \le \delta$~~

if $m \to \infty$ $\quad n >> 2\gamma$

$\le$ w/ prob $\ge 1-\delta$

$$\{\hat{\varepsilon}(\hat{h}) - \hat{\varepsilon}(\hat{h^*})\} \le \tau$$

$$\text{if } \left[\hat{\varepsilon}(h^*) - \hat{\varepsilon}(\hat{h})\right] < (n - 2\gamma)$$

w/ prob δ ~~max~~ $(\hat{\varepsilon}\,\hat{h} - \hat{\varepsilon}(\hat{h})) < \tau$ $\quad \tau < n - 2\gamma$

as $m \to \infty$

$\Rightarrow$ true w/ prob at least $1-\delta$

5c)

$$\hat{\varepsilon}(h^0) \le \varepsilon(h^0) + \gamma$$

$$\le \varepsilon(h^*) + \gamma$$

$$\le \varepsilon(\hat{h}) + \gamma$$

$$\le \hat{\varepsilon}(\hat{h}) + 2\gamma$$

w/ prob $\ge 1-\delta$

$$\Rightarrow \hat{\varepsilon}(h_0) - \hat{\varepsilon}(\hat{h}) \le 2\gamma \quad (1)$$

w/ prob $\ge 1-\delta$

$$\hat{\varepsilon}(h_0) < \hat{\varepsilon}(\hat{h}) + \eta - 2\gamma$$

$$\hat{\varepsilon}(h_0) - \hat{\varepsilon}(\hat{h}) < \eta - 2\gamma \quad (2)$$

given (1), if ~~(1) holds~~,

$$2\gamma < \eta - 2\gamma$$

$\Rightarrow$ (2) holds w/ prob $\ge 1-\delta$

4 (a)

false, in general steps under newton's gradient steps take longer, but converge to optimum in fewer steps, can't infer which method will converge in less time

(b) false, b/c stochastic gradient descent is <u>stochastic</u>, we randomly sample ~~train~~ data from training dataset, ~~so the data can be~~ different training data will produce different optimal parameters

(c) ~~if~~ $K(x,y)$ is valid kernel $\Rightarrow$
K is symmetric, PSD $\Rightarrow$ ~~$\sum_k \sum_i z$~~

True

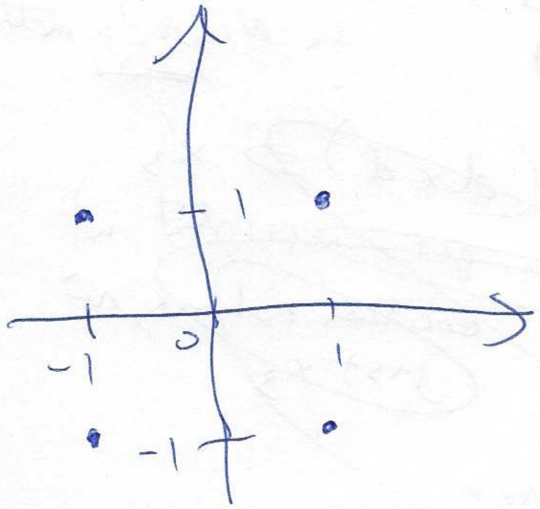
$$K_{norm}^{(x,y)} = \frac{K(x,y)}{\sqrt{K(x,x)}\sqrt{K(y,y)}}$$

$$z^T K z = \sum_k \left(\sum_i z_i \phi_k(x^{(i)}) \right)^2$$

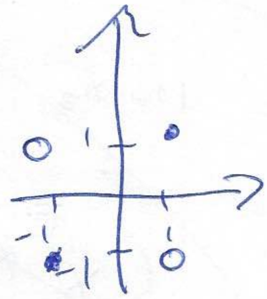
$$z^T K_{norm} z = \sum_k \left(\sum_i \tilde{z}_i \phi_k(x^{(i)}) \right)^2 \geq 0$$

$$\tilde{z}_i \triangleq \frac{z}{\sqrt{K(x,x)}}$$

6d)



no, counter example



6e) True

if $h \geq 0$, ~~$\Rightarrow$ dist~~ h is distance, or margin from separating hyperplane

if $h(x_0) = 0$, and $h(x_1) = c$

$c = h(x_1) - h(x_0) = w(x_1 - x_0)$ ~~$+b - b$~~

6f) We can't tell, b/c VC dimension just tells us size of largest dataset that classifier shatters (can classify all points), ~~but~~, but there may exist datasets of VC-dimension that are not shattered by classifier

6g) ~~True~~ ~~False~~ true

$\theta = (X^T X)^{-1} X^T y$ = Newton's method (onestep)

i) bc $\theta = (X^T X)^{-1} X^T \vec{y}$

linear invariant to Newton's method

$x_3' = 100 x_3$

$(d \times d) \Rightarrow x_3'$ gets multipled w/ other columns, not just x_3

$\Rightarrow (X'^T X')^{-1}$

~~$(d \times n) \times n (n \times d)(d \times n$~~

$(n \times d\ d \times n)$ ~~$(n \times d)$~~ $(n \times 1)$

$(d \times n\ n \times d)(d \times n)(n \times 1)$

$(d \times 1)$

6h) ii) $\nabla J = (\theta^T x - y) x$

false

if $\theta_0 = [0] \Rightarrow \nabla J_0^{(x)} = -yx$

$\nabla(J_0(x') = -y x'$

$\Rightarrow$

$\Rightarrow \theta_1 = \theta_0 - \alpha \nabla J_0$

$\theta_1(x) = \theta_0(x) - \alpha \nabla J_0(x)$

$x_0 = x_0'$

$\Rightarrow \theta_1(x) = +y x'$

$\theta x_1 = x_1'$

$\Rightarrow 100 x_2 = x_2'$

$\Rightarrow \theta_0(x) = yx$

$\Rightarrow \theta_2' = 100\theta_2$

$\Rightarrow$