Shu Bin

HW2 (I checked the solution for this HW)

1) a) $\sigma'(x) = \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right)$

$$= \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= e^{-x}(1+e^{-x})^{-2}$$

$$= e^{-x}(1+e^{-x})^{-1}(1+e^{-x})^{-1}$$

$$= \left(\frac{e^{-x}}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right)$$

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

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

b) $\nabla_\theta \, nll(\theta) = \nabla_\theta\left[-\sum_i y_i \log\sigma(\theta^T\vec{x_i}) + (1-y_i)\log(1-\sigma(\theta^T\vec{x_i}))\right]$

$$= -\sum_i y_i \frac{1}{\sigma(\theta^T\vec{x_i})}\sigma'(\theta^T\vec{x_i}) + (1-y_i)\left(\frac{-1}{1-\sigma(\theta^T\vec{x_i})}\right)\sigma'(\theta^T\vec{x_i})$$

$$= -\sum_i y_i \frac{1}{\sigma(\theta^T\vec{x_i})}\sigma(\theta^T\vec{x_i})(1-\sigma(\theta^T\vec{x_i})) + (1-y_i)\left(\frac{-1}{1-\sigma(\theta^T\vec{x_i})}\right)(1-\sigma(\theta^T\vec{x_i}))\sigma(\theta^T\vec{x_i})$$

$$= -\sum_i y_i(1-\sigma(\theta^T\vec{x_i}))\vec{x_i} + (1-y_i)(-\sigma(\theta^T\vec{x_i}))\vec{x_i}$$

$$= -\sum_i y_i\vec{x_i} - y_i\sigma(\theta^T\vec{x_i}) - \vec{x_i}\sigma(\theta^T\vec{x_i}) + y_i\sigma(\theta^T\vec{x_i})$$

$$= -\sum_i \vec{x_i}(y_i - \sigma(\theta^T\vec{x_i}))$$

$$= \sum_i \vec{x_i}(\sigma(\theta^T\vec{x_i}) - y_i)$$

let $\mu_i = \sigma(\theta^T\vec{x_i})$, then

$$= X^T(\mu - y)$$

TGabato22

C) $H_\theta = \nabla_\theta (\nabla_\theta \, n\mathcal{U}(\theta))^T$

$= \nabla_\theta (x^T(\mu - y))^T$

$= \nabla_\theta ((\mu-y)^T x)$

$= \nabla_\theta (\mu^T x - y^T x)$

$= \nabla_\theta \sigma(x\theta)^T x = x^T \operatorname{diag}(\mu(1-\mu)) x = x^T S x$

to show H_θ is positive semi-definite, we need to show S is positive semi-definite

since $\sigma(\cdot) \in (0,1) \Longrightarrow \sigma(\cdot)(1-\sigma(\cdot)) \geq 0 \Longrightarrow$ the λs of S are non-negative

$\Downarrow$

H_θ is positive semi-definite

2) $\int_{\mathbb{R}} \frac{1}{Z} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = 1 \Longrightarrow \int_{\mathbb{R}} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = Z$

$Z^2 = \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) dx\,dy$

$= \int_0^\infty \int_0^{2\pi} \exp\left(-\frac{r^2}{2\sigma^2}\right) r\, d\theta\, dr$

$= 2\pi \int_0^\infty \exp\left(-\frac{r^2}{2\sigma^2}\right) r\, dr$

$= 2\pi(-\sigma^2) \int_0^\infty \exp\left(-\frac{r^2}{2\sigma^2}\right)\left(-\frac{r}{\sigma^2}\right) dr$

$= -2\pi\sigma^2 \exp\left(-\frac{r^2}{2\sigma^2}\right)\Big|_0^\infty = -2\pi\sigma^2(0-1) = 2\pi\sigma^2$

$\Longrightarrow Z = \sigma\sqrt{2\pi}$

3) a)

$$\underset{w}{\arg\max} \sum_{i=1}^{N} \log N(y_i | w_0 + w^T \vec{x_i}, \sigma^2) + \sum_{j=1}^{D} \log N(w_j | 0, \tau^2)$$

$$= \underset{w}{\arg\max} \sum_{i=1}^{N} \log \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - w_0 - w^T\vec{x_i})^2}{2\sigma^2}\right) + \sum_{j=1}^{D} \log \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{w_j^2}{2\tau^2}\right)$$

$$= \underset{w}{\arg\max} \sum_{i=1}^{N} \left(-\frac{(y_i - w_0 - w^T\vec{x_i})^2}{2\sigma^2} - \log\sqrt{2\pi}\sigma\right) + \sum_{j=1}^{D} \left(-\frac{w_j^2}{2\tau^2} - \log\sigma\sqrt{2\pi}\right)$$

$$= \underset{w}{\arg\max} -\left((N+D)\log\sigma\sqrt{2\pi} + \sum_{i=1}^{N} (y_i - w_0 - w^T\vec{x_i})^2 (2\sigma^2)^{-1} + \sum_{j=1}^{D} \frac{w_j^2}{2\tau^2}\right)$$

then we have

$$= \underset{w}{\arg\min} \sum_{i=1}^{N} (y_i - w_0 - w^T\vec{x_i})^2 + \frac{\sigma^2}{\tau^2} \sum_{j=1}^{D} w_j^2$$

$$= \underset{w}{\arg\min} \sum_{i=1}^{N} (y_i - w_0 - w^T\vec{x_i})^2 + \lambda \|w\|_2^2 \quad \text{where } \lambda = \frac{\sigma^2}{\tau^2}$$

b) let $f = \|Ax - b\|_2^2 + \|\Gamma x\|_2^2$

$$\nabla_x f = \nabla_x\left[(Ax-b)^T(Ax-b) + (\Gamma x)^T(\Gamma x)\right]$$

$$0 = \nabla_x\left[(x^TA^T - b^T)(Ax-b) + x^T\Gamma^T\Gamma x\right]$$

$$0 = \nabla_x\left[x^TA^TAx - x^TA^Tb - b^TAx + b^Tb + x^T\Gamma^T\Gamma x\right.$$

$$0 = 2A^TAx - 2A^Tb + 2\Gamma^T\Gamma x$$

$$A^Tb = A^TAx + \Gamma^T\Gamma x$$

$$A^Tb = (A^TA + \Gamma^T\Gamma)x$$

$$x^* = (A^TA + \Gamma^T\Gamma)^{-1}A^Tb$$

let $\Gamma = \sqrt{\lambda} I$, then $x^* = (A^TA + \lambda I)^{-1}A^Tb$

c) $x^* = 8.8150$ validation set RMSE $= 0.6956$

test set RMSE $= 0.7445$

d) let $f = \|Ax + b\vec{1} - y\|_2^2 + \|\Gamma x\|_2^2$

$$= (Ax + b\vec{1} - y)^T(Ax + b\vec{1} - y) + (\Gamma x)^T(\Gamma x)$$

$$= x^TA^TAx + 2b\vec{1}^TAx - 2y^TAx - 2b\vec{1}^Ty + b^2n + y^Ty + x^T\Gamma^T\Gamma x$$

$$\nabla_x f = 2A^TAx + 2bA^T\vec{1} - 2A^Ty + 2\Gamma^T\Gamma x$$

$$\nabla_b f = 2\vec{1}^TAx - 2\vec{1}^Ty + 2bn = 0$$

set both gradients to 0: $\begin{cases} 2A^TAx + 2bA^T\vec{1} - 2A^Ty + 2\Gamma^T\Gamma x = 0 \\ 2\vec{1}^TAx - 2\vec{1}^Ty + 2bn = 0 \end{cases}$ $\Rightarrow b^* = \dfrac{\vec{1}^T(y-Ax)}{n}$

then $2A^TAx + 2\left(\dfrac{\vec{1}^T(y-Ax)}{n}\right)A^T\vec{1} - 2A^Ty + 2\Gamma^T\Gamma x = 0$

$$(A^TA + \Gamma^T\Gamma)x + \left(\frac{\vec{1}^T(y-Ax)}{n}\right)A^T\vec{1} - A^Ty = 0$$

$$(A^TA + \Gamma^T\Gamma)x + \frac{1}{n}A^T\vec{1}\vec{1}^Ty - \frac{1}{n}A^T\vec{1}\vec{1}^TAx - A^Ty = 0$$

$$\left(A^T\left(I - \frac{1}{n}\vec{1}\vec{1}^T\right)A + \Gamma^T\Gamma\right) \cdot x = A^T\left(I - \frac{1}{n}\vec{1}\vec{1}^T\right)y$$

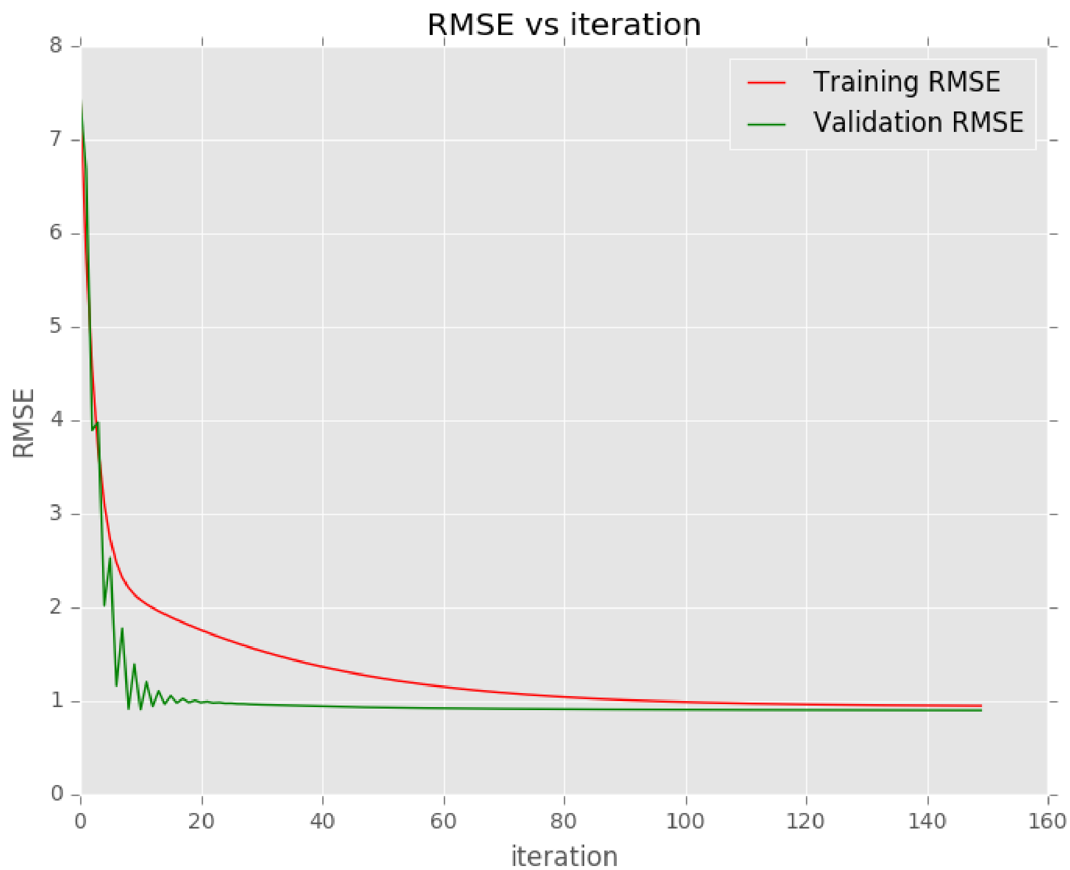
$$x^* = \left[A^T\left(I - \frac{1}{n}\vec{1}\vec{1}^T\right)A + \Gamma^T\Gamma\right]^{-1}A^T\left(I - \frac{1}{n}\vec{1}\vec{1}^T\right)y$$

diff in bias 1.9077E-10

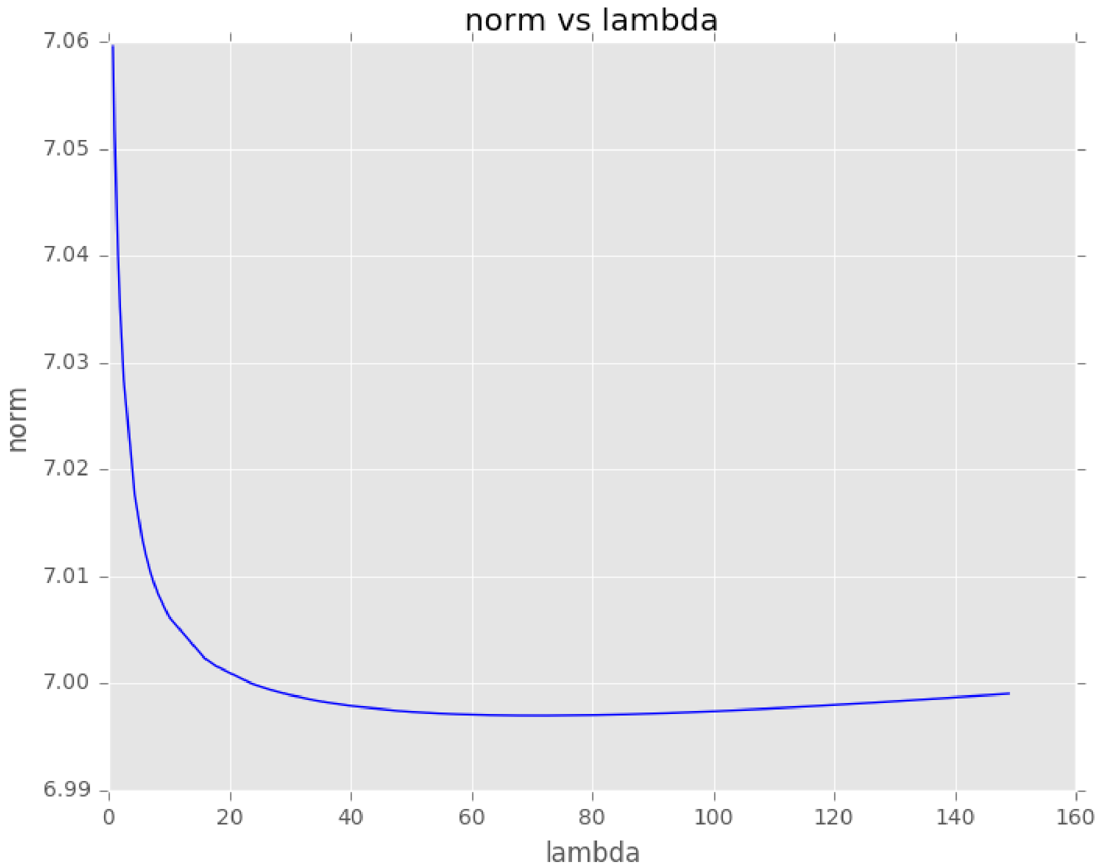
diff in weights: 2.4802E-10

e) diff in bias: 1.5386E-01

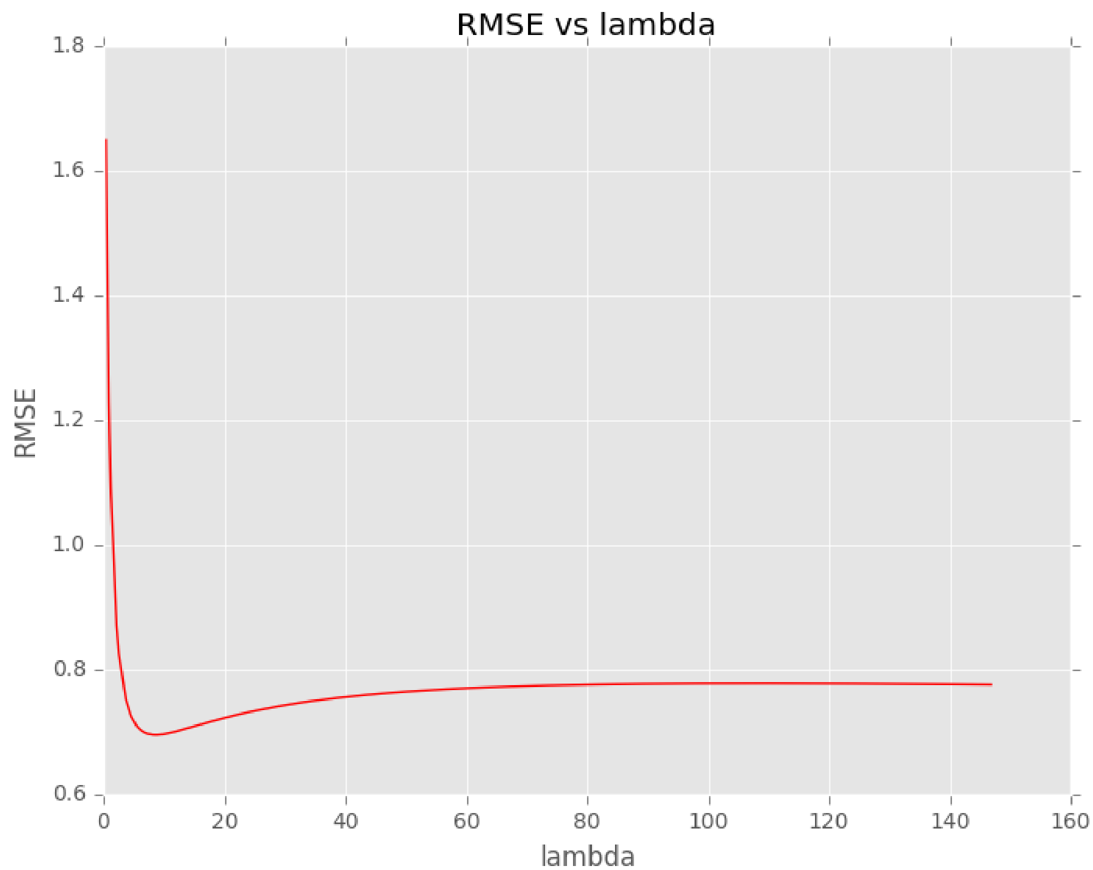
diff in weights: 7.9778E-01



RMSE vs iteration
Training RMSE
Validation RMSE
RMSE
iteration
0
1
2
3
4
5
6
7
8
0
20
40
60
80
100
120
140
160



norm vs lambda
7.06
7.05
7.04
7.03
7.02
7.01
7.00
6.99
norm
0
20
40
60
80
100
120
140
160
lambda



RMSE vs lambda
RMSE
lambda
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0
20
40
60
80
100
120
140
160