HW 2

$$1 + e^{-x} \overline{\big)\, e^{-x}}$$

1\. a) $\sigma(x) = (1 + e^{-x})^{-1} \Rightarrow \sigma'(x) = -(1 + e^{-x})^{-2} \cdot -(e^{-x})$

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

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

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

b) $NLL(\vec{w}) = -\sum_{i=1}^{N} \left[y_i \overset{\ln}{\cancel{\log}} \mu_i + (1 - y_i) \overset{\ln}{\cancel{\log}} (1 - \mu_i) \right]$

$\mu_i = w^T x_i$

$$= \sum_{i=1}^{N} y_i \overset{\ln}{\cancel{\log}} (\sigma(w^T x_i)) + (1 - y_i) \overset{\ln}{\cancel{\log}} (1 - \sigma(w^T x_i))$$

$\vec{x}_i$ represents the ith x point ~~on~~ each pt in the vector is a field

We'll optimizing the parameters given by $\vec{w}$ the gradient is steepest

$$\nabla_{\vec{w}^T} NLL(\vec{w}) = \sum_{i=1}^{N} \left[y_i \frac{\sigma'(w^T x_i)\, x_i}{\sigma(w^T x_i)} + (1 - y_i) \frac{-\sigma'(w^T x_i)\, x_i}{1 - \sigma(w^T x_i)} \right]$$

$$= \sum_{i=1}^{N} \left[y_i \frac{\cancel{\sigma(w^T x_i)}\,(1 - \sigma(w^T x_i))\, x_i}{\cancel{\sigma(w^T x_i)}} + -(1 - y_i) \frac{(\sigma(w^T x_i)\, \cancel{(1 - \sigma(w^T x_i))})\, x_i}{\cancel{1 - \sigma(w^T x_i)}} \right]$$

$$= \sum_{i=1}^{N} y_i (1 - \sigma(w^T x_i)) x_i + (y_i - 1)(\sigma(w^T x_i)) x_i$$

substitute μ_i

$$= \sum_{i=1}^{N} (y_i - y_i \mu_i + y_i \mu_i - \mu_i) x_i = \sum_{i=1}^{N} (y_i - \mu_i)\, x_i = X^T (\mu - y)$$

c) $H = \frac{d}{d\vec{w}} g(\vec{w})^T = \frac{d}{d\vec{w}} (X^T(\mu - y))^T = \frac{d}{d\vec{w}} ((\mu - y)^T X) = \frac{d}{d\vec{w}} (\mu^T X - y^T X) = \frac{d}{d\vec{w}} \mu^T X$

$$= X \frac{d}{d\vec{w}} \sum \sigma(w^T x_i) = X \sum \sigma(w^T x_i)(1 - \sigma(w^T x_i))\, x_i = X \mu (1 - \mu) X^T$$

Problem 2

$$P(x;\sigma^2) = \frac{1}{Z}\exp\left(\frac{-x^2}{2\sigma^2}\right)$$

normal distribution: $1 = \int \frac{1}{Z}\exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \frac{1}{Z}\int e^{\frac{-x^2}{2\sigma^2}} = \frac{\sqrt{2}\sigma}{Z}\int e^{-a^2}\, da$

let $a = \frac{x}{\sqrt{2}\sigma}$ $\quad da = \frac{1}{\sqrt{2}\sigma} dx \quad = \sqrt{2}\sigma\, da = dx$

let $I = \int e^{-a^2} da$

$$I^2 = \int e^{-a^2} da \int e^{-b^2} db = \iint e^{-(a^2+b^2)}\, da\, db = \int_0^{2\pi}\int_0^{\infty} r e^{-r^2}\, dr\, d\theta$$

$$= \frac{1}{2}\int_0^{2\pi} d\theta = \pi$$

$I = \sqrt{\pi}$

$$1 = \frac{\sqrt{2}\sigma\sqrt{\pi}}{Z} \Rightarrow Z = \sqrt{2\pi}\,\sigma$$

Problem 2

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

$$N(x, \mu \mid \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

thus it follows:

$$\arg\max_w \sum_{i=1}^{N} \log\left(\cancel{\exp}\left(\frac{(y_i - (w_0 + w^T x_i))^2}{2\sigma^2}\right)\cdot\frac{1}{\sqrt{2\pi}\,\sigma}\right) + \sum_{i=1}^{D}\log\left(\cancel{\exp}\left(\frac{(w_j - 0)^2}{2\tau^2}\right)\cdot\frac{1}{\sqrt{2\pi}\,\sigma}\right)$$

separate logs

$$\arg\max_w \sum_{i=1}^{N}\left[\left(\frac{(y_i - (w_0 + w^T x_i))^2}{2\sigma^2}\right) + \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)\right] + \sum_{i=1}^{D}\left[\frac{w_j^2}{2\tau^2} + \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)\right]$$

$$\arg_w\max -\frac{1}{2}(N+D)\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - (w_0 + w^T x_i))^2 + \frac{1}{2\tau^2}\sum_{i=1}^{D} w_j^2$$ * unsure how to handle term → check answer key

$$\arg\max_w = -\frac{1}{2}(N+D)\log(2\pi\sigma^2) \sum_{i=1}^{N}(y_i - w_0 + w^T x_i)^2 + \lambda \|w\|_2^2$$

ignore coefficient $= \text{argmin} = \sum_{i=1}^{N}(y_i - w_0 - w^T x_i)^2 + \lambda\|w\|_2^2$

Problem

b) minimize: $\|A\vec{x} - \vec{b}\|_2^2 + \|\Gamma\vec{x}\|_2^2$ by finding the gradient

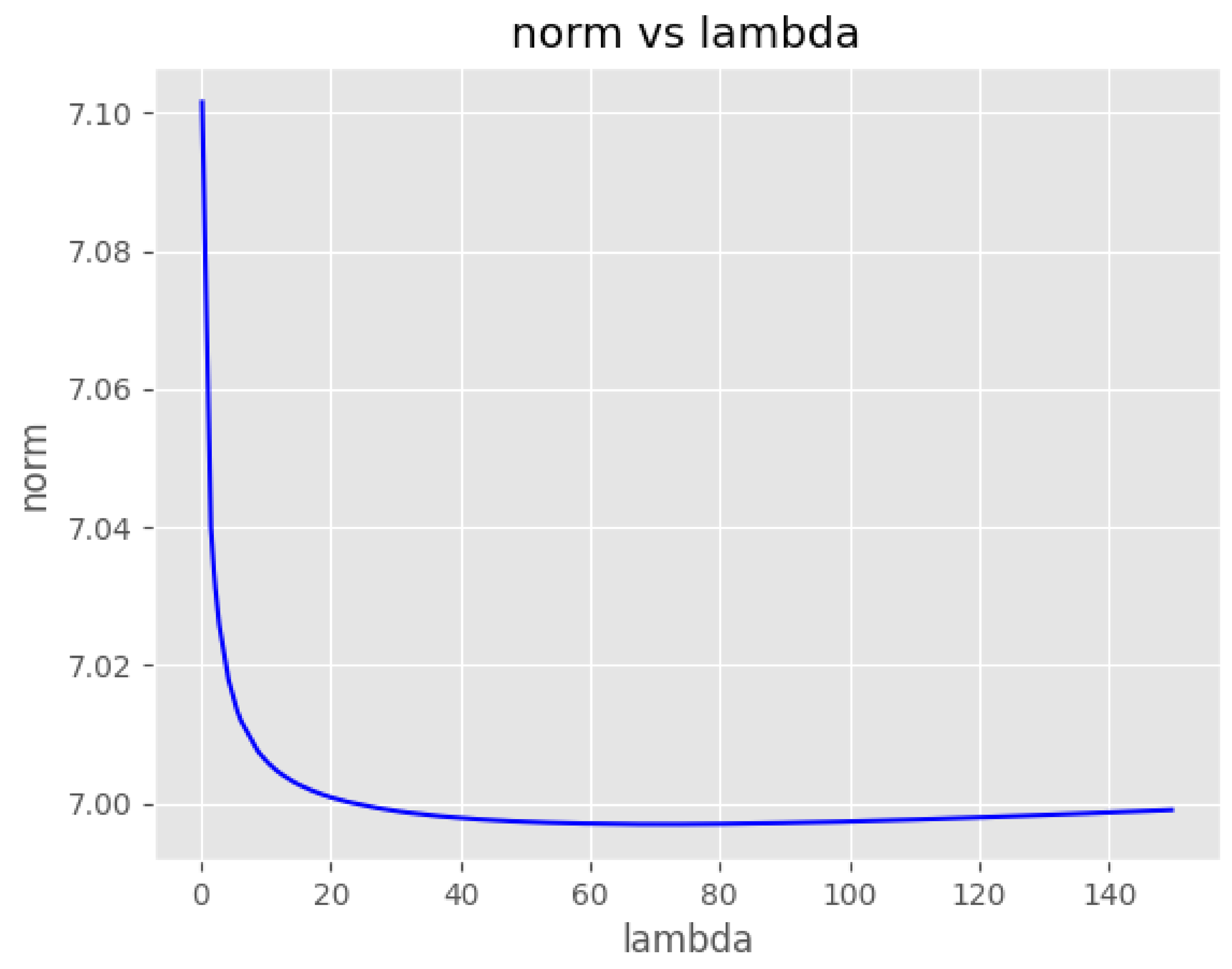
$$\nabla_x f = \nabla_x\left[(A\vec{x} - \vec{b})^T(A\vec{x} - \vec{b}) + (\Gamma\vec{x})^T(\Gamma x)\right]$$

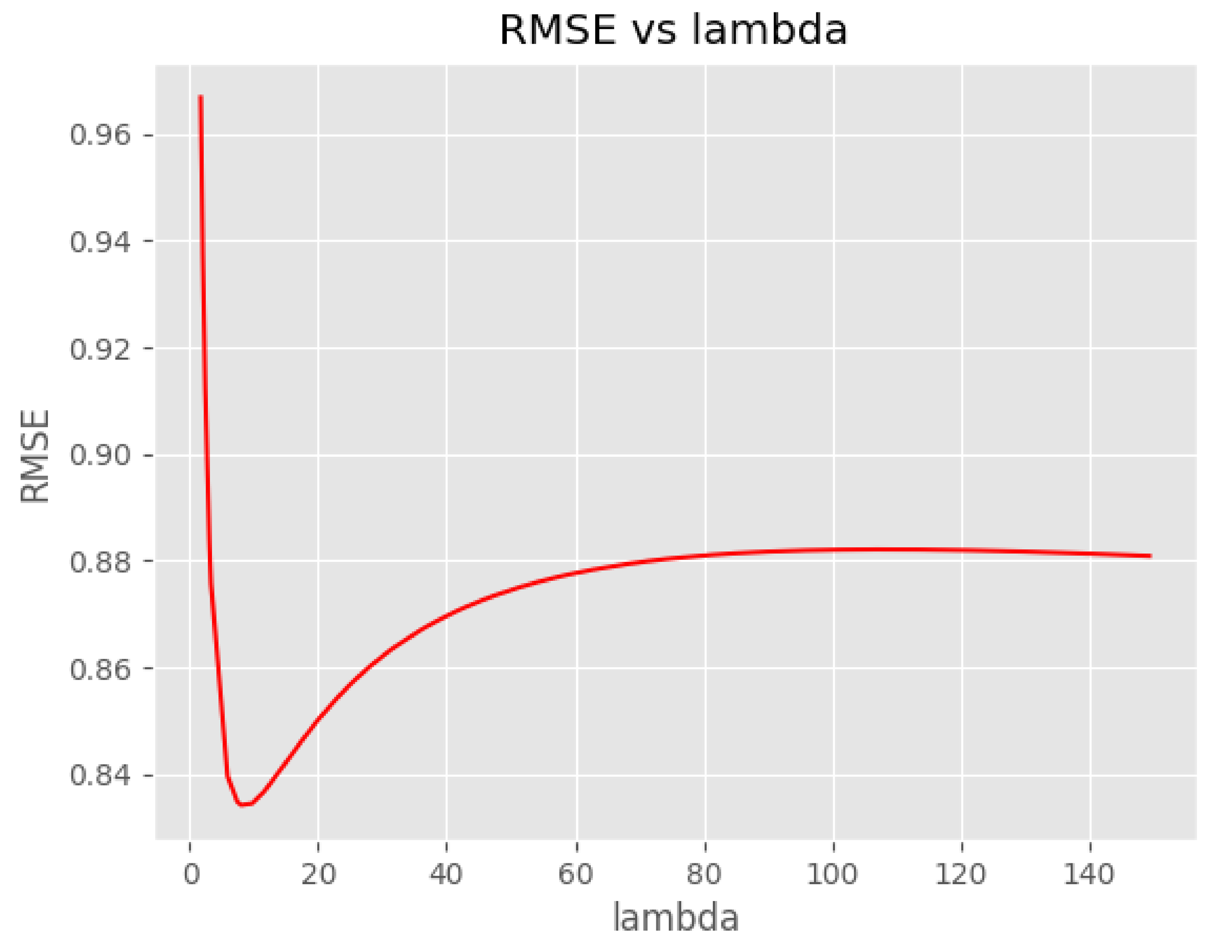
$$= \nabla_x\left[x^T A^T A\vec{x} - 2x^T A^T \vec{b} \quad b^T b \quad x^T \Gamma^T \Gamma x\right.$$

$$= 2A^T A x - 2A^T b + 2\Gamma^T\Gamma x$$

gradient $= 0$ → critical point

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





optimal regularization parameter = 0.834
RMSE valdation set = 1.1281
RSME test set = 0.8627

d) minimize : $\| Ax + b1 - y \|_2^2 + \| \Gamma x \|_2^2$

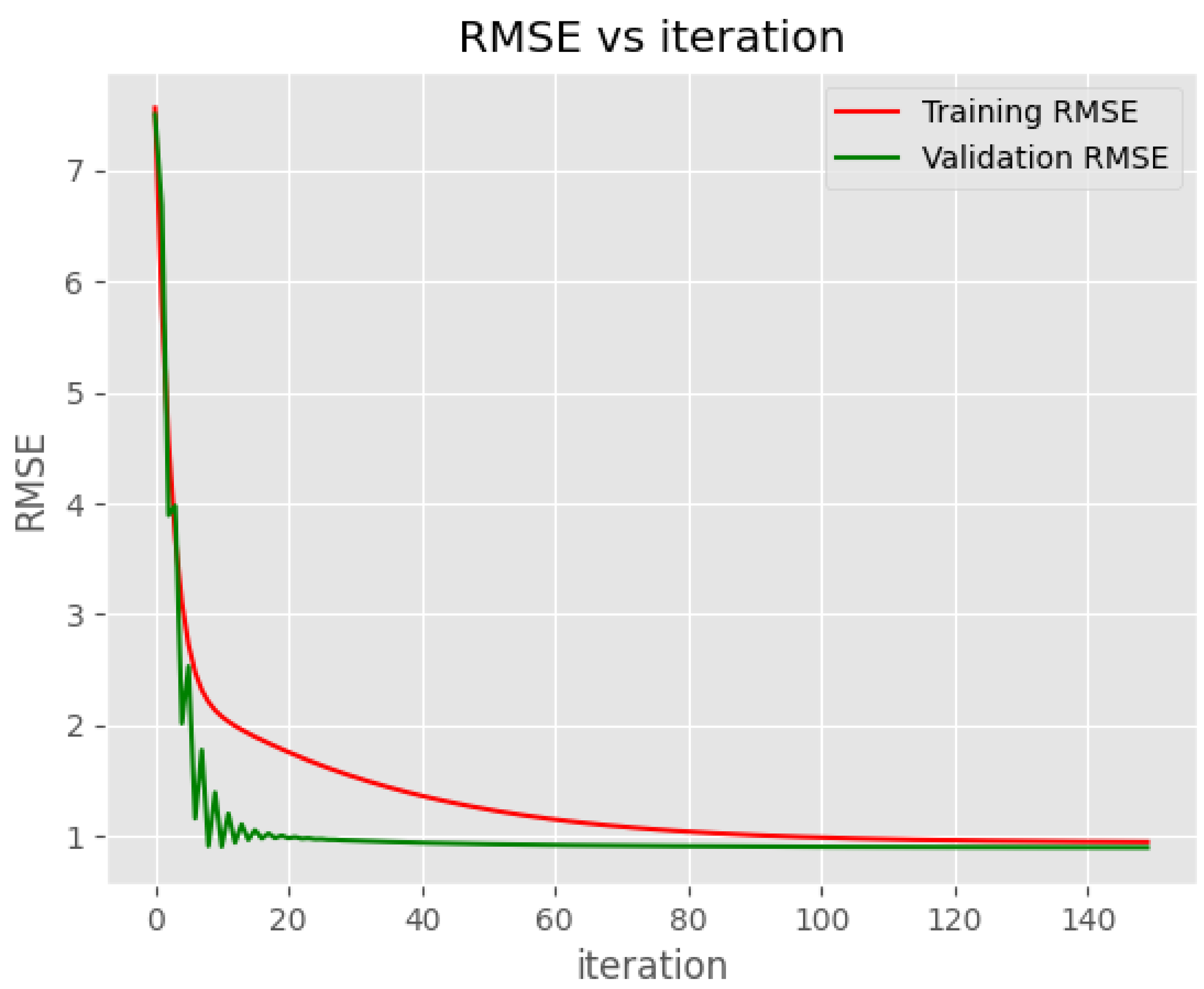
$$= (Ax + b1 - y)^T (Ax + b1 - y) + (\Gamma x)^T (\Gamma x)$$

$$= x^T A^T A x + 2b1^T - 2y^T A x - 2b1^T y + b^2 n + y^T y + x^T \Gamma^T \Gamma x$$

$$\nabla_b f = 2 1^T A x - 2 1^T y + 2bn = 0$$

$$b = \frac{1^T (y - Ax)}{n}$$

$$x^* = (A^T A + \Gamma^T \Gamma) x + \frac{1^T (y - Ax)}{n} A 1 - A^T y = 0$$



diff in bias = 0.153
diff in weights 1.148