Ahmed Bilal

HW2 Math 189r

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

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

$= \underbrace{\frac{1}{(1+e^{-x})}}_{\sigma(x)} \frac{e^{-x}}{(1+e^{-x})}$

$= \sigma(x) \frac{e^{-x}}{(1+e^{-x})}$ (checked solution here)

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

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

b) $nll(\theta) = -\sum_i y_i \log \sigma(\theta^T x_i) + (1-y_i)\log(1-\sigma(\theta^T x_i))$ ★ saw this from sol.

$\nabla_\theta nll(\theta) = -\sum (y_i) \frac{1}{\sigma(\theta^T x_i)} \underbrace{\sigma'(\theta^T x_i)}_{\text{we partial}} + (1-y_i) \frac{1}{(1-\sigma(\theta^T x_i))} \underbrace{(0-\sigma'(\theta^T x_i))}_{\text{we partial}}$

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

$= -\sum (y_i)(1-\sigma(\theta^T x_i))(x_i) - (1-y_i)(\sigma(\theta^T x_i))x_i$

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

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

$\mu_i = \sigma(\theta^T x_i)$

$= \sum_i (\mu_i - y_i)(x_i) = X^T(\mu_i - Y)$ ★ saw that step from sols.

c) $H = X^T S X$, where

$$\underbrace{H_\theta = \nabla_\theta(\nabla_\theta \, nll(\theta))^T}_{\text{* from solution}}$$

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

* saw from solution $= \nabla_\theta \, \mu^T X$

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

since $S = \operatorname{diag}(\mu_i(1-\mu_i), \ldots \mu_n(1-\mu_n))$

$$= X^T S X$$

---

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

since $\int p(x;\sigma^2)\,dx = \frac{1}{Z}\int \exp\left(-\frac{x^2}{2\sigma^2}\right)dx = 1$

$$Z = \int \exp\left(-\frac{x^2}{2\sigma^2}\right)dx$$

* saw solution $Z^2 = \int_R \exp\left(-\frac{x^2}{2\sigma^2}\right)dx \int_R \exp\left(-\frac{y^2}{2\sigma^2}\right)dy$

$$\iint_R \exp\left(\frac{-\overbrace{(x^2+y^2)}^{r^2}}{2\sigma^2}\right) dxy \quad \text{(convert to polar coords)}$$

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

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

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

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

$$Z^2 = 2\pi\sigma^2$$
$$Z = \sqrt{2\pi}\,\sigma$$

---

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

props of Gaussian: $X \sim N(\mu, \sigma^2)$

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

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

* checked solution here

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

$$= \arg\max_w -\left(\underbrace{(N+D)\log\sqrt{2\pi}(\sigma)}_{\text{constant value, can ignore its effect on w for now}} + \sum_{i=1}^N \frac{(y_i - w_0 - w^T x_i)^2}{2\sigma^2} + \sum_{j=1}^D \frac{w_j^2}{2\tau^2}\right)$$

* same from soln (max function is equivalent to minimizing the neg)

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

set $\lambda = \sigma^2/\tau^2$

$$= \arg\min_w \sum_{i=1}^N (y_i - w_0 - w^T x_i)^2 + \lambda\sum_{j=1}^D w_j^2$$

$$= \arg\min_w \sum_{i=1}^N (y_i - w_0 - w^T x_i)^2 + \lambda \|w\|_2^2$$

h) find closed form solution x^* to ridge regression problem

* saw solt. $\rightarrow \nabla_x f = \nabla_x [(Ax-b)^T(Ax-b) + (\Gamma x)^T(\Gamma x)]$

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

$$= \nabla_x [x^TA^TAx - b^TAx - x^TA^Tb + b^Tb + x^T\Gamma^T\Gamma x]$$

$$= \nabla_x [x^TA^TAx - 2x^TA^Tb + \underbrace{b^Tb}_{\text{constant, with no } x \text{ term}} + x^T\Gamma^T\Gamma x]$$

* saw soltt $\rightarrow$ $= 2A^TAx - 2A^Tb + 2\Gamma^T\Gamma x$

$$\nabla_x f = 0$$

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

$$\cancel{2}A^TAx + \cancel{2}\Gamma^T\Gamma x = \cancel{2}A^Tb$$

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

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

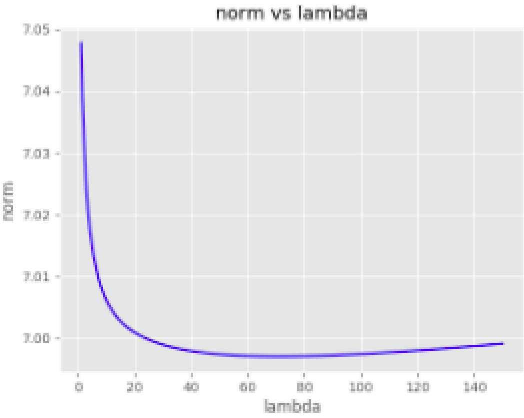
$$x^* = (A^TA + \Gamma^T\Gamma)^{-1}A^Tb$$

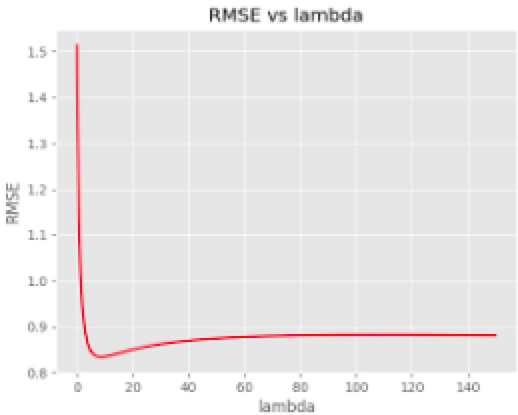
* saw sol'n $\longrightarrow$ if we let $\Gamma = \sqrt{\lambda} I$, objective then becomes:

$$\min f = \|Ax - b\|_2^2 + \lambda x^T x$$

c)

==> The optimal regularization parameter is 8.5974.
==> The RMSE on the validation set with the optimal regularization parameter is 0.8340.
==> The RMSE on the test set with the optimal regularization parameter is 0.8628.





e) was unable to get convergence plot

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

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

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

$$= x^T A^T A x + 2b\mathbb{1}^T Ax - 2y^T Ax - 2b\mathbb{1}^T y + b^2 n + y^T y + x^T \Gamma^T \Gamma x$$

$$\nabla_x f = 2A^T A x + 2bA^T\mathbb{1} - 2A^T y + 2\Gamma^T\Gamma x$$

set $\nabla_x f = 0$

* Saw from solution: $\nabla_b f = 2\mathbb{1}^T Ax - 2\mathbb{1}^T y + 2bn = 0$

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

(from x^* equation from b)

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

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

$$[A^TA + \Gamma^T\Gamma - \tfrac{1}{n}A^T\mathbb{1}\mathbb{1}^TA]x = A^Ty - \tfrac{1}{n}A^T\mathbb{1}\mathbb{1}^Ty$$

* Saw soln $[A^T(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T)A + \Gamma^T\Gamma]x = A^T(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T)y$

$$x^* = [A^T(I - \tfrac{1}{n}\mathbb{1}\mathbb{1}^T)A + \Gamma^T\Gamma]^{-1}A^T(I - \tfrac{1}{n}\mathbb{1}\mathbb{1}^T)y$$

(from sample code, difference in bias: 4.2E-10

difference in weights: 5.5E-10)

(similar to results from c)