1. (b). $\sigma'(x) = \sigma(x)\cdot(1-\sigma(x))$

2. Ridge Reg:

$$J(\theta) = \|y - X\theta\|^2 + \underline{\lambda\|\theta\|^2}$$

$$\nabla_\theta J(\theta) = \nabla_\theta\left((y-X\theta)^T(y-X\theta) + \lambda\theta^T\theta\right)$$

$$= 2(X^TX + \alpha I)\theta - 2X^Ty$$

$$\theta^* = (X^TX + \alpha I)^{-1}X^Ty$$

$$\theta^*_y = (X^TX\quad)^{-1}X^Ty$$

$$x = (f_1,\ f_2,\ f_1+f_2).$$

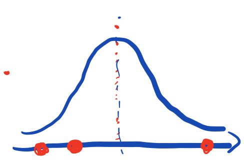
X^TX may not inversiable.

3. MAP. MLE

$$\longrightarrow \theta_{MLE} = \arg\max_\theta P(x_1, x_2, \cdots x_m \mid \theta).$$

$$= \arg\max_\theta \prod_i^m p(x_i \mid \theta).$$

$$= \arg\max_\theta \underbrace{\prod_i^m \left[\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left\{\frac{-(x_i-\theta)^2}{2\sigma^2}\right\}\right]}.$$



θ

$$\log L = \sum_i^m\left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i-\theta)^2\right)$$

$$\nabla_\theta L = 0 \implies \theta_{MLE} = \frac{1}{m}\cdot\sum_i^m x_i$$

$$P(\theta) \sim N(\nu, \mu^2)$$

$$\longrightarrow \theta_{MAP} = \arg\max_\theta P(\theta \mid x_1, x_2 \cdots x_m).$$

$$= \arg\max_\theta \frac{\underline{P(x_1, x_2 \cdots x_m \mid \theta)\, P(\theta)}}{\cancel{P(x_1, x_2 \cdots x_m)}}$$

$$= \arg\max_\theta \underline{\prod_i^m p(x_i \mid \theta)\cdot P(\theta)}.$$

$$\nabla_\theta \log Q' = 0 \implies \theta_{MAP} = \sigma^2\nu + \mu^2\cdot\sum_{i=1}^m x^i \longrightarrow \frac{\sum_i^m x^i}{\ldots}$$

$$\sigma^2 + \boxed{m}\mu^2.$$
$\frac{\quad}{m}$ θ_{MLE}

$m \to \infty$

4. Softmax Reg.

$$l(\theta) = \sum_i^m \sum_l^k \cdot 1\{y^i = l\} \cdot \log \frac{e^{\theta_l^T x^i}}{\sum_j^k e^{\theta_j^T x^i}}$$

$$1\{y^i = l\} = \begin{cases} 1, & y^i = l \\ 0. & \text{o.w.} \end{cases}$$

$$l(\theta) = \underbrace{\log \frac{e^{\theta_{y_1}^T \cdot x^1}}{\Sigma}}_{1^{st}} + \underbrace{\log \frac{e^{\theta_{y_2}^T \cdot x^2}}{\Sigma}}_{2^{nd}} + \cdots + \underbrace{\log \frac{e^{\theta_{y_m}^T \cdot x^m}}{\Sigma}}$$

$2^{nd} \swarrow y_2 = 1$

$$= \log \cdot (e^{\theta_{y_1}^T \cdot x^1}) + \log (e^{\theta_{y_2}^T \cdot x^2}) + \cdots + \quad m \cdot \log(\Sigma).$$

$$= \sum_i^m \sum_l^k \cdot 1\{y^i = l\} \cdot \theta_l^T \cdot x^i + m \cdot \log\left(\sum_j^k e^{\theta_j^T x^i}\right).$$

$$\nabla_{\theta_1} l(\theta) = \sum_i^m \nabla_{\theta_1}\left(\sum_l^k 1\{y^i = l\} \theta_l^T x^i\right).$$

$$\log(f(x)) = \frac{f'(x)}{f(x)}$$

$$= \sum_i^m \cdot (1\{y^i = 1\} \cdot x^i) + m \cdot \frac{e^{\theta_1^T x^i} \cdot x^i}{\Sigma}.$$

5. $\min_x f(x) \triangleq \|y - Ax\|^2.$ $\quad A_{m \times n}$ $\quad m \ll n.$

$x_0 = 0$ implicit bias.

$$\nabla_x f(x) = A^T(Ax - y)$$

$$x_{k+1} = x_k - 2\alpha \cdot A^T(Ax - y). = (I - 2\alpha A^T A) \cdot x_k + 2\alpha A^T y$$

$x_0 = 0 \qquad x_1 = 2\alpha A^T y$

$$x_2 = (I + I - 2\alpha A^TA)\cdot 2\alpha A^Ty$$

$$x_3 = (I + (I-2\alpha A^TA) + (I-2\alpha A^TA)^2)\cdot 2\alpha A^Ty.$$

$$\vdots$$

$$x_{k+1} = 2\alpha\left(\sum_{t=0}^{k}(I-2\alpha A^TA)^t\right)\cdot A^Ty$$

$$\sum_{t}^{\infty}(1-q)^t = \frac{1}{q}$$ A^TA may not inversiable.

$$A = U\Sigma V^T = U[\Sigma_1\ 0]\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U\Sigma_1\cdot V_1^T.$$

$$2\alpha A^TA = V\Sigma^TU^T\cdot U\Sigma V^T = V\Sigma^T\Sigma V^T\cdot 2\alpha.$$

$$I - 2\alpha A^TA = VV^T - V\Sigma^T\Sigma V^T 2\alpha = V[I-2\alpha\Sigma^T\Sigma]V^T$$

$$(I-2\alpha A^TA)^2 = V[I-2\alpha(\Sigma^T\Sigma)]V^T\cdot V[\cdots]V^T$$

$$= V[I-2\alpha(\Sigma^T\Sigma)]^2V^T$$

So $$(I-2\alpha A^TA)^t = V[I-2\alpha(\Sigma^T\Sigma)]^t\cdot V^T$$

$$x_\infty = 2\alpha\cdot V\left[\sum_{t=0}^{\infty}(I-2\alpha\Sigma^T\Sigma)^t\right]\cdot\cancel{V^TV}\Sigma^TU^T\cdot y$$

$\Sigma_1 = \Sigma_1^T$ $$\Sigma^T\Sigma = \begin{bmatrix}\Sigma_1 \\ 0\end{bmatrix}[\Sigma_1\ 0] = \Sigma_1^2.$$

$$x_\infty = 2\alpha\cdot V(2\alpha\Sigma_1^2)^{-1}\cdot\Sigma^T\cdot U^Ty.$$

$$= \left[V_1 \ V_2 \right] \cdot \Sigma_1^{[illegible]} \cdot \begin{bmatrix} [illegible] \\ 0 \end{bmatrix} \cdot U^{[illegible]} y$$

$$= \underline{\underline{V_1 \Sigma_1^{-1} U^T y}}. \quad = A^T (AA^T)^{-1} y.$$