Aaron Vo Homework 2 2/24/24

2) Prove closed-form solution for Ridge Regression
$w = (\lambda I + X^T \cdot X)^{-1} \cdot X^T \cdot y$

↳ def $E(w) = \sum_{i=1}^{m} (w^T \cdot x^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^{m} w_j^2$

↳ Differentiate $\nabla E(w)$

↳ $\nabla E(w) = 2 \sum_i (w^T \cdot x^i - y^i) x^i + 2\lambda w = 0$

$= 2 \sum_i (w^T x^i x^i - y^i x^i) + 2\lambda w$

$= \dfrac{2X^T X w - 2X^T y + 2\lambda w}{2} = \dfrac{0}{2}$

↳ $w X^T X + \lambda w = X^T y$

$\dfrac{w(X^T X + \lambda I)}{X^T X + \lambda I} = \dfrac{X^T y}{X^T X + \lambda}$

↳ $w = \dfrac{X^T y}{X^T X + \lambda I}$ $\blacksquare$

3) The posterior probability is $\hat{p}_k = \delta(s_k(x))_k = \dfrac{\exp(s_k(x))}{\sum_{j=1}^{k} \exp(s_j(x))}$

1. For the softmax regression model, we need to estimate a total of $(n+1) \cdot k$ parameters.

↳ Total parameters $= k \cdot (n+1)$

↳ k is the total # of classes

↳ n is the dimensionality of x

2. Derive $J(\theta)$ with regards to θ_k

$$J(\theta) = \frac{-1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \log(\hat{p}_k^{(i)})$$

↳ $\hat{p}_k^{(i)} = \frac{\exp(s_k(x^{(i)}))}{\sum_{j=1}^{K}}$ ex

↳ $d J(\theta) / d\theta_k = \frac{-1}{m} \sum_{i=1}^{m} (y_k^{i} - p_k^{i})$ $\blacksquare$