CS 391L Machine Learning Assignment 3

Name: Shun Zhang

Email address: `jensen.zhang@utexas.edu`

EID: `sz4554`

Problem 1

(a)

$$Cov(y) = E((y - E(y)(y - E(y)^T) \tag{1}$$
$$= E((Ax + b - E(Ax + b))(Ax + b - E(Ax + b)^T)) \tag{2}$$
$$= E((Ax - E(Ax))(Ax - E(Ax))^T) \qquad \text{Linearity of } E \tag{3}$$
$$= AE((x - E(x))(x - E(x))^T)A^T \tag{4}$$
$$= A\Sigma A^T \qquad \text{Def. of Cov.} \tag{5}$$

(b) Base case: by the defition of eigenvalue and eigenvector, $Ax = \lambda x$.

Inductive hypothesis: assume $A^k x = \lambda^k x$ for some $k \in N$. Want to show $A^{k+1}x = \lambda^{k+1}x$.

$$A^k x = \lambda^k x \qquad \text{I.H.} \tag{6}$$
$$A^{k+1}x = A\lambda^k x \tag{7}$$
$$A^{k+1}x = \lambda^k Ax \tag{8}$$
$$A^{k+1}x = \lambda^{k+1}x \qquad Ax = \lambda x \tag{9}$$

Problem 2

(a) $\tanh x = \dfrac{\sinh x}{\cosh x} = \dfrac{e^x - e^{-x}}{e^x + e^{-x}} = \dfrac{1 - e^{-2x}}{1 + e^{-2x}}$. Compared to $\dfrac{1}{1 + e^{-x}}$, tanh has a steeper slope.

To be an appropriate sigmoid function, tanh needs to be scaled to range of $[0, 1]$.

(b) $1 - \tanh^2$.

(c) We know W^k is 2×2.

$$\frac{\partial H}{\partial w_{ij}^k} = \frac{\partial}{\partial w_{ij}^k} \sum_{k=0}^{K-1} (\lambda^{k+1})^T g(W^k x^k) \tag{10}$$

$$= \frac{\partial}{\partial w_{ij}^k} (\lambda_1^{k+1} g(w_{11} x_1^k + w_{12} x_2^k) + \lambda_2^{k+1} g(w_{21} x_1^k + w_{22} x_2^k)) \tag{11}$$

So, generally,

$$\frac{\partial H}{\partial w_{ij}^k} = \lambda^{k+1} \frac{\partial}{\partial w_{ij}^k} g(W^k x^k) \tag{12}$$

$$= \lambda^{k+1} \frac{\partial}{\partial w_{ij}^k} g(\sum_l w_{il} x_l^k) \tag{13}$$

$$= \lambda^{k+1} x_j^k g'(w_i^k x^k) \tag{14}$$

Problem 3