Khai Nguyen - 915552057

1) a) $\frac{\partial a^T x}{\partial x} = \frac{\partial x^T a}{\partial x} = a$

b) $\frac{\partial (\|x\|_2^2)}{\partial x} = \frac{\partial (x'x)}{\partial x} = 2x$

c) $\frac{\partial x^T A x}{\partial x} = (A + A^T)x = 2Ax$ (Since "A" is symmetric)

d) $\frac{\partial \exp(-w^T x)}{\partial x} = \exp(-w^T x) \cdot \frac{\partial(-w^T x)}{\partial x} = -w \exp(-w^T x)$

2) Since we need a linear regression model, we can represent it as

$$\underbrace{\begin{bmatrix} 1 & x_{(1)} \\ 1 & x_{(2)} \\ \vdots & \\ 1 & x_{(n)} \end{bmatrix}}_{X\ (d+1)\ \text{dimension}} \cdot \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}}_{W\ (d+1)\ \text{dimension}} = \underbrace{\begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(n)} \end{bmatrix}}_{Y}$$

$y_{(i)} = w_0 + x_{(i)_1} w_1 + x_{(i)_2} w_2 + \dots + x_{(i)_d} w_d \quad i \in [1; n]$

or $y_{(i)} = f(x) = w' X$

For linear regression we use the Quadratic loss function

$L = \|XW - Y\|_2^2 = 0$

Or, we can find W column vector from

$XW = Y$

$\Rightarrow X'XW = X'Y$

$\Rightarrow W = (X'X)^{-1} X'Y$

3) a) Exponential loss

1) Pick a model

$$f(x) = \sum_{i=1}^{d} w_i x_i + w_0$$

2) Objective function

$$Obj = Loss(w) + Regularizer(w,b)$$

$$Obj = \sum_{i=1}^{d} \exp(-y_i(w x_i + w_0)) + \frac{\lambda}{2}\|w\|^2$$

3) Compute gradient descent

$$w_i = w_i - \eta \frac{d}{dw_i}(loss(w) + regularizer(w,b))$$

$$= w_i - \eta\left(\sum_{i=1}^{d} y_i x_i \exp(-y_i(w x_i + w_0)) + \eta \lambda w_i\right)$$

b) Hinge loss

1) Pick a model

$$f(x) = \sum_{i=1}^{d} w_i x_i + w_0$$

2) Objective function

$$Obj = \sum_{i=1}^{d} \max\left[0, 1 - (w_i x_i + w_0) y_i\right] + \frac{\lambda}{2}\|w\|^2$$

3) Comput gradient descent

$$w_i = \begin{cases} w_i - \eta\left((-x_i y_i) + \lambda w_i\right) & \text{if } (w_i x_i + w_0) y_i < 1 \\ w_i - \lambda w_i & \text{if } (w_i x_i + w_0) y_i \geq 1 \end{cases}$$

4. a) TP = 2

b) FP = 3

		Predict	
		Yes (pos)	No (neg)
Actual	Yes	TP = 2	FN = 1
	No	FP = 3	TN 94

c) Precision

$$P = \frac{TP}{TP + FP} = \frac{2}{2+3} = \frac{2}{5} = 40\%$$

d) Recall:

$$R = \frac{TP}{TP + FN} = \frac{2}{2+1} = \frac{2}{3} = 66.7\%$$

e) $$F1_A = 2\frac{P \cdot R}{P + R} = \frac{2 \cdot \left(\frac{2}{5} \cdot \frac{2}{3}\right)}{\frac{2}{5} + \frac{2}{3}} = \frac{1}{2}$$

$F1_B$

		Predict	
		Yes	No
Actual	Yes	TP = 1	FN = 2
	No	FP = 1	TN = 96

$$P_B = \frac{TP}{TP + FP} = \frac{1}{1+1} = \frac{1}{2}$$

$$R_B = \frac{TP}{TP + FN} = \frac{1}{1+2} = \frac{1}{3}$$

$$\Rightarrow F1_B = \frac{2 \cdot \left(\frac{1}{2} \cdot \frac{1}{3}\right)}{\frac{1}{2} + \frac{1}{3}} = \frac{2}{5}$$

$\Rightarrow$ A has better performance than B

5) PLA:

$$W = W + y_n x_n$$

For regular gradient descent:

We have:

$$W \leftarrow W - \eta \frac{\partial l(y,y')}{\partial w}$$

If we take loss function $= -yy'_i = -y_i(wx_i + w_0)$

$$\Rightarrow \frac{\partial l}{\partial w} = -y_i x_i$$

$$\Rightarrow w \leftarrow w - (\eta y_i x_i)$$

By picking $\eta = 1$, we have the update rule for PLA:

$$w \leftarrow w + y_n x_n$$

Therefore PLA is a gradient descent algorithm