COMP 9417 — Machine Learning
Tutorial - Classification

Question 2

(C)

Email	a?	b?	c?	Class
e_1	0	1	0	+
e_2	0	1	1	+
e_3	1	0	0	+
e_4	1	1	0	+
e_5	1	1	0	—
e_6	1	0	1	—
e_7	1	0	0	—
e_8	0	0	0	—

Then the bit vector is $(2,3,1)$ and $(3,1,1)$

adding 2 pseudo-documents, one containing each word and one cotaining none of them.

Then the parameter vector is $\hat{\theta}_{\oplus} = (\frac{3}{6}, \frac{4}{6}, \frac{2}{6}) = (0.5, 0.67, 0.33)$

$\hat{\theta}_{\ominus} = (\frac{4}{6}, \frac{2}{6}, \frac{2}{6}) = (0.67, 0.33, 0.33)$

For $e^* = abbdebb$

The bit vector is $X = (1, 1, 0)$

The likelihoods is $P(x|\oplus) = 0.5 \times 0.67 \times (1-0.33) = 0.222$

$P(x|\ominus) = 0.67 \times 0.33 \times (1-0.33) = 0.148$

Then the likelihood ratio is $\frac{P(x|\oplus)}{P(x|\ominus)} = \frac{0.222}{0.148} = \frac{3}{2} > 1$

Because $\frac{P(\oplus)}{P(\ominus)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$

Then $\frac{P(\oplus|x)}{P(\ominus|x)} = \frac{P(x|\oplus)}{P(x|\ominus)} \cdot \frac{P(\oplus)}{P(\ominus)} = \frac{3}{2} \times 1 = \frac{3}{2} > 1$

Therefore, ML classification of x is spam.

Question 3

(a)

Because the sigmoid function is close to $(0,1)$ piecewise and continuously differentiable.

Linear regression does not apply to classification problems because the regression line changes when a new sample is added.

(b)

Because $y|x \sim \text{Bernoulli}(p^*)$, $p^* = \sigma(x^T w^*)$, $\sigma(z) = \frac{1}{1-e^{z}}$

Then $L(w^*) = \prod_{i=1}^{n} \sigma(x^T w^*)^{x_i} [1-\sigma(x^T w^*)]^{1-x_i}$

$$\log L(w^*) = \sum_{i=1}^{n} x_i \log \sigma(x^T w^*) + (1-x_i)\log[1-\sigma(x^T w^*)]$$

$$= n\bar{x}\log[\sigma(x^T w^*)] + n(1-\bar{x})\log[1-\sigma(x^T w^*)]$$

$$= n\bar{x}\log\left[\frac{1}{1+e^{-x^T w^*}}\right] + n(1-\bar{x})\log\left[1-\frac{1}{1+e^{-x^T w^*}}\right]$$

$$= -n\bar{x}\log(1+e^{-x^T w^*}) - n(1-\bar{x})[x^T w + \log(1+e^{-x^T w^*})]$$

$$\frac{\partial \log L(w^*)}{\partial w^*} = -n\bar{x}\frac{x e^{-x^T w^*}}{1+e^{-x^T w^*}} - n(1-\bar{x})\left[x + \frac{x e^{-x^T w^*}}{1+e^{-x^T w^*}}\right]$$

Let $\frac{\partial \log L(w^*)}{\partial w^*} = 0$

then $\bar{x}\frac{x e^{-x^T w^*}}{1+e^{-x^T w^*}} = (\bar{x}-1)x + (\bar{x}-1)\frac{x e^{-x^T w^*}}{1+e^{-x^T w^*}}$

$$\frac{e^{-x^T w^*}}{1+e^{-x^T w^*}} = \bar{x}-1$$

$$\frac{1}{1+e^{-x^T w^*}} - 1 = 1-\bar{x}$$

$$e^{-x^T w^*} = \frac{1}{2-\bar{x}} - 1$$

$$\log e^{-x^T w^*} = \log\frac{\bar{x}-1}{2-\bar{x}}$$

$$x^T w^* = \log(2-\bar{x}) - \log(\bar{x}-1)$$

$$w^* = (x^T)^{-1}[\log(2-\bar{x}) - \log(\bar{x}-1)]$$