Dylan Tan
1004385

Question 1.1

a) T
b) F
c) T
d) F

Question 1.2

$h_\Theta(x)$ is the estimated probability that y = 1 on input x.

Therefore, $h_\theta(x) = 0.28$ would imply that $p(y = 1 \mid x; \theta) = 0.28$. This means that (3) applies and (4) would not apply.

Since $p(y = 1 \mid x; \theta) + p(y = 0 \mid x; \theta) = 1$,

$0.28 + p(y = 0 \mid x; \theta) = 1$
$p(y = 0 \mid x; \theta) = 1 - 0.28$
$\therefore p(y = 0 \mid x; \theta) = 0.72$.

Hence, this means that (2) applies and (1) does not apply.

In conclusion, (1) and (4) do not apply and (2) and (3) apply.

Question 1.3

Given $h_\Theta(x) = g(\theta_0 + x_1\theta_1 + x_2\theta_2)$ and $\theta = [3\ {-3}\ 0]^T$,

$\theta^T x = 3 - 3x_1$

We know that y=1 if $\theta^T x >= 0$ and y=0 if $\theta^T x < 0$,

$\therefore$ y = 1 if $3 - 3x_1 >= 0$ and y = 0 if $3 - 3x_1 < 0$

and thus, y=1 if $x_1 <= 1$ and y=0 if $x_1 > 1$

Question 1.4

Given $h_\Theta(x) = g(\theta_0 + x_1\theta_1 + x_2\theta_2 + x_1^2\theta_3 + x_2^2\theta_4)$ and $\theta = [-64\ 0\ 0\ 1\ 1]^T$,

$\theta^T x = -64 + x_1^2 + x_2^2$

We know that y=1 if $\theta^T x >= 0$ and y=0 if $\theta^T x < 0$,

$\therefore$ y = 1 if $-64 + x_1^2 + x_2^2 >= 0$ and y = 0 if $-64 + x_1^2 + x_2^2 < 0$

and thus, y=1 if $x_1^2 + x_2^2 >= 64$ and y=0 if $x_1^2 + x_2^2 < 64$ (A circle with radius of 8)

Question 1.5

If we compute the likelihood of data as in equation 1, because it is a multiplication of probabilities, where probability is a value between 0 and 1, the value of the likelihood will approach a very small value that eventually, if a computer is used to compute it, may run out of floating point precision to represent it, resulting in what we call an underflow. Hence, if the likelihood value becomes small enough, it will be represented as 0 instead of the actual value, that is close to 0 but NOT 0.

Therefore, by using log-likelihood instead, we may be able to avoid this problem, which is more computationally “convenient”.