Problem Set 2
Solutions to Written Problems

1. Written problem 1 (Logistic sigmoid function)
SOLUTION:

Answer: From logistic sigmoid function, we have:

$$1-\sigma(a) = 1 - \frac{1}{1+exp(-a)} = \frac{1+e^{-a}-1}{1+e^{-a}} = \frac{e^{-a}}{1+e^{-a}} = \frac{1}{e^{a}+1} = \sigma(-a)$$

The inverse of the logistic sigmoid is easily found as follows:

$$y = \sigma(a) = \frac{1}{1+e^{-a}} \Rightarrow \frac{1}{y} - 1 = e^{-a} \rightarrow\Rightarrow \ln\left\{\frac{1-y}{y}\right\} \Rightarrow \ln\left\{\frac{y}{1-y}\right\} = a = \sigma^{-1}(y)$$

2. Written problem 2 (Derivation of LDA)
SOLUTION:

Answer: If we expand and substitute the correct probability distributions into the above formula, we get

$\ln\frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)} = \ln p(x|C_1) + \ln p(C_1) - \ln p(x|C_2) - \ln p(C_2)$

$= -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - \ln\left((2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}\right) + \ln\pi_1 + \frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu^2) + \ln\left((2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}\right) - \ln\pi_2$

$= \frac{1}{2}\left(-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + (x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right) + \ln\pi_1 - \ln\pi_2$

Note that we can simplify

$-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) = -x^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}\mu_1$

and similarly

$(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) = x^T\Sigma^{-1}x - \mu_2^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_2 + \mu_2^T\Sigma^{-1}\mu_2$

so the quadratic terms $x^T\Sigma^{-1}x$ cancel, leaving only terms linear in x. After some grouping some terms (not strictly necessary to prove the function is linear), we arrive at

$\ln\frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)} = \ln\frac{\pi_1}{\pi_2} - \frac{1}{2}(\mu_2+\mu_1)^T\Sigma^{-1}(\mu_1-\mu_2) + x^T\Sigma^{-1}(\mu_1-\mu_2) = x^T\theta + \theta_0$

where $\theta = \Sigma^{-1}(\mu_1-\mu_2)$ and $\theta_0 = \ln\frac{\pi_1}{\pi_2} - \frac{1}{2}(\mu_2+\mu_1)^T\Sigma^{-1}(\mu_1-\mu_2)$.

3. Written problem 3 (LR classification with label noise)
SOLUTION:

(a):

Answer: Using the sum rule,
$p(t=1|x) = p(t=1, y=1|x) + p(t=1, y=0|x)$
$= p(t=1|y=1,x)p(y=1|x) + p(t=1|y=0,x)p(y=0|x) = \tau p(y=1|x) + (1-\tau)(1-p(y=1|x))$
Here we used the fact that $p(t=1|y=1,x)$ is the probability of the label being correct and $p(t=1|y=0,x)$ is the probability of the label being incorrect.

(b):

Answer: Substituting the expression for $p(t=1|x)$ from (a) gives the final cost

$$\begin{aligned}
-\ln p(D|\theta) = &-\sum_{i=1}^{m} t^{(i)} \ln\left[\tau p(y=1|x) + (1-\tau)(1-p(y=1|x))\right] \\
&+(1-t^{(i)}) \ln\left(1 - \left[\tau p(y=1|x) + (1-\tau)(1-p(y=1|x))\right]\right) \\
= &-\sum_{i=1}^{m} t^{(i)} \ln\left[\tau\sigma(\theta^T x^{(i)}) + (1-\tau)\left(1-\sigma(\theta^T x^{(i)})\right)\right]\Big) \\
&+(1-t^{(i)}) \ln\left(1 - \left[\tau\sigma(\theta^T x^{(i)}) + (1-\tau)(1-\sigma(\theta^T x^{(i)}))\right]\right)
\end{aligned}$$

This is okay, but we can also simplify further:

$$\begin{aligned}
-\ln p(D|\theta) = &-\sum_{i=1}^{m} t^{(i)} \ln\left[\sigma\left(\theta^T x^{(i)}\right)(2\tau-1) - \tau + 1\right] \\
&+(1-t^{(i)}) \ln(-\sigma\left(\theta^T x^{(i)}\right)(2\tau-1) + \tau)
\end{aligned}$$