3.

a) Bernoulli distribution

b) The value of x is classified as class 1 if:

$\hat{y} = \arg\max_k P(y=k|x) = \arg\max_k P(y=k)P(x|y=k)$

Since $P(y=0)=P(y=1)=\frac{1}{2}$, x is classified as class 1 if:

$P(x|y=1) > P(x|y=0) \Rightarrow \lambda_1 e^{-\lambda_1 x} > \lambda_0 e^{-\lambda_0 x} \Rightarrow x < \frac{\ln\lambda_1 - \ln\lambda_0}{\lambda_0 - \lambda_1}$

4.

As w becomes larger, the predicted probabilities for corretly classified points approach 1, resulting in the likelihood approaching its maximum value. This causes the magnitude $\|w\|$ to diverge to infinity. We can add L_2 regularization term, e.g. Ridge Regression, which penalize large value of w, thereby preferring a solution with a finite magnitude.

5.

The Softmax function is: $P(y=1|x) = \frac{e^{w^Tx}}{e^{w^Tx}+e^0} = \frac{e^{w^Tx}}{e^{w^Tx}+1} = \frac{1}{1+e^{-w^Tx}}$

which is equivalent the sigmoid function: $\sigma(w^Tx) = \frac{1}{1+e^{-w^Tx}}$

6.

We have: $\sigma(a) = \frac{1}{1+e^{-a}}$, $1-\sigma(a) = \frac{e^{-a}}{1+e^{-a}}$

$$\frac{\partial\sigma(a)}{\partial a} = \frac{\partial}{\partial a}\left(\frac{1}{1+e^{-a}}\right) = \frac{\partial}{\partial(1+e^{-a})}(1+e^{-a})^{-1} \cdot \frac{\partial(1+e^{-a})}{\partial a}$$

$$= (-1)\frac{1}{(1+e^{-a})^2} \cdot (-1) \cdot e^{-a} = \frac{e^{-a}}{(1+e^{-a})^2}$$, which is equivalent to $\sigma(a)\cdot(1-\sigma(a))$

7.

$\phi(x_1, x_2) = x_1^2 + x_2^2$

8.

Bayes' Theorem:

$p(y=1|x) \propto p(y=1)\,p(x|y=1), \quad p(y=0|x) \propto p(y=0)\,p(x|y=0)$

and $\dfrac{p(y=1)\,p(x|y=1)}{p(y=0)\,p(x|y=0)} = 1$

take logarithm: $\log \pi_1 - \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) - \frac{1}{2}\log|\Sigma_1|$

$= \log \pi_0 - \frac{1}{2}(x-\mu_0)^T \Sigma_0^{-1}(x-\mu_0) - \frac{1}{2}\log|\Sigma_0|$

$\Rightarrow -\frac{1}{2}(x^T\Sigma_1^{-1}x - 2x^T\Sigma_1^{-1}\mu_1 + \mu_1^T\Sigma_1^{-1}\mu_1) + \log\pi_1 - \frac{1}{2}\log|\Sigma_1|$

$= -\frac{1}{2}(x^T\Sigma_0^{-1}x - 2x^T\Sigma_0^{-1}\mu_0 + \mu_0^T\Sigma_0^{-1}\mu_0) + \log\pi_0 - \frac{1}{2}\log|\Sigma_0|$

$\Rightarrow x^T\left(\frac{1}{2}(\Sigma_0^{-1} - \Sigma_1^{-1})\right)x + \left((\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0)\right)^T x + \text{const.} = 0$

and $\text{const.} = \log\pi_1 - \log\pi_0 + \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0 - \frac{1}{2}\mu_1^T\Sigma_1^{-1}\mu_1 - \frac{1}{2}\log\frac{|\Sigma_1|}{|\Sigma_0|}$

$\Rightarrow$ is in the form: $\Gamma = \{x \mid x^TAx + b^Tx + c = 0\}$

with: $A = \frac{1}{2}(\Sigma_0^{-1} - \Sigma_1^{-1})$

$b = \Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0$

$c = \log\pi_1 - \log\pi_0 + \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0 - \frac{1}{2}\mu_1^T\Sigma_1^{-1}\mu_1 - \frac{1}{2}\log\frac{|\Sigma_1|}{|\Sigma_0|}$