The task is to prove that classical GDA has a linear decision boundary, given that target distribution follows Bernoulli distribution $y \sim \mathcal{B}(\phi)$ and each of two classes follows the Gaussian distribution with **same** covariance matrix $p(x|y=0) \sim \mathcal{N}(\mu_0, \Sigma)$ and $p(x|y=1) \sim \mathcal{N}(\mu_1, \Sigma)$

$$p(y) = \begin{cases} \phi & \text{if } y = 1 \\ 1-\phi & \text{if } y = 0 \end{cases} \tag{1}$$

$$p(x|y=0) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right] \tag{2}$$

$$p(x|y=1) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right] \tag{3}$$

Technically we need to derive the posterior distribution $p(y=1|x)$ and proove that it is in the form analogous to logistic regression, but with different coefficients $\theta_1 \in \mathbb{R}^n,\ \theta_0 \in \mathbb{R}$, which are functions of $\phi,\ \mu_0,\ \mu_1,\ \Sigma$.

$$p(y=1|x) = \frac{1}{1+\exp[-(\theta_1^T x + \theta_0)]}$$

First, let's denote multidimensional Gaussian distributions in a simplified manner

$$p(x|y=1) = \frac{1}{c}\exp\big[f(x,\mu_1)\big] \qquad \text{where } f(x,\mu_1) = -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

We will use the Bayes rule and law of total probability for $p(x)$

$$p(y=1|x) = \frac{p(x|y=1)p(y=1)}{p(x)} = \frac{p(x|y=1)p(y=1)}{p(x|y=1)p(y=1)\ +\ p(x|y=0)p(y=0)} =$$

$$= \frac{1}{1+\frac{p(x|y=0)p(y=0)}{p(x|y=1)p(y=1)}} = \frac{1}{1+\frac{\exp(f(x,\mu_0))}{\exp(f(x,\mu_1))}\frac{1-\phi}{\phi}} = \frac{1}{1+\exp\Big[-\Big(f(x,\mu_1)-f(x,\mu_0)+log(\frac{\phi}{1-\phi})\Big)\Big]}$$

So we would need to simplify $f(x,\mu_1) - f(x,\mu_0)$ part. First lets consider that covariance matrix Σ and it's inverse Σ^{-1} are symmetrical and lets open the brackets for the quadratic form and take a closer look at two inner terms

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = x^T\Sigma^{-1}x - \underbrace{\mu^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu}_{} + \mu^T\Sigma^{-1}\mu$$

As every symmetric matrix S by spectral decomposition theorem can be represented as product of three matrices, where Q - orthonormal ($Q^TQ = I$), and Λ - diagonal, and therefore always exists a square root of symmetric matrix S such that $\sqrt{S}\sqrt{S} = S$

$$S = Q\Lambda Q^{-1} = Q\Lambda Q^T$$

$$\sqrt{S}\sqrt{S} = Q\sqrt{\Lambda}Q^T \cdot Q\sqrt{\Lambda}Q^T = Q\sqrt{\Lambda}\sqrt{\Lambda}Q^T = Q\Lambda Q^T = S$$

Now lets consider two terms $\mu^T Sx$ and $x^T S\mu$, taking into account that $a^T x = x^T a$ and that $\sqrt{S}$ is also symmetric

$$\mu^T Sx = \mu^T\sqrt{S}\sqrt{S}x = (\sqrt{S}\mu)^T(\sqrt{S}x) = (\sqrt{S}x)^T(\sqrt{S}\mu) = x^T\sqrt{S}\sqrt{S}\mu = x^T S\mu$$

Given that result we can simplify the quadratic form

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = x^T\Sigma^{-1}x - 2x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu$$

So the second order terms would cancel out in case of identical covariance matrix for each class.

$$f(x,\mu_1) - f(x,\mu_0) = -\frac{1}{2}\Big[\cancel{x^T\Sigma^{-1}x} - 2x^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}\mu_1 \tag{4}$$

$$-\cancel{x^T\Sigma^{-1}x} + 2x^T\Sigma^{-1}\mu_0 - \mu_0^T\Sigma^{-1}\mu_0\Big] = \tag{5}$$

$$= x^T\Sigma^{-1}(\mu_1-\mu_0) - \frac{1}{2}(\mu_1-\mu_0)^T\Sigma^{-1}(\mu_1+\mu_0) =$$

$$= x^T\Sigma^{-1}\Delta\mu - \underbrace{\frac{1}{2}\Delta\mu^T\Sigma^{-1}(\mu_1+\mu_0)}_{\text{const}}, \quad \Delta\mu = \mu_1 - \mu_0$$

And now we can derive the complete equation for posterior distribution $p(y=1|x)$

$$p(y=1|x) = \frac{1}{1+\exp\Big[-\Big(f(x,\mu_1) - f(x,\mu_0) + log(\frac{\phi}{1-\phi})\Big)\Big]} \tag{6}$$

$$= \frac{1}{1+\exp\Big[-\Big(x^T\underbrace{\Sigma^{-1}\Delta\mu}_{\theta_1} \underbrace{- \frac{1}{2}\Delta\mu^T\Sigma^{-1}(\mu_1+\mu_0) + log(\frac{\phi}{1-\phi})}_{\theta_0}\Big)\Big]} \tag{7}$$

Right now we can also show that if we assume that two classes have different covariance matrices the decision boundary would be second order hypersurface.

$$p(y=1|x) = \frac{1}{1+\exp(-(x^T\theta_2 x + x^T\theta_1 + \theta_0))}, \quad \theta_2 = \Sigma_1 - \Sigma_0 \in \mathbb{R}^{n\times n}$$