1) Using the result of problem 2 compute the maximum likelihood estimate for the prior probabilities. The result is the same as of last week. Problem 2 shows that the MLE for the prior = $\frac{c_j}{n}$, c_j is the histogram of the sample observations, thus is the number of observations in BG and FG respectively. And n is the sum of all the observations. Thus, we get the same answer:

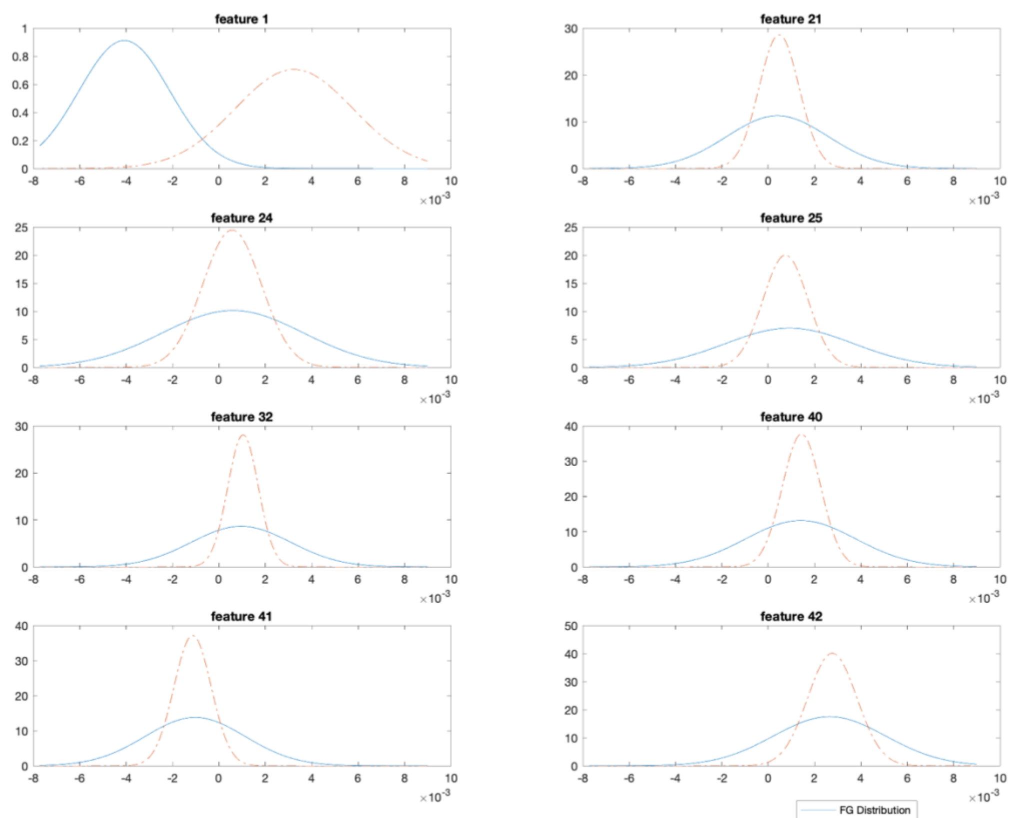
$$P_Y(grass) = 0.8081 \quad P_Y(cheetah) = 0.1919$$

2) First, I compute the mean and variance for each of the 64 columns for BG and FG respectively, using the formulas and then compute the normal densities and compare them under each feature to get a 64-plot. Then, I pick the best and worst features depending on their deviation of means and distribution dispersions.

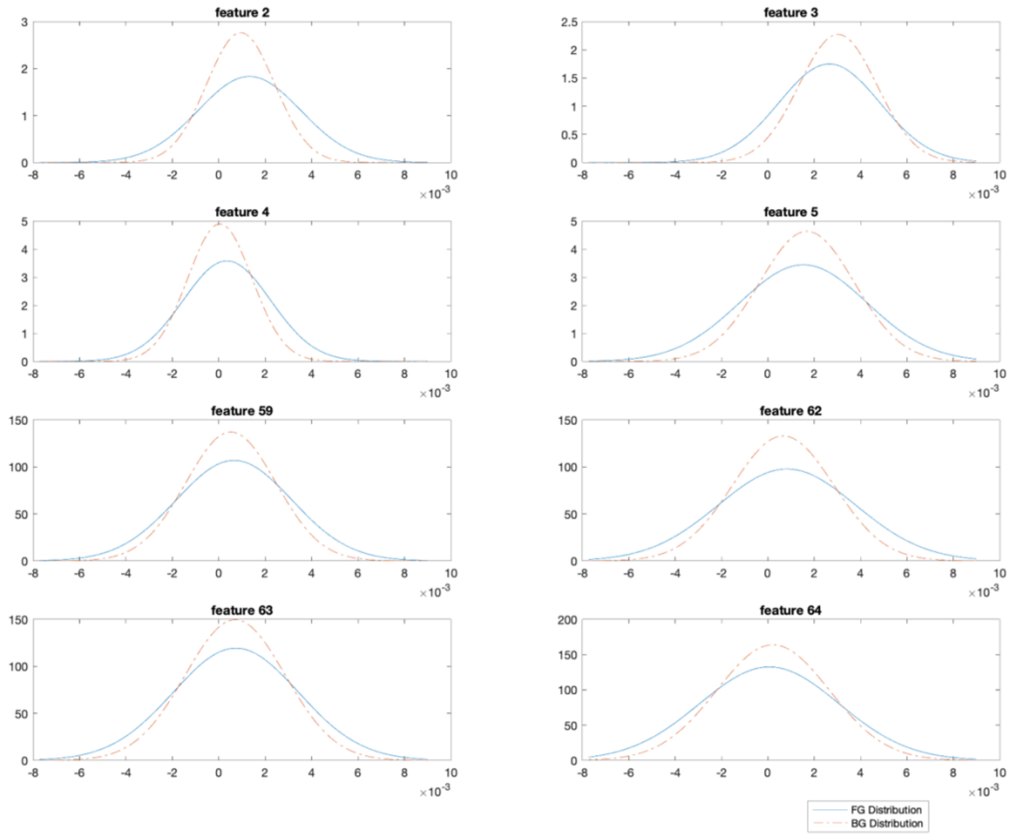
$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})^2.$$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

The best 8 plots of the marginal densities: feature 1,21,24,25,32,40,41,42.



The worst 8 plots of the marginal densities: feature 2,3,4,5,59,62,63,64.



3) The 64-dimensional Gaussians

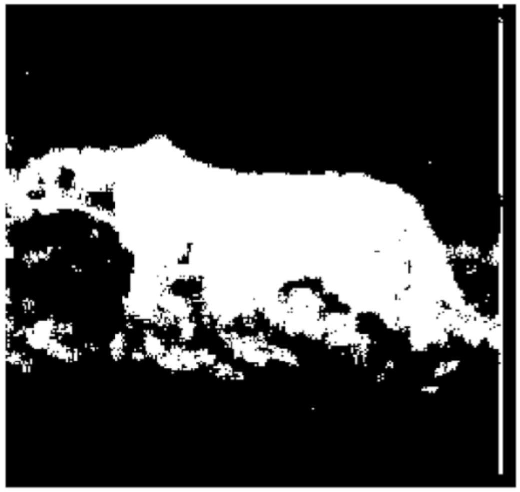
To calculate the posterior probabilities of cheetah and grass using 64 features (d=64), first I convert the image into 8x8 matrix as I did in BDR_ImageProcessing and match with the zig zag position to get the features matching with training data.

I use the Multi-variate Gaussian mean and covariance formula to get the mean vectors (1x64) and covariance matrices (64x64) for BG and FG.

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k \qquad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^t.$$

Then, I calculate the log of multivariate Gaussian density function to get the posterior probabilities and then apply the discriminant function to get the state variable based on the decision rules: $P_{X|Y}(x|cheetah) > P_{X|Y}(x|grass),\ A = 1, else\ A = 0.$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$



a) Mask for the 64-dimensional Gaussian is above.
   Detection rate: $P_{X|Y}(g(x) = cheetah|cheetah) \; = \; 0.9272$
   False alarm rate: $P_{X|Y}(g(x) = cheetah|grass) \; = \; 0.0938$
   Error rate: $P_{X|Y}(g(x) = cheetah|grass) \times P_Y(grass) +$
   $P_{X|Y}(g(x) = grass|cheetah) \; \times \; (cheetah) \; = \; 0.0898$

b) Mask for the 8-dimensional Gaussian is above.
   Detection rate: 0.9001
   False alarm rate: 0.0420
   Error rate: 0.0531

   The result of using 8 features is better than using 64 features, because in the 64 feature case we also include distributions which are similar and cannot used to distinguish grass and cheetah pattern.