Homework 2

Wenjun Zhang

A53218995

a) According the result of problem 2, the maximum likelihood estimate for the prior probabilities can be expressed as:

$$\pi_j^* = \frac{c_j}{n}$$

where c_j is the number of observations and n is the total number of observations.
Using the training data in TrainingSamplesDCT_8_new.mat, the histogram of observations is shown in Figure.1, where the first one is the grass and the second one is the cheetah. And using the equation above, we can compute the maximum likelihood estimate for the prior probabilities as:

$$P_Y(cheetah) = \frac{250}{250 + 1053} = 0.1919$$

$$P_Y(grass) = \frac{1053}{250 + 1053} = 0.8081$$

and show them in Figure.2, where the first and the second one are the grass and cheetah.

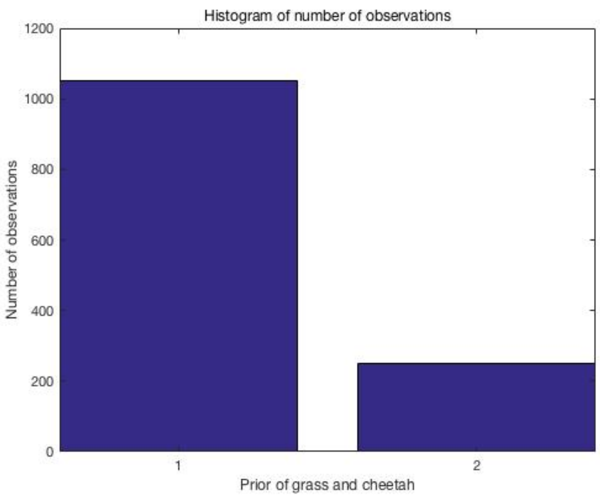

Figure. 1 Histogram of observations

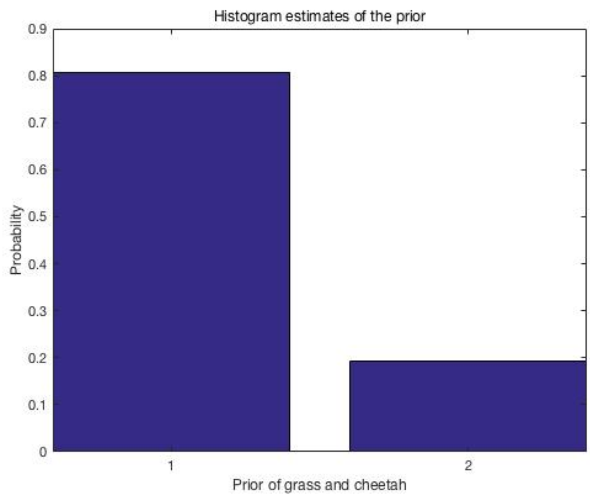

Figure.2 Histogram estimate of the prior

This essentially is the same as what we did in last week, where we compute the probabilities using the number of observations divided by the total amount of observations.

b) Computing the mean and variance for each dimension of features in TrainingSamplesDCT_8_new.mat, we can plot the marginal densities for the two classes $P_{X_k|Y}(x_k|cheetah)$ and $P_{X_k|Y}(x_k|grass)$ under Gaussian assumption. We plot the 64 plots in Figure.3 ~ Figure.6.

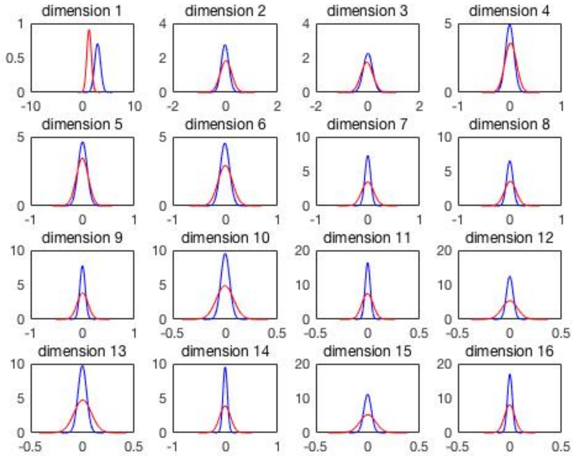

Figure.3 Marginal densities (1)

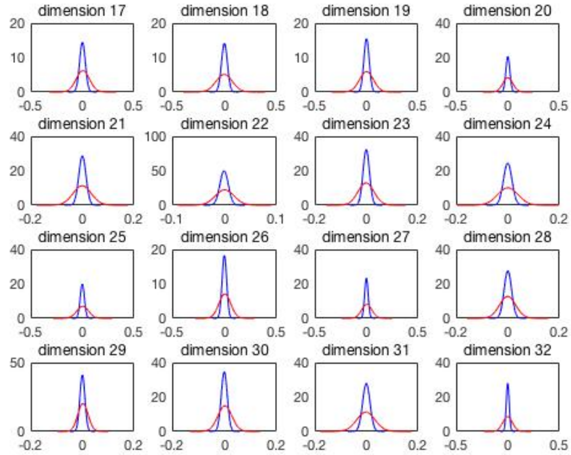

Figure.4 Marginal densities (2)

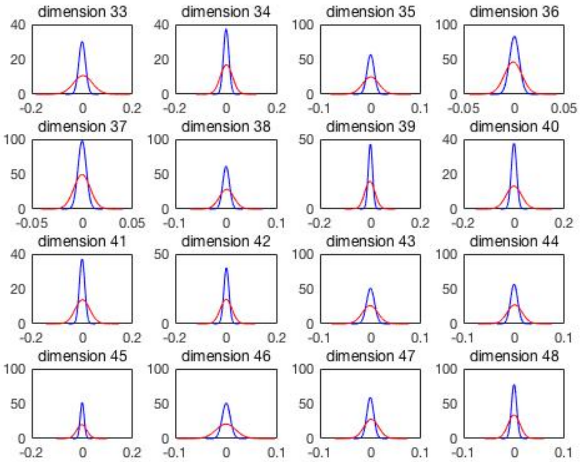

Figure.5 Marginal densities (3)

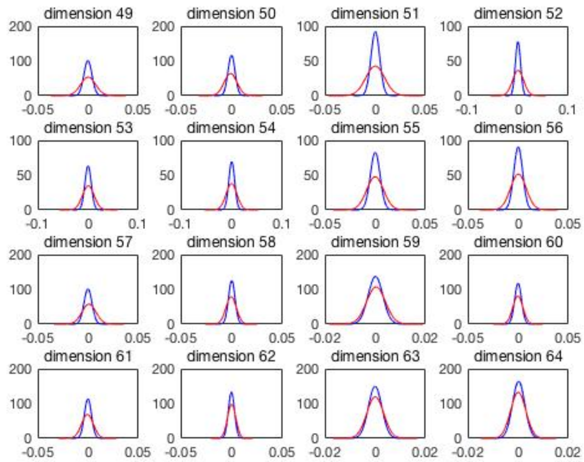

Figure.6 Marginal densities (4)

By visual inspection, we select [1,11,14,17,23,26,32,40] as the best 8 features and [3,4,5,59,60,62,63,64] as the worst features, the marginal densities of them are shown in Figure.7 and Figure.8.

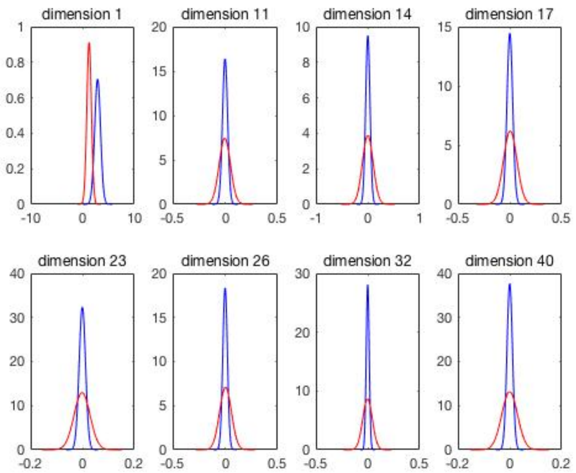

Figure.7 The 8 best features

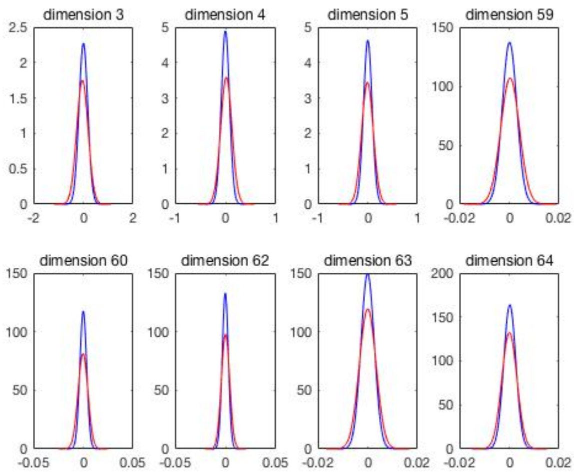

Figure.8 The 8 worst features

c) We can compute $P_{X|Y}(x|cheetah)$ and $P_{X|Y}(x|grass)$ using the equation below:

$$P_{X|Y}(x|i) = \frac{1}{\sqrt{(2\pi)^d|\Sigma_i|}} \exp\{-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\}$$

Then we can use the sigmoid function below:

$$g_0(x) = \frac{1}{1+\exp\{d_0(x-\mu_0)-d_1(x-\mu_1)+\alpha_0-\alpha_1\}}$$

where

$$\begin{cases} d_i(x,y) = (x-y)^T\Sigma_i^{-1}(x-y) \\ \alpha_i = \log(2\pi)^d|\Sigma_i| - 2logP_Y(i) \end{cases}$$

And choose the pixel to be "cheetah" if

$$g_{grass}(x) < 0.5$$

Using the 64-dimensional Gaussians, the classification masks is shown in Figure.9. The decision rate $P_{X|Y}(g(x) = cheetah|cheetah) = 0.9319$, and the false alarm rate $P_{X|Y}(g(x) = cheetah|grass) = 0.0924$. So the probability error is

$$P_E = (1-0.9319)*0.1919 + 0.0924*0.8081 = 0.0877$$

Using the best 8 features, the classification masks is shown in Figure.10. The decision rate $P_{X|Y}(g(x) = cheetah|cheetah) = 0.9638$, and the false alarm rate $P_{X|Y}(g(x) = cheetah|grass) = 0.0961$. So the error of probability is

$$P_E = (1-0.9638)*0.1919 + 0.0961*0.8081 = 0.0477$$

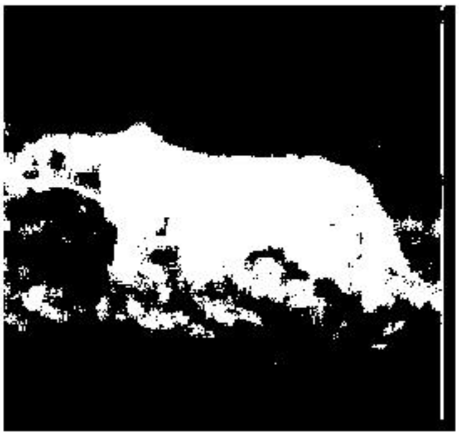

Figure.9 Classification mask with 64 features

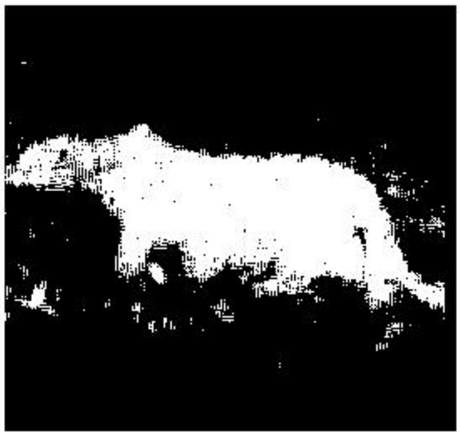

Figure.10 Classification mask with 8 best features

We can see from the probability error that the best 8 features beat the 64 features. The reason for this is that the 64 features include some features which share similar densities between classes. So while the number of features increase, the model for the class conditional densities become worse due to these features. Thus, the accuracy of the model decreases.