Gaussian Classifier

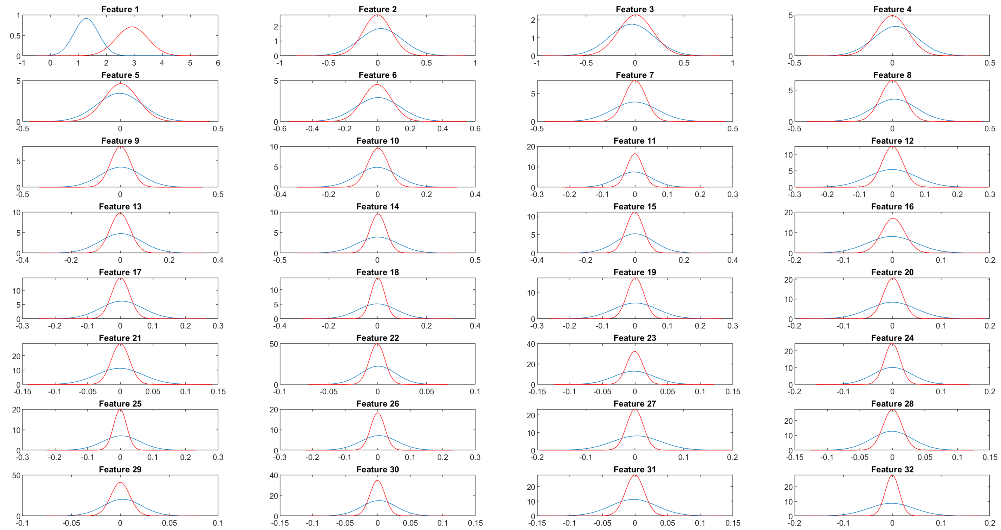
Saurabh Mirani

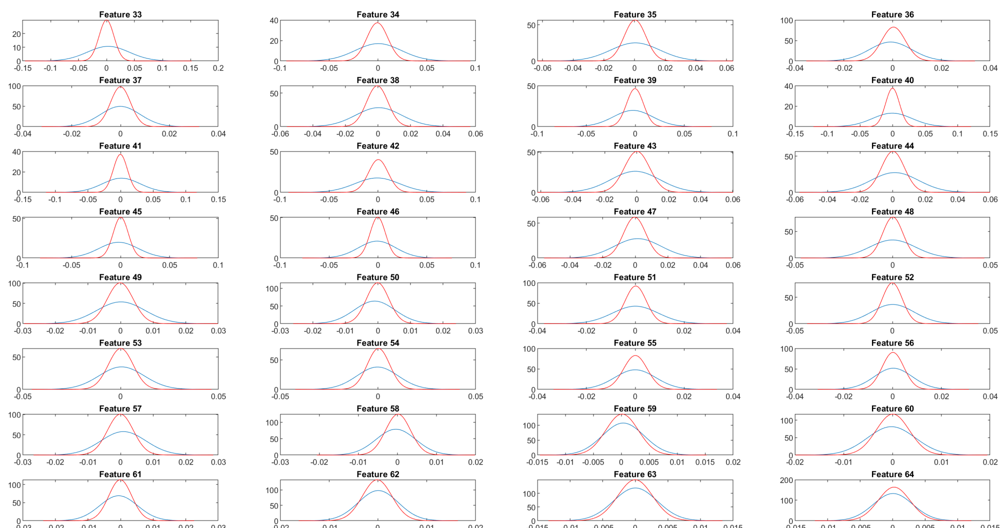
October 2019

1. **6. (a)** In the first assignment, prior probabilities were calculated using the number of samples from cheetah and grass.
   Here,
   $$P_Y(i) = \frac{c_i}{n} \tag{1}$$
   where,
   c_i is is the number of times that the observed value is i
   and n is the number of independent observations from Y
   Hence, the prior probabilities remain the same in both the cases.

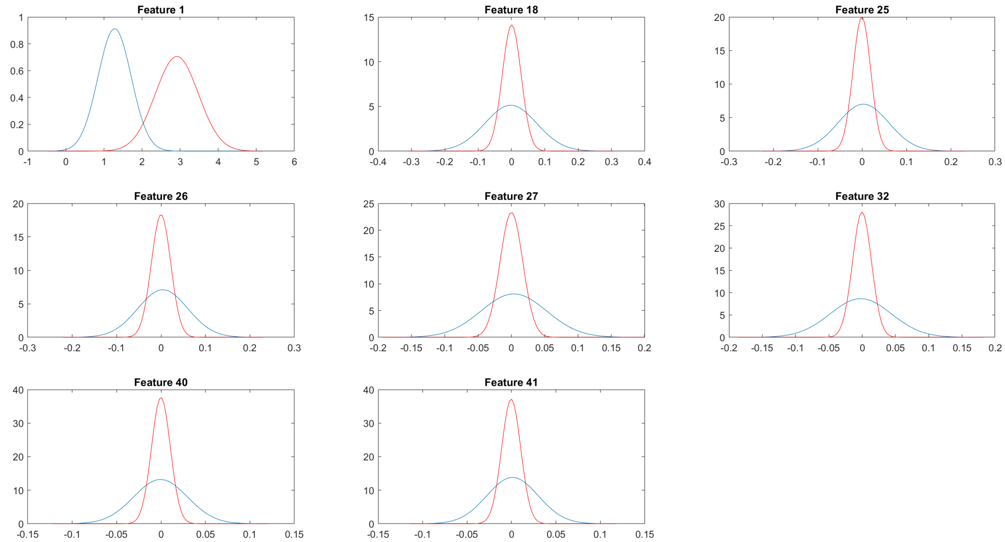
2. **6 (b)** The class conditional probability distribution can be given by,
   $$P_{X|Y}(X|i) = \frac{1}{\sqrt{(2\pi)^d|\Sigma_i|}} exp\{-\frac{(x-\mu_i)^T\Sigma_i(x-\mu_i)}{2}\} \tag{2}$$
   where,
   $$\mu_i = \frac{1}{n}\sum_j x_j^{(i)} \;\;\&\;\; \Sigma_i = \frac{1}{n}\sum_j (x_j^{(i)}-\mu_i)(x_j^{(i)}-\mu_i)^T$$
   Consider the case of single variable Gaussian,
   Let
   $$g = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x-\mu}{\sigma}^2}$$
   $$\frac{\partial g}{\partial \mu} = \frac{\Sigma(x-\mu)}{\sigma^2}$$
   $$\frac{\partial g}{\partial \sigma} = \frac{-N}{\sigma} + \frac{\Sigma}{\sigma^3}$$
   $$\frac{\partial^2 g}{\partial \mu^2} = \frac{-N}{\sigma^2}$$
   $$\frac{\partial^2 g}{\partial \sigma^2} = \frac{-2N}{\sigma^2}$$
   $$\frac{\partial^2 g}{\partial \mu \partial \sigma} = \frac{\partial^2 g}{\partial \sigma \partial \mu} = \frac{-2\Sigma(x-\mu)}{\sigma^3}$$
   $H = \begin{bmatrix} \frac{-N}{\sigma^2} & \frac{-2\Sigma(x-\mu)}{\sigma^3} \\ \frac{-2\Sigma(x-\mu)}{\sigma^3} & \frac{-2N}{\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{-N}{\sigma^2} & 0 \\ 0 & \frac{-2N}{\sigma^2} \end{bmatrix}$
   Now, $\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \frac{-N}{\sigma^2} & 0 \\ 0 & \frac{-2N}{\sigma^2} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = -\frac{Na^2}{\sigma^2} - \frac{2Nb^2}{\sigma^2} < 0$
   Hence, Hessian is negative definite.

The plots of the marginal densities for the 64 features are:

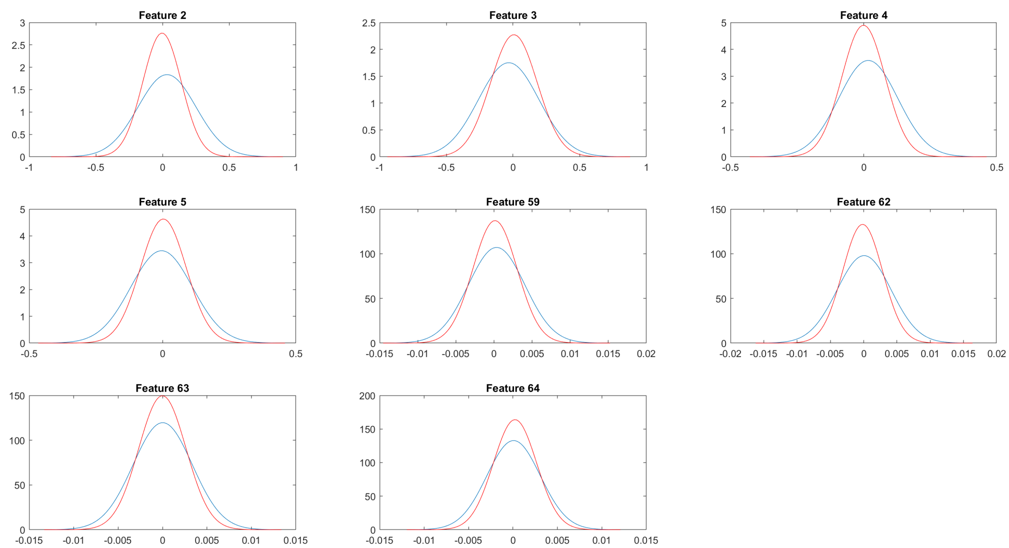




We get the best 8 features by visualization. The best feature can be visualized as the plot with different probability distribution. More the difference better is the feature. Hence, by visualization the best 8 features are: [1, 18, 25, 26, 27, 32, 40, 41] The plots of the marginal densities for the best 8 features are:



Similarly, the worst 8 features will have almost same probability distributions. By visualization the worst 8 are: [2, 3, 4, 5, 59, 62, 63, 64] The plots of the marginal densities for the worst 8 features are:



3. **6 (c)**

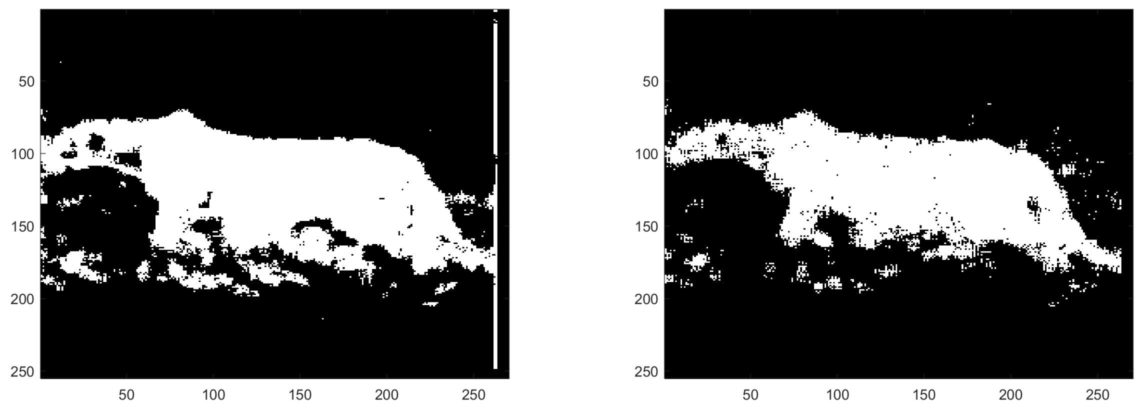
The Gaussian classifier is given by,

$$i^*(x) = \underset{i}{\operatorname{argmin}}[d_i(x, \mu_i) + \alpha_i] \tag{3}$$

where,
$d_i(x, y) = (x - y)^T \Sigma_i^{-1}(x - y)$ is the Mahalanobis distance,

$\alpha_i = log(2\pi)^d|\Sigma_i| - 2logP_Y(i)$

The classification masks are:



(a) 64-Dimension

(b) 8-Dimension

Figure 1: Classification masks

The probability of error is given by

$$P_{error} = \frac{No.\ of\ misclassified\ pixels\ FG}{Total\ true\ no.\ of\ FG\ pixels} * P_Y(Cheetah) + \frac{No.\ of\ misclassified\ pixels\ BG}{Total\ true\ no.\ of\ BG\ pixels} * P_Y(Grass)$$

$$P_{error}64D = \frac{962}{13209} * 0.1919 + \frac{5220}{55641} * 0.8081 = 0.089$$

$$P_{error}8D = \frac{1443}{13209} * 0.1919 + \frac{2369}{55641} * 0.8081 = 0.055$$

Hence, for 64D P_{error} is 0.089 and for 8D P_{error} is 0.055.

We can see that the P_{error} for 8D is less than that of the 64D. This is because in 64D there are features which are bad and result in misclassification. Since, the bad features have similar marginal densities for the two classes it is difficult to determine which class. However, this is not the case with 8D. All the bad features are removed and only the best ones are used. It is easier to determine the class. Hence, the chances of committing an error is quite low as compared with the case when all features are used.