Gaussiana no espaço R²

Hernane Braga Pereira - 2014112627

1. Introdução

Este relatório tem como objetivo exemplificar a obtenção das densidades de probabilidades na discriminação de dados binários, havendo, ou não, correlação entre os mesmos.

2. Densidade de probabilidades para dados com correlação nula

A probabilidade de um ponto x pertencer à classe C1, ou C2, dado que elas são independentes, é expressa através das equações:

$$P(\mathbf{x}|C_1) = \frac{1}{2\pi\sigma_{11}\sigma_{12}} e^{-\left(\frac{(x_1-\mu_{11})^2}{2\sigma_{11}^2} + \frac{(x_2-\mu_{21})^2}{2\sigma_{21}^2}\right)}$$

$$P(\mathbf{x}|C_2) = \frac{1}{2\pi\sigma_{12}\sigma_{22}} e^{-\left(\frac{(x_1-\mu_{12})^2}{2\sigma_{12}^2} + \frac{(x_2-\mu_{22})^2}{2\sigma_{22}^2}\right)}$$

Equação 1. Cálculo de densidade de probabilidades para dados com correlação nula

Para este primeiro exemplo, um conjunto de dados foi gerado com a classe C1 (pontos vermelhos) possuindo média em torno do ponto (2,2) e para a classe C2 (pontos azuis) com média em torno do ponto (4,4). Ambas as classes possuem desvio padrão $\sigma = 0,6$.

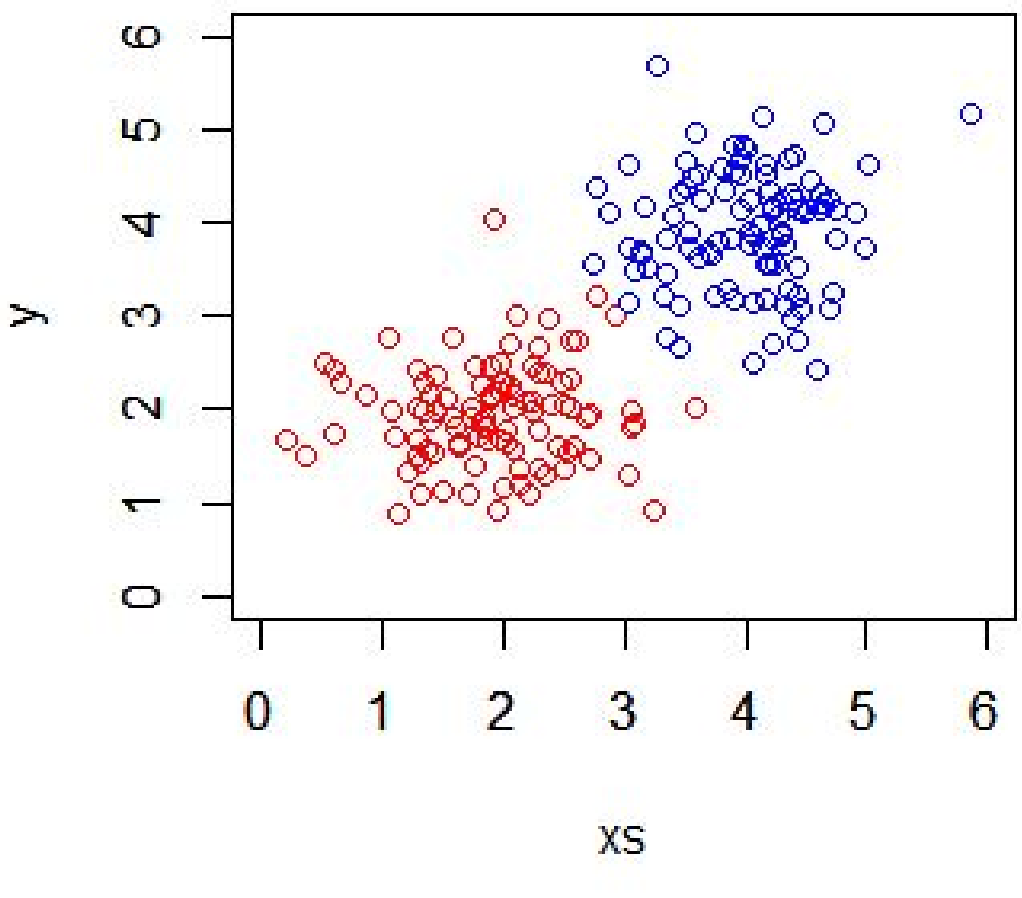

Figura 1. Dados com $\sigma = 0,6$

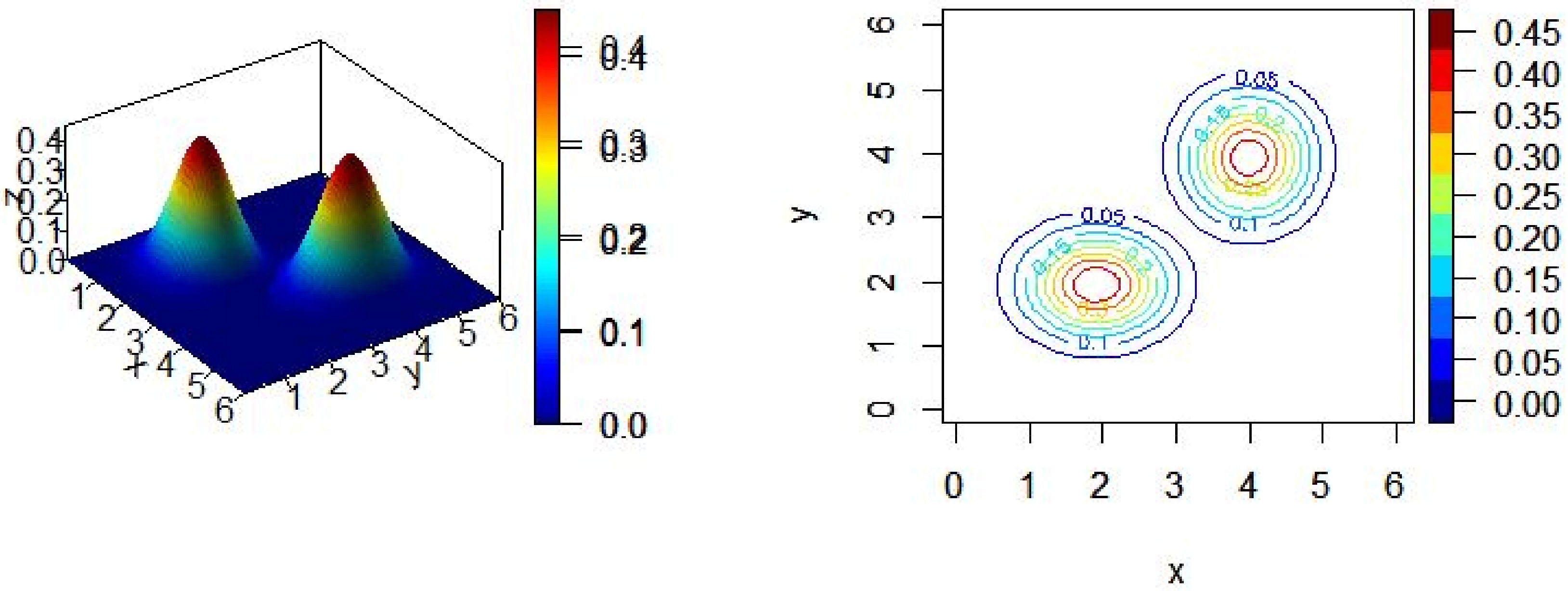

Figura 2. Densidade de probabilidade (à esquerda) e superfície de contorno (à direita) das classes com coeficiente de correlação nulo

Ao aplicar a equação 1 nos dados, conclui-se que foi possível separar as classes C1 e C2, não havendo interseção entre as gaussianas das mesmas.

3. Densidade de probabilidades para dados com correlação

Para dados que não são independentes, é preciso levar em consideração a covariância entre as variáveis. Para tal, utiliza-se a seguinte equação:

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \left(\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)}$$

Equação 2. Cálculo de densidade de probabilidades para dados que não sejam independentes

Para este segundo exemplo, as mesmas classes C1 e C2 foram geradas com as mesmas médias, porém com desvio padrão unitário $\sigma = 1$ e com o coeficiente de correlação p sendo variado.

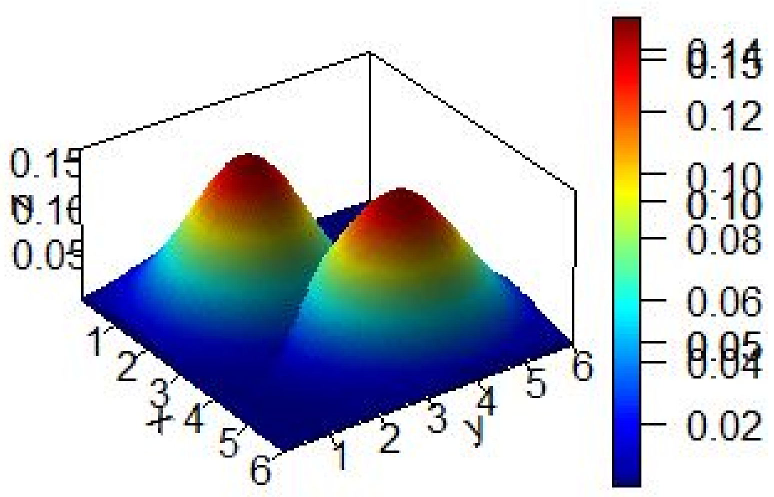

Figura 3. Densidade de probabilidade com $\sigma = 1\ e\ p = 0,1$

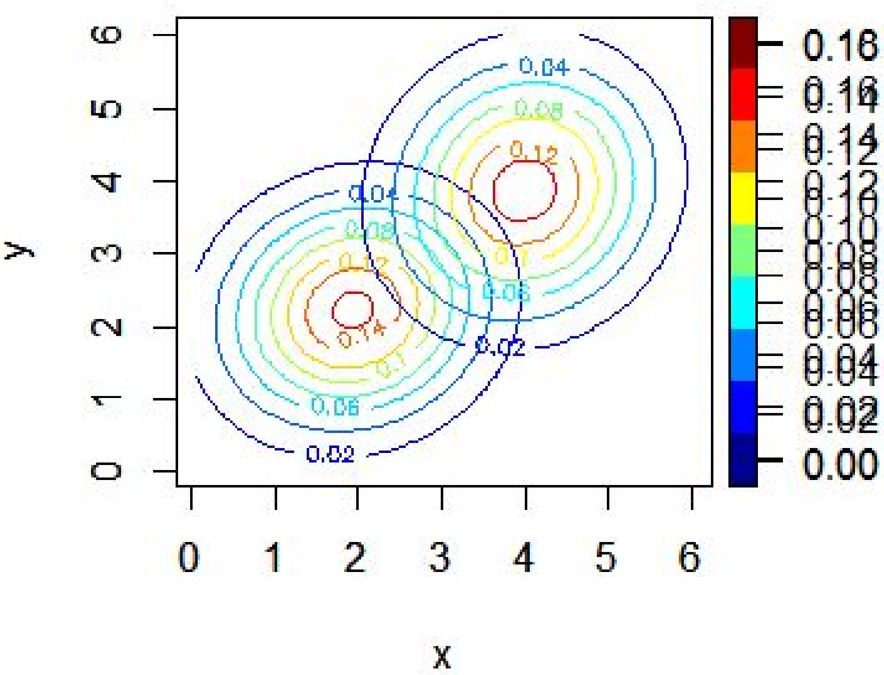

Figura 4. Superfície de separação com $\sigma = 1\ e\ p = 0,1$

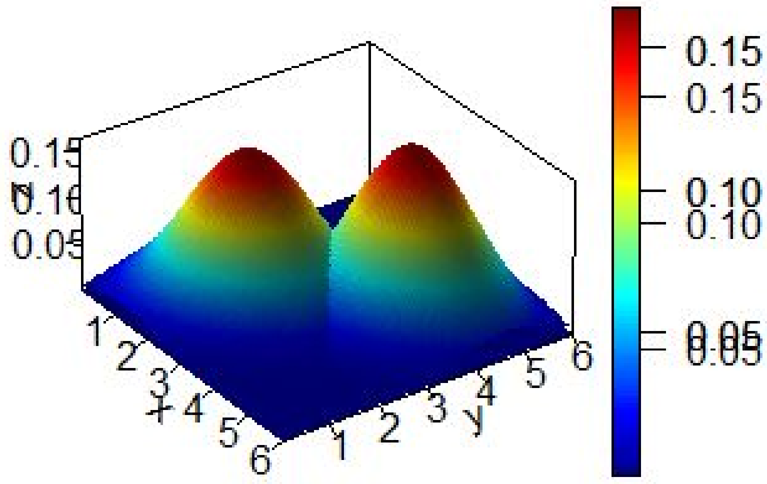

Figura 5. Densidade de probabilidade com $\sigma = 1\ e\ p = 0,3$

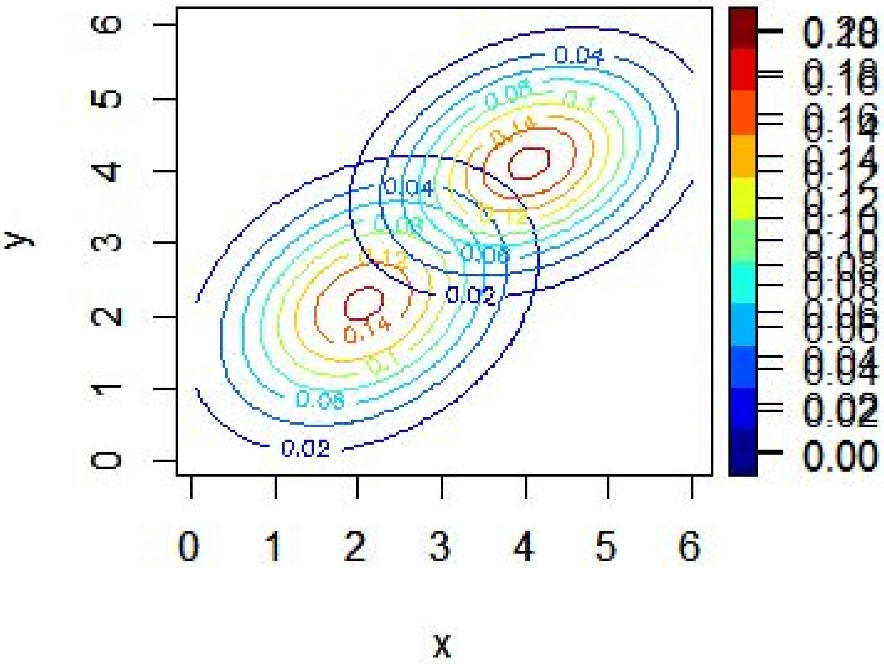

Figura 6. Superfície de separação com $\sigma = 1\ e\ p = 0,3$

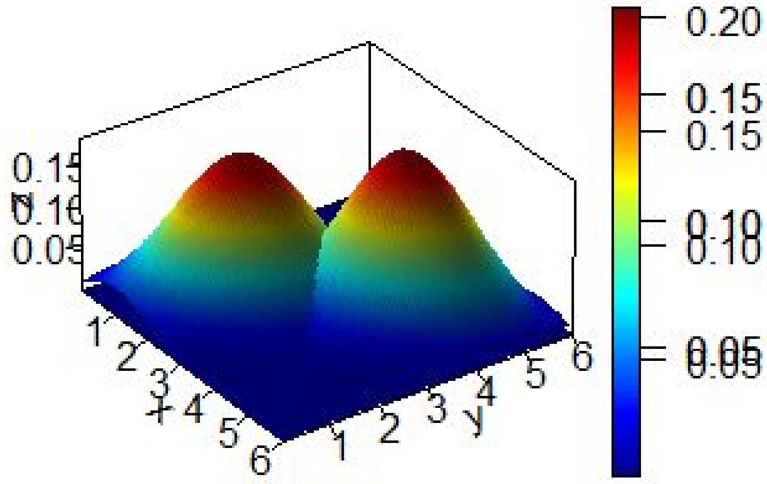

Figura 7. Densidade de probabilidade com $\sigma = 1\ e\ p = 0,5$

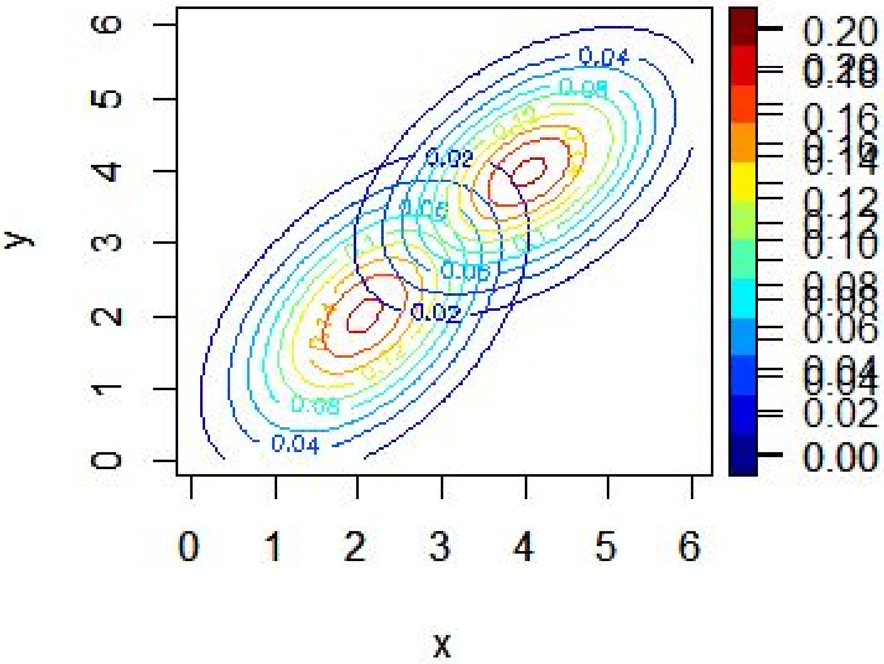

Figura 8. Superfície de separação com $\sigma = 1\ e\ p = 0,5$

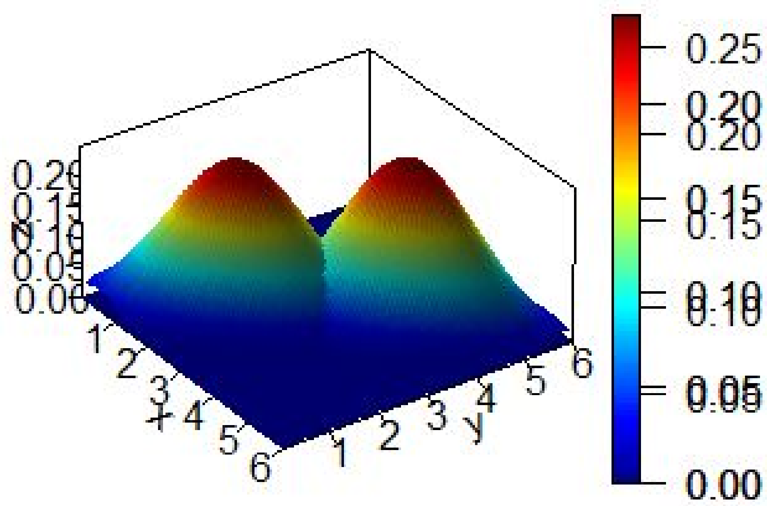

Figura 9. Densidade de probabilidade com $\sigma = 1\ e\ p = 0,7$

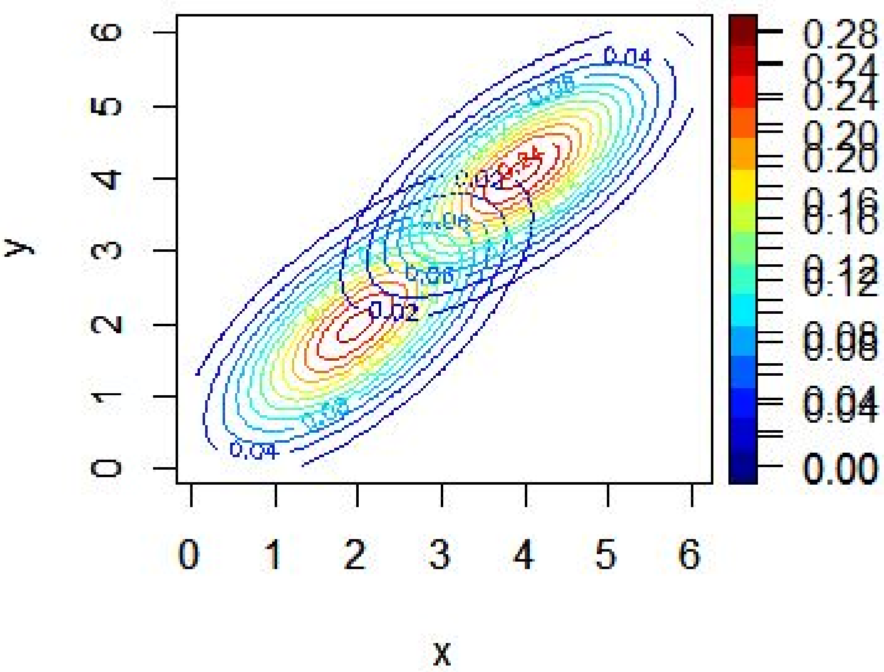

Figura 10. Superfície de separação com $\sigma = 1\ e\ p = 0,7$

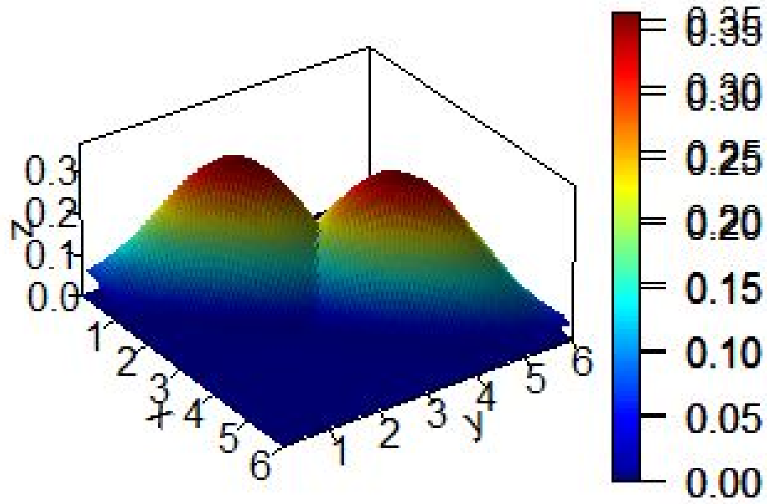

Figura 11. Densidade de probabilidade com $\sigma = 1 \; e \; p = 0,9$

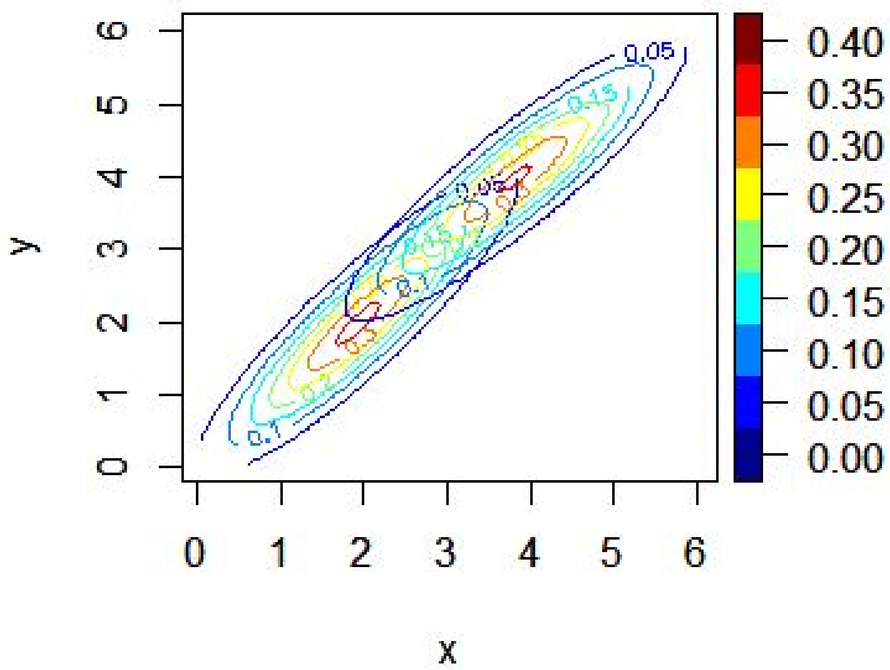

Figura 12. Superfície de separação com $\sigma = 1 \; e \; p = 0,9$

Ao analisar os resultados gráficos, nota-se que o aumento da correlação entre as variáveis dificulta a separação de classes. Com isto, conclui-se que o método de estimativa de densidade de probabilidade não é o mais recomendável quando as variáveis possuem correlação entre si.

3. Referências

[1] Estimativa de densidade, Notas de aula, agosto de 2019.