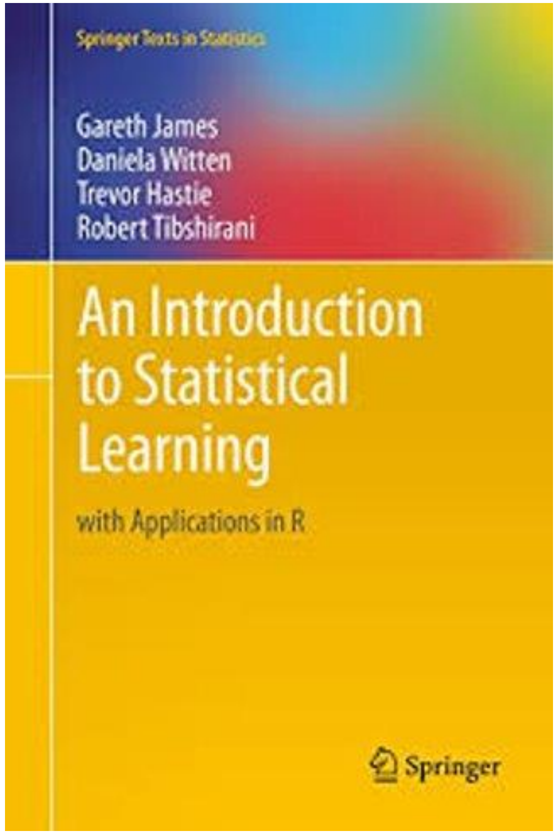
Capítulo 4 -Classificação

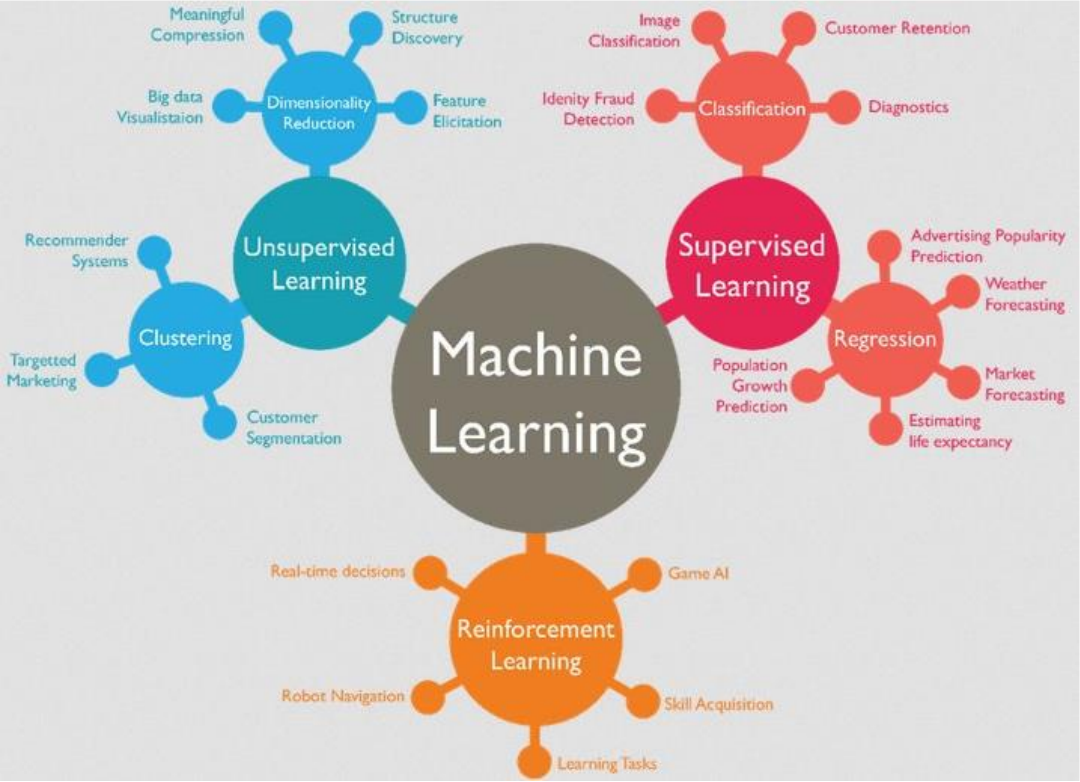
Marlete Silva

18/09/2020

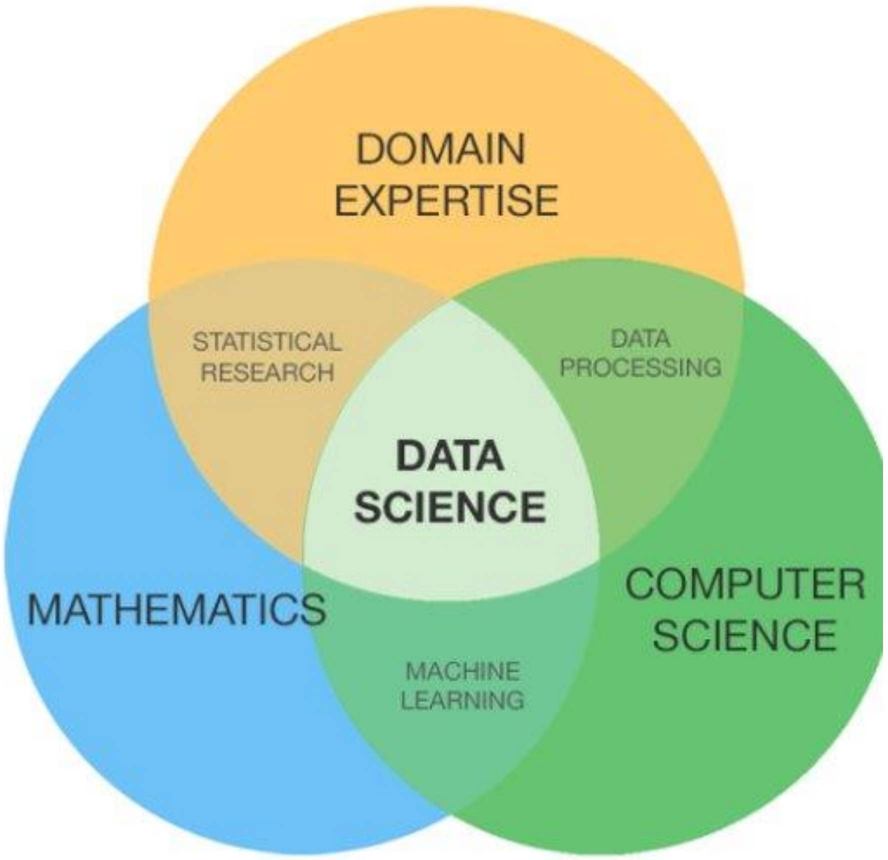


Aprendizagem Supervisionada

- **Regressão**: previsão de resultados quantitativos (numéricos)
- **Classificação**: previsão de resultados qualitativos (discretos)



- O modelo de regressão linear discutido no Capítulo 3 assume que a resposta a variável Y é quantitativa, podendo ser qualitativa;
- Por exemplo, a cor dos olhos é qualitativa, tomando qualitativo em valores de azul, marrom ou verde;
- Muitas vezes, as variáveis qualitativas são referidas como categórico;
- Neste capítulo, estudamos abordagens para prever respostas qualitativas, um processo que é conhecido como classificação;



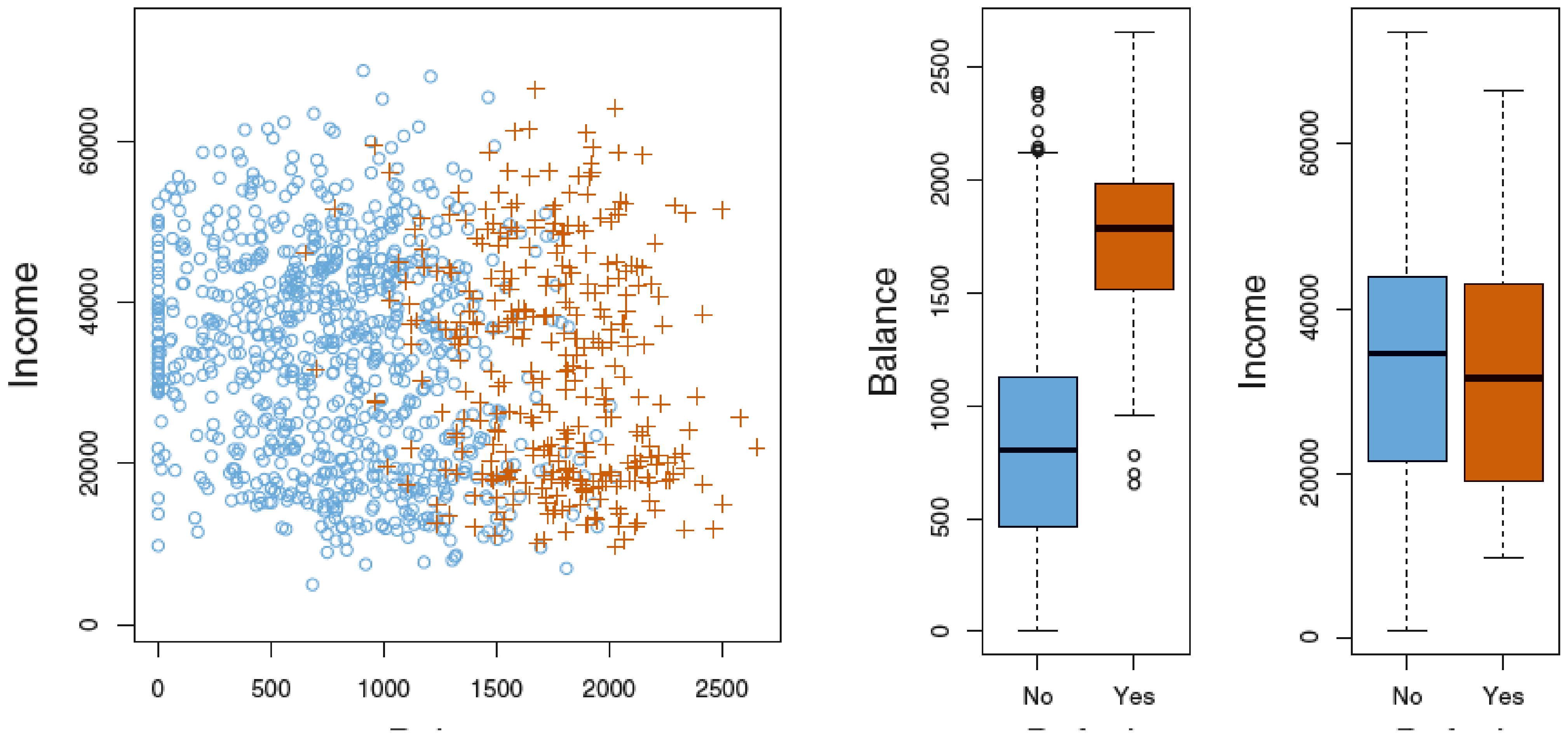
- Neste capítulo, discutimos três dos mais classificadores amplamente utilizados:
  - ❑regressão linear
  - ❑análise discriminante linear
  - ❑ regressão logística
  - ❑ Naive bayes

Problemas de classificação

- 1. Uma pessoa chega ao pronto-socorro com um conjunto de sintomas que possivelmente pode ser atribuído a uma das três condições médicas.
  - Qual das três condições o indivíduo tem?
- 2. Um serviço de banco online deve ser capaz de determinar se ou não uma transação realizada no site é fraudulenta, com base do endereço IP do usuário, histórico de transações anteriores e assim por diante.
- 3. Com base nos dados da sequência de DNA de uma série de pacientes com e sem uma determinada doença, um biólogo gostaria de descobrir quais mutações no DNA são deletérias (causadoras de doenças) e não são.

- Assim, na configuração de classificação temos um conjunto de observações de treinamento (x1, y1),. . . , (xn, yn) que podemos usar para construir um classificador.

- Na Figura 4.1 Estamos interessados em prever se um indivíduo ficará inadimplente em seu pagamento com cartão de crédito, com base em renda anual e saldo mensal do cartão de crédito.



O painel esquerdo da Figura 4.1 exibe indivíduos que estão inadimplentes em um determinado mês em laranja, e aqueles quem não estão inadimplentes em azul.

4.4 Análise Discriminante Linear

- A regressão logística envolve modelar diretamente Pr (Y = k | X = x) usando o função logística, para o caso de duas classes de resposta.
- No jargão estatístico, modelamos a distribuição condicional da resposta Y, dado o (s) preditor (es) X.
- Por que precisamos de outro método, quando temos regressão logística?
  - Quando as classes são bem separadas, o parâmetro estima para o modelo de regressão logística são surpreendentemente instáveis.

    Na Discriminante linear a análise não sofre deste problema.
  - Se n for pequeno e a distribuição dos preditores X for aproximadamente normal em cada uma das classes, o modelo discriminante linear é novamente mais estável do que o modelo de regressão logística.
  - A análise discriminante linear é popular quando temos mais de duas classes de resposta.

- Suponha que desejamos classificar uma observação em uma das K classes, onde K ≥ 2

$$\Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)}. \tag{4.10}$$

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right), \tag{4.11}$$

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_l)^2\right)}. \tag{4.12}$$

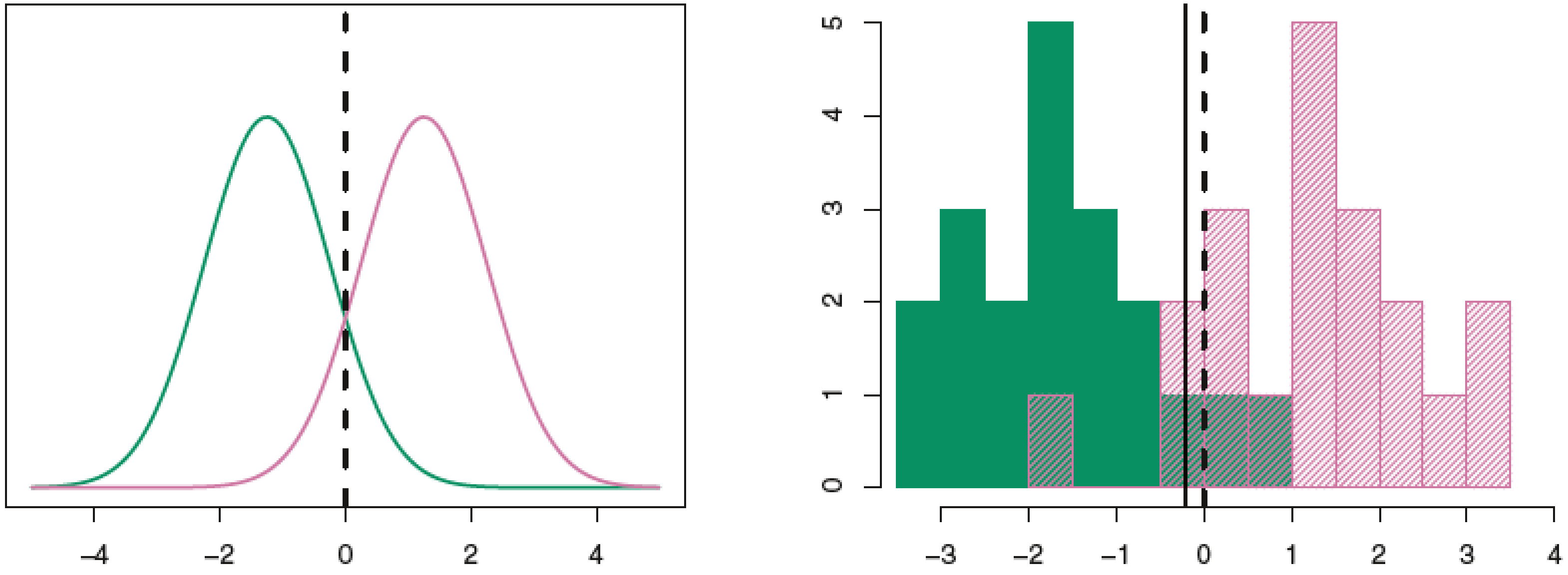


FIGURA 4.4. Esquerda: Duas funções unidimensionais de densidade normal são mostradas. A linha vertical tracejada representa o limite de decisão de Bayes. Direita: 20 observações foram retirados de cada uma das duas classes e são mostrados como histogramas. O limite de decisão de Bayes é novamente mostrado como uma linha vertical tracejada. O sólido a linha vertical representa o limite de decisão LDA estimado a partir do treinamento dados.

- Análise Discriminante Linear 141 (LDA) aproxima o classificador Bayes.

$$\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k) \tag{4.13}$$

$$x = \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} = \frac{\mu_1 + \mu_2}{2}. \tag{4.14}$$

$$\begin{aligned} \hat{\mu}_k &= \frac{1}{n_k} \sum_{i:y_i=k} x_i \\ \hat{\sigma}^2 &= \frac{1}{n-K} \sum_{k=1}^{K} \sum_{i:y_i=k} (x_i - \hat{\mu}_k)^2 \end{aligned} \tag{4.15}$$

$$\hat{\pi}_k = n_k/n. \tag{4.16}$$

$$\hat{\delta}_k(x) = x \cdot \frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k) \tag{4.17}$$

- Às vezes temos conhecimento das probabilidades de associação de classe $\pi 1,. . . , \pi K$, que pode ser usada diretamente. Na ausência de qualquer informações, LDA estima πk usando a proporção das observações de treinamento que pertencem à classe k.
- palavra linear no nome do classificador deriva do fato que as funções discriminantes $\hat{}\delta k$ (x) em (4.17) são funções lineares de x (como discriminante oposta a uma função mais complexa de x).
- O painel direito da Figura 4.4 exibe um histograma de um amostra de 20 observações de cada classe. Para implementar LDA, começamos estimando πk, μk e $\sigma 2$ usando (4.15) e (4.16).
- Em seguida, calculamos o limite de decisão, mostrado como uma linha sólida preta, que resulta da atribuição uma observação para a classe para a qual (4.17) é o maior.

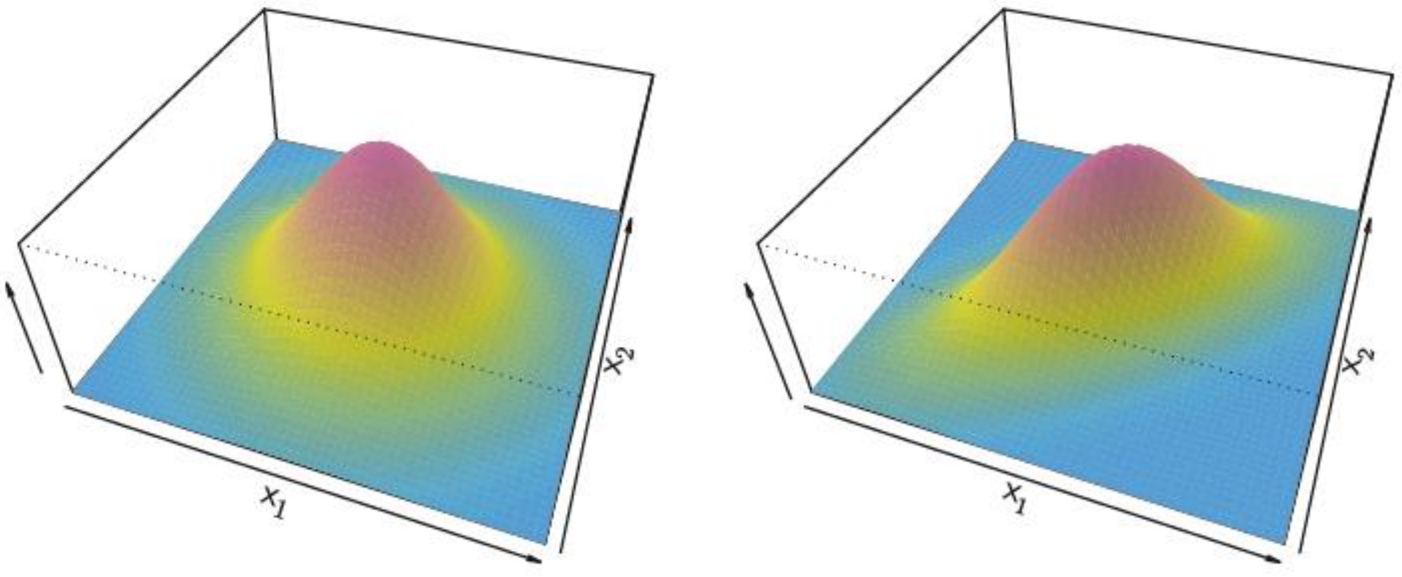


FIGURA 4.5. Duas funções multivariadas de densidade gaussiana são mostradas, com p = 2. Esquerda: Os dois preditores não estão correlacionados. Direita: As duas variáveis têm uma correlação de 0,7. A altura da superfície em qualquer ponto particular representa a probabilidade de que ambos X1 e X2 caem em uma pequena região em torno desse ponto.

A figura indica que o limite de decisão LDA está ligeiramente à esquerda de o limite de decisão de Bayes ideal, que em vez disso é igual a (μ1 + μ2) / 2 = 0. Qual é o desempenho do classificador LDA com esses dados? Uma vez que este é dados simulados, podemos gerar um grande número de observações de teste para calcular a taxa de erro de Bayes e a taxa de erro do teste LDA.

4.4.3 Análise Discriminante Linear para p> 1

- A distribuição multivariada de Gauss assume que cada preditor individual segue uma distribuição normal unidimensional, como em (4.11), com alguma correlação entre cada par de preditores.

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right), \qquad (4.11)$$

Formalmente, a densidade multivariada de Gauss é definido como

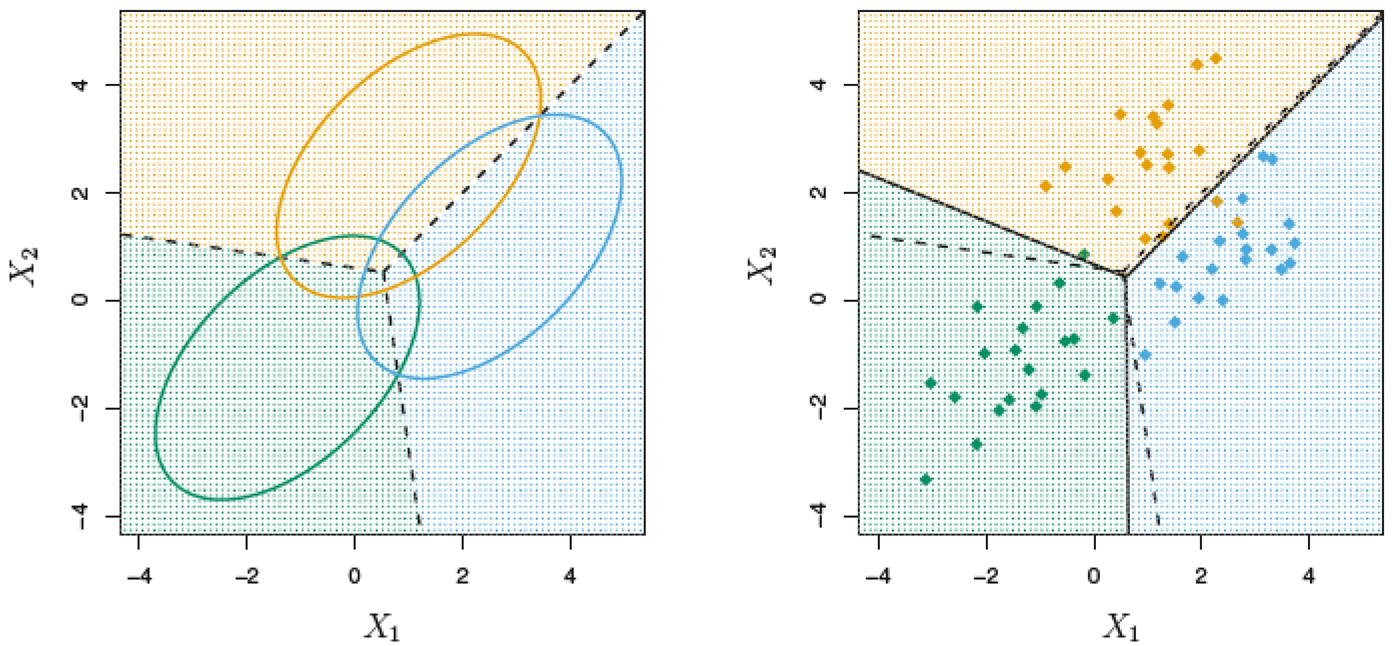


FIGURA 4.6. Um exemplo com 3 classes. As observações de cada classe são retirados de uma distribuição gaussiana multivariada com p = 2, com uma classe específica vetor médio e uma matriz de covariância comum. Esquerda: bolinhas que contêm 95% da probabilidade de cada uma das três classes são mostrados. As linhas tracejadas são os limites de decisão de Bayes. À direita: 20 observações foram geradas a partir de cada classe, e os limites de decisão LDA correspondentes são indicados usando linhas pretas sólidas. Os limites de decisão de Bayes são mais uma vez mostrados com linhas tracejadas.

As taxas de erro de teste para os classificadores Bayes e LDA são 0,0746 e 0,0770, respectivamente.

- As três elipses representam regiões que contêm 95% da probabilidade para cada uma das três classes. As linhas tracejadas são os limites de decisão de Bayes. Em outras palavras, eles representam o conjunto de valores x para os quais δk (x) = δ (x); ou seja Para k = l. (O termo logπk de (4.19) desapareceu porque cada um dos as três classes têm o mesmo número de observações de treinamento; ou seja, πk é o mesmo para cada classe.)

$$\delta_k(x) = x^T \mathbf{\Sigma}^{-1} \mu_k - \frac{1}{2} \mu_k^T \mathbf{\Sigma}^{-1} \mu_k + \log \pi_k \qquad (4.19)$$

$$x^T \mathbf{\Sigma}^{-1} \mu_k - \frac{1}{2} \mu_k^T \mathbf{\Sigma}^{-1} \mu_k = x^T \mathbf{\Sigma}^{-1} \mu_l - \frac{1}{2} \mu_l^T \mathbf{\Sigma}^{-1} \mu_l \qquad (4.20)$$

- O modelo LDA se ajusta aos resultados de 10.000 amostras de treinamento em uma taxa de erro de treinamento de 2,75%.

		True default status		
		No	Yes	Total
Predicted default status	No	9,644	252	9,896
	Yes	23	81	104
	Total	9,667	333	10,000

		True default status		
		No	Yes	Total
Predicted	No	9,432	138	9,570
default status	Yes	235	195	430
	Total	9,667	333	10,000

TABELA 4.5. Uma matriz de confusão compara as previsões do LDA ao verdadeiro padrão status para as 10.000 observações de treinamento no conjunto de dados padrão, usando um valor limite modificado que prevê o padrão para quaisquer indivíduos cuja posterior a probabilidade de inadimplência excede 20%.

- O classificador Bayes funciona atribuindo uma observação à classe para em que a probabilidade posterior pk (X) seja maior. No caso de duas classes, equivale a atribuir uma observação à classe padrão se

$$\Pr(\texttt{default} = \texttt{Yes}|X = x) > 0.5. \tag{4.21}$$

$$P(\texttt{default} = \texttt{Yes}|X = x) > 0.2. \tag{4.22}$$

- Usando um limite de 0,5, como em (4.21), minimiza a taxa de erro geral, mostrada como uma linha sólida preta. Este é esperado, uma vez que o classificador de Bayes usa um limite de 0,5 e é conhecido por ter a menor taxa de erro geral.

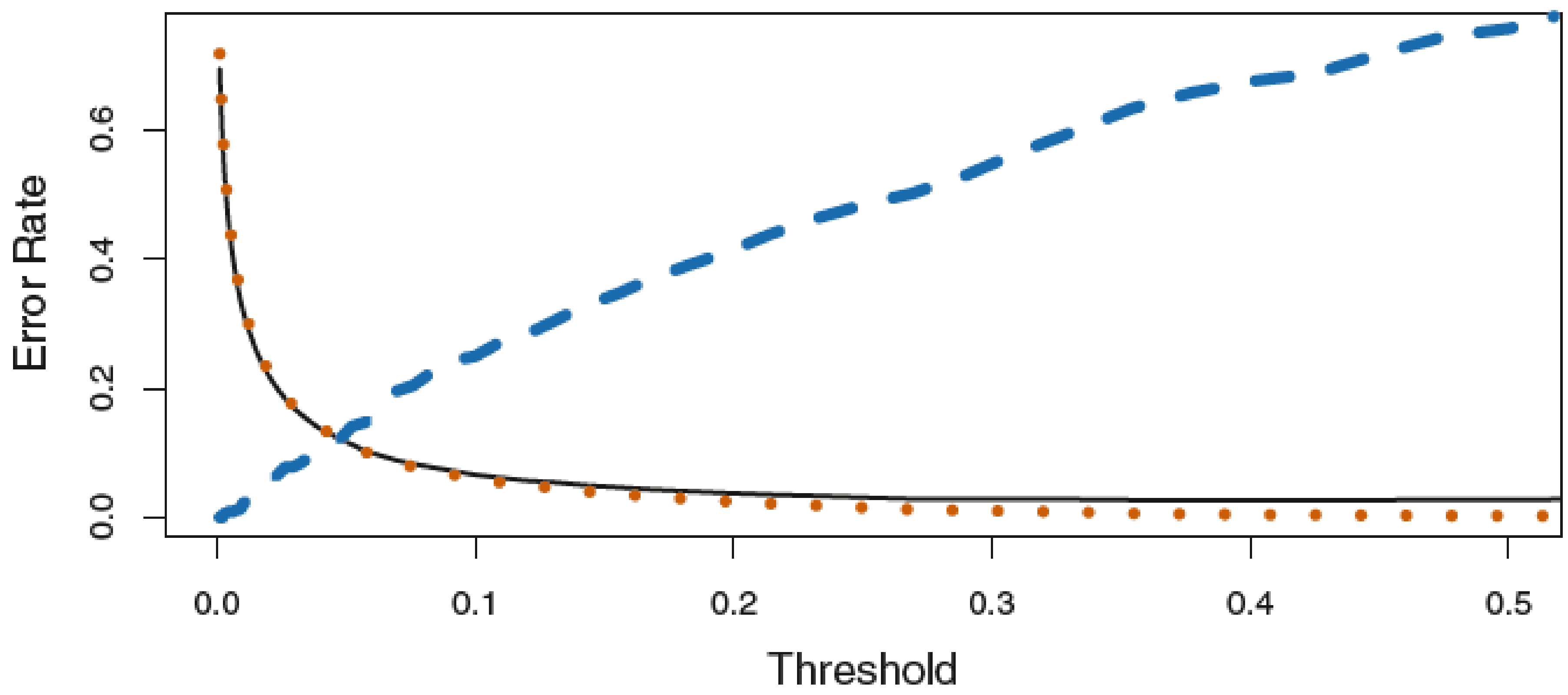


FIGURA 4.7. Para o conjunto de dados padrão, as taxas de erro são mostradas como uma função de valor limite para a probabilidade posterior que é usado para realizar a atribuição.
A linha sólida preta exibe a taxa de erro geral. A linha tracejada azul representa a fração de clientes inadimplentes que são classificados incorretamente, e a linha laranja pontilhada indica a fração de erros entre os não inadimplentes clientes.

A CURVA ROC (Receiver Operating Characteristic)

- A curva ROC é um gráfico popular para exibir dois tipos de erros para todos os limites possíveis. O nome “ROC” é histórico, e vem da teoria da comunicação. É um acrônimo para Receptor de Características Operacionais para avaliar a qualidade de saída do classificador

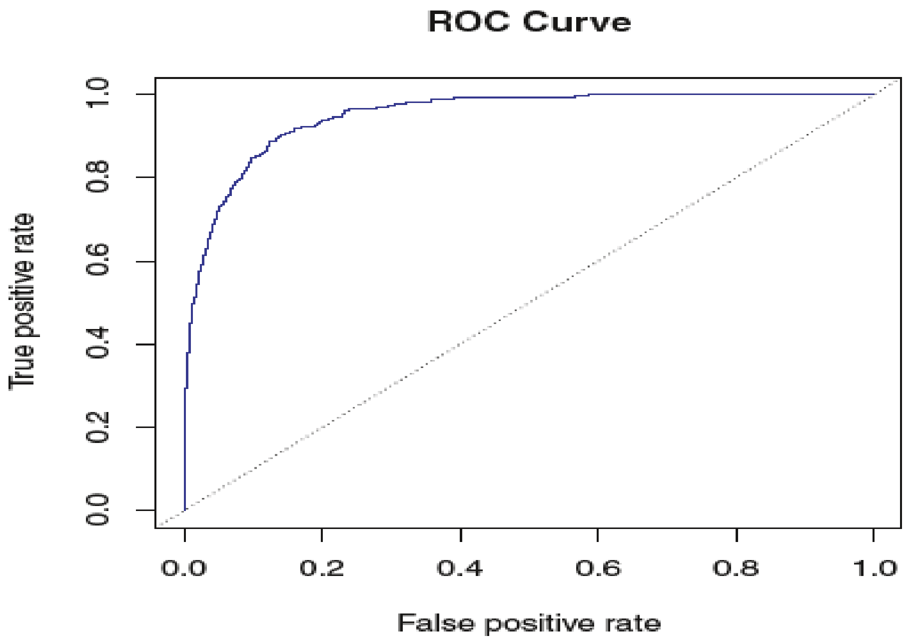


FIGURA 4.8. Uma curva ROC para o classificador LDA nos dados padrão.

4.4.4 Análise Discriminante Quadrática

- A análise discriminante quadrática (QDA) fornece uma alternativa quadrático discriminante.
- Como LDA, o classificador QDA resulta da suposição de que o as observações de cada classe são extraídas de uma distribuição gaussiana, e plugando estimativas para os parâmetros no teorema de Bayes, a fim de realizar predição.

$$
\begin{aligned}
\delta_k(x) &= -\frac{1}{2}(x-\mu_k)^T\mathbf{\Sigma}_k^{-1}(x-\mu_k)+\log\pi_k \\
&= -\frac{1}{2}x^T\mathbf{\Sigma}_k^{-1}x+x^T\mathbf{\Sigma}_k^{-1}\mu_k-\frac{1}{2}\mu_k^T\mathbf{\Sigma}_k^{-1}\mu_k+\log\pi_k \quad (4.23)
\end{aligned}
$$

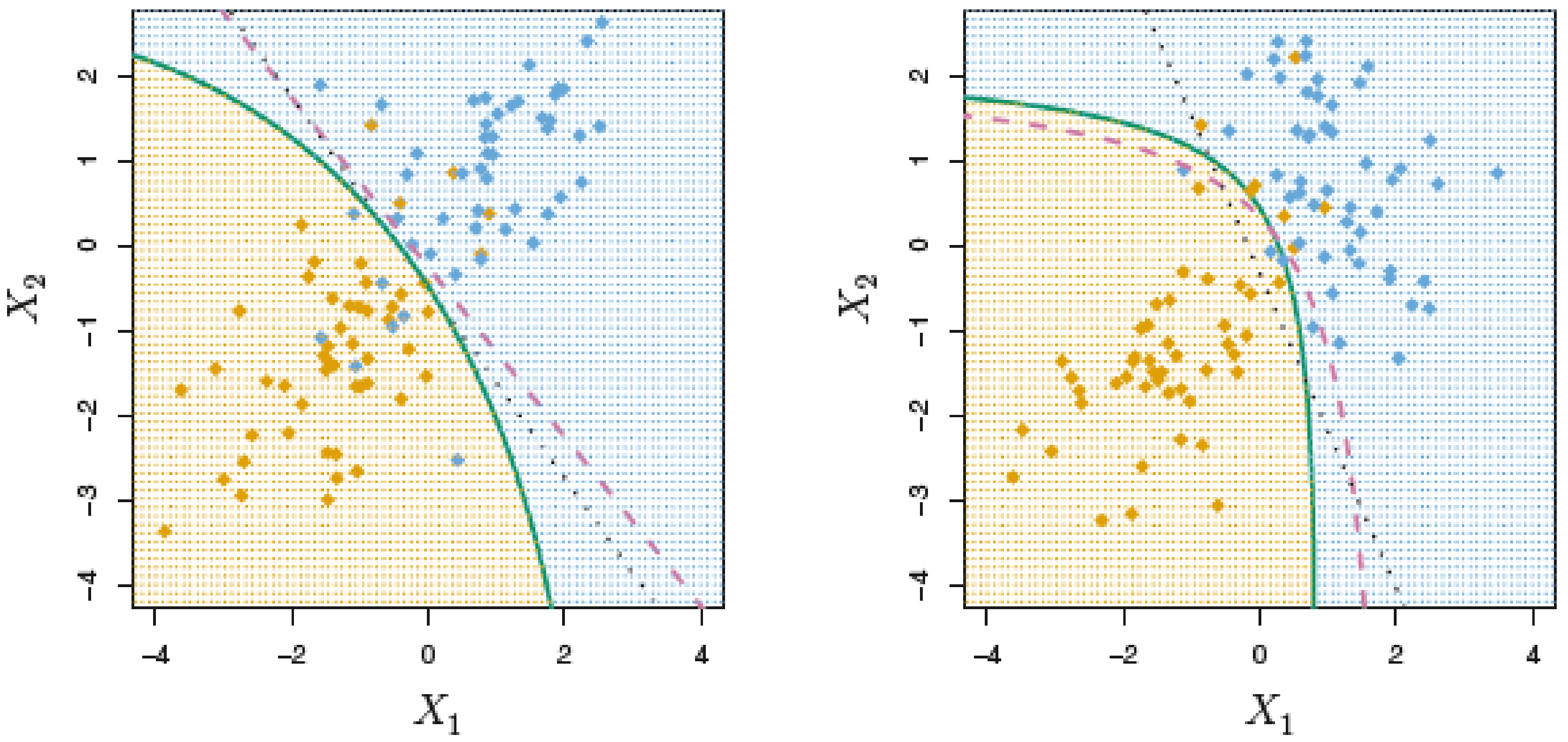


FIGURA 4.9. Esquerda: The Bayes (tracejado roxo), LDA (pontilhado preto) e QDA (sólido verde) limites de decisão para um problema de duas classes com Σ1 = Σ2. O sombreado indica a regra de decisão QDA.

4.5 A Comparison of Classification Methods

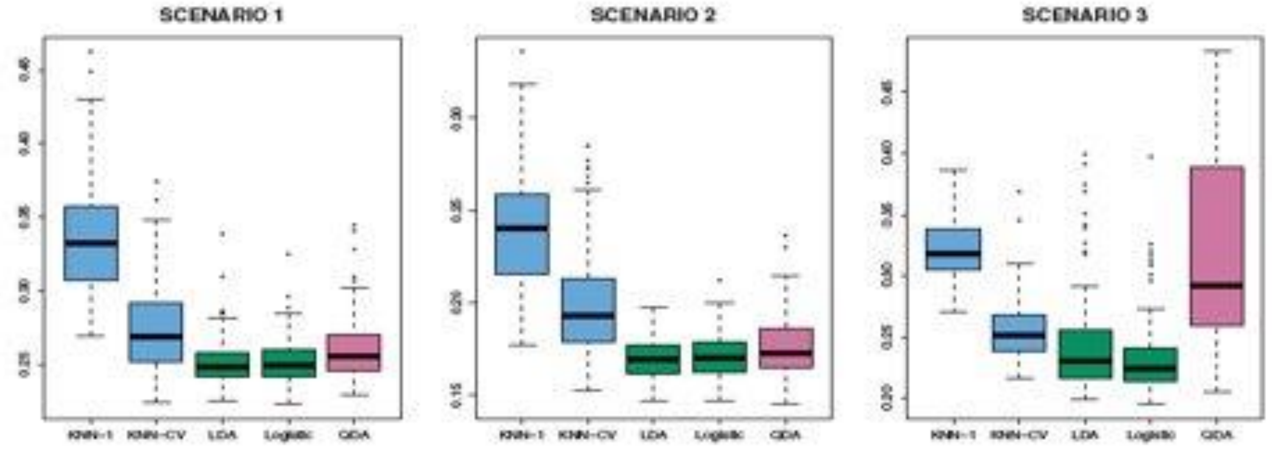


FIGURE 4.10. *Boxplots of the test error rates for each of the linear scenarios described in the main text.*

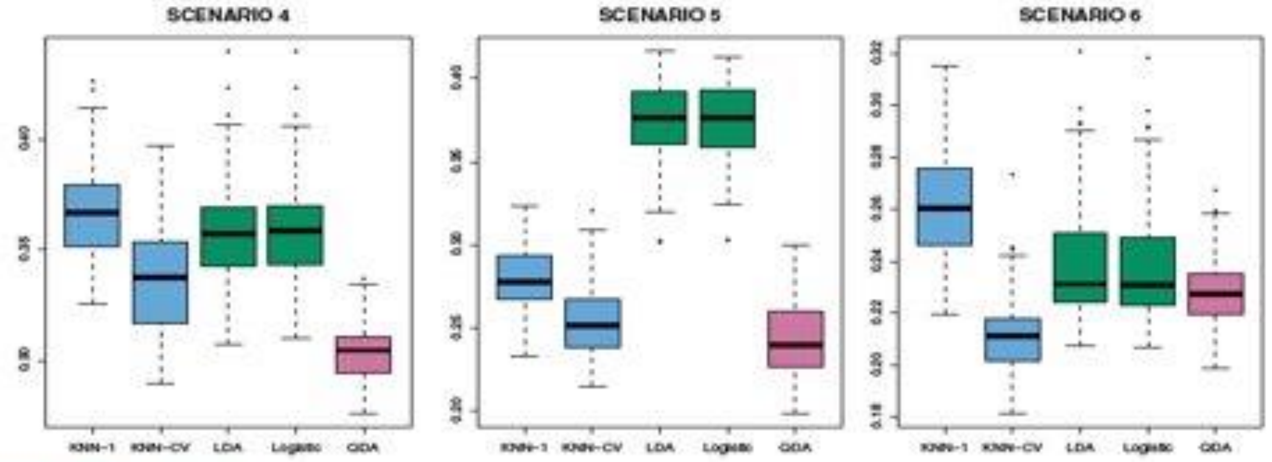


FIGURE 4.11. *Boxplots of the test error rates for each of the non-linear scenarios described in the main text.*

4.6 Lab: Logistic Regression, LDA, QDA, and KNN

Obrigada!!

- Marlete_eng@outlook.com