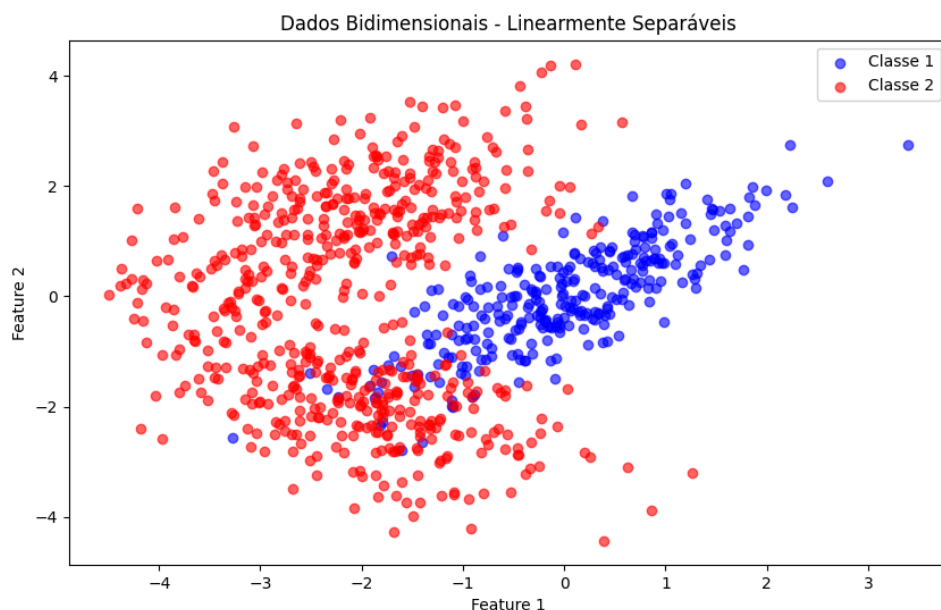


EXERCÍCIO  
TREINAMENTO E COMPARAÇÃO DE CLASSIFICADORES LDA E QDA

ALUNO: FELIPPE VELOSO MARINHO  
MATRÍCULA: 2021072260  
DISCIPLINA: APRENDIZADO DE MÁQUINA

**Gerando os Dados:**

A geração dos dados é feita definindo os valores de médias e matriz de covariância para as duas classes em antecedência. A variável  $n$  define o número de amostras, com a classe 2 tendo o dobro de amostras que a classe 1.



Com isso é feita a divisão de dados em treino e teste com aproximadamente 70% pra treino e 30% para teste e importados os modelos de discriminação quadrática.

**Diferença entre LDA e QDA:**

- **LDA (Discriminação Linear):** Este método assume que cada classe possui uma distribuição normal multivariada com a mesma matriz de covariância para todas as classes, resultando em uma fronteira de decisão linear.
- **QDA (Discriminação Quadrática):** Esse método relaxa a suposição de covariância igual entre as classes, permitindo que cada classe tenha sua própria matriz de covariância.

Então basicamente esperamos que o LDA funcione melhor em conjuntos de dados onde as classes são bem separadas linearmente. Já o QDA, esperamos que funcione melhor em dados onde a relação entre as classes é mais complexa e não-linear.

Através da acurácia podemos prever previsões corretas em relação ao total de previsões realizadas.

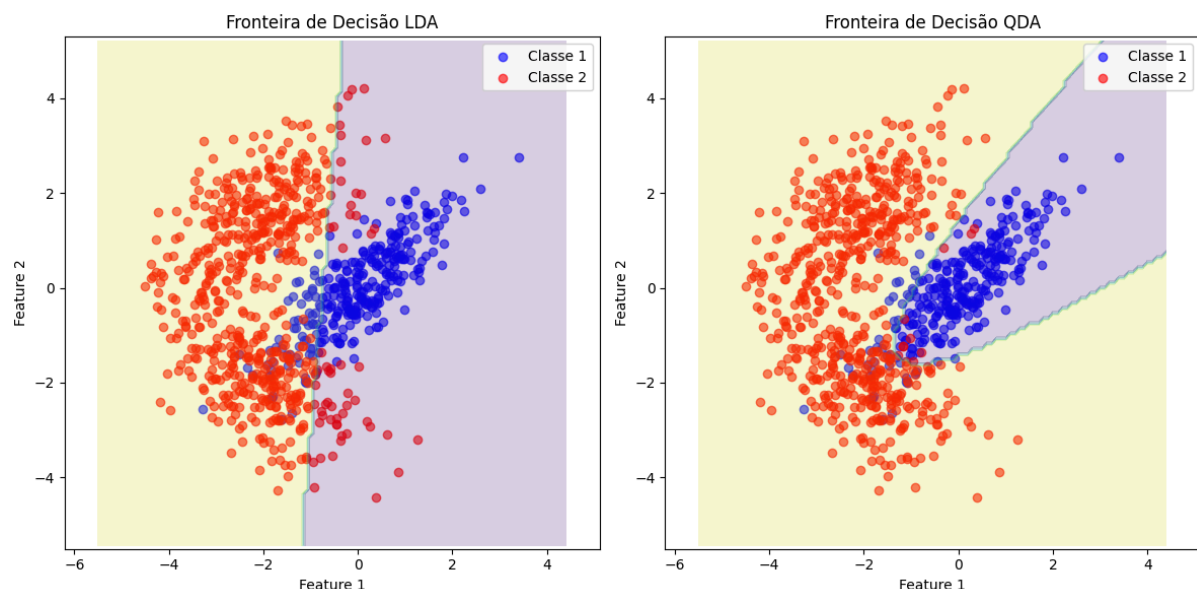
$$\text{Acurácia} = \frac{\text{Número de Previsões Corretas}}{\text{Número Total de Previsões}}$$

As precisões(acurácia) calculadas foram as seguintes:

Precisão LDA: 0.84

Precisão QDA: 0.94

É possível que a acurácia do QDA se apresenta melhor, justamente devido a dispersão dos nossos dados demonstradas nas fronteiras de decisão abaixo:



Além da observação visual podemos questionar a confiabilidade desses nossos preditores. Seria a acurácia uma boa métrica para esse problema ou a existência de classes desbalanceadas poderiam nos criar uma falsa interpretação da análise do nosso conjunto de dados?

Analisando a distribuição de classes e o `classification_report` importado diretamente da `sklearn` vemos que há uma

```
Distribuição de classes no conjunto de teste: {0.0: 83, 1.0: 187}
```

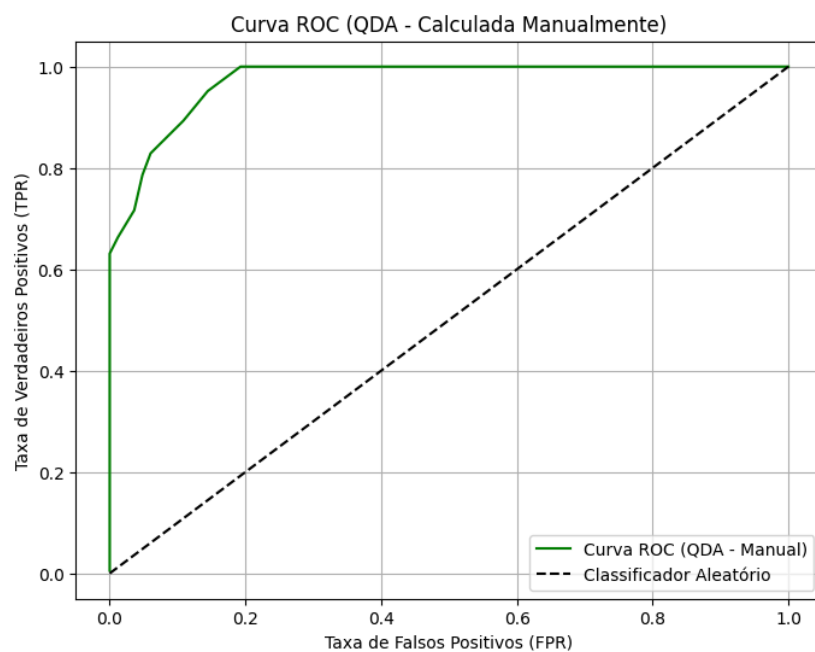
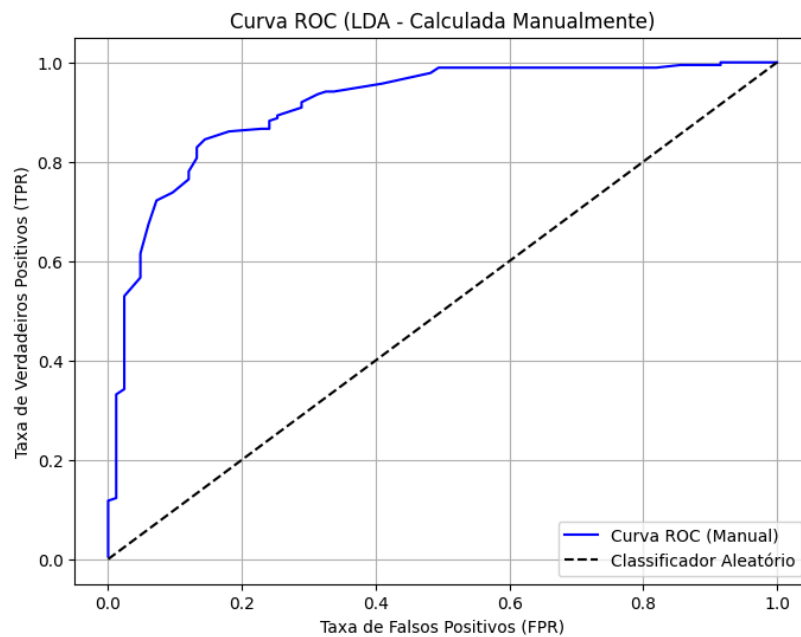
Calculando de acordo com `n`, vemos que temos aproximadamente 27,67% na classe 0 e 62,33% na classe 1. Esse desbalanceamento não é extremo, mas pode impactar o desempenho de modelos que favoreçam a classe majoritária. Ainda assim, ambos os modelos (LDA e QDA) apresentam métricas relativamente equilibradas entre as classes.

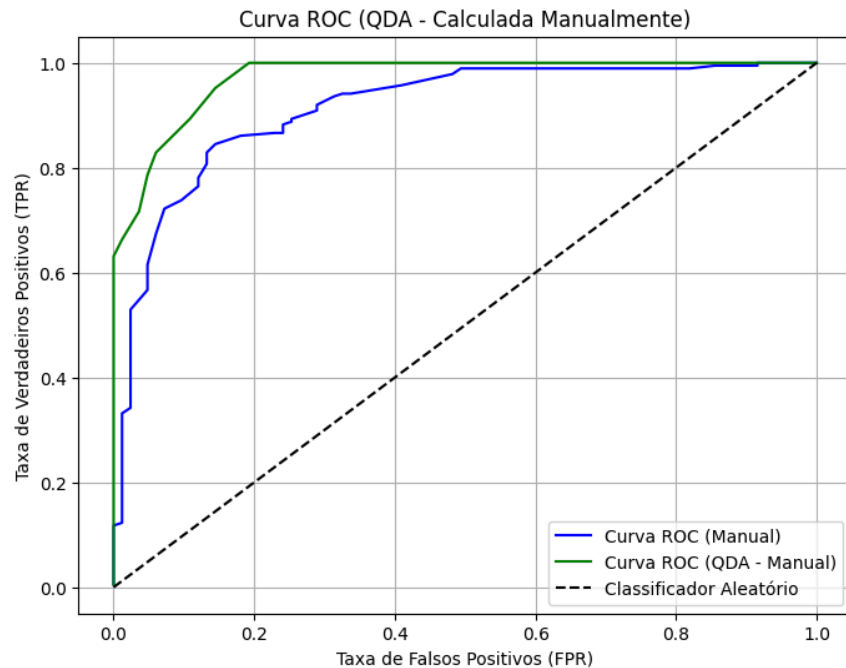
## Análise e Discussão:

Após a análise das fronteiras de decisão e acurácias, foi utilizada a curva ROC.

A **curva ROC** é uma ferramenta de visualização que ilustra a performance de um classificador binário, mostrando a relação entre a **taxa de verdadeiros positivos (TPR)** e a **taxa de falsos positivos (FPR)** em diferentes limiares de decisão.

- **Taxa de Verdadeiros Positivos (TPR):** Indica a proporção de exemplos da classe positiva que foram corretamente classificados.
- **Taxa de Falsos Positivos (FPR):** Indica a proporção de exemplos da classe negativa que foram incorretamente classificados como positivos.





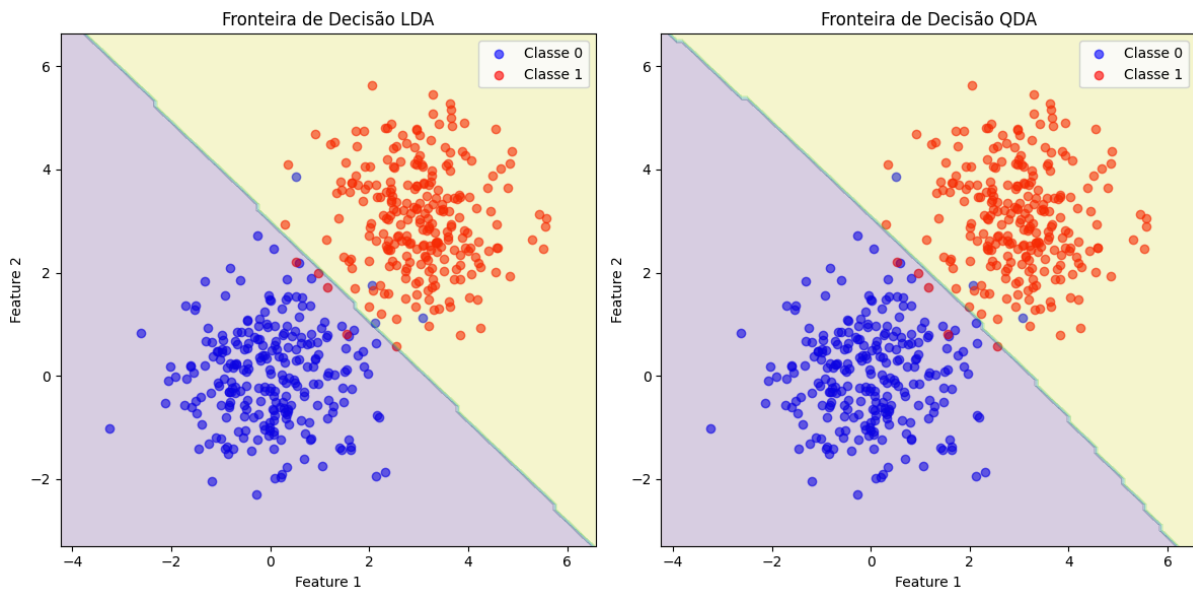
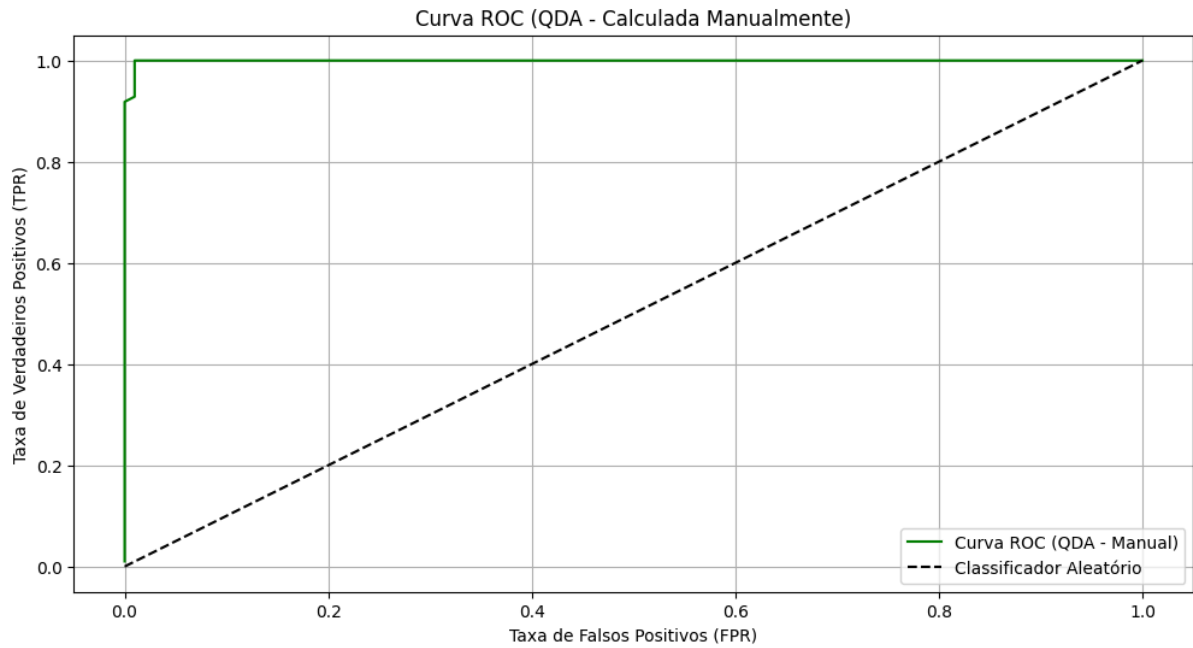
Ao observar a **curva ROC e a AUC** (área sob a curva ROC), geralmente o QDA apresentará uma AUC maior que o LDA em cenários onde os dados não são linearmente separáveis, indicando uma melhor **capacidade de discriminação**.

O QDA parece melhor ajustado aos dados devido à sua capacidade de modelar fronteiras de decisão não lineares. Como ele permite a criação de contornos de decisão quadráticos, o QDA é capaz de capturar padrões mais complexos na distribuição dos dados, representando mais fielmente as classes.

Já o LDA, com suas fronteiras lineares, pode subestimar a complexidade dos dados, sendo mais adequado para casos onde as classes são linearmente separáveis. Como o QDA permite maior flexibilidade, ele tende a se ajustar melhor, especialmente quando os dados mostram uma separação não linear.

### **Alterando o centro da classe**

Ao ajustar o centro da classe 1 para (3,3) temos uma acurácia muito maior



A acurácia do LDA foi de 1.0, enquanto a do QDA ficou em 0.98. Isso indica que ambos os classificadores tiveram excelente desempenho. Isso se deve a nova organização do nosso conjunto de dados que apresentou uma melhor separação de maneira linear.

Observando as curvas ROC, ambas mostram excelente desempenho, com altas taxas de verdadeiros positivos e baixas taxas de falsos positivos ao longo dos thresholds.

Portanto, a mudança no centro da classe 1 torna os dados mais bem distribuídos, favorecendo classificações tanto lineares (LDA) quanto quadráticas (QDA). Isso reflete-se na alta acurácia dos dois modelos. Com isso, o LDA parece ser o mais adequado pela sua simplicidade e precisão, enquanto o QDA ainda funciona muito bem, embora possa ser um pouco mais sensível a mudanças na variância e covariância dos dados.