

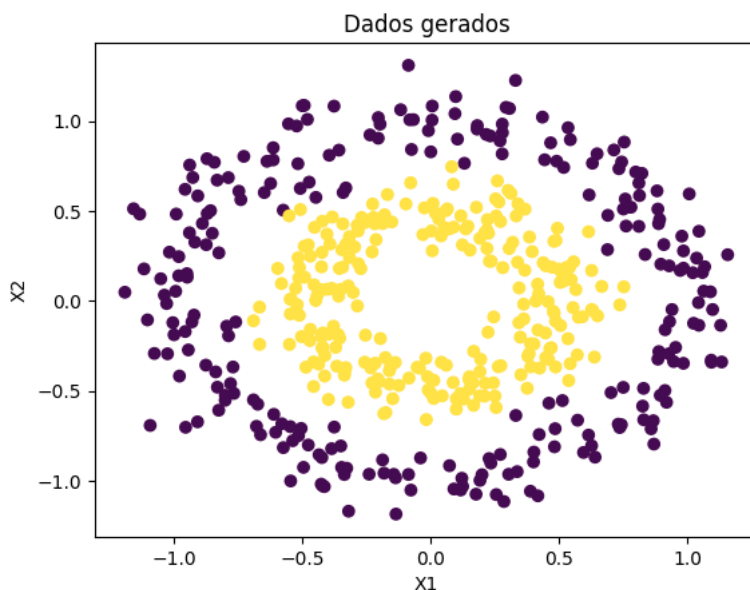
**Exercício 1.** Investigue as referências mais importantes sobre as técnicas baseadas em árvores de decisão e discuta as principais similaridades e diferenças das árvores de classificação, bagging, boosting e florestas aleatórias.

**Exercício 2.** Elabore um estudo para comparar o tempo computacional em um problema de classificação com  $k$  classes de árvores de classificação, bagging, boosting e florestas aleatórias.

**Exercício 3.** Qual a diferença entre os métodos Naive Bayes e análise discriminante linear? Elabore um estudo para comparar as previsões obtidas em um problema de classificação com  $k$  classes utilizando um ou outro método. Compare os tempos computacionais utilizando pacotes prontos em Python ou R.

**Exercício 4.** Sobre as máquinas de suporte vetorial (*support vector machine* ou SVM), discuta características como facilidade de implementação, tempo computacional, eficiência em grandes bases de dados, necessidade do uso de funções *kernel*, suposições de distribuição de probabilidade, uso para regressão ou classificação e outros aspectos que julgar necessário.

**Exercício 5.** Considere a figura a seguir, obtida de dados simulados em que as classes estão representadas em roxo e amarelo, assinale todas as alternativas corretas e justifique a sua escolha para cada item.



- (1) É possível afirmar previamente à modelagem que existe um único kernel adequado para este problema de classificação.
- (2) O problema não pode ser resolvido utilizando máquinas de suporte vetorial.
- (3) O SVM não-linear via kernels aumenta a dimensão do espaço dos dados por meio de funções das preditoras e possibilita a busca por um hiperplano separador no espaço de dimensões mais altas.
- (4) É possível utilizar SVM para obter um classificador de margem não linear, utilizando funções kernel.