

Notas:

- Este documento inclui alguns exemplos de perguntas que poderão sair nos testes da UC de Aprendizagem Automática (2023/2024).
- É de notar que nem todos os tipos de perguntas aqui representadas poderão aparecer nos testes.
- O docente reserva-se no direito de também incluir tipos de perguntas não representadas aqui.
- A quantidade de perguntas neste documento não é representativo do que irá acontecer nos testes.

Tipo 1: Verdadeiro e Falso com ou sem Justificação

1.) Para duas execuções do algoritmo de clustering K-Means, é esperado obter os mesmos resultados? Justifica.



a) Verdadeiro

b) Falso

Tipo 2: Escolha Múltipla

2.) O que acontece quando se diminui a dimensionalidade de um dataset usando o PCA?

- 1) As features irão manter interpretabilidade;
- 2) As features irão perder alguma interpretabilidade;
- 3) As features irão conter toda a informação dos dados;
- 4) As features irão perder alguma informação dos dados.

a) 1 e 3

b) 1 e 4

c) 2 e 3

d) 2 e 4

Tipo 3: Avaliação de desenho experimental

3.) Para cada uma das descrições listadas abaixo, circula se o desenho experimental é adequado ou problemático. Se considerar que é problemático, indique brevemente os problemas com a abordagem.

3.1) Uma equipa de trabalho reporta um erro de treino baixo e afirma que a sua abordagem é boa.

a) Adequado

b) Problemático



Tipo 4: Circular respostas certas

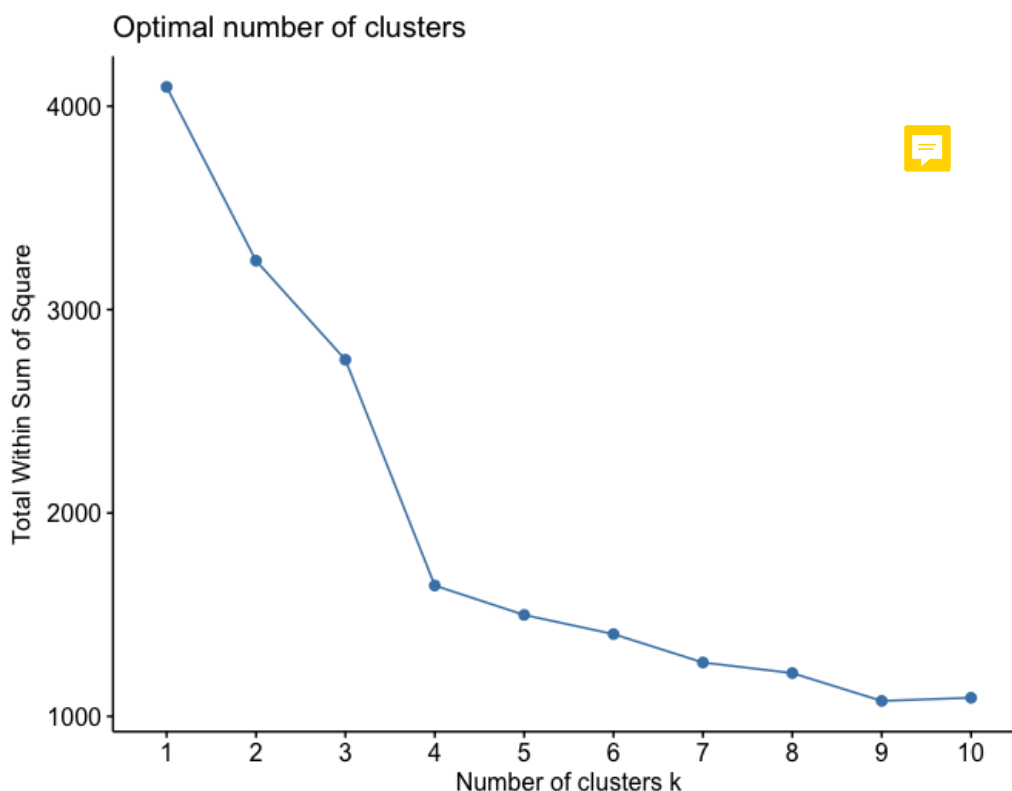
4.) Foram estudados vários métodos para controlar o overfitting para diversos classificadores. Abaixo, encontram-se listados vários classificadores e ações que podem afetar o seu bias e variância. Indique (circulando) como o bias e a variância mudam em resposta à ação:

4.1) Reduzir a profundidade (depth) máxima de uma árvore de decisão:

Bias	Variância
Diminuir	Diminuir
Aumentar	Aumentar
Permanecer inalterado	Permanecer inalterado

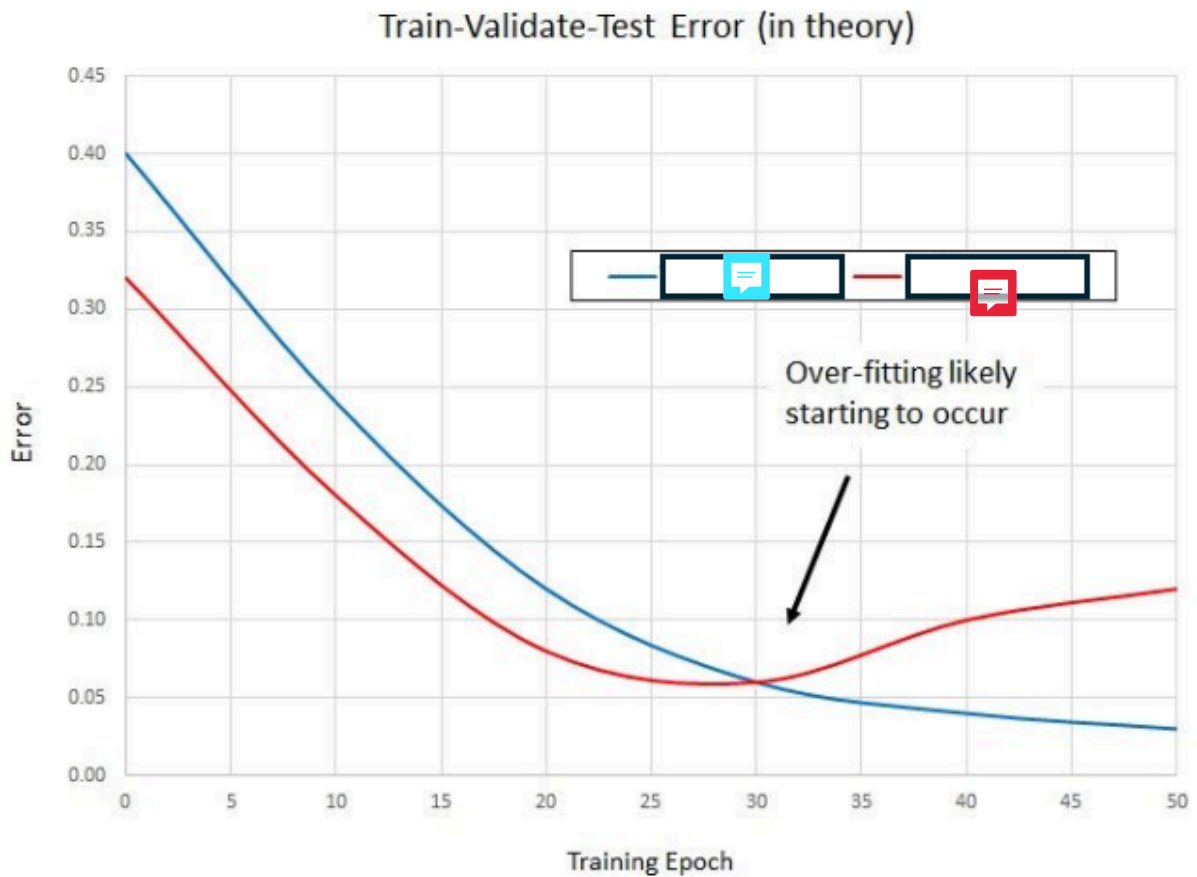
Tipo 5: Análise de Imagens/Gráficos

5.) Dado o seguinte knee plot, qual seria o numero de clusters (k) a usar no algoritmo de clustering k-means?



Tipo 6: Legendar figuras

6.) Legenda a figura seguinte usando: “training error” e “validation error”.



Tipo 7: Perguntas de desenvolvimento

7.) Como lidar com valores omissos (missing values) num conjunto de dados antes de treinar um modelo de machine learning? Descreva pelo menos duas técnicas comuns de tratamento de valores omissos e discuta suas vantagens e desvantagens.

Tipo 8: Cálculo matemático

8.) Dada a seguinte matriz de confusão calcula a accuracy do modelo.

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Tipo 9: Analisar trechos de código

9.) O trecho de código seguinte apresenta vários problemas/erros. Tenta identificar o maior número possível sabendo que existem no total 4 problemas/erros.

```
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

# Gerar dados de exemplo
X = np.array([[1, 2], [3, 4], [5, 6], [7, 8]])
y = np.array([7.2, 6.3, 6.9, 5.8, 5.2, 4.9])

# Escalar os dados
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Separar os dados em treino e teste
X_train, y_test, X_test, y_train = train_test_split(
    *arrays: X_scaled, y, test_size=0.2, random_state=42)

# Criar modelo de regressão logística
model = LogisticRegression()

# Treinar o modelo
model.fit(X_train, y_train)

# Fazer previsões
predictions = model.predict(X_test)

# Calcular a accuracy
acc = accuracy_score(y_test, predictions)
print("Accuracy:", acc)
```