



Nome do Aluno: Luan Damato		T.I.A.: 31817051
Nota:	Visto:	

Questão 01. (1,5 ponto) considere o algoritmo k-means exibido abaixo:

1. Selecione aleatoriamente k centros de grupos
2. Repita até convergir:
 - a. *Passo-E*: associe os pontos ao centro de grupo mais próximo
 - b. *Passo-M*: atualize o centro do grupo como o ponto médio dos seus membros

O passo 2 do algoritmo apresenta um comando de repetição que será executado até que se atinja uma situação de convergência. A convergência, e o final das execuções do comando repetitivo, indica que foi obtida a solução ótima do problema sendo resolvido pelo algoritmo? Justifique cuidadosamente sua resposta.

Não, pois pode haver uma convergência não ideal devido ao posicionamento incorreto dos centros, isto é, como os k centros são escolhidos aleatoriamente, o algoritmo separa os dados conforme a distância deles com esse centro, podendo não corresponder a um bom centro de distribuição de dados.

Questão 02. (1,5 ponto) Quais são os três tipos gerais de aprendizagem de máquina? Descreva a ideia geral de cada um deles.

- **Aprendizagem supervisionada:** Tem como objetivo principal a aprendizagem de um modelo a partir de um conjunto de exemplos rotulados com informações relevantes a seu respeito.

- **Aprendizagem não-supervisionada:** Tem como objetivo principal a aprendizagem de um modelo sem o auxílio um conjunto de exemplos rotulados.

- **Aprendizagem por reforço:** Tem como objetivo principal a adaptação da máquina em diferentes cenários através de um sistema de recompensa e punição, baseados de acordo com a resposta retornada e suas circunstâncias.

Questão 03. (1,0 pontos) descreva, com detalhes, o conceito de dimensão de dados, que se trata de um elemento importante a ser considerado na aprendizagem de máquina.

Dimensões são categorias de informações, um conjunto de características (features) pelos quais você pode agrupar os dados manipulados. Como exemplo, em uma base de dados de celulares temos marca, modelo e cor constituem a dimensão desses dados.

Questão 04. (3,0 pontos) Descreva brevemente cada uma das tarefas de aprendizagem de máquina apresentadas abaixo. Para cada uma delas, apresente um problema prático específico que pode ser resolvido com ela e como a tarefa pode ser usada para resolvê-lo.

a) Agrupamento

Tipo de aprendizagem não-supervisionada, usada para organizar um conjunto de dados em subgrupos (clusters). Cada é um conjunto de dados com semelhanças e que são, de certa forma, distintos dos dados pertencentes a outros grupos. Exemplo: classificação de clientes com base nos seus interesses para direcionar estratégias de marketing personalizadas para cada grupo.

b) Classificação

Tipo de aprendizagem supervisionada com objetivo de rotular novas instâncias, sem uma ordem definida, usando como base observações passadas. No exemplo do filtro de e-mails, temos um caso de classificação binária.

c) Redução de dimensionalidade

Tipo de aprendizagem não-supervisionada com objetivo de eliminar ruído sem perder a qualidade dos dados manipulados, mantendo suas características relevantes. Um exemplo é dados meteorológicos.

Questão 05. (2,0 pontos) na construção de uma árvore de decisão, podemos optar pela altura máxima que a árvore poderá ter. No entanto, é possível que a árvore obtida não atinja a altura especificada. Descreva brevemente o que pode causar esta situação.

Quando a altura máxima de uma árvore de decisão não for atingida, o índice de gini não pode ser diminuído, assim, se houver qualquer aumento da altura retorna o mesmo resultado e, como consequência, não vai melhorar a classificação e/ou qualidade da árvore, retornando impurezas.

Questão 06. (1,0 ponto) Descreva, brevemente e com suas próprias palavras, a diferença entre a inteligência artificial "clássica" e a aprendizagem de máquina.

A inteligência artificial "clássica" busca simular a inteligência humana, e, que sistema tem um comportamento "inteligente" a partir de dados anteriores, já na aprendizagem de máquina, buscamos a capacidade de se adaptar, modificar e melhorar seu comportamento e suas respostas sempre, não estagnando em um ponto e fazendo com que novos dados em conjunto com os já associados gerem novos aprendizados.