

1 K-means

1. Comente sobre algumas aplicações de algoritmos de agrupamento.
2. Descreva intuitivamente/graficamente o funcionamento do algoritmo de K-means.
3. Apresente o pseudo-código dos passos do K-means.
4. Descreva matematicamente a função de custo do K-means.
5. Explique por que o algoritmo de K-means é do tipo “coordenada descendente”.
6. A função de custo do K-means é convexa? Se não, descreva um procedimento para amenizar o problema com mínimos locais.
7. Com se dá a escolha do número ideal de grupos pelo método do “cotovelo”?
8. Qual a prática mais comum quando nenhum exemplo é atribuído a um determinado grupo (além possivelmente do próprio centroide)?

2 PCA

1. Quais os usos mais comuns para algoritmos de redução de dimensionalidade?
2. Descreva com suas palavras a intuição usada por PCA para reduzir a dimensionalidade do dado.
3. Quais as semelhanças e diferenças entre PCA e regressão linear?
4. Apresente, usando pseudo-código e fórmulas matemáticas, os passos do algoritmo de PCA.
5. Qual é o problema de otimização por trás do algoritmo de PCA? Como ele se relaciona com auto-vetores?
6. Qual a diferença entre a projeção e a reconstrução do dado original? Ilustre graficamente e mostre a fórmula para o cálculo de cada uma delas.
7. Como se define o erro de projeção e como ele é usado para obter o número de componentes principais? Como a matriz S da decomposição SVD pode ser usada para este mesmo propósito?
8. Como PCA pode ser combinado com métodos supervisionados e em que circunstâncias isso é interessante? Quais cuidados devem ser tomados para que se evite o vazamento de dados do treino para o teste nesta situação?
9. PCA é recomendado na prevenção de *overfitting*? Discuta a respeito.