

# Segundo teste de AA / AA second test.

## Pergunta 1; Question 1

### Português

**Atenção:** a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

Considere a equação abaixo, relacionando, com probabilidade 1 menos delta, o limite superior do erro verdadeiro da hipótese de menor erro empírico com o seu erro empírico e um termo adicional cuja magnitude é da ordem da expressão indicada, em função da dimensão Vapnik-Chervonenkis, delta e o tamanho do conjunto de treino

$$E(\hat{h}) \leq \hat{E}(\hat{h}) + \mathcal{O} \left( \sqrt{\frac{VC(\mathcal{H})}{m}} \ln \frac{m}{VC(\mathcal{H})} + \frac{1}{m} \ln \frac{1}{\delta} \right)$$

Assumindo um classificador linear, explique como esta equação nos permite prever o efeito que uma expansão não linear dos atributos para um espaço com mais dimensões vai afectar o sobreajustamento (*overfitting*). E se a transformação para um espaço com mais dimensões for linear?

Deve focar os seguintes aspectos por ordem decrescente de prioridade:

- A relação entre esta equação e o sobreajustamento (*overfitting*)
- A relação entre a expansão não linear dos atributos e esta equação.
- O que a equação nos permite prever no caso da expansão não linear dos atributos.
- O que a equação nos permite prever no caso desta expansão ser linear.

### English:

**Attention:** your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Consider the equation below, relating, with a probability of  $1 - \delta$ , the upper bound for the true error of the hypothesis with the lowest empirical error to its empirical error plus a term whose magnitude is on the order of the given expression, depending on the Vapnik-Chervonenkis dimension,  $\delta$  and the size of the training set:

$$E(\hat{h}) \leq \hat{E}(\hat{h}) + \mathcal{O} \left( \sqrt{\frac{VC(\mathcal{H})}{m} \ln \frac{m}{VC(\mathcal{H})} + \frac{1}{m} \ln \frac{1}{\delta}} \right)$$

Assuming a linear classifier, explain how this equation allows us to predict the effect a nonlinear expansion of the attributes to a higher dimensional space will have on overfitting. And what if the expansion is done with a linear transformation?

You should focus the following aspects in decreasing order of importance:

- How this equation relates to overfitting
- The relation between the nonlinear expansion of the attributes and this equation.
- What the equation allows us to predict in the case of the nonlinear expansion of the attributes.
- What the equation allows us to predict if this expansion is linear.

## Pergunta 2; Question 2

### Português

**Atenção:** a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

Considere os algoritmos de Análise de Componente Principal (*Principal Component Analysis*, PCA) e t-SNE (*t-distributed stochastic neighbor embedding*). Qual dos dois seria melhor para visualizar a duas dimensões um conjunto de dados com muitos atributos? Justifique a sua resposta focando os seguintes aspectos por ordem decrescente de prioridade:

- Explicar em traços gerais como funciona cada algoritmo.
- Explicar resumidamente o que cada um preserva da estrutura dos dados originais.
- Explicar como isso torna um melhor que o outro para visualizar dados a duas dimensões.

### English:

**Attention:** your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Consider the Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) algorithms. Which of these two would be better to visualize in two dimensions a data set with many attributes? Justify your answer covering these aspects in decreasing order of priorities:

- Explain in general terms how each algorithm works.
- Briefly explain what aspects of the structure of the original data each of these algorithms best preserves.
- Explain how this makes one better than the other for visualizing the data in two dimensions.

### Pergunta 3; Question 3

#### Português

**Atenção:** a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

Quantização vectorial é o processo de converter vectores de valores contínuos num número mais pequeno e pré-determinado de vectores representantes do conjunto. Explique como isso pode ser feito com aprendizagem não supervisionada explicando as propriedades do algoritmo que utilizaria e explicando que valores atribuiria ao parâmetro ou parâmetros que tivesse de determinar manualmente. Explique também se podemos, ou não, obter a quantização vectorial de um conjunto de dados usando aglomeração baseada em densidade.

Deve focar os seguintes aspectos por ordem decrescente de prioridade:

- Explicar o algoritmo escolhido para esta tarefa.
- Explicar como é que este algoritmo permite fazer quantização vectorial.
- Explicar como determinar o ou os parâmetros que tenham de ser decididos manualmente.
- Explicar porque é que a aglomeração baseada em densidade é, ou não é, adequada neste caso.

#### English:

**Attention:** your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Vector quantization is the process of converting vectors with continuous values in a smaller and pre-determined number of vectors representative of the data set. Explain how this can be done with unsupervised learning, explaining the algorithm you would use and how to choose the parameter values you would need to specify manually. Finally, explain whether density-based clustering would, or would not, be a good way to solve the vector quantization problem.

You should focus these aspects in decreasing order of priorities:

- Explain the algorithm you chose for this task.
- Explain how this algorithm allows us to compute a vector quantization.
- Explain how to determine the parameter or parameters that must be manually decided.
- Explain why density based clustering is, or is not, appropriate for this problem.

## Pergunta 4; Question 4

### Português

**Atenção:** a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

Três metralhadoras foram testadas apontando-as para o mesmo alvo e dando aproximadamente cem tiros com cada uma. No final do teste, no alvo e na parede circundante temos cerca de 300 marcas de balas. Considere que as balas saem de cada metralhadora com alguma aleatoriedade numa distribuição aproximadamente normal, que a dispersão e o ponto médio destas distribuições podem ser diferentes para as 3 metralhadoras e que não sabemos à partida a que metralhadora cada bala pertence nem o número exacto de balas disparadas.

Explique que algoritmo usaria para estimar o ponto central da distribuição de balas e a correspondente dispersão para os três conjuntos de balas. Explique também que problemas haveria se tentássemos fazer isto com o algoritmo K-means.

Deve focar os seguintes aspectos por ordem decrescente de prioridade:

- Explicar em traços gerais o algoritmo que escolheu.
- Justificar a escolha desse algoritmo relacionando-o com as características do problema.
- Explicar como é que obteria, no final, os três valores de ponto médio e variância para os três conjuntos de balas.
- Explicar que problemas surgiriam se tentássemos calcular isto com o algoritmo K-means
- Com base nestes resultados, conseguia saber a que metralhadora correspondia cada valor de ponto médio e variância obtido? Justifique a resposta.

### English:

**Attention:** your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Three machine guns were tested aiming them at the same target and shooting approximately one hundred rounds with each machine gun. Assume that the bullets leave each machine gun with a partly random trajectory with an approximately normal distribution, that the mean point and variance of these distributions may differ between the three guns and that we do not know which bullet hole corresponds to each gun nor the exact number of rounds fired by each gun.

Explain which algorithm you would use to estimate the mean point and and variance of the distributions for the three sets of bullets. Also explain what problems would arise if we tried to solve this with the K-means algorithm.

You should focus these aspects in decreasing order of priorities:

- Explain in general terms the algorithm you selected.
- Justify your choice relating the algorithm with the characteristics of this problem.
- Explain how you would obtain the three sets of mean point and variance for the three sets of bullets.
- Explain what problems would arise if you tried to do this with the K-means algorithm.
- From these results, can you tell which machine gun corresponds to each set of mean point and variance obtained? Justify your answer.

## Pergunta 5; Question 5

### Português

**Atenção:** a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

*Expectation maximization* é uma abordagem muito usada em problemas de aprendizagem não supervisionada. Explique esta abordagem e como é usada no algoritmo de Baum-Welch focando os seguintes aspectos por ordem decrescente de prioridade:

- Brevemente, explique em que consiste esta abordagem.
- Explique que tipo de problemas ajuda a resolver, e porquê.
- Explique como o algoritmo de Baum-Welch recorre a esta abordagem de *Expectation maximization* para treinar modelos ocultos de Markov.

### English:

**Attention:** your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Expectation maximization is a useful approach in unsupervised learning. Explain this approach and how it is used in the Baum-Welch algorithm, focusing the following aspects by decreasing order of priority:

- Briefly explain this approach.
- Explain the type of problems it helps solve and why.
- Explain how the Baum-Welch algorithm uses this approach of Expectation maximization to train hidden Markov models.