
Lista 2 - Exercícios Geral:

Prof. Dr. Jodavid Ferreira

Discente:

* Gabriel D'assumpção de Carvalho

Data: 27/07/2024

1. **Quais são os três principais conceitos do Machine Learning e como funciona cada um deles?**

A. Aprendizado Supervisionado: O aprendizado supervisionado é um processo de aprendizado em que os dados estão no formato estruturado, portanto se tem o conhecimento da característica de todas as variáveis de entrada (X) e de saída (y). Portanto, o modelo vai ser treinado com os dados de entrada e saída $D = \{X, y\}$ e o algoritmo vai tentar encontrar uma função em $g \sim f(X, y)$

B. Aprendizado Não Supervisionado: Diferente do modelo supervisionado, no não supervisionado os dados de entrada e saída não são estruturados, por tanto a ideia do modelo é criar grupos das observações com base na semelhança entre os dados. Entretanto, o algoritmo no caso do não supervisionado vai tentar agrupar os dados de entrada X (features) em grupos (clusters).

C. Aprendizado Por Reforço: Já no aprendizado por reforço se tem um modelo onde vamos ter uma função objetivo e a cada iteração do modelo com o ambiente pode se ter uma recompensa ou perda na função objetivo.

2. **O que é um modelo de Machine Learning?** O modelo de Machine Learning é um algoritmo que tem a capacidade de aprender as relações entre os dados, então se temos uma grande quantidade de dados o algoritmo tem uma grande capacidade de aprender o padrão existente entre eles.

3. **O que é um conjunto de treinamento?** O conjunto de treinamento é um conjunto de dados onde o modelo vai ser treinado e o algoritmo vai tentar estimar uma função onde com a inputação dos features (X) o algoritmo vai nos retornar a saída (y) no caso dos modelos supervisionados.

4. **O que é um conjunto de teste?** O conjunto de teste é a base de dados que vai avaliar o desempenho do modelo com base na estimativa realizada através do conjunto de

treinamento, essa separação é feita para que o modelo se torne mais preciso e tenha uma maior capacidade de generalizar, sendo capaz de prever novos dados.

5. **Considerando que estamos trabalhando com dados estruturados, como é a estrutura de um conjunto de dados para um estudo de Análise Supervisionada?** Cite exemplos de problemas que podem ser aplicados Análise Supervisionada. Para um dados estruturados o conjunto de dados vão ser separado em features (**X**) e targets (**y**), onde os features pode ser idade, peso, altura, raça, circunferência da cintura e etc, e o target pode ser a probabilidade da pessoa ter diabetes ou nao.

6. **Quais são as etapas do Processo de Aprendizagem de um Modelo?** O processo de aprendizado deve ter as seguintes etapas:

- A. Coleta dos Dados
- B. Analise Exploratoria dos Dados
- C. Preparação dos Dados
- D. Separação dos Dados de Treinamento e Teste ou Treinamento, Validação e Teste.
- E. Treinamento do Algoritmo ou Treinamento e Validação do Algoritmo
- F. Teste do Algoritmo
- G. Avaliação do Modelo

7. **O que é Underfitting e Overfitting?** Como podemos identificar esses problemas em um modelo de Machine Learning?

- **Underfitting:** É o nome que se da quando o modelo não tem a capacidade de compreender o padrão ou relação existente aos dados, sendo um modelo muito simples.
- **Overfitting:** Ao contrario do under, o overfitting se da quando o algoritmo é bem complexo e ele aprende demasiadamente o padrão dos dados do conjunto de treinamento e perde a capacidade de generalizar para novos dados.

8. **Como podemos definir o erro total de previsão de um modelo?** O erro total da previsão de um modelo pode ser dado por meio da seguintes estatísticas:

A. **Distância Euclidiana:**

$$d(y, \hat{y}) = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

B. **Erro Absoluto Total:**

$$EAT = \sum_{i=1}^n |y_i - \hat{y}_i|$$

C. **Erro Absoluto Médio:**

$$EAM = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

D. **Média do Resíduo ao Quadrado:**

$$\text{MSR} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

9. Qual a regra geral de um bom modelo?
10. Cite duas métricas e coloque a equação descrevendo as variáveis de solução que podemos utilizar para avaliar quantitativamente o desempenho do modelo de Regressão?
11. Cite duas métricas e coloque a equação descrevendo as variáveis de solução que podemos utilizar para avaliar quantitativamente o desempenho do modelo de Classificação?