**Lista de Exercícios – Introdução ao Aprendizado de Máquina Departamento de Computação - Universidade Federal de São Carlos**

1 – Quais são os dois tipos principais de aprendizado na área de Aprendizado de Máquina?

**Os dois tipos principais de aprendizado na área de Aprendizado de Máquina é o aprendizado Supervisionado e Não Supervisionado.**

2 – O que é *overfitting* e como ele pode ser evitado?

**Overfitting é quando o modelo utilizado está tão complexo e conhece tão bem os dados que adapta seu resultado para coincidir com os resultados, e isto é maléfico para o algoritmo, visto que não permite que novos dados sejam inseridos e resultados corretos sejam obtidos a partir do modelo. Para evitar, deve-se construir um algoritmo que atinja um ponto óptimo de complexidade, em que o viés e a variância se coincidam.**

3 – O que é ‘conjunto de treino’ e ‘conjunto de teste’ em Aprendizado de Máquina? Tendo um conjunto de dados inicial, quanto desse conjunto você alocaria para o ‘conjunto de treino’ e para o ‘conjunto de teste’?

-----------------------------------

4 – Liste e explique pelo menos quatro aplicações onde pode-se utilizar Aprendizado de Máquina Supervisionado.

**Aprendizado de máquina supervisionado, pode ser utilizado em várias aplicações, mas quatros delas são:**

1. **KNN: Esse algoritmo tem como intuito descobrir o k vizinhos mais próximos de todos os pontos do conjunto de dados, e pode ser utilizado para fazer alguma decisão com base na distância dos pontos**
2. **Uma aplicação possível é a classificação de emails em spam ou não spam.**
3. **Rede Neurais**
4. **Indução de árvores de decisão**

5 – Liste e explique duas técnicas de Aprendizado de Máquina Não Supervisionado.

Aprendizado de máquina não supervisionado, pode ser utilizado em várias aplicações, **mas duas delas são:**

1. **KMeans: Calcula a k-ésima média de um conjunto de dados**
2. **DBSCAN/HDBSCAN: Algoritmos de agrupamentos, que funcionam com base na densidade dos dados, ou seja, na densidade do conjunto de dados, e baseado nessa densidade, o algoritmo agrupa os dados em conjuntos de dados.**

6 – Qual a diferença entre Aprendizado de Máquina Supervisionado e Não Supervisionado?

**A principal diferença entre o aprendizado supervisionado e o não supervisionado é que o aprendizado supervisionado funciona com base na posse dos resultados esperados, ou seja, o aprendizado supervisionado sabe os resultados esperados, e calcula seus resultados com base nos resultados esperados, já o aprendizado não supervisionado não tem os resultados esperados, e obtem seus resultados com base na entrada. Assim a principal diferença é que o aprendizado supervisionado, aprende com base em um conjunto de regras ou informações, enquanto o aprendizado não supervisionado constrói seus resultados com base na entrada.**

7 – Qual a diferença entre aprendizado indutivo e dedutivo? Qual o mais utilizado em Aprendizado de Máquina?

**A diferença entre aprendizado indutivo e dedutivo é que o aprendizado intuitivo aprende fazendo generalizações sobre casos específicos, ou seja, dado um conjunto de exemplos, o aprendizado deduz hipóteses e conclusões sobre esse conjunto, enquanto o aprendizado dedutivo utiliza de um modelo para deduzir qual será o dado obtido. Ou seja, o intuitivo induz do conjunto de exemplos conclusões e hipóteses, e o dedutivo deduz por um modelo as conclusões e hipóteses.**

8 – Como escolher um algoritmo de Aprendizado de Máquina para um determinado problema de classificação de dados?

**Para escolher um algoritmo de aprendizado de máquina para um determinado problema de classificação de dados, o mais importante é entender o problema que está sendo tratado, e escolher um algoritmo de classificação que melhor resolva este algoritmo.**

9 – Para resolver um determinado problema, quando devemos usar regressão e quando devemos usar classificação? Forneça pelo menos três exemplos de problemas de classificação e três exemplos de problemas de regressão.

**A regressão é quando se utiliza uma função para determinar um único valor real de algo, enquanto a classificiação utiliza uma função para prever a qual classe o ponto pertence.**

**Classificação:**

**-Classificar se um email é spam ou não**

**-Classifica um animal em uma das classe de animais**

**-Definir se um paciente está saudável ou não**

**Regressão:**

**-Prever o valor de uma apólice de seguros**

**-Prever o valor de mercado de um imovel**

**-Prever o tempo de internação de um paciente**

10 – Como você usaria Aprendizado de Máquina para implementar um filtro de spam? Explique os processos de desenvolvimento.

--------------------------------------------------------------------------------------------

11 – Considere que você tem uma lista de algoritmos de Aprendizado de Máquina. Dado um conjunto de dados, como você decidiria qual algoritmo utilizar?

**Aprendizado de máquina é uma área muito vasta. Visto isso, para decidir qual algoritmo utilizar, basicamente iria entender o problema que quero resolver, e decidir qual o melhor algoritmo, comparando com as funções e o que o algoritmo soluciona. Por exemplo, se me problema é agrupar um conjunto de dados, em K grupos, que não são conhecidos, o mais provável é utilizar algoritmos de agrupamento.**

12 – O que é viés e variância em Aprendizado de Máquina?

**Melhor Explicação:** <https://www.mastersindatascience.org/learning/difference-between-bias-and-variance/>

**Variância é o valor que mostra o quão o resultado estimado irá ser alterado se o conjunto de dados for alterado, ou seja, tiver novos valores inserios. Já viés é o quão um valor estimado difere do valor real em um conjunto de dados, ou seja, o quão a estimativa tem de erro.**

13 – Explique o que é a relação viés-variância em Aprendizado de Máquina?

**Melhor Explicação:** <https://www.mastersindatascience.org/learning/difference-between-bias-and-variance/>

**Essa é uma relação muito importante, visto que com o aumento da variância o viés diminui, e vice, e versa, ou seja, há uma troca, caso a variância aumenta o viés diminui, por isso deve haver um resultado ótimo entre esses dois valores, para obter um modelo correto, e que não contenha inconsistência.**

14 – Explique como funciona um algoritmo de Agrupamento de Dados. Forneça pelo menos três exemplos de algoritmos de agrupamento.

**O algoritmo de agrupamento de dados utiliza um conjunto de dados, e com ele, agrupa os pontos em K grupos, que não são informados pelo projetista, dessa forma, o algoritmo pega a similaridade entre os pontos e com isso agrupa os pontos em “clusters”.**

**Exemplo:**

1. **DBSCAN**
2. **HDBSCAN**
3. **K means**

15 – Qual a diferença entre Mineração de Dados e Aprendizado de Máquina?

**Mineração de dados, constrói modelos a partir de um conjunto de dados, lida com dados, e representam o conhecimento induzido e principalmente, utiliza algoritmos de aprendizado de máquina. Aprendizado de máquina, utiliza os dados, e modelos para obter alguma conclusão**

16 – Qual é a diferença entre Inteligência Artificial e Aprendizado de Máquina?

**Podemos dizer que aprendizado de máquina é um ramo da Inteligência aritifical, logo a comparação não existe muito, visto que a IA meio que é “pai” do AM. Mas o IA, engloba muita coisa além do aprendizado de máquina, e o AM utiliza algumas técnicas da IA para funcionar.**