

Exercícios

- Machine Learning

1) “Diz-se que um programa de computador aprende com a experiência **E** com relação a alguma classe de tarefas **T** e medida de desempenho **P**, se seu desempenho nas tarefas **T**, medido por **P**, melhora com a experiência **E**.”

Essa definição foi cunhada por Tom M. Mitchel. Defina um problema onde *machine learning* possa ser aplicada (**T**), a medida de desempenho (**P**) e a experiência (**E**) que pode ser passada para uma máquina de aprendizado entender esse problema.

Resposta

T: Classificar e-mails como Spam e Não-Spam

P: Número de E-mails corretamente classificados

E: Dataset rotulado com características de e-mails que são e que não são Spam.

2) Quais as características de um problema de Classificação e Regressão?

Resposta:

Problemas de classificação possuem um target (variável objetiva) com valores em um intervalo discreto. Essa variável representa classes (rótulos) de um problema, podendo ser numérica ou categórica (nominal). Em um problema de regressão clássico a variável target é numérica e seus valores estão em um intervalo contínuo, ou seja, essencialmente cada instância do dataset possui um valor que representa uma aferição, uma medida.

3) Dado um determinado problema, quais as características que são necessárias para a aplicação de *Machine Learning*?

Resposta:

Para o uso de aprendizado de máquina o problema precisa, essencialmente, possuir dados e que esses dados possuam algum padrão. Com isso, os algoritmos entenderão esses dados para treinar uma função que representa

melhor o problema. Se a função já é conhecida, é desnecessário treinar algum algoritmo.

4) Quais as diferenças de um aprendizado supervisionado e não-supervisionado?

Resposta:

No aprendizado supervisionado os algoritmos são treinados utilizando a variável objetivo (target), Os parâmetros dos algoritmos são ajustados sempre olhando para a variável objetivo. No aprendizado não-supervisionado a variável objetivo não é utilizada e essencialmente o treinamento ocorre tentando encontrar similaridades entre as instâncias do conjunto de treinamento.

5) Ler o texto do link abaixo:

<https://maquinasqueaprendem.com/2019/11/07/quando-usar-aprendizagem-de-maquina/>

Resposta:

Leiam!

- Árvores de Decisão

6) Qual a definição e utilidade do conceito de Entropia para a construção de uma árvore de decisão?

Resposta:

A entropia refere-se a “impureza” de um determinado atributo. Mede a capacidade que um atributo possui de discriminar as classes. Quanto mais próximo de 1 pior é o atributo. Valores próximos de zero indicam que o atributo é menos impuro, ou seja, maior capacidade tem de seus valores discriminarem as classes.

7) Dada a base de treinamento abaixo, calcule a Entropia para cada um dos atributos categóricos:

Classe	Febre	Enjoo	Manchas	Dor	Temperatura
Doente	Sim	Sim	Pequenas	Sim	40.0
Saudável	Sim	Sim	Grandes	Não	38.0
Doente	Não	Não	Grandes	Não	37.0
Doente	Não	Sim	Pequenas	Sim	37.5
Saudável	Sim	Não	Grandes	Sim	38.0
Doente	Não	Não	Grandes	Sim	37.0
Doente	Sim	Não	Pequenas	Não	41.0
Saudável	Não	Sim	Grandes	Sim	36.0
Doente	Sim	Sim	Pequenas	Sim	40.0
Saudável	Sim	Sim	Grandes	Não	38.0

Doente	Não	Não	Grandes	Não	37.0
Doente	Não	Sim	Pequenas	Sim	37.5
Saudável	Sim	Não	Grandes	Sim	38.0
Doente	Não	Não	Grandes	Sim	37.0
Doente	Sim	Não	Pequenas	Não	41.0
Saudável	Não	Sim	Grandes	Sim	36.0

Resposta:

Entropia – Febre:

$$\text{Ent}(\text{CLASSE}, \text{Febre}=\text{Sim}) = -3/8 \log_2 (3/8) - 5/8 \log_2 (5/8) = \mathbf{0,954}$$

$$\text{Ent}(\text{CLASSE}, \text{Febre}=\text{Não}) = -6/8 \log_2 (6/8) - 2/8 \log_2 (2/8) = \mathbf{0,811}$$

Entropia – Enjoo:

$$\text{Ent}(\text{CLASSE}, \text{Enjoo}=\text{Sim}) = -4/8 \log_2 (4/8) - 4/8 \log_2 (4/8) = \mathbf{1}$$

$$\text{Ent}(\text{CLASSE}, \text{Enjoo}=\text{Não}) = -6/8 \log_2 (6/8) - 2/8 \log_2 (2/8) = \mathbf{0,811}$$

Entropia – Manchas:

$$\text{Ent}(\text{CLASSE}, \text{Manchas}=\text{Pequenas}) = -6/6 \log_2 (6/6) - 0/6 \log_2 (0/6) = \mathbf{0}$$

$$\text{Ent}(\text{CLASSE}, \text{Manchas}=\text{Grandes}) = -4/10 \log_2 (4/10) - 6/10 \log_2 (6/10) = \mathbf{0,973}$$

Entropia – Dor:

$$\text{Ent}(\text{CLASSE}, \text{Dor}=\text{Sim}) = -6/10 \log_2 (6/10) - 4/10 \log_2 (4/10) = \mathbf{0,973}$$

$$\text{Ent}(\text{CLASSE}, \text{Dor}=\text{Não}) = -4/6 \log_2 (4/6) - 2/6 \log_2 (2/6) = \mathbf{0,918}$$

8) Usando a tabela da questão anterior, calcule a entropia do atributo numérico “Temperatura”.

Resposta:

Primeiramente precisamos ordenar (ordem crescente ou decrescente) a coluna temperatura. Dica: Use o Excel ou outra planilha!

Com a tabela já ordenada, escolhi o corte entre os valores: 37,5 e 38

Sendo assim, a média entre esses valores é: 37,75. Agora vamos calcular a entropia a partir desse corte.

$$\text{Ent}(\text{CLASSE}, \text{Temp}>37,75) = -4/8 \log_2 (4/8) - 4/8 \log_2 (4/8) = \mathbf{1}$$

$$\text{Ent}(\text{CLASSE}, \text{Temp}<37,75) = -6/8 \log_2 (6/8) - 2/8 \log_2 (2/8) = \mathbf{0,811}$$

9) Calcule o ganho de informação para cada um dos atributos das questões (7) e (8). Qual desses atributos você escolheria para iniciar a criação da árvore? Por que?

Resposta:

Para calcular o Ganho de Informação precisamos calcular a entropia total da base.

$$\text{Ent}(\text{CLASSE}) = -10/16 \log_2 (10/16) - 6/16 \log_2 (6/16) = 0,424 + 0,531 = \mathbf{0,955}$$

Agora precisamos calcular a Informação para cada atributo:

$$\text{Info}(\text{Febre}) = 8/16 * 0,954 + 8/16 * 0,811 = \mathbf{0,8825}$$

$$\text{Info}(\text{Enjoo}) = 8/16 * 1 + 8/16 * 0,811 = \mathbf{0,9055}$$

$$\text{Info}(\text{Manchas}) = 6/16 * 0 + 10/16 * 0,973 = \mathbf{0,6081}$$

$$\text{Info}(\text{Dor}) = 10/16 * 0,973 + 6/16 * 0,918 = \mathbf{0,9523}$$

$$\text{Info}(\text{Temperatura}) = 8/16 * 1 + 8/16 * 0,811 = \mathbf{0,9055}$$

$$\text{Ganho}(\text{Febre}) = 0,955 - 0,8835 = \mathbf{0,0715}$$

$$\text{Ganho}(\text{Enjoo}) = 0,955 - 0,9055 = \mathbf{0,0495}$$

$$\text{Ganho}(\text{Manchas}) = 0,955 - 0,6081 = \mathbf{0,3469}$$

$$\text{Ganho}(\text{Dor}) = 0,955 - 0,9523 = \mathbf{0,0027}$$

$$\text{Ganho}(\text{Temperatura}) = 0,955 - 0,9055 = \mathbf{0,0495}$$

Para iniciar a criação da árvore o atributo escolhido seria o “Manchas”. Esse atributo é o que possui o maior valor de ganho de informação e consequentemente o que melhor discrimina as classes pelos seus valores.

10) O que é “*overfitting*”? Como pode ocorrer?

Resposta:

Overfitting é quando o modelo treinado fica muito específico para um dado conjunto de treinamento. O modelo fica muito assertivo para dados do conjunto de treinamento mas perde sua capacidade de generalização para os dados do conjunto de teste.

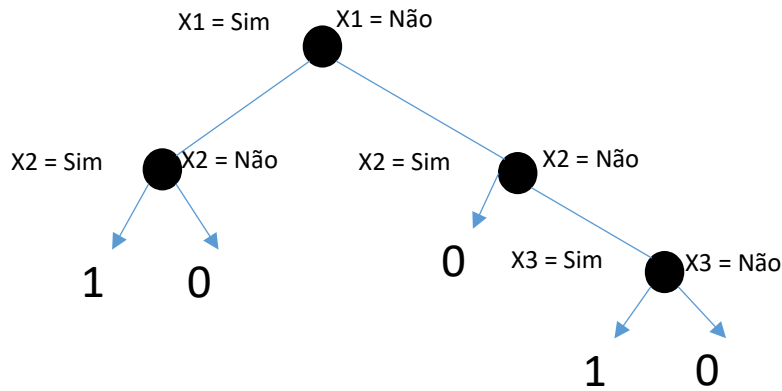
Com árvores de decisão pode ocorrer *overfitting* quando é gerada uma árvore muito profunda, ou seja, uma árvore muito especializada nos dados de treinamento.

11) Quais as soluções para evitar o *overfitting*?

Resposta:

Com árvores de decisão podemos evitar o *overfitting* encerramento mais cedo o treinamento da árvore, podendo a árvore antes de chegar a níveis mais profundos ou fazer pós-pruning, poda depois que a árvore já foi criada.

12) Dada a árvore de decisão binária abaixo, criar regras para ela.



Resposta:

se $x_1 = \text{Sim}$ e $x_2 = \text{Sim}$, então 1
 se $x_1 = \text{Sim}$ e $x_2 = \text{Não}$, então 0
 se $x_1 = \text{Não}$ e $x_2 = \text{Sim}$, então 0
 se $x_1 = \text{Não}$ e $x_2 = \text{Não}$ e $x_3 = \text{Sim}$, então 1
 se $x_1 = \text{Não}$ e $x_2 = \text{Não}$ e $x_3 = \text{Não}$, então 0

- KNN

- Dados os conjuntos abaixo, responda as questões subsequentes.

Treinamento

classe (y)	x1	x2	x3
1	5.1	3.5	1.4
1	4.9	3.0	1.4
2	7.0	3.2	4.7
1	4.7	3.2	1.3
2	6.4	3.2	4.5
1	4.6	3.1	1.5
1	5.0	3.6	1.4
2	6.9	3.1	4.9
2	5.5	2.3	4.0
2	6.5	2.8	4.6

Teste

x1	x2	x3
5.3	3.7	1.5
5.7	2.8	4.1
5.5	3.3	1.4

13) Dado os dados de treinamento e teste, classifique cada um dos padrões de teste usando a distância euclidiana com 3-NN.

Resposta:

Teste1 – Classe 1
 Teste2 – Classe 2
 Teste3 – Classe 1

14) Classifique os dados de treinamento utilizando 3-NN e a distância

$$\text{Manhatan} = \sum_{i=1}^k |a_i - b_i|$$

Resposta:

Teste1 – Classe 1

Teste2 – Classe 2

Teste3 – Classe 1

Mesmo mudando a medida de distância, o resultado foi o mesmo da questão anterior.

15) Use o 3-NN com a distância euclidiana normalizada e compare os resultados com a questão (13)

Resposta:

Façam esse!

16) Dadas as bases de treinamento e teste abaixo, calcule os resultados para os padrões de teste, usando 3-NN com distância euclidiana.

Treinamento

y	x1	x2	x3
2.3	5.1	3.5	1.4
2.7	4.9	3.0	1.4
1.8	7.0	3.2	4.7
2.4	4.7	3.2	1.3
1.5	6.4	3.2	4.5
1.4	4.6	3.1	1.5
2.7	5.0	3.6	1.4
1.6	6.9	3.1	4.9
2.0	5.5	2.3	4.0
3.5	6.5	2.8	4.6

Teste

x1	x2	x3
5.3	3.7	1.5
5.7	2.8	4.1
5.5	3.3	1.4

Resposta:

Questão igual a (13)

- Redes Neurais

17) O que são redes neurais artificiais? Explique com suas palavras o processo de aprendizado.

Resposta:

São modelos de machine learning inspirados no funcionamento das redes neurais biológicas dos seres humanos.

O aprendizado das RNA consiste em ajustar os pesos da rede. Tanto ajustar os pesos das conexões como ajustar os pesos dos neurônios (threshold). O conhecimento de uma rede neural é representado pelos seus pesos.

18) Qual a diferença entre “função de ativação” e “função de saída”?

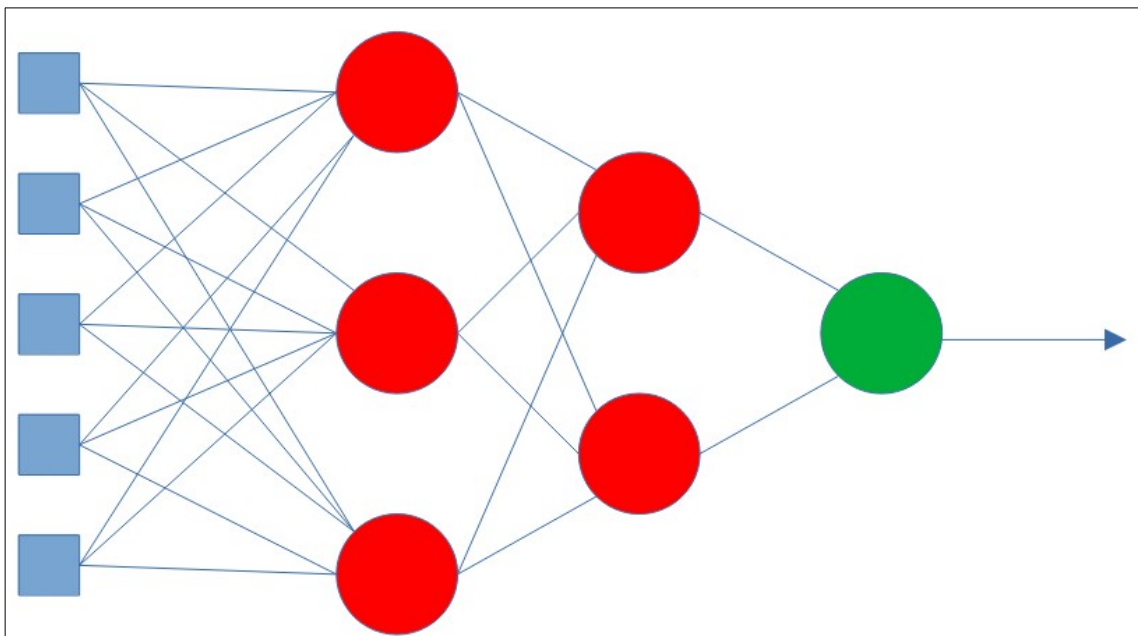
Resposta:

A função de ativação é a função que ajusta o valor calculado no neurônio via entradas vezes os pesos. Esse ajuste corresponde a uma transformação não-linear do somatório das entradas pelos pesos.

A função de saída ajusta o valor de saída da rede neural.

19) Monte uma rede neural com a seguinte Arquitetura:

- A base utilizada tem 5 atributos de entrada;
- A rede tem 2 camadas escondidas com 3 e 2 neurônios, respectivamente;
- A rede é completamente conectada e é *feedforward*;
- Possui 1 neurônio de saída.



20) O Perceptron é uma rede neural que possui limitações. Quais seriam essas limitações? Explique.

Resposta:

O perceptron representa uma linha no espaço de características (atributos),

assim sendo, ele só é efetivo quando o problema (dataset) é linearmente separável entre suas classes, ou seja, é possível traçar uma linha reta separando dados de 2 classes.

21) Explique com suas palavras o processo de aprendizado de uma rede neural.

Resposta:

Dizer que uma rede neural aprendeu, é basicamente dizer que seus pesos foram ajustados. O aprendizado de uma RN é ajustar seus pesos. A cada ciclo do algoritmo de aprendizado, os padrões do conjunto de treinamento são apresentados, um a um e a rede, e seus pesos são ajustados, assim como também o BIAS de cada neurônio.

22) Quais pré-processamentos são necessários a um dataset antes de trabalharmos com redes neurais artificiais? Explique.

Resposta:

Redes neurais artificiais não aceitam dados categóricos, que devem ser transformados em dados numéricos através de alguma transformação, como a OneHot Encoder. A coluna target também deve ser codificada para numérico. Também, as RNAs são sensíveis aos dados quando esses estão em escalas diferentes, ou seja, é interessante que todas as colunas sejam normalizadas antes do treinamento.

23) As Redes MLP vieram para resolver algumas limitações dos Perceptrons, mas para isso algumas alterações devem ser feitas na sua arquitetura e funções de ativação. Quais são essas mudanças?

Resposta:

Com redes MLP, problemas não linearmente separáveis podem agora ser resolvidos. É recomendado usar funções de ativação não lineares, como a Tangente Hiperbólicas e Sigmoid Logística. Sobre a arquitetura...o aumento de camadas em uma rede neural traz uma melhor separação dos dados no espaço de características, formando figuras mais complexas que podem ajudar a separar os dados.