

# Universidade Federal de Mato Grosso do Sul

Faculdade de Computação  
Inteligência Artificial  
Prof. Edson Takashi Matsubara

## Lista de Exercícios 1

1. Por que a indução nem sempre gera hipóteses verdadeiras?
2. Na década de 80, a inteligência artificial era representado principalmente pelos sistemas especialistas e sistemas baseados em conhecimento. Nesses sistemas um especialista de domínio trabalhava em conjunto com um engenheiro de conhecimento que em conjunto trabalhavam para representar o conhecimento. O que muda com o uso de aprendizado de máquina?
3. Qual é a diferença entre aprendizado supervisionado e não-supervisionado?
4. Qual é a diferença entre classificação e regressão?
5. Qual a diferença entre conjunto de treinamento e teste?
6. O que é melhor, obter um algoritmo que seja melhor no conjunto de treinamento ou no conjunto de teste? Por quê? Dê um exemplo real na sua justificativa.
7. O uso de  $k = 1$  pode levar à classificações incorretas caso existam exemplos com ruído no conjunto de treinamento. Explique por quê?
8. Por que a normalização dos valores é de grande importância para o K-NN? Dê um exemplo que se torna evidente a necessidade de normalização dos valores.
9. Por que deve-se transformar atributos nominais em atributos numéricos quando se utiliza K-NN? Explique uma maneira de realizar a conversão de atributos nominais em atributos numéricos.
10. Apresente situações onde é evidente que o K-NN com pesos é melhor que o K-NN tradicional.
11. Por que KNN não é indicado para ser utilizado em conjunto de dados com muitos atributos?
12. Por que KNN também não é indicado para ser utilizado com conjunto com muitos exemplos de treinamento? Como solucionar esse problema?
13. O que é a estimação de máxima verossimilhança?
14. Calcule

$$\theta = \arg \max_{\theta} \theta^h (1 - \theta)^{m-h} \quad (1)$$

15. Calcule o gradiente das seguintes funções:

- (a)  $f(x) = x^2 + 2$
- (b)  $f(x) = (x - 2)^2$
- (c)  $f(x, y, z) = x^3 + y^2 + z$
- (d)  $g(z) = \frac{1}{1+e^{-z}}$

16. Faça 3 iterações do método de gradiente descendente para cada função utilizando valores arbitrários para  $x, y$  e  $z$ .

17. Dado que

$$h_{\theta}(x^{(i)}) = \sum_{k=0}^n \theta_k x_k^{(i)} \quad (2)$$

e

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (3)$$

calcule  $\frac{\partial}{\partial \theta_j} J(\theta)$  e mostre a regra de atualização de  $\theta_j$ .

18. Considere os pontos (0,2) e (5,4). Encontre a reta utilizando regressão linear que fique próxima aos pontos.

19. Considere a Tabela 1. Utilizando regressão logística encontre a reta que

Table 1: Pesos e alturas

altura	peso	sexo
1.50	40	feminino
1.51	42	masculino
1.60	60	feminino
1.61	62	masculino

separa homens e mulheres.

20. Mostre a equivalência do perceptron e o algoritmo de regressão logística.

21. Faça duas iterações do algoritmo de perceptron na tabela verdade do AND.

22. Faça duas iterações do algoritmo de perceptron na tabela verdade do OR.

23. Explique a função do *bias* no perceptron.

24. Explique porque a regressão logística não consegue aprender a tabela verdade do XOR.

25. Considere o conjunto de dados palestra apresentado na Tabela 2.

Induza uma árvore de decisão utilizando ganho de informação. Mostre todas as contas e detalhes da indução desta árvore.

Table 2: Conjunto de dados palestra

<i>palestra</i>	<i>relevancia</i>	<i>comida</i>	<i>distancia</i>	<i>brinde</i>	<i>classe</i>
1	alta	sim	perto	sim	sim
2	alta	não	perto	não	sim
3	alta	sim	perto	sim	sim
4	alta	não	perto	não	sim
5	alta	sim	perto	sim	sim
6	alta	não	perto	não	sim
7	alta	sim	perto	sim	não
8	alta	não	longe	não	sim
9	alta	sim	longe	sim	não
10	alta	não	longe	não	não
11	baixa	sim	perto	sim	sim
12	baixa	sim	longe	não	sim
13	baixa	sim	perto	sim	sim
14	baixa	sim	longe	não	não
15	baixa	não	perto	sim	não
16	baixa	não	longe	não	não
17	baixa	não	perto	sim	não
18	baixa	não	longe	não	não
19	baixa	não	perto	sim	não
20	baixa	não	longe	não	não

26. Qual é a vantagem de se utilizar razão de ganho?
27. Transforme a árvore induzida em um conjunto de regras.
28. Qual é a importância de realizar poda em árvores de decisão?
29. Ainda utilizando o mesmo conjunto de dados, induza um conjunto de regras utilizando a ordered rules e non-ordered rules.
30. Por que no aprendizado de regras se faz necessário o uso de uma regra padrão (regra default)?
31. Por que pode-se converter uma árvore de decisão em regras mas não se pode converter regras em árvores de decisão? Dê exemplos.
32. Para que serve a correção de laplace? Dê exemplos.
33. Considere uma regra  $R$  que cobre 10 exemplos dos quais 9 exemplos estão classificados corretamente. Agora considere uma poda nesta mesma regra que faz a regra cobrir 20 exemplos dos quais 17 deles estão corretos. Calcule o erro pessimista (perr) de ambos os casos e veja se a poda pode melhorar ou piorar a classificação dos exemplos.
34. Quais são as propriedades para uma função  $d : X \times X \rightarrow R$  seja considerado uma medida de distância.
35. Verifique se as funções abaixo podem ser consideradas medidas de distância:
  - (a) produto escalar de dois vetores;

- (b) coseno entre dois vetores;
  - (c) seno entre dois vetores;
  - (d) a diferença de número de caracteres entre duas cadeias de caracteres;
  - (e) o módulo da diferença de número de caracteres entre duas cadeias de caracteres.
36. Considere a Tabela 3 Faça o agrupamento hierárquico utilizando a distância

Table 3: Conjunto agrupamento

id	$x_1$	$x_2$
A	10	11
B	11	11
C	20	15
D	21	17
E	15	15
F	40	5

de manhattan com ligação simples, completa e média de grupo.

37. Qual é o provável outlier do deste conjunto de dados?
38. Faça agrupamento por K-médias utilizando a Tabela 3. Faça para  $K = 2, 3, 4$
39. O agrupamento por K-médias consegue representar agrupamentos não esféricos? E o hierárquico? Explique e dê exemplos.
40. Utilize a regra do cotovelo para verificar qual é a provavel quantidade de clusters utilizando agrupamento hierárquico e agrupamento por k-médias.