

# Inteligência Artificial – Lista 4

Prof. Murilo Coelho Naldi

## Aprendizado de Máquina

Dado o conjunto de treinamento a seguir com informações sobre o consumo de bebidas de seis clientes:

Cliente	Estado Civil	Ganha	Emprego	Consumo
1	solteiro	bem	diurno	baixo
2	casado	bem	noturno	baixo
3	casado	mal	diurno	alto
4	solteiro	bem	noturno	alto
5	casado	bem	diurno	baixo
6	solteiro	mal	diurno	alto

Induza uma árvore de decisão capaz de classificar clientes segundo o seu **consumo** (use o consumo como a classe). A árvore deve apresentar o maior ganho possível para cada nó, segundo uma das medidas abaixo, **a sua escolha**:

$$\text{Entropia}(t) = - \sum_{i=1}^c p(i|t) \log_2 p(i|t)$$

$$\text{Gini\_Index}(t) = 1 - \sum_{i=1}^c [p(i|t)]^2$$

$$\text{Erro\_Class}(t) = 1 - \max_{i \in \{1, \dots, c\}} [p(i|t)]$$

onde:

$p(i|t)$  = fração de dados pertencente à classe  $i$  em um nó  $t$

$c$  = número de classes

$0 \log_2 0 = 0$

Faça a construção nó-a-nó da árvore e todos os cálculos feitos, justificando a escolha dos atributos em cada nível. Escolha sempre o atributo com maior ganho ou menor impureza em cada nível.

Em seguida, classifique passo-a-passo os objetos de teste abaixo segundo a árvore de decisão obtida. Calcule a precisão ( $VP+VN / VP+VN+FN+FP$ ) e a taxa de erro ( $FN+FP / VP+VN+FN+FP$ ) do classificador obtido.

Cliente	Estado Civil	Ganha	Emprego	Consumo
7	solteiro	bem	noturno	baixo
8	casado	mal	diurno	alto
9	solteiro	mal	diurno	alto

Em seguida, calcule a similaridade entre os clientes da tabela (1-6) e os clientes novos (7-9). Basicamente, conte quantas características são iguais. Quanto maior o número de características iguais, maior a similaridade. A partir das similaridades obtidas, classifique os clientes novos utilizando a técnicas dos  $k$  vizinhos mais próximos. Utilize  $k=1$  e depois  $k=3$ . Compare os resultados com os obtidos pela árvore de decisão.

Para mais exercícios, utilize o livro STEINBACH, M., KUMAR, V. TAN, P. Introdução ao Data Mining (Mineração de Dados). Edição 1. Ciência Moderna 2009. ISBN 9788573937619.