



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Instituto de Ciências Exatas e de Informática

Marcos Antonio Lommez Candido Ribeiro¹

Lista #1

Inteligência Artificial

¹Aluno de Graduação em Ciência da Computação – tonilommez@hotmail.com

1) Leia os slides “MODULO / Introdução à Machine Learning / Introdução à Machine Learning.PDF” que estão no CANVAS, e explique:

1) Os tipos de problemas normalmente resolvidos por aprendizado de máquina

O aprendizado de máquina consegue resolver problemas formalmente classificados como:

1. Classificação
2. Regressão
3. Agrupamento (clusterização)
4. Regras de associação

Que são problemas que nascem da necessidade de extrair uma informação a partir de uma base de dados para prever novas informações baseado naquele treinamento. Como a partir de uma base de dados o algoritmo aprender a classificar de uma determinada ação vai ou não ocorrer, ou se a partir de determinados dados de entrada descobrir o preço de um determinado produto. Mas também existem leves variações como o caso do agrupamento onde a ideia é dado uma base de dados descobrir grupos que se distinguem dos outros de alguma maneira. Por fim, as Regras de associação é próximo ao agrupamento, mas com a ideia central em descobrir padrões e semelhanças entre grupos.

2) Os tipos de aprendizado de máquina existentes

Os tipos de aprendizado de máquina se diferem pela forma como são treinados (e consequentemente sua finalidade também acaba mudando).

1. Supervisionado
2. Não supervisionado
3. Semi supervisionado
4. Reforço
5. Deep Learning

Os do tipo Supervisionado são aqueles onde se possui o dado de saída para validação do modelo. Os não supervisionados são aqueles onde esse dado não existe ou não há necessidade de tê-los, como em problemas de agrupamento. Semi supervisionados sendo um meio-termo entre os 2 anteriores. Por fim, o aprendizado por reforço onde entram

as redes neurais, e sua melhoria que pode ser considerada um tipo de aprendizado que é o deep learning.

3) O funcionamento do processo de KDD

Knowledge Discovery in Databases é um processo para mineração de dados criado com o intuito de identificar padrões uteis em uma base de dados, criando assim uma nova fonte de informações que possa ser usada.

Para esse processo existe uma sequencia de etapas a ser seguidas. Sendo:

1. Entendimento da base de dados e metas
2. Seleção de valores que serão utilizados
3. Pre-processamento e limpeza dos dados
4. Transformação dos dados
5. Uso de algoritmos de mineração
6. Avaliação do desempenho do modelo e sua interpretação
7. Descoberta do conhecimento e sua visualização

2) Considerando-se a base de dados sobre “Esperar ou não pelo restaurante” (verificar base de dados ‘Restaurante.csv’ disponibilizada no CANVAS), pede-se:

Exemplo	Alternativo	Bar	Sex/Sab	fome	Ciente	Preço	Chuva	Res	Tipo	Tempo	conc
X1	Sim	Não	Não	Sim	Alguns	RRR	Não	Sim	Francês	0-10	Sim
x2	Sim	Não	Não	Sim	Cheio	R	Não	Não	Tailandês	30-60	Não
x3	Não	Sim	Não	Não	Alguns	R	Não	Não	Hamburger	0-10	Sim
x4	Sim	Não	Sim	Sim	Cheio	R	Sim	Não	Tailandês	out/30	Sim
X5	Sim	Não	Sim	Não	Cheio	RRR	Não	Sim	Francês	>60	Não
X6	Não	Sim	Não	Sim	Alguns	RR	Sim	Sim	Italiano	0-10	Sim
X7	Não	Sim	Não	Não	Nenhum	R	Sim	Não	Hamburger	0-10	Não
X8	Não	Não	Não	Sim	Alguns	RR	Sim	Sim	Tailandês	0-10	Sim
X9	Não	Sim	Sim	Não	Cheio	R	Sim	Não	Hamburger	>60	Não
X10	Sim	Sim	Sim	Sim	Cheio	RRR	Não	Sim	Italiano	out/30	Não
X11	Não	Não	Não	Não	Nenhum	R	Não	Não	Tailandês	0-10	Não
X12	Sim	Sim	Sim	Sim	Cheio	R	Não	Não	Hamburger	30-60	Sim

Figura 1: Esperar ou não pelo restaurante

Entropia de Atributo

$$H(A) = p(v_1) \cdot H(C|v_1) + p(v_2) \cdot H(C|v_2) + \dots + p(v_n) \cdot H(C|v_n)$$

Entropia de Classe

$$H(C) = -(p_1 \log_2(p_1) + p_2 \log_2(p_2) + \dots + p_n \log_2(p_n))$$

Ganho de Informação:

$$IG(A) = H(S) - H(A)$$

1) Calcular o ganho de informação de cada atributo. Que atributo é a raiz da árvore?

Entropia de Classe

$$H(C) = -\left(\frac{6}{12} \log_2\left(\frac{6}{12}\right) + \frac{6}{12} \log_2\left(\frac{6}{12}\right)\right) = 1.0$$

Alternativo

$$\begin{aligned} H(A) &= \frac{6}{12} \left(-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right) + \\ &\quad \frac{6}{12} \left(-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right) = 1.0 \\ IG &= 1.0 - 1.0 = 0.0 \end{aligned}$$

Bar

$$\begin{aligned} H(A) &= \frac{6}{12} \left(-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right) + \\ &\quad \frac{6}{12} \left(-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right) = 1.0 \\ IG &= 1.0 - 1.0 = 0.0 \end{aligned}$$

Sex/Sab

$$\begin{aligned} H(A) &= \frac{7}{12} \left(-\frac{4}{7} \log_2\left(\frac{4}{7}\right) - \frac{3}{7} \log_2\left(\frac{3}{7}\right) \right) + \\ &\quad \frac{5}{12} \left(-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right) = 0.979 \\ IG &= 1.0 - 0.979 = 0.020 \end{aligned}$$

Fome

$$\begin{aligned} H(A) &= \frac{7}{12} \left(-\frac{5}{7} \log_2 \left(\frac{5}{7} \right) - \frac{2}{7} \log_2 \left(\frac{2}{7} \right) \right) + \\ &\quad \frac{5}{12} \left(-\frac{1}{5} \log_2 \left(\frac{1}{5} \right) - \frac{4}{5} \log_2 \left(\frac{4}{5} \right) \right) = 0.804 \\ IG &= 1.0 - 0.804 = 0.195 \end{aligned}$$

Cliente

$$\begin{aligned} H(A) &= \frac{4}{12} \left(-\frac{4}{4} \log_2 \left(\frac{4}{4} \right) \right) + \\ &\quad \frac{6}{12} \left(-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right) + \\ &\quad \frac{2}{12} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) = 0.459 \\ IG &= 1.0 - 0.459 = 0.540 \end{aligned}$$

Preço

$$\begin{aligned} H(A) &= \frac{3}{12} \left(-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right) + \\ &\quad \frac{7}{12} \left(-\frac{4}{7} \log_2 \left(\frac{4}{7} \right) - \frac{3}{7} \log_2 \left(\frac{3}{7} \right) \right) + \\ &\quad \frac{2}{12} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) = 0.804 \\ IG &= 1.0 - 0.804 = 0.195 \end{aligned}$$

Chuva

$$\begin{aligned} H(A) &= \frac{7}{12} \left(-\frac{3}{7} \log_2 \left(\frac{3}{7} \right) - \frac{4}{7} \log_2 \left(\frac{4}{7} \right) \right) + \\ &\quad \frac{5}{12} \left(-\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right) = 0.979 \\ IG &= 1.0 - 0.979 = 0.021 \end{aligned}$$

Res

$$\begin{aligned} H(A) &= \frac{5}{12} \left(-\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right) + \\ &\quad \frac{7}{12} \left(-\frac{4}{7} \log_2 \left(\frac{4}{7} \right) - \frac{3}{7} \log_2 \left(\frac{3}{7} \right) \right) = 0.979 \\ IG &= 1.0 - 0.979 = 0.021 \end{aligned}$$

Tipo

$$\begin{aligned} H(A) &= \frac{2}{12} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \\ &\quad \frac{4}{12} \left(-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) + \\ &\quad \frac{4}{12} \left(-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) + \\ &\quad \frac{2}{12} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) = 1.0 \\ IG &= 1.0 - 1.0 = 0.0 \end{aligned}$$

Tempo

$$\begin{aligned} H(A) &= \frac{6}{12} \left(-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right) + \\ &\quad \frac{2}{12} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \\ &\quad \frac{2}{12} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \\ &\quad \frac{2}{12} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) = 0.792 \\ IG &= 1.0 - 0.792 = 0.208 \end{aligned}$$

O atributo que será colocado na raiz da árvore será o atributo **Cliente**.

2) Que atributo estará no segundo nível da árvore. Faça os cálculos e apresente a árvore gerada até o segundo nível da árvore.

Entropia de Classe

$$H(C) = - \left(\frac{4}{6} \log_2 \left(\frac{4}{6} \right) + \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right) = 0.918$$

Alternativo

$$H(A) = \frac{5}{6} \left(-\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right) + \frac{1}{6} \left(-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right) = 0.809$$

$$IG = 0.918 - 0.809 = 0.109$$

Bar

$$H(A) = \frac{3}{6} \left(-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right) + \frac{3}{6} \left(-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right) = 0.918$$

$$IG = 0.918 - 0.918 = 0.000$$

Sex/Sab

$$H(A) = \frac{1}{6} \left(-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right) + \frac{5}{6} \left(-\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{5} \right) \right) = 0.809$$

$$IG = 0.918 - 0.809 = 0.109$$

Fome

$$H(A) = \frac{4}{6} \left(-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) + \frac{2}{6} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) = 0.667$$

$$IG = 0.918 - 0.667 = 0.251$$

Cliente

$$H(A) = \frac{6}{6} \left(-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right) = 0.918$$

$$IG = 0.918 - 0.918 = 0.0$$

Preço

$$\begin{aligned} H(A) &= \frac{4}{6} \left(-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) + \\ &\quad \frac{2}{6} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) = 0.666 \\ IG &= 0.918 - 0.666 = 0.251 \end{aligned}$$

Chuva

$$\begin{aligned} H(A) &= \frac{4}{6} \left(-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right) + \\ &\quad \frac{2}{6} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) = 0.874 \\ IG &= 0.918 - 0.874 = 0.044 \end{aligned}$$

Res

$$\begin{aligned} H(A) &= \frac{4}{6} \left(-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) + \\ &\quad \frac{2}{6} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) = 0.666 \\ IG &= 0.918 - 0.666 = 0.251 \end{aligned}$$

Tipo

$$\begin{aligned} H(A) &= \frac{2}{6} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \\ &\quad \frac{1}{6} \left(-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right) + \\ &\quad \frac{2}{6} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \\ &\quad \frac{1}{6} \left(-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right) = 0.667 \\ IG &= 0.918 - 0.667 = 0.251 \end{aligned}$$

Tempo

$$\begin{aligned} H(A) &= \frac{2}{6} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \\ &\quad \frac{2}{6} \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \\ &\quad \frac{2}{6} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) = 0.667 \\ IG &= 0.918 - 0.667 = 0.251 \end{aligned}$$

Neste caso o valor a ser escolhido poderá variar entre os atributos **Fome**, **Preço**, **Res**, **Tipo** e **Tempo**. Neste caso escolheremos **Fome**.

Arvore de decisão formada:

