

Introdução ao aprendizado de máquina

Definições - ML:

- Programas que aprendem a melhorar o desempenho por meio de experiência.
- O aprendizado de máquina é o treinamento de um modelo a partir de dados que generalizam uma decisão contra uma medida de desempenho.

Estrutura de dados:

The diagram illustrates the structure of a dataset. It features a table with 6 columns and 15 rows. Above the table, two labels with arrows point to specific columns: 'Atributos de entrada' points to the first four columns (Aparência, temperatura, Umidade, Ventando), and 'Atributo de classificação' points to the last column (Jogar). To the left of the table, the label 'Instâncias' has two arrows pointing to the first and last rows (d1 and d14), indicating that each row represents an instance of the data.

	Atributos de entrada				Atributo de classificação
	Aparência	temperatura	Umidade	Ventando	Jogar
d1	Sol	Quente	Alta	Não	Não
d2	Sol	Quente	Alta	Sim	Não
d3	Nublado	Quente	Alta	Não	Sim
d4	Chuva	Agradável	Alta	Não	Sim
d5	Chuva	Fria	Normal	Não	Sim
d6	Chuva	Fria	Normal	Sim	Não
d7	Nublado	Fria	Normal	Sim	Sim
d8	Sol	Agradável	Alta	Não	Não
d9	Sol	Fria	Normal	Não	Sim
d10	Chuva	Agradável	Normal	Não	Sim
d11	Sol	Agradável	Normal	Sim	Sim
d12	Nublado	Agradável	Alta	Sim	Sim
d13	Nublado	Quente	Normal	Não	Sim
d14	Chuva	Agradável	Alta	Sim	Não

- Os atributos de uma estrutura de dados podem ser nominais ou numéricos.

Tipos de problemas:

- **Classificação:**

Nesse caso, queremos prever ou descrever uma classe. A atributo de classificação nominal.

Exemplo	Alternativo	Bar	Sex/Sab	fome	Cliente	Preço	Chuva	Res	Tipo	Tempo	Conclusão (Vai esperar)?
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	10-30	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	10-30	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

- **Regressão:**

Semelhante ao de classificação, porém o atributo de classe é numérico.

Exemplo de situação:

- Prever na base do restaurante o valor da comida, com base nos atributos.
- Prever na temperatura (se fosse numérico), com base nos outros atributos da base.
- Quanto tempo de vida a pessoa terá.

- **Agrupamento (Clusterização):**

O objetivo é agrupar as instâncias, de acordo com os atributos de entrada. Não é conhecido o atributo de classificação.

Exemplo:

- Identificar perfis de usuários em redes sociais.

- **Regras de associação:**

Busca de semelhança/associações entre os elementos.

Exemplo:

- Em um sistema de e-commerce → Quem comprou o produto A também comprou o produto B.

Tipos de aprendizagem:

- **Aprendizagem supervisionado:**

Indutor recebe um conjunto de exemplos na forma (entrada, rotulo_desejado)

Técnicas:

- Redes Neurais do tipo Multilayer Perceptron
- Máquina de Vetores Suporte
- Árvores de Decisão

- **Aprendizado não supervisionado:**

Indutor recebe apenas atributos de entrada. Com o objetivo de encontrar agrupamentos.

Técnicas:

- Redes Neurais do tipo mapas auto-organizáveis
- Algoritmo K-Médias

- **Aprendizado semi-supervisionado:**

A aprendizagem semi-supervisionada utiliza dados rotulados e não rotulados para aprender melhor o modelo.

Técnicas:

- SVM

- **Aprendizado por reforço:**

Concentra-se na maximização das recompensas do resultado.

Técnicas:

- Rede Neural

- **Deep Learning:**

Rede neural com muitas camadas. Para cada reconhecendo situações com ordem de complexidade maiores.

Técnicas:

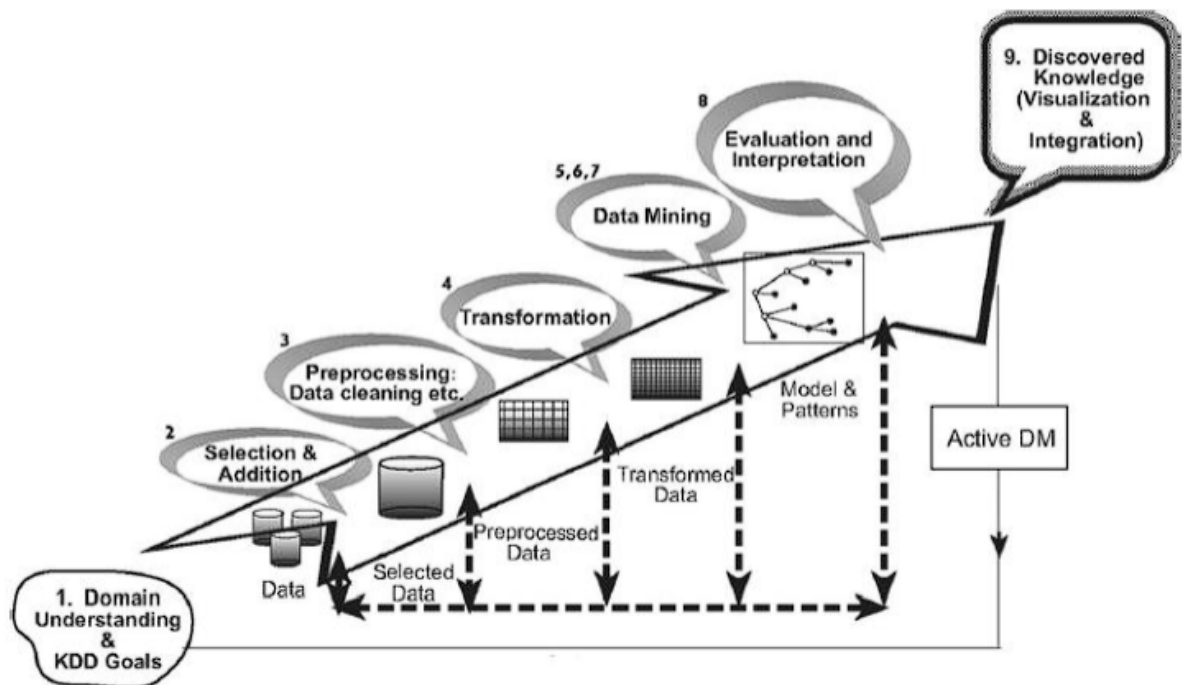
- Rede Neurais

Descoberta de conhecimento:

Descobrir conhecimento significa identificar, receber informações relevantes, e poder processá-las e agregá-las ao conhecimento prévio de seu usuário, mudando o estado de seu conhecimento atual, a fim de que determinada situação ou problema possa ser resolvido.

Mineração de dados:

É um termo utilizado para descrever o processo de descoberta de padrões interessantes e desconhecidos em grandes banco de dados. A mineração de dados é uma parte do processo geral de descoberta de conhecimento em banco de dados (KDD - Knowledge Discovery in Databases).



Visão geral:

