

APRENDIZADO DE MÁQUINA

Cristiane Neri Nobre

Aprendizado de Máquina

Definições

- Construção de programas de computador que melhoram seu desempenho por meio de experiência
- Um programa aprende a partir da experiência **E**, em relação a uma classe de tarefas **T**, com medida de desempenho **P**, se seu desempenho em **T**, medido por **P**, melhora com **E** (Tom.M. Mitchell, 1997)
- O aprendizado de máquina é o treinamento de um modelo a partir de dados que generalizam uma decisão contra uma medida de desempenho (Jason Brownlee)

Aprendizado de Máquina

Estrutura de dados

Atributos de entrada

Atributo de classificação

Dia	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

Instâncias

- Os atributos podem ser nominais ou numéricos

Aprendizado de Máquina

Tipos de problemas

Os algoritmos de aprendizagem de máquina são classificados quanto ao **tipo de problemas** de aprendizagem:

- Classificação
- Regressão
- Agrupamento (clusterização)
- Regras de associação

Aprendizado de Máquina

Tipos de problemas

Classificação

- É um dos tipos de problemas mais utilizados.
- Neste caso, queremos **prever ou descrever** uma classe.
 - A atributo de classificação é **nominal**

Aprendizado de Máquina

Tipos de problemas

Classificação - Exemplos

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

Aprendizado de Máquina

Tipos de problemas

Classificação - Exemplos

Dia	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

Aprendizado de Máquina

Tipos de problemas

Regressão

Semelhante ao problema anterior. Porém o atributo de classe é **numérico**

Imagine as seguintes situações:

Prever na base do restaurante o valor da comida, com base nos atributos

Prever a temperatura (se fosse numérico), com base nos outros atributos da base

Aprendizado de Máquina

Tipos de problemas

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

Aprendizado de Máquina

Tipos de problemas

Regras de associação

Busca semelhança/associações entre os elementos

Exemplo:

Em sistemas e-commerce:

Quem comprou o produto A também comprou o produto B

Aprendizado de Máquina

Tipos de aprendizagem

Os algoritmos de aprendizagem de máquina são classificados quanto ao tipo **de aprendizagem**:

- Supervisionado
- Não supervisionado
- Semi supervisionado
- Reforço
- Deep Learning

Aprendizado de Máquina

Tipos de aprendizagem

Aprendizado Supervisionado

- Indutor recebe conjunto de exemplos na forma (entrada, rótulo_desejado)

Técnicas:

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

Aprendizado de Máquina

Tipos de aprendizagem

Aprendizado Não-supervisionado

- Indutor recebe apenas atributos de entrada
- Encontrar agrupamentos

Técnicas:

- Redes Neurais do tipo mapas auto-organizáveis
- Algoritmo k-médias

Aprendizado de Máquina

Tipos de aprendizagem

Aprendizado Semi-supervisionado

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

Técnicas:

- SVM

Aprendizado de Máquina

Tipos de aprendizagem

Aprendizado por reforço

- Concentra-se na maximização das recompensas do resultado.

Técnicas:

- Rede neural

Aprendizado de Máquina

Tipos de aprendizagem

Deep learning

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

Técnicas:

- Rede neural

Aprendizado de Máquina

Qualidade do modelo

Como avaliar a qualidade do modelo?

- Depende do tipo de aprendizado!
 - Em problemas de **classificação**, avaliar pela **matriz de confusão**
 - Em problemas de **agrupamento**, avaliar utilizando-se métricas específicas tal como silhouette index
 - Em problemas de regressão, utilizam-se as métricas: Mean squared error (MSE), Mean absolute error (MAE), Normalized MSE and MAE

Aprendizado de Máquina

Qualidade do modelo

Como avaliar a qualidade do modelo pela matriz de confusão?

Taxas de VP, VN, FP, FN, precisão, sensibilidade, F-measure, etc

Aprendizado de Máquina

Qualidade do modelo

Como avaliar a qualidade do modelo pela matriz de confusão?

Precisão é a taxa de instâncias **corretamente** classificadas como pertencentes a classe em questão dentre todos os que foram classificados na classe em questão.

$$Pr = \frac{VP}{VP + FP}$$

Sensibilidade ou **recall** é a taxa de instâncias **corretamente** classificadas como pertencentes a classe em questão dentre todos os que realmente são da classe em questão

$$Sens = \frac{VP}{VP + FN}$$

Aprendizado de Máquina

Qualidade do modelo

Como avaliar a qualidade do modelo pela matriz de confusão?

Acurácia é a taxa total instâncias corretamente classificadas.

$$Ac = \frac{VP + VN}{VP + VN + FP + FN}$$

F-Measure é uma média harmônica entre precisão e sensibilidade.

$$FM = \frac{2 * \textit{precisão} * \textit{sensibilidade}}{\textit{precisão} + \textit{sensibilidade}}$$

Aprendizado de Máquina

Ferramentas

Que ferramentas podem ser utilizadas?

Weka - <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>

Orange - <https://orange.biolab.si/>

R, Python, Knime, dentre outras...

