

Introdução ao Aprendizado de Máquina

Alessandro Copetti

alessandro.copetti@gmail.com

Objetivos

- Introdução aos conceitos básicos de Aprendizado de Máquina
- Apresentação prática do Weka

Livro

- Rezende, S. O. (2003). **Sistemas inteligentes: fundamentos e aplicações**. Editora Manole
 - Cap.4: Conceitos sobre aprendizado de máquina. Maria Carolina Monard e José Augusto Baranauskas



Livro

- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). **Data Mining: Practical machine learning tools and techniques**. Morgan Kaufmann.

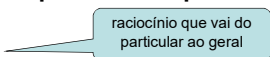
<https://www.cs.waikato.ac.nz/ml/weka/book.html>



Aprendizado de Máquina

- É uma área de IA
- Objetivo:
 - desenvolver técnicas computacionais sobre o aprendizado
 - construir sistemas capazes de adquirir conhecimento de forma automática
- Sistema de aprendizado
 - é um programa de computador que toma decisões baseado em experiências acumuladas por meio da solução bem-sucedida de problemas anteriores

Hierarquia do Aprendizado

- Indução 
 - é a forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto particular de exemplos
 - as hipóteses geradas através da inferência indutiva podem ou não preservar a verdade
 - se o no. de exemplos for insuficiente, ou se não forem bem escolhidos, as hipóteses obtidas podem ser inconsistentes
 - é um dos principais métodos para derivar conhecimento novo e prever eventos futuros

Aprendizado Indutivo

- **Supervisionado**
 - é fornecido ao algoritmo de aprendizado um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido
- **Não-supervisionado**
 - o algoritmo analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira
 - formação de agrupamentos ou clusters

Aprendizado Supervisionado

- Todo exemplo possui um atributo especial, denominado rótulo ou classe
 - é o que se deseja aprender
- O objetivo do algoritmo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados

Aprendizado Supervisionado – Classificação

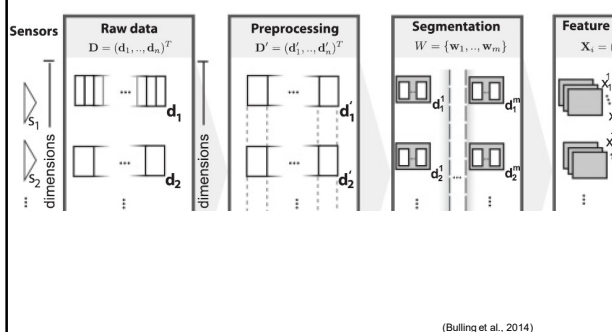
- **Classificador ou Hipótese**
 - dado um conjunto de exemplos de treinamento, um indutor gera como saída um classificador (hipótese)
 - dado um novo exemplo, o classificador tenta prever com a maior precisão possível sua classe
- **Taxa de erro**
 - é obtida utilizando uma equação que compara a classe verdadeira de cada exemplo com o rótulo atribuído pelo classificador

Aprendizado Supervisionado

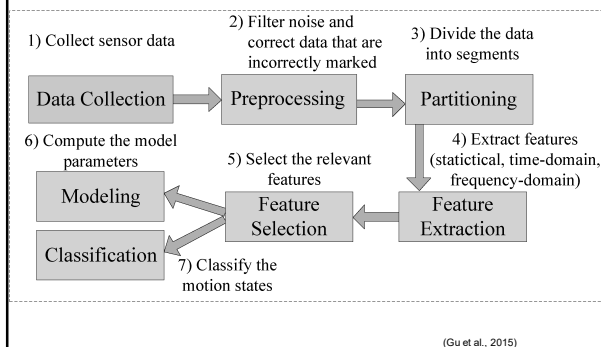
- **Conjunto de exemplos**
 - conjunto de treinamento
 - usado para o aprendizado do conceito
 - conjunto de teste
 - usado para medir o grau de efetividade do conceito aprendido

A habilidade de generalização de classificadores se refere a seu desempenho ao classificar padrões de teste que não foram utilizados durante o treinamento

Reconhecimento de atividades – etapas



Reconhecimento de atividades – etapas



WEKA



Weka

- Formato arquivo entrada
 - .arff (attribute-relation file format)
- Exemplo de:
 - Pré-processamento
 - Classificador
 - Clusterização
 - Regras de associação
 - Seleção de atributos
- Geração de código

```
weather.nominal - WordPad
Arquivo  Editar  Exibir  Inserir  Formatar  Ajuda

@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```

```
weather - WordPad
Arquivo  Editar  Exibir  Inserir  Formatar  Ajuda

@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Undo Edit... Save...

Filter: Choose None Apply

Current relation: weather
Instances: 14
Attributes: 5

Attributes: All None Invert

No.	Name
1	outlook
2	temperature
3	humidity
4	windy
5	play

Selected attribute: outlook
Name: outlook
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

Label	Count
sunny	5
overcast	4
rainy	5

Class: play (Nom) Visualize All

Status: OK Log

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier: Choose J48 -C 0.25 -M 2

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds: 10
- ☐ Percentage split %: 66

 More options...

(Nom) play

Start Stop

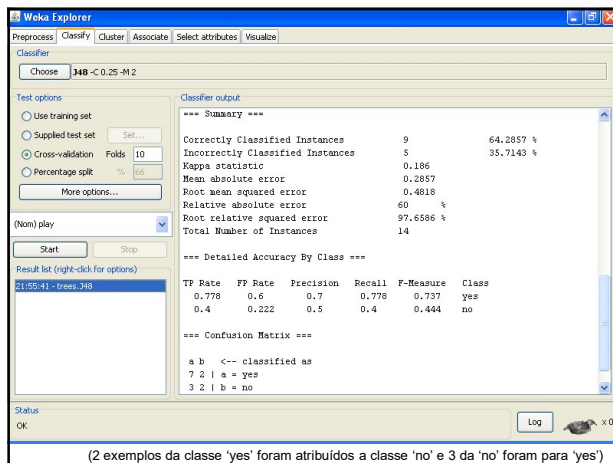
Result list (right-click for options): 31.55:41 - trees_J48

Classifier output:

```
=== Classifier model (full training set) ===
J48 pruned tree
-----
outlook = sunny
| humidity <= 75: yes (2.0)
| humidity > 75: no (3.0)
outlook = overcast: yes (4.0)
outlook = rainy
| windy = TRUE: no (2.0)
| windy = FALSE: yes (3.0)

Number of Leaves : 5
Size of the tree : 8
Time taken to build model: 0.03 seconds
=== Stratified cross-validation ===
```

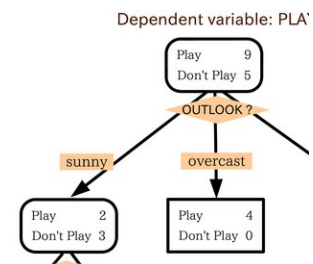
Status: OK Log



- The **True Positive (TP)** rate is the proportion of examples which were classified as class x , among all examples which truly have class x , i.e. how much part of the class was captured. It is equivalent to **Recall**. In the confusion matrix, this is the diagonal element divided by the sum over the relevant row, i.e. $7/(7+2)=0.778$ for class yes and $2/(3+2)=0.4$ for class no in our example.
- The **False Positive (FP)** rate is the proportion of examples which were classified as class x , but belong to a different class, among all examples which are not of class x . In the matrix, this is the column sum of class x minus the diagonal element, divided by the rows sums of all other classes; i.e. $3/5=0.6$ for class yes and $2/9=0.222$ for class no.
- The **Precision** is the proportion of the examples which truly have class x among all those which were classified as class x . In the matrix, this is the diagonal element divided by the sum over the relevant column, i.e. $7/(7+3)=0.7$ for class yes and $2/(2+2)=0.5$ for class no.
- The **F-Measure** is simply $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$, a combined measure for precision and recall.

Árvore de Decisão

Árvore de Decisão - Exemplo



(Wikipedia)

Árvore de Decisão - Representação

- Cada nó de decisão contém um atributo
- Cada ramo descendente corresponde a um possível valor deste atributo
- Cada folha está associada a uma classe
- Cada percurso na árvore (da raiz a folha) corresponde a uma regra de classificação

Árvore de Decisão - Algoritmo

- Árvores de Decisão classificam instâncias ordenando-as árvore abaixo, a partir da raiz até alguma folha
- A função aprendida é representada por uma Regra de Decisão
- Algoritmo base: ID3 e seu sucessor, o C4.5. O ID3 começa com a questão: "Qual atributo deve ser testado na raiz da árvore?"
- Cada atributo da instância é avaliado usando um teste estatístico para determinar quão bem ele sozinho classifica os exemplos de treinamento

Árvore de Decisão

- Vantagens:
 - Estrutura de fácil manipulação
 - Produzem modelos que podem ser facilmente interpretados por humanos
- Desvantagens:
 - Pouca robustez a dados de grande dimensão
 - Acurácia afetada por atributos pouco relevantes
 - Dificuldade em lidar com dados contínuos

Conceitos

Aprendizado Supervisionado – Classificação

- Matriz de confusão
 - medida do modelo de classificação
 - mostra o número de classificações corretas versus as classificações preditas para cada classe
 - o número de acertos, para cada classe, se localiza na diagonal principal da matriz

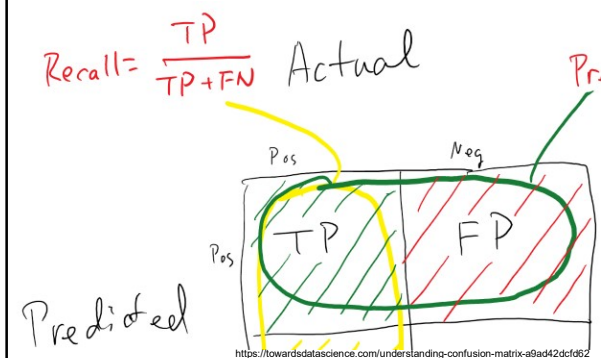
Matriz de confusão

Predictive Model: Evaluation

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

		actual class	
		yes	no
predictive result / classification	yes	tp (true positive)	fp (false positive)
	no	fn (false negative)	tn (true negative)

Matriz de confusão



Matriz de confusão

- ♦ Várias medidas de desempenho podem ser obtidas da matriz de confusão:

- Taxa de acerto $Ac = \frac{TP + TN}{n}$
- Sensibilidade ou revocação (recall) $Sen = \frac{TP}{TP + FN}$
- Especificidade (precision) $Esp = \frac{TN}{FP + TN}$
- F-measure $F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})}$

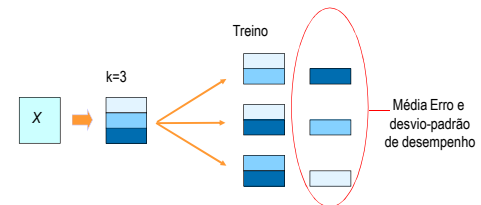
(<http://www.cin.ufpe.br/~igcf/>)

Validação cruzada

- Método de amostragem cross-validation (CV)
 - em *R-fold* cross-validation, os exemplos são aleatoriamente divididos em r partições mutuamente exclusivas (*folds*) de tamanho aproximadamente igual a n/r exemplos
 - os exemplos nos $(r-1)$ *folds* são usados para treinamento e a hipótese induzida é testada no *fold* remanescente
 - este processo é repetido r vezes, cada vez considerando um *fold* diferente para teste
 - o erro na CV é a média dos erros calculados em cada um dos r *folds*

(p. 110)

Validação cruzada – k-fold



- Quando $k = n^\circ$ dados $\Rightarrow$ leave-one-out

(http://www.cin.ufpe.br/~igcf)

Aprendizado Supervisionado – Classificação

- Poda
 - técnica de lidar com ruído e overfitting
 - Pré-poda: durante a geração da hipótese, alguns exemplos de treinamento são deliberadamente ignorados
 - Pós-poda: a hipótese é generalizada por meio da eliminação de algumas partes (ex: corte e de alguns ramos em uma árvore de decisão)

Aprendizado Supervisionado – Classificação

- *Overfitting (sobreajuste)*
 - quando a hipótese é muito específica para o conjunto de treinamento utilizado
 - a hipótese ajusta-se em excesso ao conjunto de treinamento
- *Underfitting*
 - quando a hipótese ajusta-se muito pouco ao conjunto de treinamento

Aprendizado de máquina aplicado no reconhecimento de atividades humanas

Bulling, A., Blanke, U., & Schiele, B. (2014). **A tutorial on human activity recognition using body-worn inertial sensors**. ACM Computing Surveys (CSUR), 46(3), 33.

Reconhecimento de atividades

Table I. Main Characteristics of Human Activity Recognition

Type	Characteristic	
Execution	Offline	The system records the sensor data while the activity is performed afterwards. Typical applications such as health monitoring.
	Online	The system acquires sensor data while the activity is performed. Typically used for activity recognition applications in human-computer interaction.
Generalisation	User independent	The system is optimised for a large number of users.
	User dependent	The system is optimised for a specific user.

(Bulling et al., 2014)

Recognition	Continuous	The system automatically processes or gestures in the stream.
	Isolated (Segmented)	The system assumes that gestures are segmented at the start and end. It only classifies the sensor data into activity classes. The oracle performs cross-modality segmentation and assessing classification performance.
Activities	Periodic	Activities or gestures exhibit a periodic pattern, such as running, rowing, biking, etc. and frequency-domain features are used for classification.
	Sporadic	The activity or gesture occurs sporadically among other activities or gestures. The system isolates the subset of data for classification.
	Static	The system deals with the static pointing gestures.

(Bulling et al., 2014)

Pré-processamento

- Normalização $\tilde{x}_i^j = \frac{x_i^j - \min}{\max - \min}$
 - ex: min-max. todas características ficam no [0, 1]
- Discretização
 - atributo numérico (contínuo) -> discreto (categórico)
- Limpeza
 - dados podem apresentar problemas provenientes da coleta; tratamento de atributos incompletos, remover ruídos

Pré-processamento

- Redução de dados
 - pode ser realizada de três formas:
 - número de exemplos
 - número de atributos
 - número de valores
 - Discretização
 - Suavização

Pré-processamento

- Resample
 - Produces a random subsample of a dataset using either sampling with replacement or without replacement.
 - The original dataset must fit entirely in memory. The number of instances in the generated dataset may be specified. The dataset must have a nominal class attribute. If not, use the unsupervised version. The filter can be made to maintain the class distribution in the subsample, or to bias the class distribution toward a uniform distribution.

Extração de características

Feature	
Minimum	min
Maximum	max
Range	range
Sum	sum
Mean	μ
Variance	$\sigma^2 = \frac{1}{n} \sum (x_i - \mu)^2$
Standard deviation	$\sigma = \sqrt{\sigma^2}$
Upper/Lower bounds (95%)	U/L
Skewness	SKEW : $\frac{1}{n} \sum \frac{(x_i - \mu)^3}{\sigma^3}$

Medidas da forma da distribuição:
Skewness = assimetria
Kurtosis = gordura das caudas

(Sells et al., 2018)

Referências

- Bulling, A., Blanke, U., & Schiele, B. (2014). **A tutorial on human activity recognition using body-worn inertial sensors**. ACM Computing Surveys (CSUR), 46(3), 33.
- Sells, V., & Marshall, A. (2018). **A Classification-Based Algorithm to Detect Forged Embedded Machines in IoT Environments**. IEEE Systems Journal, (99), 1-11.
- Powers, David. (2007). **Evaluation: From precision, recall and fmeasure to roc, informedness, markedness and correlation**. Journal of Machine Learning Technologies. 2. 37-63.

Referências

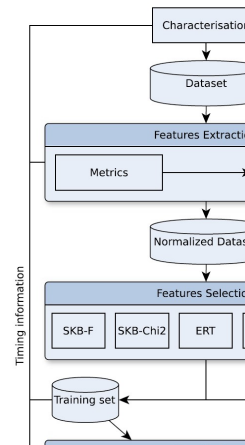
- Gu, F., Kealy, A., Khoshelham, K., & Shang, J. (2015). **User-independent motion state recognition using smartphone sensors**. *Sensors*, 15(12), 30636-30652.

Referências

- UC Irvine Machine Learning Repository: <http://archive.ics.uci.edu/ml/>
- <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
- Notas de aula. Prof. Ivan Gesteira Costa Filho. Centro de Informática. Universidade Federal de Pernambuco

FIM

alessandro.copetti@gmail.com



(Selis et al., 2018)