

Classificadores

Prof. Flávio Varejão

Departamento de Informática

Universidade Federal do Espírito Santo

Conteúdo

- Introdução
- Classificação
 - ZeroR
 - OneR
 - Regressão Logística
 - Árvores de Decisão e Regras
 - Vizinho Mais Próximo
 - Naive Bayes
 - Redes Neurais
- Considerações Finais

Introdução

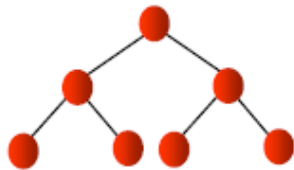
Condicionantes

- Tarefa Definida
 - Classificação
 - Regressão
 - Associação
 - Agrupamento
- Dados Preparados
 - Limpeza
 - Integração
 - Transformação
 - Redução

Modelos

- Aplicação do Método a Dados produz Modelo
 - Árvores de Decisão
 - Redes Neurais
- Método e Dados podem ser o próprio Modelo
 - Vizinho Mais Próximo

Representação de Modelos



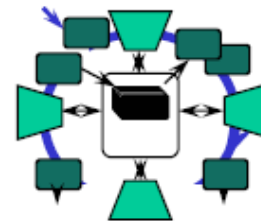
Árvores de decisão

```
If a = 2
    then classe=bom
If b = 2 and c = quente
    then classe=ruim
```

Regras



Modelos não
lineares



Modelos baseados em
distâncias (CBR & k-NN)

Métodos de Aprendizado de Classificadores

Classificador ZeroR

ZeroR

- ZeroR (Zero Regras)
 - Classifica sempre na classe majoritária
 - Serve como classificador baseline
 - Somente com bases razoavelmente balanceadas
 - Não deve ser aplicado em bases extremamente desbalanceadas
 - Detecção de anomalias
 - Quase todos normais
 - ZeroR tem acurácia alta mas não é informativo

Classificador OneR


OneR

- OneR (Regras de Uma Característica)
 - Usa característica com maior valor preditivo para classificar
 - Serve como classificador baseline

Which one is the best predictor ?				
Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

OneR

- Selecione a característica com maior acurácia

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

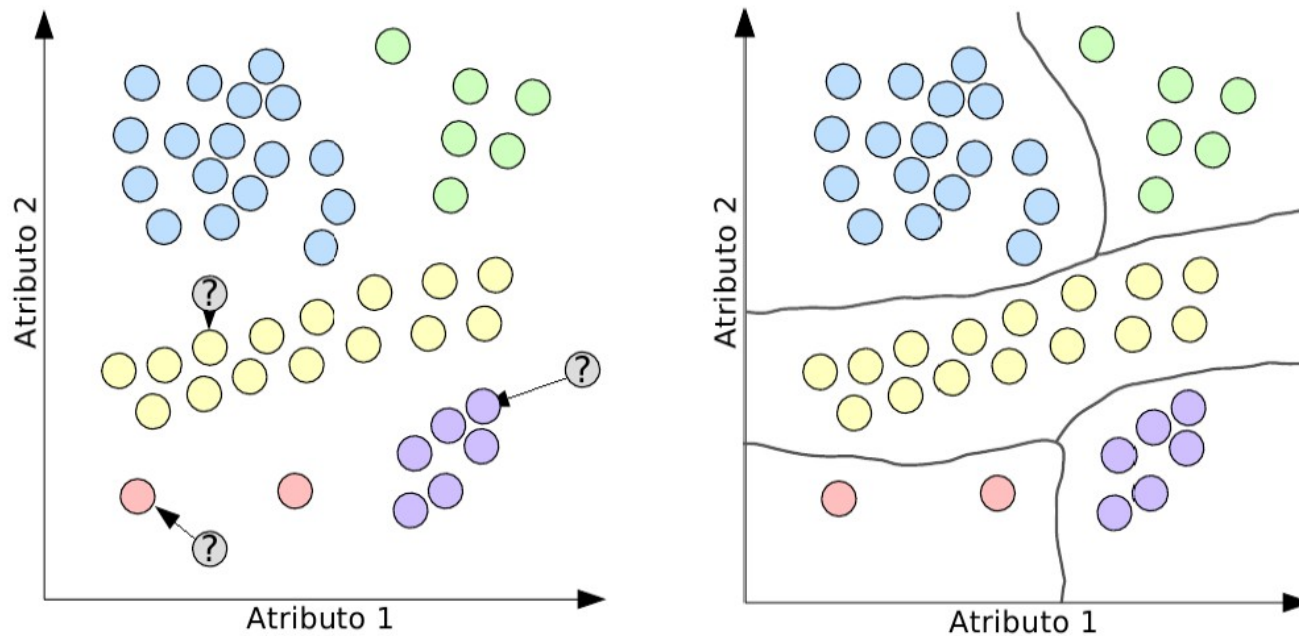
IF Outlook = Sunny THEN PlayGolf = Yes
IF Outlook = Overcast THEN PlayGolf = Yes
IF Outlook = Rainy THEN PlayGolf = No

Vizinho Mais Próximo

Vizinho Mais Próximo

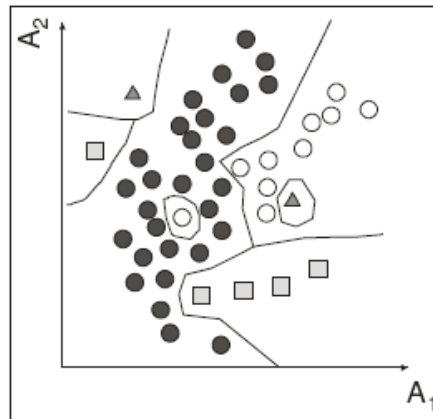
- NN (Nearest Neighbor)
- Bastante Intuitivo
 - Se uma instância de classe desconhecida estiver bem próxima de uma de classe conhecida, as classes devem ser as mesmas.
- Não há etapa de geração de modelos
 - Modelo são os dados e o método
- Necessária medida de distância
 - Normalmente a Euclidiana

Vizinho Mais Próximo



Vizinho Mais Próximo

- Vantagens
 - Dispensa fase de treinamento
 - Aplicável a qualquer tipo de distribuição
- Desvantagem
 - Muito sujeito a influência de ruído
 - Outliers



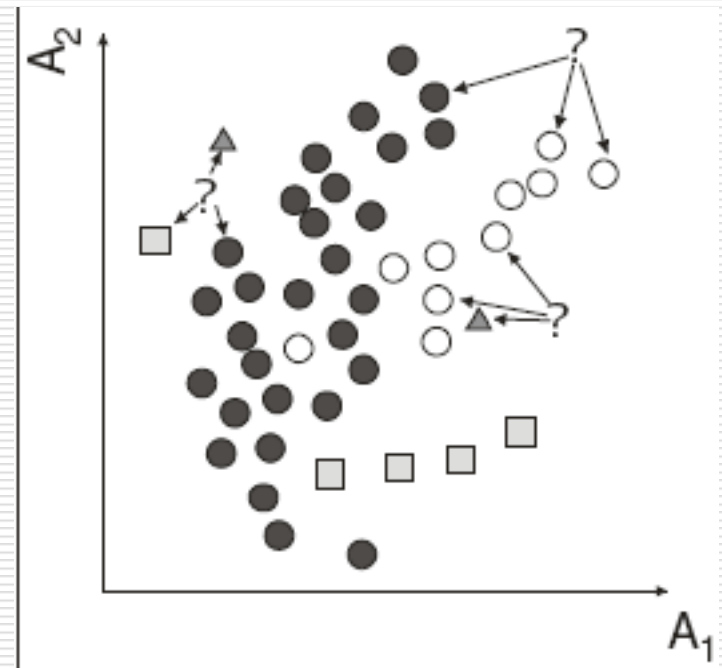
K-Vizinhos Mais Próximos (KNN)

- Extensão do conceito de vizinho mais próximo
 - Ao invés de considerar *um* vizinho mais próximo, considera K vizinhos
- Classificação
 - Seleciona-se as K amostras mais próximas e escolhe-se a classe da maioria
- Vantagem
 - Reduz a influência de ruído

K-Vizinhos Mais Próximos

□ Exemplo

- $K = 3$
- Outliers ($K=1$) deixam de influenciar na classificação
- Superfície de separação de classes é “suavizada”



Questões KNN

□ K

– Definição do Valor

- Fixo
- Variável
 - Mútuo: só conta os ($n < k$) exemplos que são vizinhos entre si
 - Não Mútuo: conta os k vizinhos + m de quem é vizinho

Questões KNN

- Distância
 - Métrica
 - Euclidiana
 - Normalização
 - Características Nominais
 - » 0 ou 1
 - Hamming
 - Só quando todas características nominais
 - Soma das diferenças
 - Manhattan
 - Minkowski

Questões KNN

□ Classificação

- Votação por maioria simples
 - Empate
 - Aleatório
 - Aumento de K (1 a 1 até desempate)
- Proximidade
 - Peso maior para mais próximos
 - $1/d$ ou $1 - d$
 - Distância d absoluta ou normalizada

Regressão Logística

Classificador Regressão Logística

