

Classificação

Eduardo Raul Hruschka

Agenda:

- Conceitos de Classificação
- Técnicas de Classificação
 - *One Rule* (1R)
 - Naive Bayes (com seleção de atributos)
 - Árvores de Decisão
 - K-Vizinhos Mais Próximos (K-NN)
- Super-ajuste e validação cruzada
- Combinação de Modelos

Classificação

- Tarefa: dado um conjunto de exemplos pré-classificados, induzir um modelo/classificador para novos casos.
- Aprendizado Supervisionado: classes são conhecidas para os exemplos usados para construir o modelo/classificador
- Um classificador pode ser um conjunto de regras lógicas, uma árvore de decisão, um modelo Bayesiano, uma rede neural, etc.
- Aplicações típicas: aprovação de crédito, marketing direto, detecção de fraude, ...

Abordagem prática:

- Algoritmos simples frequentemente funcionam muito bem na prática. Além disso:
 - Menor tempo de construção do modelo;
 - Combinação (*ensembles*) de algoritmos simples;
 - *Baselines*.
- Sugestão:
 - Usar um único atributo (melhor discriminador – 1R);
 - Usar todos os atributos, assumindo independência condicional (*Naive Bayes*);
 - Árvores de Decisão (interpretabilidade)
 - Regressão Logística;
 - K-NN;
 - Modelos/Algoritmos mais sofisticados.
- Sucesso de cada algoritmo é dependente do domínio de aplicação (*No Free Lunch Theorems*)

Exemplo Pedagógico:

Outlook	Temperature	Humidity	Windy	Play
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
rainy	63	84	true	?

*Weather Data** :

Considerando-se dados históricos, construir um modelo para os valores do atributo meta *play*.

Alternativas:

- Encontrar uma função discriminadora, $f(\mathbf{x})$, que mapeia cada \mathbf{x} em um rótulo de classe. Ex: $f(\mathbf{x})=0$ para C_1 e $f(\mathbf{x})=1$ para C_2 .
- Modelar a distribuição de probabilidades *a posteriori* de classes, $P(C_k/\mathbf{x})$, diretamente, usando modelos discriminativos.
- Inicialmente encontrar as densidades condicionais de classe, $P(\mathbf{x}/C_k)$, bem como $P(C_k)$, individualmente para cada classe, e depois usar o Teorema de Bayes:

$$P(C_k | \mathbf{x}) = \frac{P(\mathbf{x} | C_k)P(C_k)}{P(\mathbf{x})}$$

- Equivalente a encontrar $P(\mathbf{x}, C_k)$ – modelos geradores.

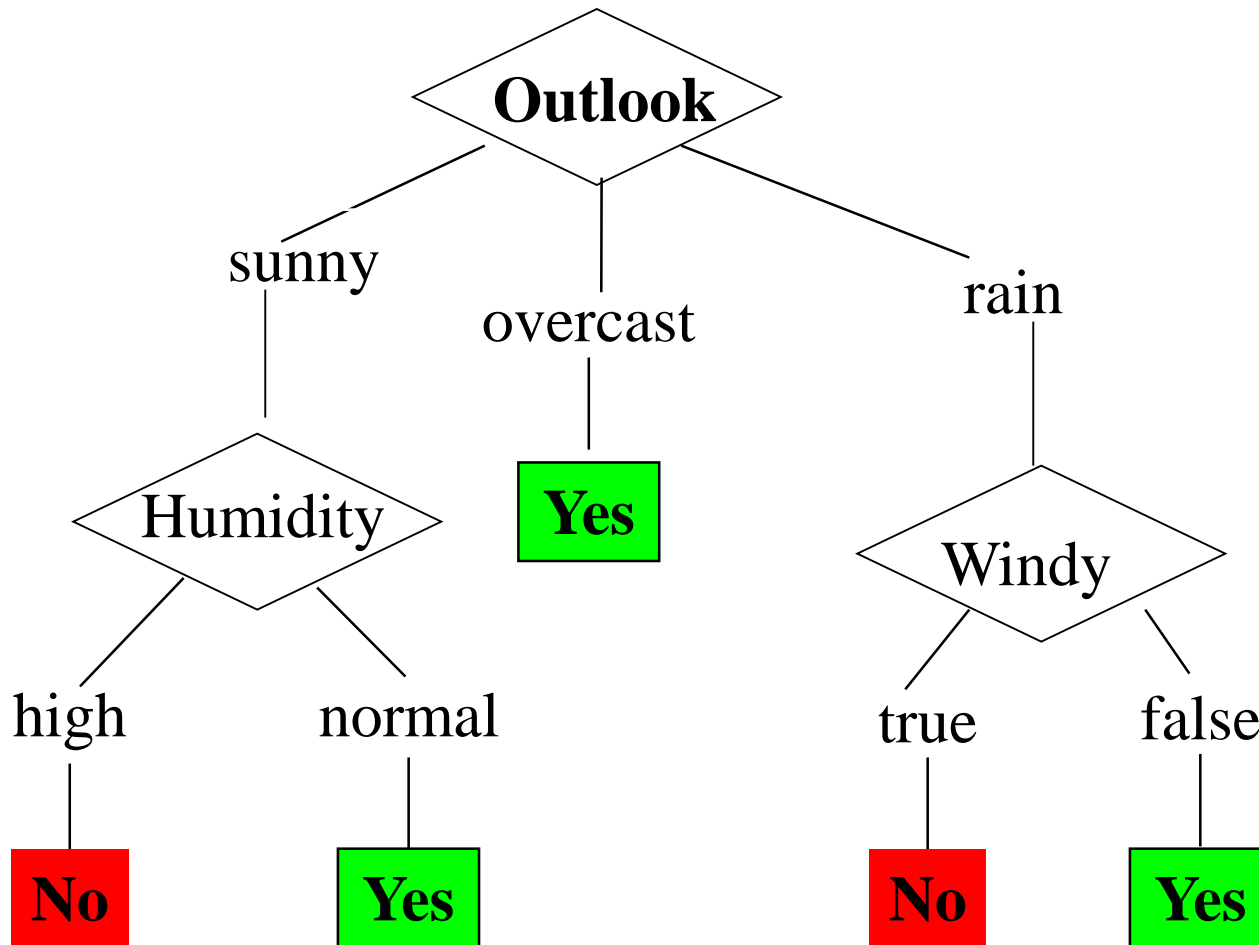
Méritos relativos:

- Modelos geradores:
 - Computacionalmente pesada e, se \mathbf{x} possui alta dimensionalidade, precisaremos de grandes amostras;
 - Permite estimar a densidade marginal dos dados, $P(\mathbf{x})$, que é útil para detectar novos dados que possuem baixa probabilidade dado o modelo (*outlier detection, novelty detection*).
- Modelos discriminativos:
 - Particularmente interessante se somente estamos interessados em $P(C_k/\mathbf{x})$, e não em $P(\mathbf{x}, C_k)$.
- Função discriminadora:
 - Alternativa mais simples, mas que causa perda considerável de informação (*e.g., reject option, combinação de modelos, etc.*)

Classificador 1R:

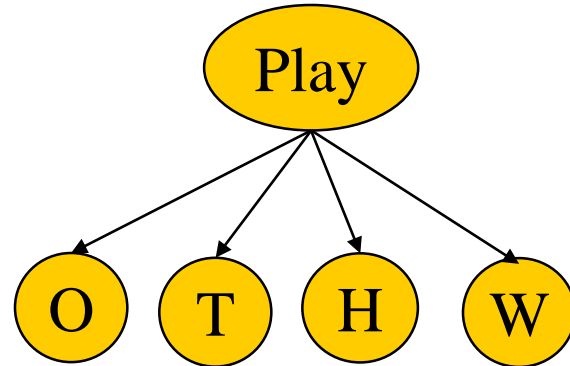
Atributo	Regra	Erro Específico	Erro Total
Outlook	Sunny → No	2/5	4/14
	Overcast → Yes	0/4	
	Rainy → Yes	2/5	
Temperature	$\leq 77.5 \rightarrow$ Yes	3/10	5/14
	$> 77.5 \rightarrow$ No	2/4	
Humidity	$\leq 82.5 \rightarrow$ Yes	1/7	3/14
	> 82.5 and $\leq 95.5 \rightarrow$ No	2/6	
	$> 95.5 \rightarrow$ Yes	0/1	
Windy	False → Yes	2/8	5/14
	True → No	3/6	

Árvore de Decisão:



- Algoritmo baseado em *Teoria da Informação*;
- Classificador resultante usualmente fornece boa interpretabilidade.

Naïve Bayes:



- Modelo Gráfico Probabilístico (Rede Bayesiana):
 - Grafo acíclico direcionado:
 - Nós representam variáveis aleatórias;
 - Arestas representam “influências”
- Interpretável?

Qual classificador escolher?

- Existem centenas de algoritmos disponíveis em softwares livres e proprietários;
- Escolha do melhor classificador para um novo problema é, usualmente, uma questão empírica e extremamente dependente do problema em si;
- Experimentar com classificadores variados;
- Por onde começar?
 - Análise exploratória de dados (*Estatística Descritiva*);
 - Algoritmos mais simples antes e, se necessário, aplicar algoritmos mais complexos (e mais caros computacionalmente).
 - Problemas difíceis normalmente requerem soluções sofisticadas: adaptação e/ou desenvolvimento de novos algoritmos particularmente adaptados ao problema que se tem em mãos.
 - Princípio *KISS* (*Keep it stupid simple*) fornece *baselines* como efeito colateral positivo e, não raramente, economiza tempo de modelagem.

Voltando à nossa agenda...

- Conceitos de Classificação
- Técnicas de Classificação
 - *One Rule* (1R)
 - Naive Bayes (com seleção de atributos)
 - Árvores de Decisão
 - K-Vizinhos Mais Próximos (K-NN)
- Super-ajuste e validação cruzada
- Combinação de Modelos