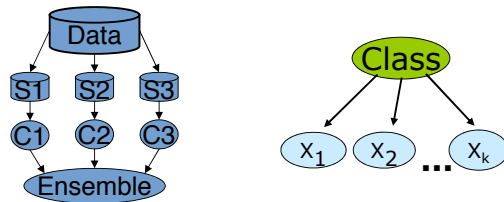


Classificadores Bagging, Boosting e Naïve Bayes

Stanley Robson de M. Oliveira



Bagging/Boosting

Naïve Bayes

Resumo da Aula

Bagging e Boosting:

- Comitês de classificadores;
- Bagging com Árvores de Decisão;
- Boosting;
- RandomForest.

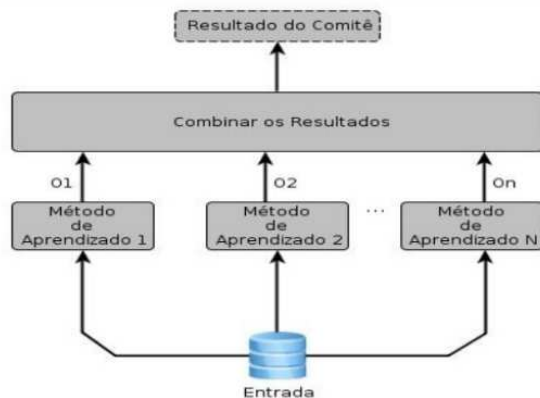
Classificação Bayesiana:

- Arcabouço probabilístico;
- Classificador Naïve Bayes.

Medidas de Avaliação de Modelos:

- Hold-out, cross-validation, percentage split;
- Ajuste de hiperparâmetros;
- Medidas Clássicas.

Comitês de Classificadores



Bagging e Boosting

■ Métodos de **aprendizado estatístico** para solucionar problemas de classificação.

■ **Comitês de classificadores**: vários classificadores ajustados aos dados e combinados para aumentar poder de predição.

■ **Bagging**: implementação da estratégia de reamostragem bootstrap; um novo classificador por amostra.

■ **Boosting**: classificadores ajustados em versões diferentes do conjunto de dados originais.

Por que funciona?

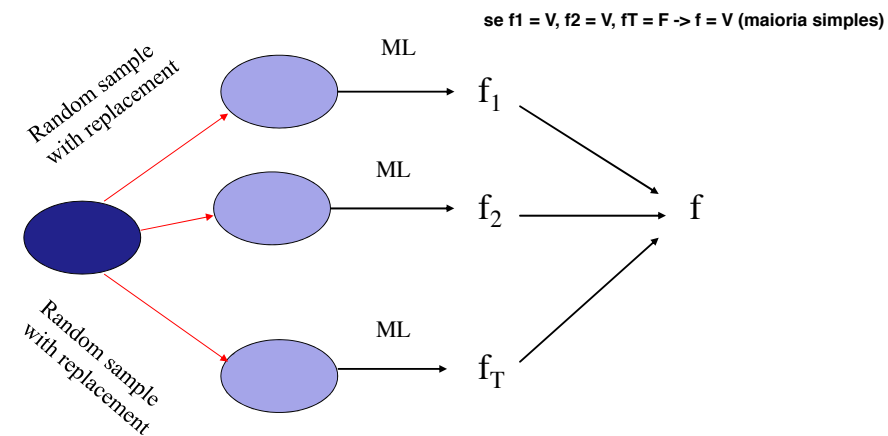
- Suponha que existam 25 classificadores base.
- Cada classificador com taxa de erro $t = 0,35$.
- Assuma **independência** entre os classificadores.
por exemplo, se eu disser um número, a próxima pessoa a chutar não vai tentar deixar próximo do meu
- Se a decisão do comitê é por **maioria simples**:

qual a probabilidade de 13,14,15,..., 25 (a maioria) errarem?

$$\sum_{i=13}^{25} \binom{25}{i} t^i (1-t)^{25-i} = 0,06$$

- A probabilidade da **maioria errar** é menor do que cada classificador individualmente.

Bagging



Bagging ...

- Bagging = **Bootstrap aggregating** (Breiman, 1996).
- Procedimento para **reduzir a variância** de um método de aprendizado de máquina.
- Algoritmo:**
 1. Criar M **amostras bootstrap** dos dados.
 2. Ajustar um modelo para **cada amostra**.
 3. Tirar a média das previsões (ou **votos na classificação**).
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 26(2), 123-140.

Variância versus Viés

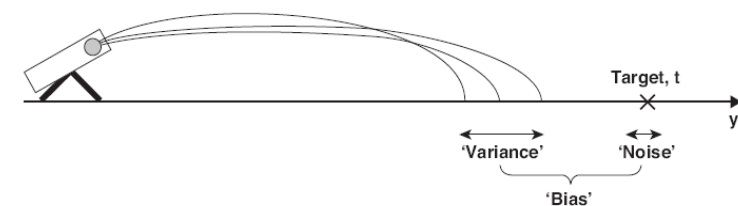


Figura: Decomposição viés/variância

- Viés (**Bias**): maior parte é erro de treinamento; modelos mais complexos têm menor viés.
- Variância**: variação em torno da média.
- Bagging**: diminui somente a **variância**.

Bagging ...

Amostra com reposição:

Dados	1	2	3	4	5	6	7	8	9	10
Amostra 1	7	9	3	9	9	1	10	7	7	5
Amostra 2	2	5	8	10	2	3	6	3	7	7
Amostra 3	1	6	4	2	4	1	10	2	5	1

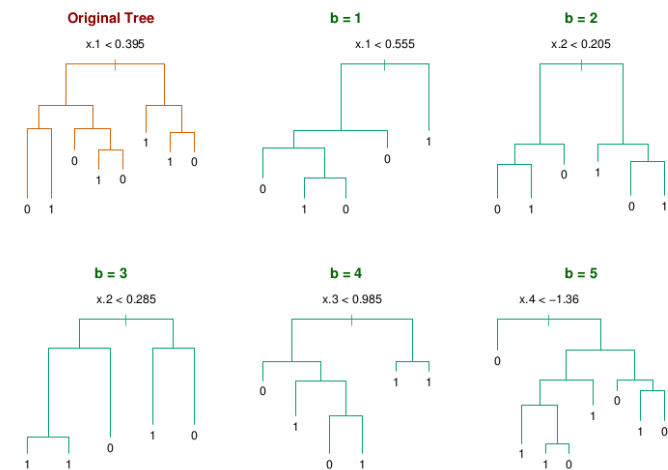
■ Cada valor tem probabilidade $(1 - \frac{1}{n})^n$ de **não** ser selecionado.

■ Conjunto de dados = $1 - (1 - \frac{1}{n})^n$ dos dados originais.

$$\lim_{n \rightarrow \infty} 1 - (1 - \frac{1}{n})^n \approx 1 - e^{-1} = 0.632$$

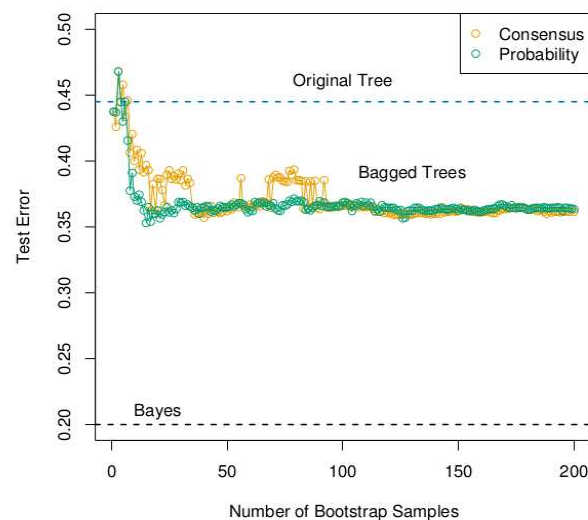
■ Isto é, em geral uma **amostra bootstrap** corresponde a aproximadamente 63% do conjunto original.

Bagging: exemplo

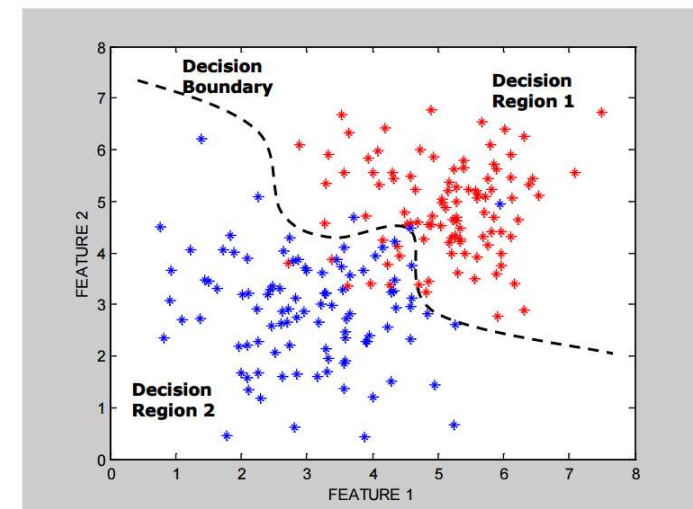


Cada **modelo** em cada **amostra** gera classificadores diferentes.

Taxa de Erro vs No. de Modelos

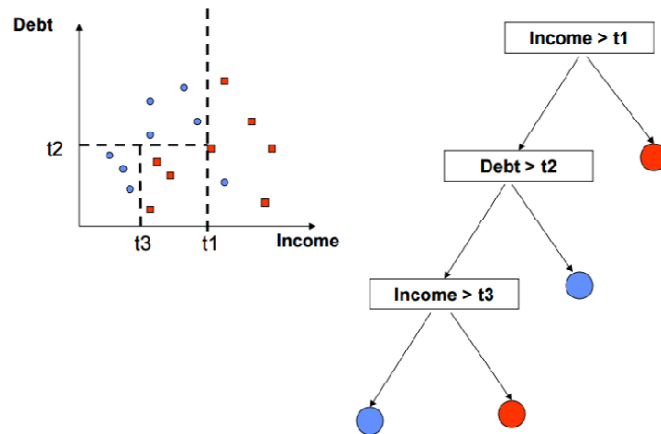


Fronteiras de Decisão



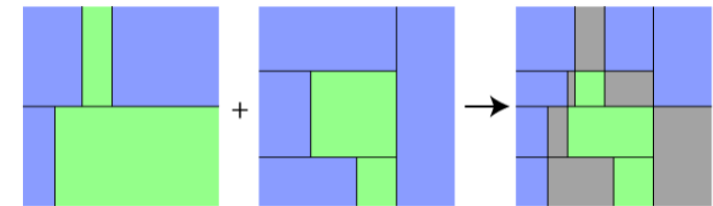
Árvore de Decisão

Exemplo: Árvore de Decisão



Combinando Árvores

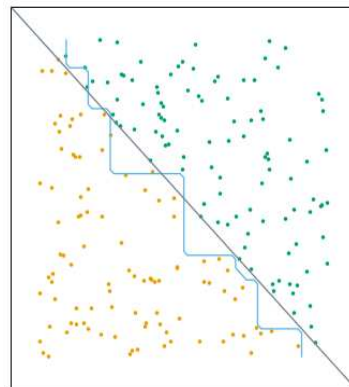
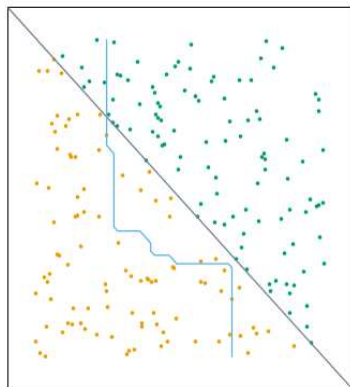
- Azul: acordo (A)
- Verde: acordo (B)
- Cinza: empate



Fronteiras de Decisão ...

Bagged Decision Rule

Boosted Decision Rule



Linear

Bagging/Boosting

Por que com árvores de decisão?

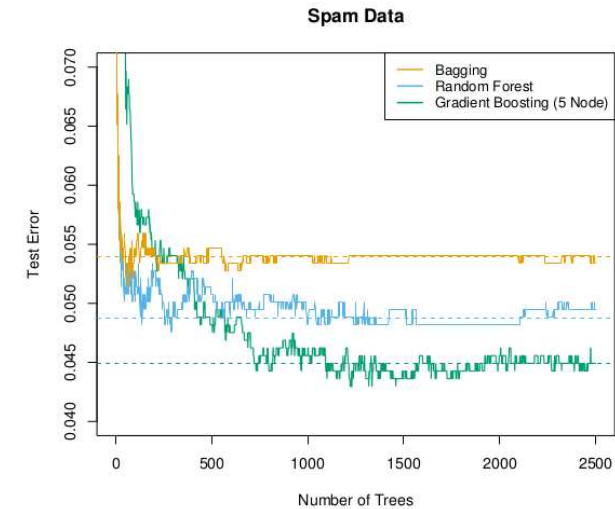
TABLE 10.1. Some characteristics of different learning methods. Key: ▲ = good, ◆ = fair, and ▼ = poor.

Characteristic	Neural Nets	SVM	Trees	MARS	k-NN, Kernels
Natural handling of data of "mixed" type	▼	▼	▲	▲	▼
Handling of missing values	▼	▼	▲	▲	▲
Robustness to outliers in input space	▼	▼	▲	▼	▲
Insensitive to monotone transformations of inputs	▼	▼	▲	▼	▼
Computational scalability (large N)	▼	▼	▲	▲	▼
Ability to deal with irrelevant inputs	▼	▼	▲	▲	▼
Ability to extract linear combinations of features	▲	▲	▼	▼	◆
Interpretability	▼	▼	◆	▲	▼
Predictive power	▲	▲	▼	◆	▲

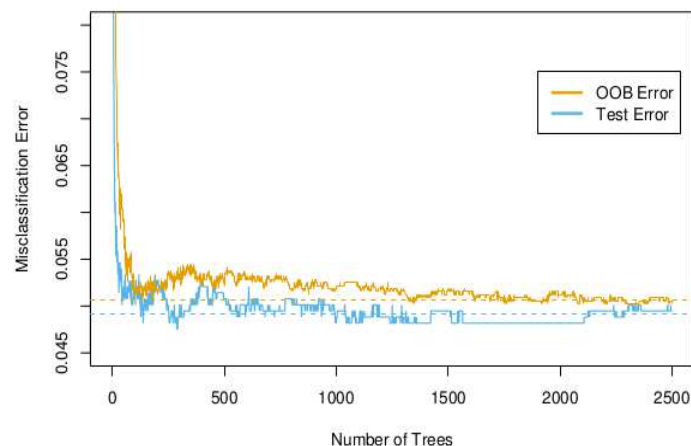
RandomForest

- Bagging é **geral**, mas o RandomForest é **específico** para **árvores de decisão**.
- Introduz **duas fontes de variabilidade**: “Bagging” e “seleção aleatória de atributos”.
- Bagging**: cada árvore é ajustada em uma **amostra bootstrap**.
- Seleção aleatória de atributos: **em cada nó**, o melhor split é escolhido de uma **amostra aleatória de m** atributos ao invés de todos.

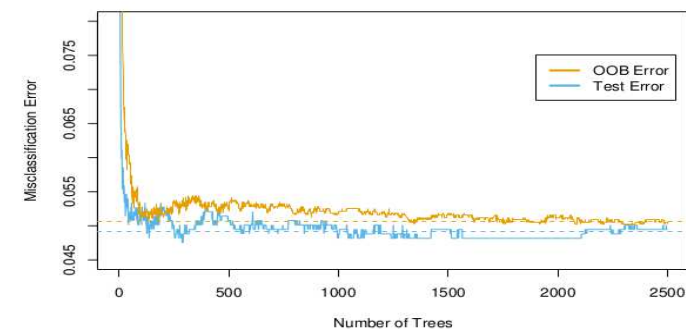
RandomForest vs Bagging



Taxa de Erro vs OOB



Taxa de Erro vs OOB ...

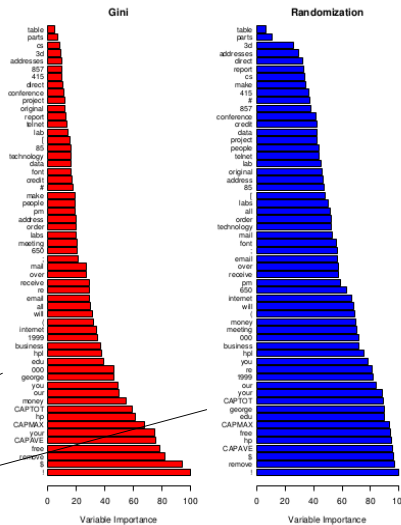


- OOB = Out Of the Bag**. Em cada amostra bootstrap, algumas observações (1/3) ficam de fora.
- Estas **observações** podem ser usadas para avaliar o modelo, com acurácia próxima de um **conjunto de teste**.

Importância das Variáveis

- Por meio das **múltiplas árvores** ajustadas é possível determinar as **variáveis** mais **importantes**.

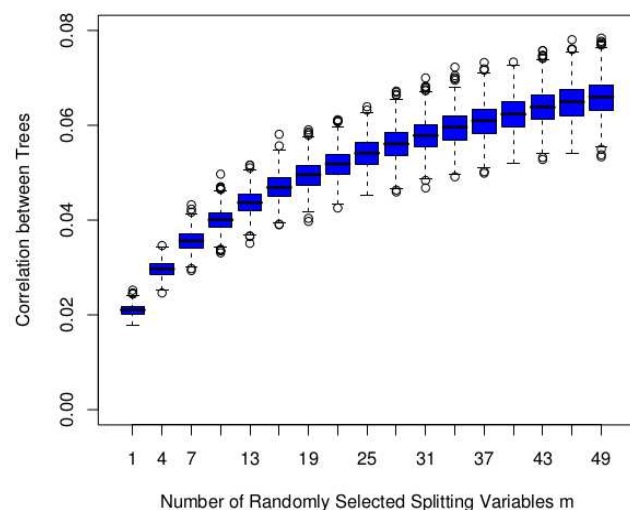
- **Índice Gini**
- **Entropia**



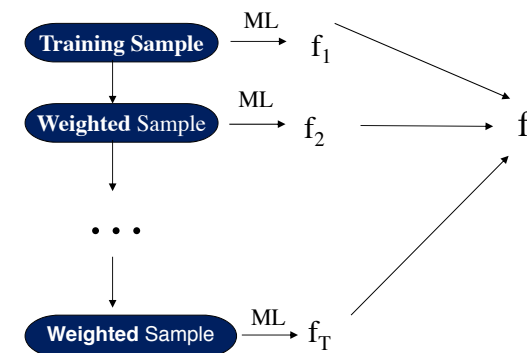
Hiperparâmetros

- Há dois parâmetros basicamente: **número de árvores** e **m** (variáveis aleatoriamente selecionadas para determinar o **split** no nó).
- Número de árvores ≥ 1000 já garante bons resultados.
- **m** pode ser utilizado como: $m = \sqrt{p}$ onde **p** é o número de atributos.
- **m** pode também pode ser determinado por **validação cruzada**.

Hiperparâmetros ...



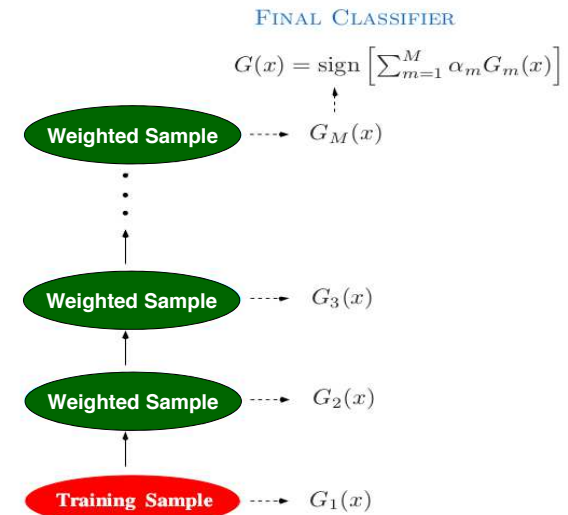
Boosting



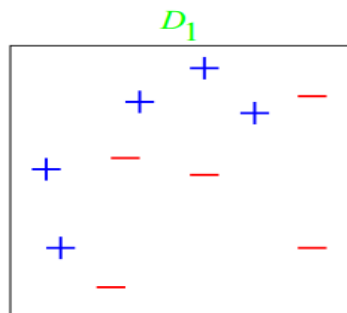
Boosting

- ❑ **Ideia:** transformar múltiplos classificadores **ruins** em um **único muito bom**.
- ❑ Um classificador é considerado **ruim** se for levemente melhor que o **chute**.
- ❑ Modifica a **distribuição dos dados**.
- ❑ Em cada rodada, as classificações mais **difíceis** ganham **maior peso**.

Boosting ...

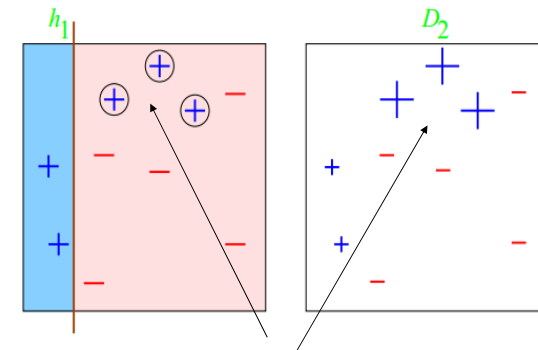


Boosting: Exemplo



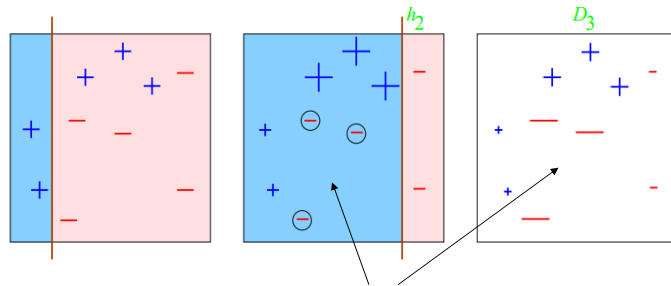
- ❑ Considere como **classificador base**: retas verticais ou horizontais.

Boosting: Exemplo ...



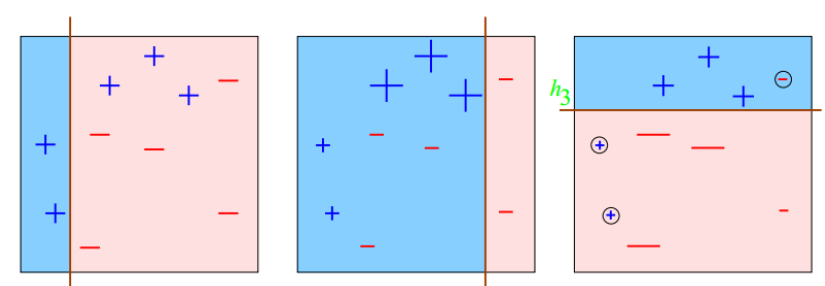
- ❑ Os **pontos classificados errados** têm seus pesos **aumentados** para a próxima fase.

Boosting: Exemplo ...



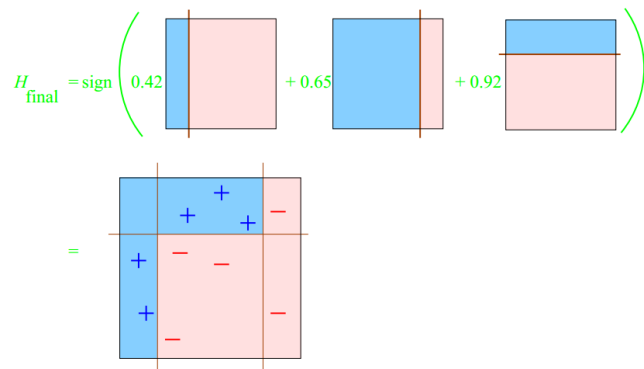
- Na **nova fase de classificação**, devido aos pesos dos erros, a reta vertical foi deslocada, dando origem a novos erros.

Boosting: Exemplo ...



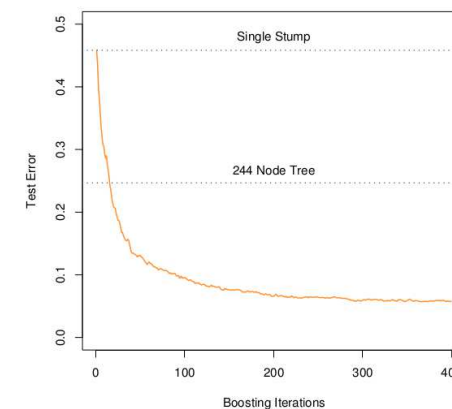
- Na terceira tentativa, **uma reta horizontal** foi escolhida, dessa vez com **menos erros**.

Boosting: Exemplo ...



- Por fim, a **combinação dos classificadores** gera um **classificador global** muito superior, atribuindo pesos distintos a cada classificador.

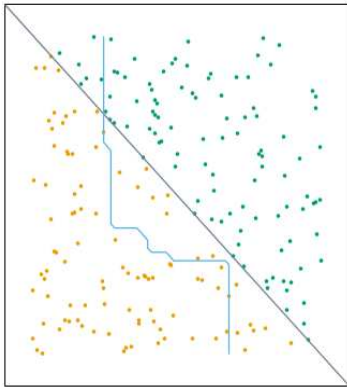
Boosting ...



- Após várias interações com stumps (**árvores com um nó**) a taxa de erro cai e se estabiliza.

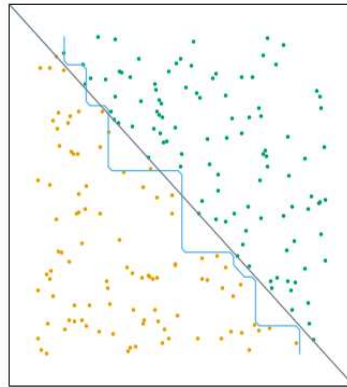
Fronteiras de Decisão

Bagged Decision Rule



Linear

Boosted Decision Rule

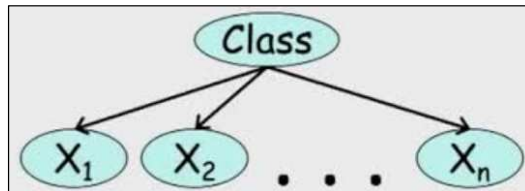


Bagging/Boosting

Boosting ...

- ❑ **Boosting** tem um dos melhores desempenhos dentre os classificadores.
- ❑ Necessita de poucos ajustes.
- ❑ Procedimento embutido para **seleção de variáveis**.
- ❑ Tem um desempenho, em geral, superior ao **RandomForest**.
- ❑ Quando o **RandomForest** se torna melhor? Facilmente paralelizável (**Big Data**).

Naïve Bayes



$$P(x_1, \dots, x_k | C) = P(x_1 | C) \times P(x_2 | C) \times \dots \times P(x_k | C)$$

Classificação Bayesiana ...

- ❑ Um **arcabouço probabilístico** para solucionar problemas de classificação.
- ❑ **Predição probabilística**: Prediz hipóteses múltiplas que são ponderadas por suas probabilidades.
- ❑ **Classificador Bayesiano**: prediz a probabilidade que uma dada amostra pertence a uma determinada classe.
- ❑ **Incremental**: Cada amostra no conjunto de treinamento pode aumentar/diminuir a probabilidade de que uma hipótese é correta.

Classificador "Naïve Bayes"

- ❑ **Suposição:** independência de atributos

$$P(x_1, \dots, x_k | C) = P(x_1 | C) \times P(x_2 | C) \times \dots \times P(x_k | C)$$

- ❑ Se o i-ésimo atributo for **nominal**:
 $P(x_i | C)$ é estimado como a **frequência relativa** das amostras no i-ésimo atributo (x_i) que pertencem à classe C.
- ❑ Se o i-ésimo atributo for **numérico**:
 $P(x_i | C)$ é estimado através de uma **função de densidade Gaussiana (Normal)**.
- ❑ Computacionalmente **fácil** em ambos os casos.

Classificação Bayesiana

- ❑ **Na Prática:** requer conhecimento inicial de muitas probabilidades \Rightarrow **significativo custo computacional**.

- ❑ **Probabilidade Condicional** : $P(C | A) = \frac{P(A, C)}{P(A)}$

$$P(A | C) = \frac{P(A, C)}{P(C)}$$

- ❑ **Teorema de Bayes** :

$$P(C | A) = \frac{P(A | C)P(C)}{P(A)}$$

Classificador "Naïve Bayes": Exemplo

- ❑ **Classes:**

compra_computador = 'sim'
compra_computador = 'nao'

- ❑ **Amostra:**

X = (Idade <= 30,
Renda = media,
Aluno = 'sim'
Credito = 'normal') ?

Idade	Renda	Aluno	Credito	Classe
<=30	alta	nao	normal	nao
<=30	alta	nao	excelente	nao
31...40	alta	nao	normal	sim
>40	media	nao	normal	sim
>40	baixa	sim	normal	sim
>40	baixa	sim	excelente	nao
31...40	baixa	sim	excelente	sim
<=30	media	nao	normal	nao
<=30	baixa	sim	normal	sim
>40	media	sim	normal	sim
<=30	media	sim	excelente	sim
31...40	media	nao	excelente	sim
31...40	alta	sim	normal	sim
>40	media	nao	excelente	nao

Classificador "Naïve Bayes": Exemplo

$P(C_i)$: $P(\text{compra_computador} = \text{"sim"}) = 9/14 = 0.643$
 $P(\text{compra_computador} = \text{"nao"}) = 5/14 = 0.357$

Cálculo de $P(X|C_i)$ para cada classe:

$P(\text{Idade} = \text{"<=30"} | \text{compra_computador} = \text{"sim"}) = 2/9 = 0.222$

$P(\text{Idade} = \text{"<= 30"} | \text{compra_computador} = \text{"nao"}) = 3/5 = 0.6$

$P(\text{Renda} = \text{"media"} | \text{compra_computador} = \text{"sim"}) = 4/9 = 0.444$

$P(\text{Renda} = \text{"media"} | \text{compra_computador} = \text{"nao"}) = 2/5 = 0.4$

$P(\text{Aluno} = \text{"sim"} | \text{compra_computador} = \text{"sim"}) = 6/9 = 0.667$

$P(\text{Aluno} = \text{"sim"} | \text{compra_computer} = \text{"nao"}) = 1/5 = 0.2$

$P(\text{Credito} = \text{"normal"} | \text{compra_computador} = \text{"sim"}) = 6/9 = 0.667$

$P(\text{Credito} = \text{"normal"} | \text{compra_computador} = \text{"nao"}) = 2/5 = 0.4$

$X = (\text{Idade} \leq 30, \text{Renda} = \text{media}, \text{Aluno} = \text{'yes'}, \text{Credito} = \text{'normal'}) ??$

$P(X|C_i)$: $P(X|\text{compra_computador} = \text{"sim"}) = 0.222 \times 0.444 \times 0.667 \times 0.667 = 0.044$

$P(X|\text{compra_computador} = \text{"nao"}) = 0.6 \times 0.4 \times 0.2 \times 0.4 = 0.019$

$P(X|C_i) \cdot P(C_i)$: $P(X|\text{compra_computador} = \text{"sim"}) \cdot P(\text{compra_computador} = \text{"sim"}) = 0.028$

$P(X|\text{compra_computador} = \text{"nao"}) \cdot P(\text{compra_computador} = \text{"nao"}) = 0.007$

Portanto, X pertence a classe ("compra_computador = sim").

Classificador "Naïve Bayes": Exemplo

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: attributes
M: mammals
N: non-mammals

multiplica todas as probabilidades de todos os atributos

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

O valor do atributo
"Class" é **mammals** ou
non-mammals ?

Como estimar Probabilidades ?

Tid	Retorno	Estado Civil	Renda Anual	Mentiu
1	Sim	Solteiro	125K	Nao
2	Nao	Casado	100K	Nao
3	Nao	Solteiro	70K	Nao
4	Sim	Casado	120K	Nao
5	Nao	Divorciado	95K	Sim
6	Nao	Casado	60K	Nao
7	Sim	Divorciado	220K	Nao
8	Nao	Solteiro	85K	Sim
9	Nao	Casado	75K	Nao
10	Nao	Solteiro	90K	Sim

Distribuição Normal :

$$P(A_i | c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Para cada par (A_i, c_i) .

X(Renda=120; Classe=Nao)?

Se Classe = Nao

Média da amostra = 110

Variância da amostra = 2975

$$P(Renda = 120 | Nao) = \frac{1}{\sqrt{2\pi(54.54)}} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$

Example of Naïve Bayes Classifier

Classe = (Retorno = Nao, Estado_Civil = Casado, Renda = 120K) ?

Classificador Naïve Bayes

$P(\text{Refund}=\text{Yes}|\text{No}) = 3/7$
 $P(\text{Refund}=\text{No}|\text{No}) = 4/7$
 $P(\text{Refund}=\text{Yes}|\text{Yes}) = 0$
 $P(\text{Refund}=\text{No}|\text{Yes}) = 1$
 $P(\text{Marital Status}=\text{Single}|\text{No}) = 2/7$
 $P(\text{Marital Status}=\text{Divorced}|\text{No}) = 1/7$
 $P(\text{Marital Status}=\text{Married}|\text{No}) = 4/7$
 $P(\text{Marital Status}=\text{Single}|\text{Yes}) = 2/7$
 $P(\text{Marital Status}=\text{Divorced}|\text{Yes}) = 1/7$
 $P(\text{Marital Status}=\text{Married}|\text{Yes}) = 0$

For taxable income:

If class=No: sample mean=110
sample variance=2975

If class=Yes: sample mean=90
sample variance=25

- $P(X|\text{Classe}=\text{Nao}) = P(\text{Retorno}=\text{Nao}|\text{Classe}=\text{Nao})$
 $\times P(\text{Casado}|\text{Classe}=\text{Nao})$
 $\times P(\text{Renda}=120\text{K}|\text{Classe}=\text{Nao})$
 $= 4/7 \times 4/7 \times 0.0072 = 0.0024$

- $P(X|\text{Classe}=\text{Yes}) = P(\text{Retorno}=\text{Nao}|\text{Classe}=\text{Sim})$
 $\times P(\text{Casado}|\text{Classe}=\text{Sim})$
 $\times P(\text{Renda}=120\text{K}|\text{Classe}=\text{Sim})$
 $= 1 \times 0 \times 1.2 \times 10^{-9} = 0$

Como $P(X|\text{Nao})P(\text{Nao}) > P(X|\text{Sim})P(\text{Sim})$

Portanto $P(\text{Nao}|X) > P(\text{Sim}|X)$

=> Classe = **Nao**.

Exercício – Naïve Bayes

- Calcular a probabilidade, utilizando o Classificador **Naïve Bayes**, para o problema de previsão de tempo: **jogar tênis**.
X (Aspecto = Sol, Temp = Fria, Umidade = Elevada, Vento = Forte) = ?

Dia	Aspecto	Temp	Umidade	Vento	Joga Tênis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Amena	Elevada	Fraco	Sim
D5	Chuva	Fria	Normal	Fraco	Sim
D6	Chuva	Fria	Normal	Forte	Não
D7	Nuvens	Fria	Normal	Fraco	Sim
D8	Sol	Amena	Elevada	Fraco	Não
D9	Sol	Fria	Normal	Fraco	Sim
D10	Chuva	Amena	Normal	Forte	Sim
D11	Sol	Amena	Normal	Forte	Sim
D12	Nuvens	Amena	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Amena	Elevada	Forte	Não

O problema da Probabilidade Nula

- ❑ O **Classificador Naïve Bayes** requer que cada probabilidade condicional seja **não nula**.

$$P(X | C_i) = \prod_{k=1}^n P(x_k | C_i)$$

- ❑ Caso uma das probabilidades seja **nula**, usa-se a **correção de Laplace**:

$$\text{Original : } P(A_i | C) = \frac{N_{ic}}{N_c}$$

c: número de classes

$$\text{Laplace : } P(A_i | C) = \frac{N_{ic} + 1}{N_c + c}$$

p: probabilidade a priori

$$\text{m - estimate : } P(A_i | C) = \frac{N_{ic} + mp}{N_c + m}$$

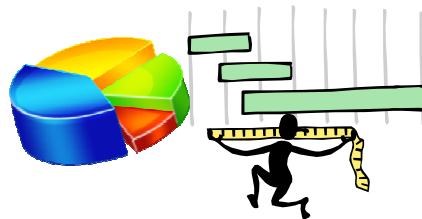
m: parâmetro

Classificador “Naïve Bayes” ...

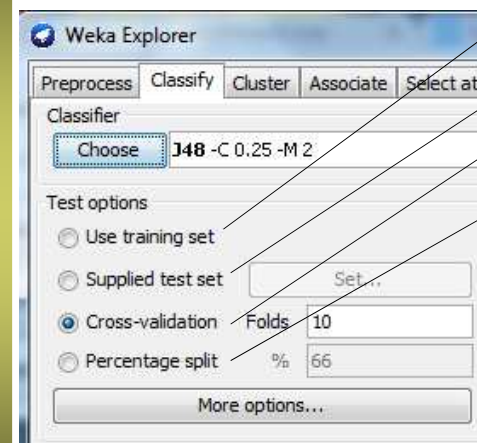
- ❑ Robusto com relação a ruídos (**pontos isolados**).
- ❑ Capacidade de lidar com valores “**missing**” ignorando a observação durante o cálculo de estimativa de probabilidade.
- ❑ Robusto com relação aos **atributos irrelevantes**.
- ❑ Na prática, se um modelo possui **atributos independentes**, o **classificador bayesiano** pode superar as **árvores de decisão**.
- ❑ Por outro lado, a **suposição de independência** pode não funcionar bem para alguns domínios.
- ❑ **Solução**: uso de outras técnicas, tais como Redes de Crença Bayesiana (**BBN**).

Medidas para Avaliação de Modelos de Classificação

Stanley Robson de M. Oliveira

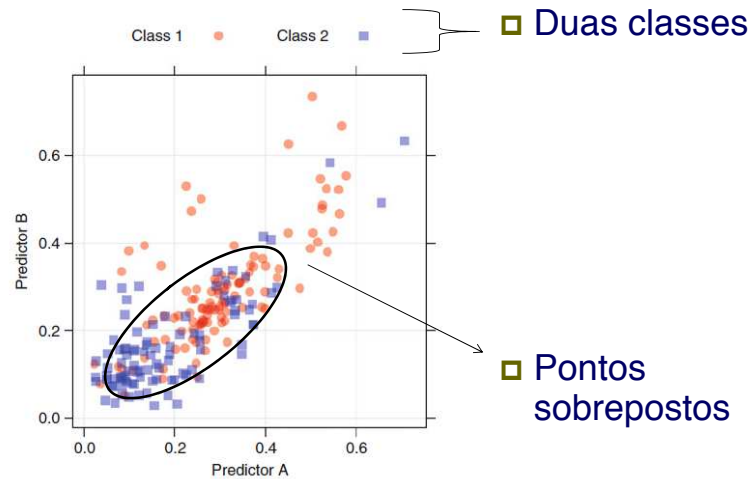


Avaliação de Modelos - WEKA

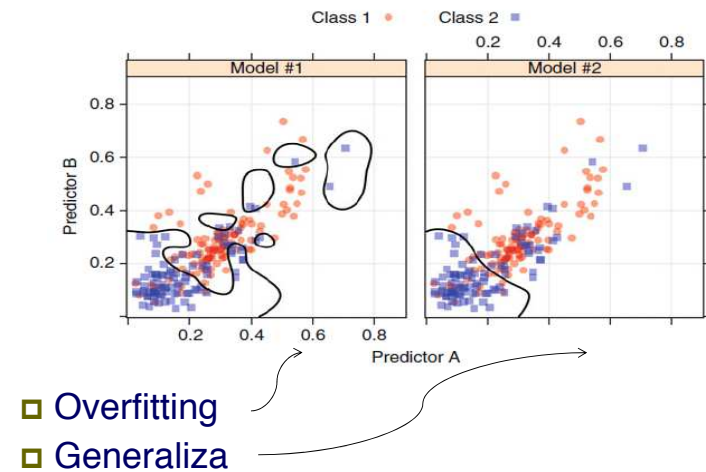


- ❑ **Subestima Erro**
- ❑ **Hold-out**
- ❑ **Validação Cruzada**
- ❑ **Split**

Problema do Overfitting



Problema do Overfitting ...



O que fazer?



- Técnicas de **reamostragem**.
- Busca de medida **confiável** da acurácia.
- Ajuste de **hiperparâmetros**.

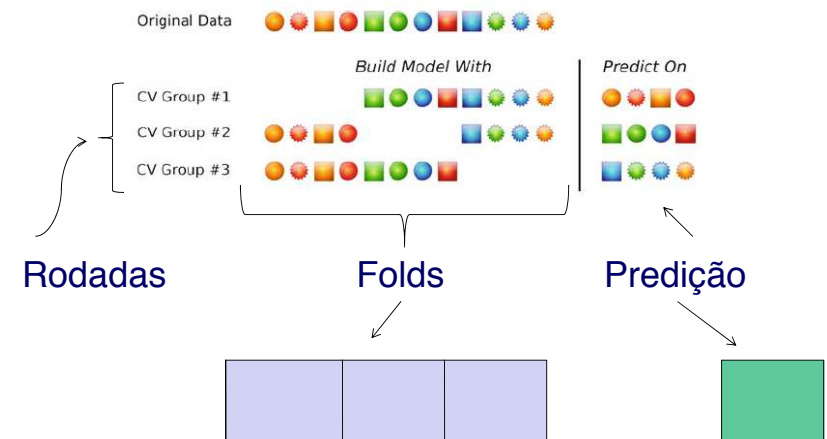
Validação Cruzada

- Técnica para avaliar a **capacidade** de **generalização** de um modelo, a partir de um **conjunto de dados**.
- Há diversas formas de particionamento: **holdout**, **k-fold** e **leave-one-out**.
- Também utilizada para escolha de **hiperparâmetros**.

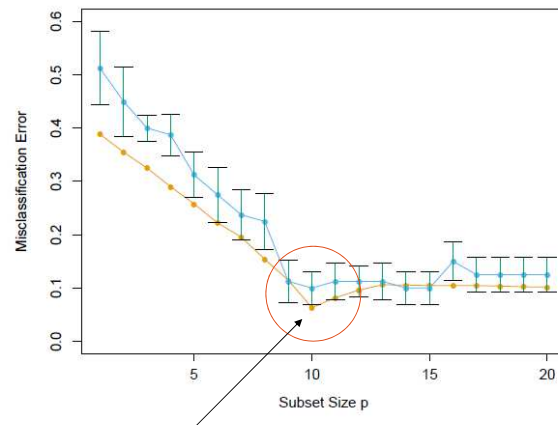
Validação Cruzada (holdout)

- Consiste em **dividir** o conjunto de dados em **dois conjuntos** mutuamente exclusivos.
- Um conjunto para **treinamento** (estimação de **parâmetros**) e outro para **teste** (validação).
- Uma **proporção** muito **comum** é considerar **2/3 para treinamento** e **1/3 para teste**.
- Esta **abordagem** é indicada quando está disponível uma **grande quantidade** de dados.
- Caso o **conjunto de dados** seja **pequeno**, o erro calculado na predição pode **sofrer muita variação**.

Validação Cruzada (k-fold)



Validação Cruzada (k-fold)...

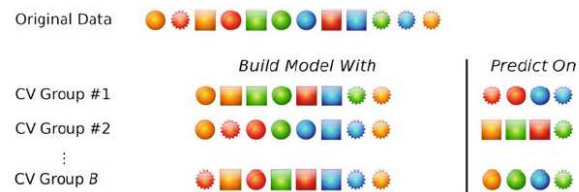


- A partir de 10 observações por **fold**, a estimativa por **Validação Cruzada** se estabiliza.

Validação Cruzada (leave-one-out)

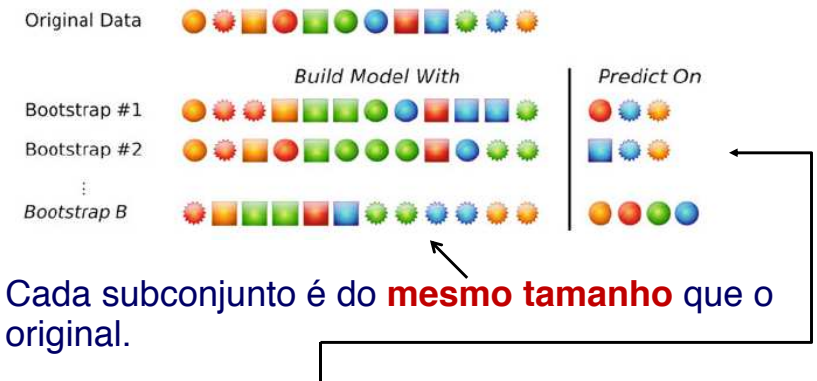
- É um caso específico do **k-fold**, com **k** igual ao número total de dados **N**.
- Apesar de apresentar uma **investigação completa** sobre a **variação do modelo**, possui um **alto custo computacional**, sendo **indicado** para aplicações com **poucos dados**.
- KOHAU, R. **A study of cross-validation and bootstrap for accuracy estimation and model selection**. In International joint Conference on Artificial Intelligence, 1995. v.14, p.1137-1145.

Percentage Split



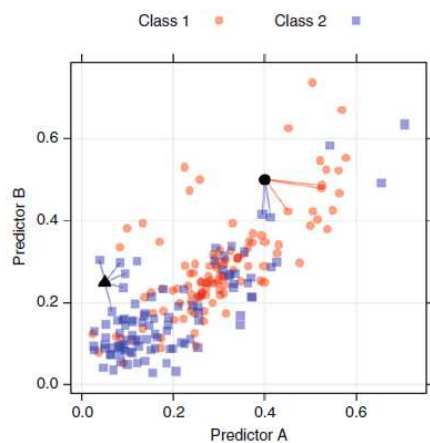
- Observações podem aparecer **múltiplas vezes**, no **treino e no teste**.
- Geralmente são feitas **mais repetições** que **k-fold cross-validation**.

Bootstrap



- Cada subconjunto é do **mesmo tamanho** que o original.
- Amostras **não selecionadas** são utilizadas para avaliar a performance do modelo.

Ajuste de Hiperparâmetros



- Qual é o melhor número de vizinhos mais próximos?
- Se k pequeno: **overfitting**.
- Se k grande: modelo **não sensível**, muito geral.

Ajuste de Hiperparâmetros - WEKA

- No **Weka**, selecione o conjunto de dados **iris.arff**.
- Use o metaclassificador: **CVParameterSelection**:
 - No parâmetro **CVParameters** edite os valores do parâmetro **K**, que será ajustado por **validação cruzada**.
 - No parâmetro **classifier**, selecione o classificador **K-NN (Ibk)** – classificador base.
- Execute o ajuste do parâmetro **K** utilizando a opção: **Cross-validation (Folds = 10)**.
- Qual foi o valor de **K** ajustado por validação cruzada?

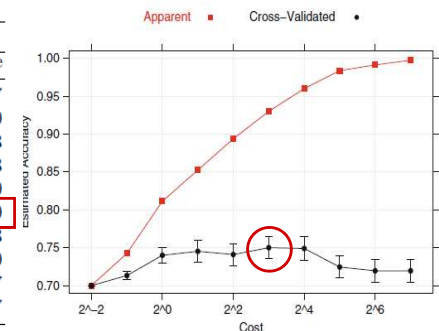
Ajuste de Hiperparâmetros ...

□ Método:

- 1 - Defina os **hiperparâmetros** candidatos.
- 2 - Para cada conjunto de **hiperparâmetros** faça:
 - Reamostre os dados (**CV**, **Bootstrap**, etc);
 - Ajuste um modelo;
 - Obtenha predição e acurácia.
- 3 - Agregue os resultados de acurácia (**ex**: média)
- 4 - Com os **hiperparâmetros** escolhidos, ajuste um modelo usando **todo** o conjunto de dados.

Ajuste de Hiperparâmetros ...

Resampled accuracy (%)			
Cost	Mean	Std. error	% Tolerance
0.25	70.0	0.0	-6.67
0.50	71.3	0.2	-4.90
1.00	74.0	0.5	-1.33
2.00	74.5	0.7	-0.63
4.00	74.1	0.7	-1.20
8.00	75.0	0.7	0.00
16.00	74.9	0.8	-0.13
32.00	72.5	0.7	-3.40
64.00	72.0	0.8	-4.07
128.00	72.0	0.8	-4.07



Recomendações

- Nenhum método de reamostragem é **sempre** o melhor.
- Se o **conjuntos de dados é pequeno**:
 - 1) Avaliação da performance: **CV repetido**.
 - 2) Escolha de modelos: Bootstrap (**baixa variância**).
- Se o conjuntos de dados é grande: **não há grandes diferenças** entre **bootstrap** e **CV**. Por eficiência computacional, escolha CV.

Medidas para Avaliação de Modelos

- **Matriz de Confusão**: um dos resultados de um classificador, após realizar o treinamento e o teste do modelo.
- **Matriz de Confusão** para um problema de duas classes:
 - Classe positiva (C_+)
 - Classe negativa (C_-)

		Predita		
		C_+	C_-	Total
Verdadeira	C_+	VP	FN	P
	C_-	FP	VN	N
Total		P'	N'	P + N

Os bons modelos apresentam **altos valores** na **diagonal principal** da Matriz (VP e VN) e **baixos valores** na **diagonal secundária** (FN e FP).

Medidas para Avaliação de Modelos

1. Taxa de Acerto (Acurácia)

- Porcentagem de exemplos que foram classificados corretamente pelo classificador:

$$\text{taxa de acerto} = \frac{VP + VN}{P + N}$$

	Predita			
	C ₊	C ₋	Total	
Verdadeira	C ₊	VP	FN	P
	C ₋	FP	VN	N
Total	P'	N'	P + N	

Medidas para Avaliação de Modelos

2. Taxa de Erro: 1- Taxa de Acerto.

- Definida como o complemento da **Taxa de Acerto**:

$$\text{taxa de erro} = \frac{FP + FN}{P + N}$$

	Predita			
	C ₊	C ₋	Total	
Verdadeira	C ₊	VP	FN	P
	C ₋	FP	VN	N
Total	P'	N'	P + N	

Medidas para Avaliação de Modelos

3. Sensitividade (Taxa de Verdadeiros Positivos)

- Precisão da Classe C₊:

$$\text{sensitividade} = \frac{VP}{P}$$

	Predita			
	C ₊	C ₋	Total	
Verdadeira	C ₊	VP	FN	P
	C ₋	FP	VN	N
Total	P'	N'	P + N	

Medidas para Avaliação de Modelos

4. Especificidade

- Precisão da Classe C₋: Taxa de Verdadeiros Negativos.

$$\text{especificidade} = \frac{VN}{N}$$

	Predita			
	C ₊	C ₋	Total	
Verdadeira	C ₊	VP	FN	P
	C ₋	FP	VN	N
Total	P'	N'	P + N	

Medidas para Avaliação de Modelos

5. Estatística Kappa

- Mede o **desempenho** do classificador [0, 1].
- É uma **medida de concordância** entre as classes **preditas** e **observadas**, que deduz o número esperado de acerto do classificador.

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

$$p_o = \frac{VP + VN}{P + N} \quad p_e = \frac{(P'P) + (N'N)}{(P + N)^2}$$

Medidas para Avaliação de Modelos

5. Estatística Kappa

- Uma possível interpretação do desempenho dos modelos, a partir da Estatística Kappa, foi introduzida por **Landis e Koch (1977)**:

Estatística Kappa	Qualidade
< 0,00	Péssima
0,00 – 0,20	Ruim
0,21 – 0,40	Razoável
0,41 – 0,60	Boa
0,61 – 0,80	Muito Boa
0,81 – 1,00	Excelente

Medidas para Avaliação de Modelos

6. Taxa de Verdadeiros Positivos

- True Positive Rate (**TPR**).
- Semelhante à medida **Sensitividade**.

$$TPR = \frac{VP}{P}$$

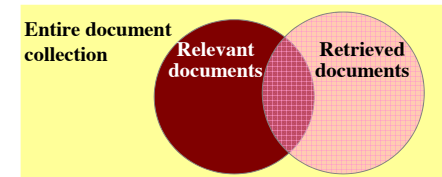
7. Taxa de Falsos Positivos

- False Positive Rate (**FPR**).
- FPR = 1 – **Especificidade**.

$$FPR = \frac{FP}{N}$$

Medidas para Avaliação de Modelos

8. Recall



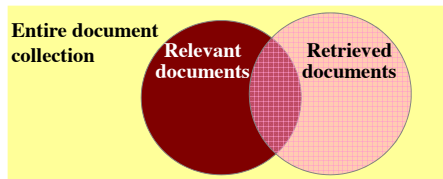
$$recall = \frac{\text{número de documentos recuperados que são relevantes}}{\text{Número total de documentos que são relevantes}}$$

$$recall = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{total relevant documents}\}|}$$

$$recall = TPR = \frac{VP}{P}$$

Medidas para Avaliação de Modelos

9. Precision



$$precision = \frac{\text{número de documentos recuperados que são relevantes}}{\text{Número total de documentos que são recuperados}}$$

$$precision = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

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

Exercício: Dataset Pedra no Rim

- ❑ Considere a **matriz de confusão** abaixo, para **avaliação** de um **classificador binário** gerado.

		Predita		Total
		Sim	Não	
Verdadeira	Sim	13	5	18
	Não	3	29	32
	Total	16	34	50

- ❑ **Calcular:** Acurácia, Erro, Sensitividade, Especificidade, Kappa, Taxa de TP e Taxa de FP para cada Classe, Precisão para cada classes (Positiva e Negativa).

Matriz de Confusão para 3 Classes

- ❑ Utilizando o software **WEKA**, selecione o dataset **IRIS**. Em seguida, construa uma **árvore de decisão** usando o algoritmo **J48**. Depois tente entender como foi feito o cálculo para as medidas a seguir:
- Acurácia;
 - Erro;
 - Kappa;
 - Taxa de TP para cada Classe;
 - Taxa de FP para cada Classe;
 - Precisão para cada Classe.