



Universidade do Minho
Departamento de Informática

Avaliação de Modelos e Métricas de Qualidade

ADI³ - LEI/MiEI @ 2024/2025, 2º sem



- Classificação e Regressão
- Avaliação de Modelos
- Métricas de Qualidade





Classificação

- *“Classification in machine learning is a supervised learning approach in which we learn from the data given to it and make new observations or classifications.”*
- Para uma coleção de dados (registos/conjunto de treino)
- Cada registo é caracterizado por uma tuplo (x, y) , onde x é o conjunto de atributos e y é a classe ou categoria atribuída:
 - x : atributo, preditor, variável independente, entrada
 - y : classe, resposta, variável dependente, saída
- Tarefa:
 - Aprender um modelo que mapeia cada conjunto de atributos x em um das classes predefinidas de y

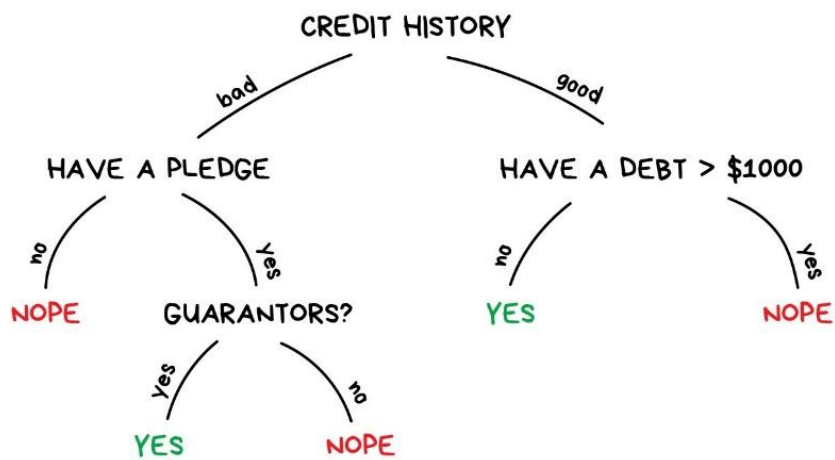




Classificação

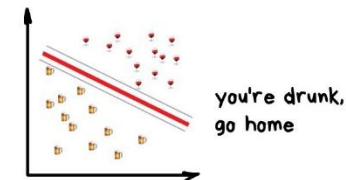
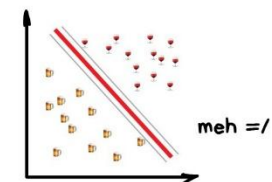
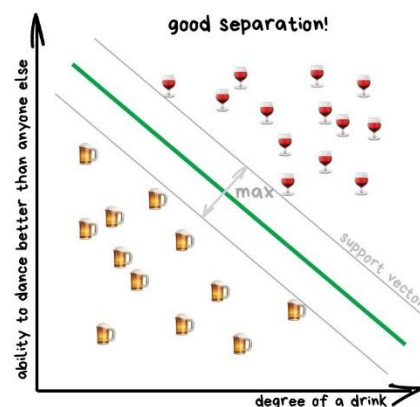


GIVE A LOAN?



DECISION TREE

SEPARATE TYPES OF ALCOHOL

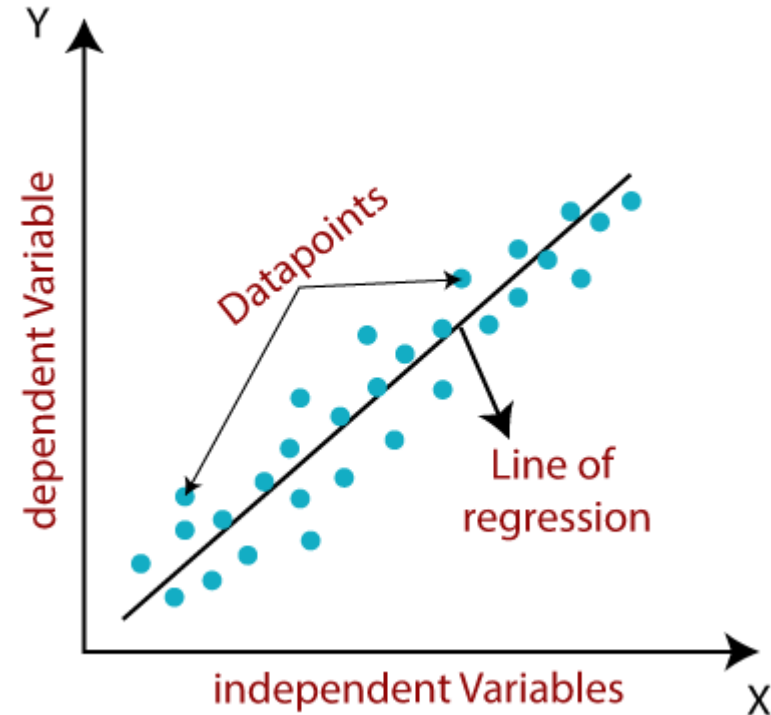


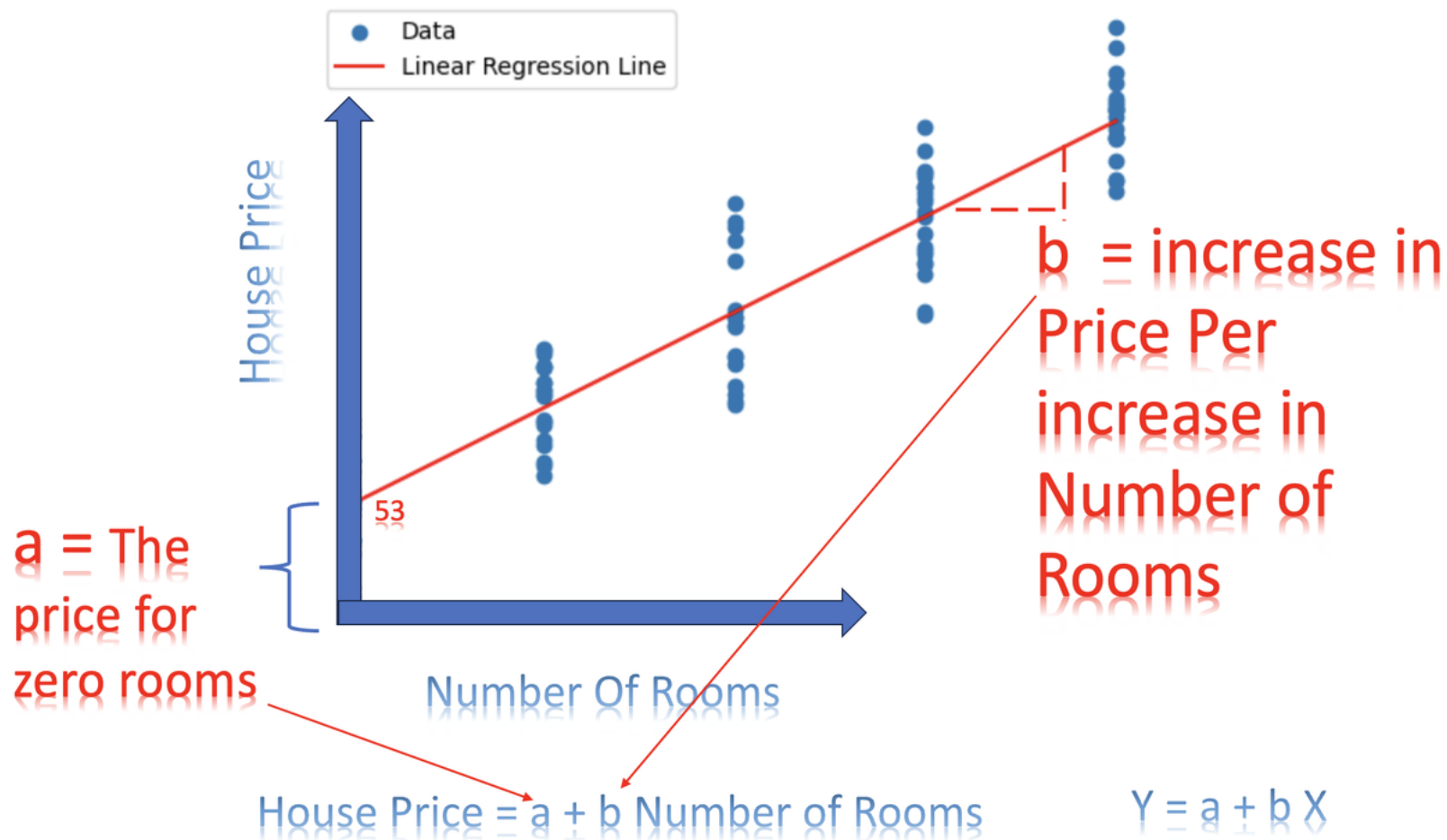
SUPPORT VECTOR MACHINE



Regressão

- *“Regression is used to determine how well certain independent variables predict a dependent variable.”*
- A regressão é uma técnica que calcula a equação de reta que melhor se adapta a um conjunto específico de dados.
- Tarefa:
 - Aprender uma equação de reta que analisa variáveis independentes (preço do gás, preço do dólar, custos de transporte), para prever o comportamento de uma variável dependente (preço do petróleo).

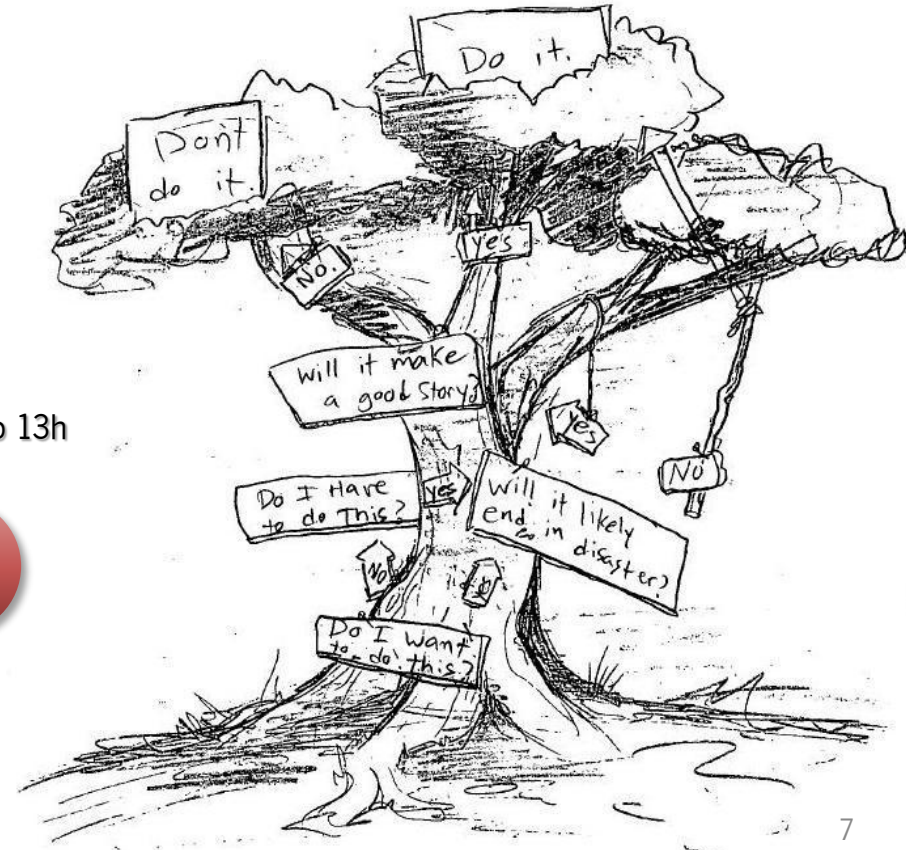
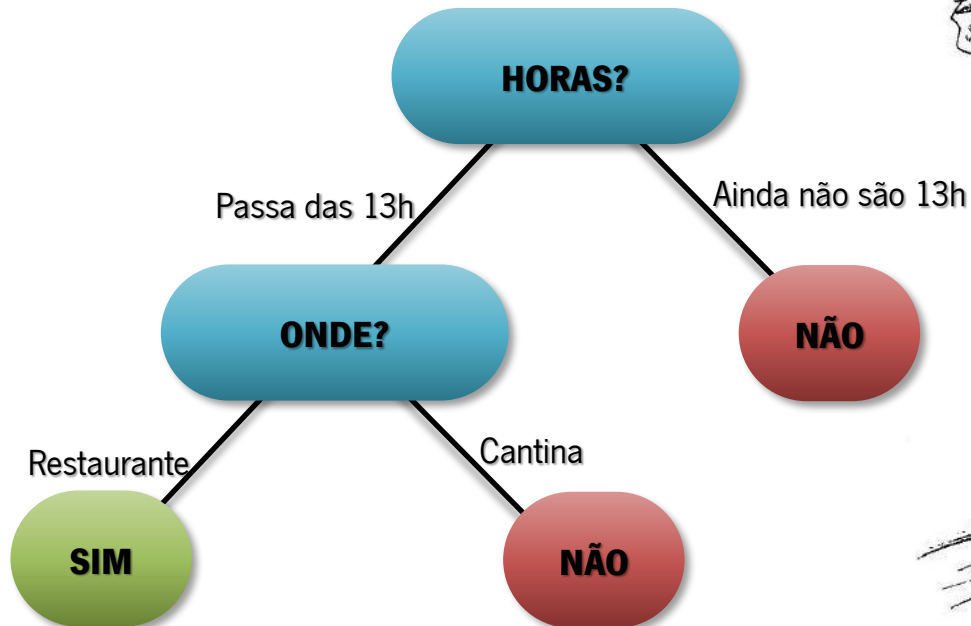






Árvores de Decisão

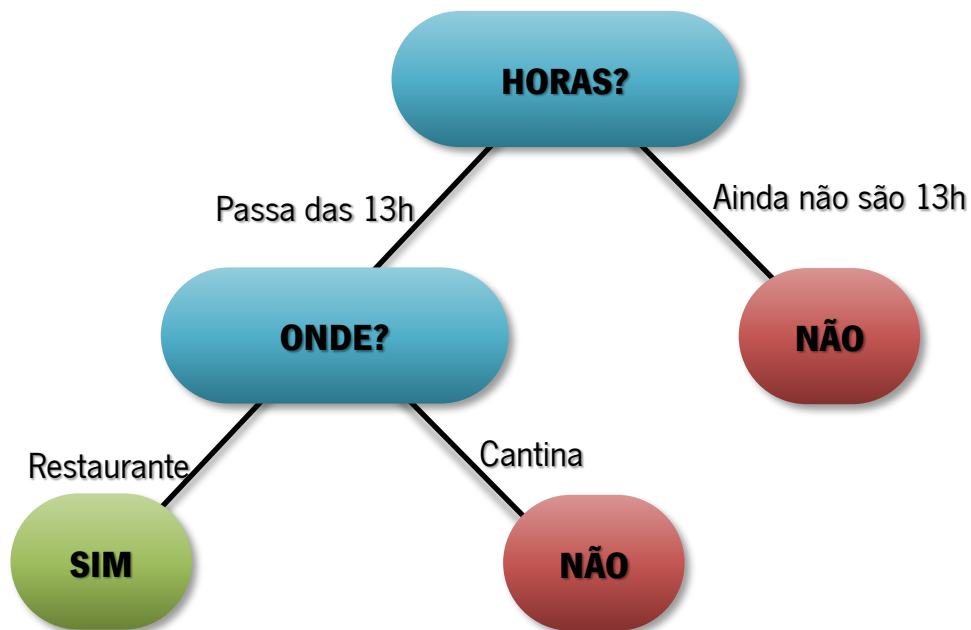
- Uma Árvore de Decisão é um grafo hierarquizado (árvore!) em que:
 - Cada ramo representa a seleção entre um conjunto de alternativas;
 - Cada folha representa uma decisão;





Árvores de Decisão Classificação

- Uma Árvore de Decisão é um grafo hierarquizado (árvore!) em que:
 - Cada nodo interno testa um atributo do *dataset*;
 - Cada ramo identifica um valor (ou conjunto de valores) do nodo testado;
 - Cada folha representa uma decisão;



HORAS	ONDE	ALMOÇAR
12h30	Cantina	NÃO
13h15	Cantina	NÃO
13h10	Restaurante	SIM
11h00	Restaurante	NÃO
13:30	Cantina	NÃO



Árvores de Decisão Classificação

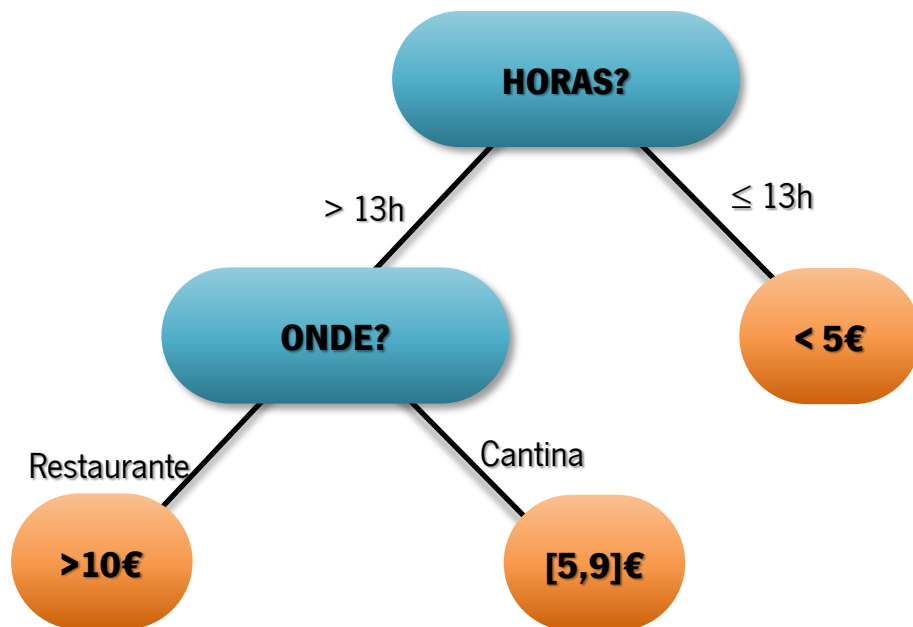
- Uma Árvore de Decisão pode ser utilizada para fazer **classificação**:
 - Decidir sobre se ou onde almoçar: classificação binária (SIM/NÃO)
 - Prever quem sobreviveu ao acidente do Titanic: classificação binária (SIM/NÃO)
 - Classificar um conjunto de imagens: classificação múltipla (laranja, kiwi, romã, ...)





Árvores de Decisão Regressão

- Uma Árvore de Decisão é um grafo hierarquizado (árvore!) em que:
 - Cada nodo interno testa um atributo do *dataset*;
 - Cada ramo identifica um valor (ou conjunto de valores) do nodo testado;
 - Cada folha representa uma decisão;

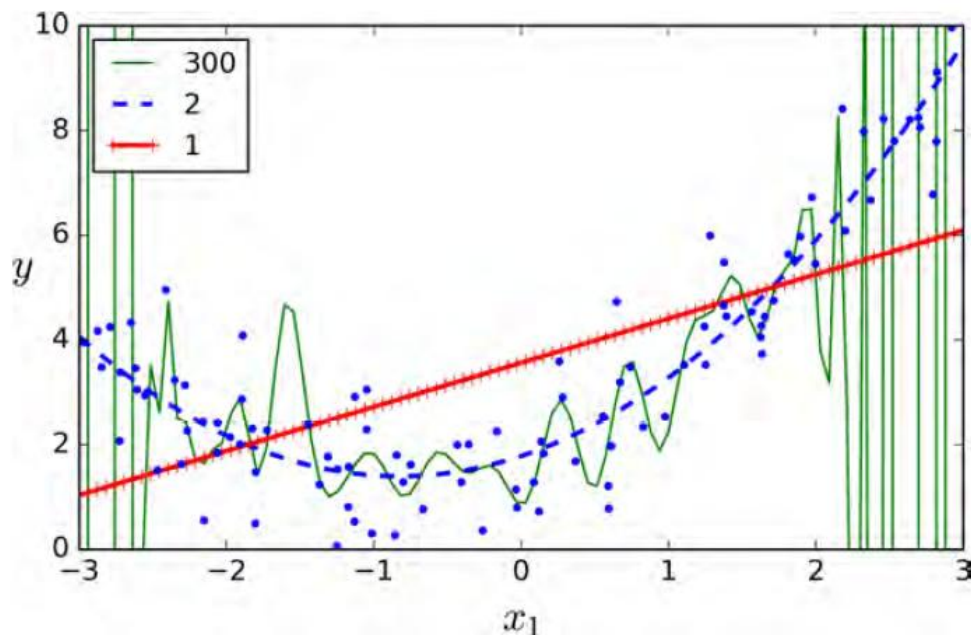


HORAS	ONDE	CUSTO
12h30	Cantina	< 5€
13h15	Cantina	> 5€ , < 9€
13h10	Restaurante	> 10€
11h00	Restaurante	< 5€
13:30	Cantina	> 5€ , < 9€



Árvores de Decisão Regressão

- Uma Árvore de Decisão pode ser utilizada para fazer **regressão**:
 - Regressão linear, polinomial, múltipla, entre outras;
 - Prever o preço do petróleo/gás/combustíveis: escala contínua ou real, em € ou \$
 - Estimar a temperatura para o dia de amanhã: escala contínua, em °C ou °F





Universidade do Minho
Departamento de Informática

Avaliação de Modelos



Avaliação de Modelos

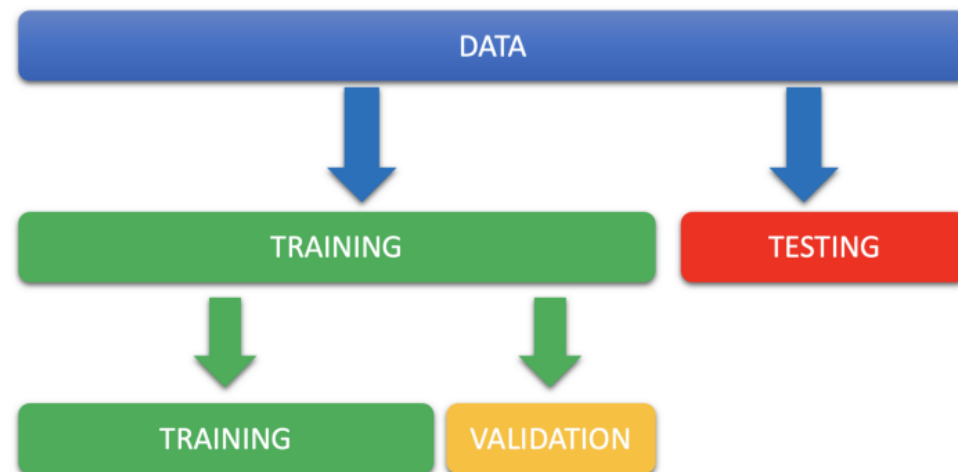
- Após a criação (treino) de um modelo usando uma técnica de aprendizagem (*machine learning*), é necessário avaliar o seu desempenho;
- A medição do desempenho de um modelo é feita com dados não apresentados durante o treino;





Avaliação de Modelos

- Dados de treino:
 - Conjunto de dados usado para ajustar o modelo;
- Dados de teste:
 - Conjunto de dados usado para fornecer uma avaliação imparcial de um modelo final ajustado ao conjunto de dados de treino.
- Dados de validação:
 - Conjunto de dados usado para fornecer uma avaliação imparcial de um ajuste do modelo, no conjunto de dados de treino;

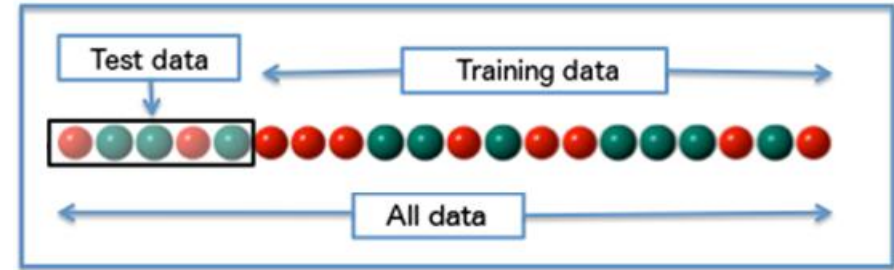




Avaliação de Modelos

Hold-out Validation

- Método de particionamento de dados;
- Divide o conjunto de dados em dados de treino e dados de teste;

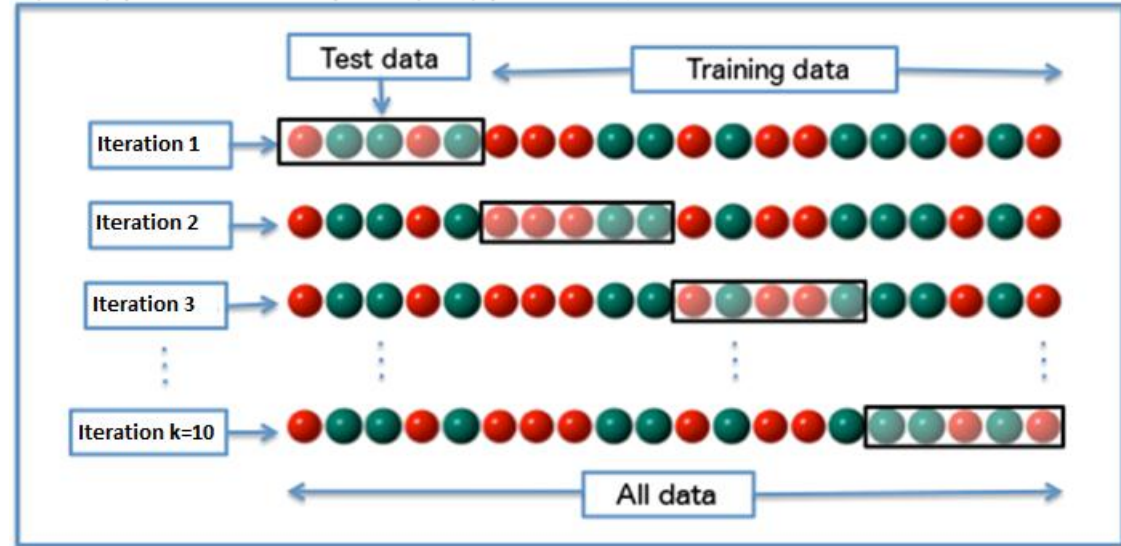


- Separa-se uma parte (*hold-out*) do conjunto de dados para treino/teste (80/20; 75/25; ...)



Avaliação de Modelos *Cross Validation*

- Método de validação por cruzamento de dados;
- Consiste em dividir o conjunto de dados em k partes (k *folds*);
 - A cada iteração, o método utiliza $k-1$ partes (*folds*) para treino e 1 parte (*fold*) para teste;
 - O processo repete-se durante k vezes;



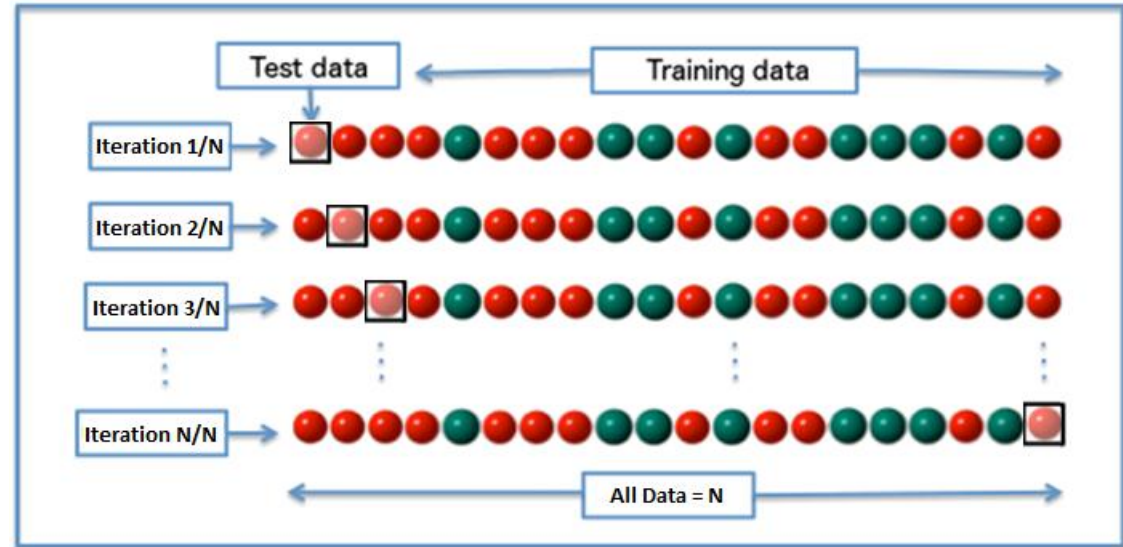
- O erro final é dado pela média dos valores parciais dos erros.



Avaliação de Modelos

Leave-one-out Cross Validation

- Método de validação por cruzamento de dados;
- Caso particular em que o número de casos N é igual ao número de *folds* k ;





Avaliação de Modelos

Cross Validation

- Qual o número ideal para k (*folds*)?
- Se o *dataset* for grande, um valor pequeno para k pode ser suficiente, uma vez que teremos uma quantidade grande de dados para treino;
- Se o *dataset* for pequeno, um valor grande de $k \approx N$ pode revelar-se mais adequado para maximizar a quantidade de dados para treino;
- Quanto maior a quantidade de *folds*, melhor a estimativa do erro, mais baixo será o viés(*) (*bias*) e menor será o sobreajuste (*overfitting*);
- De facto, o valor de k depende do valor de N !

(*) viés = distorção
enviesar = entortar



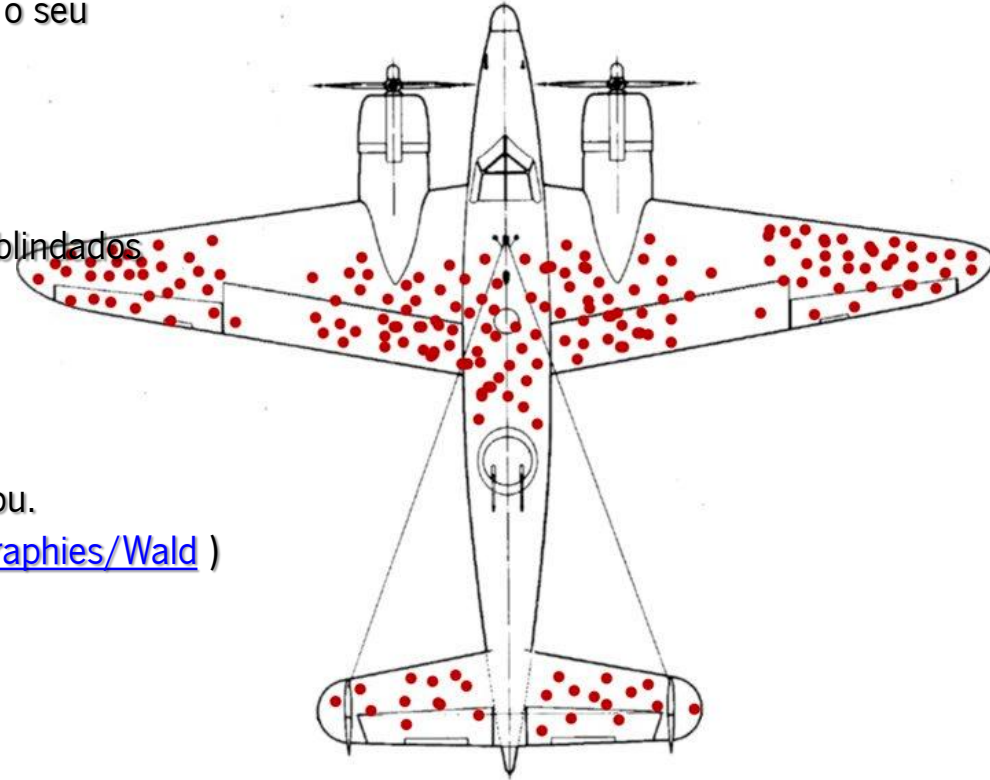


Universidade do Minho
Departamento de Informática

Métricas de Qualidade



- Durante a Segunda Guerra Mundial, a Marinha dos EUA pretendia determinar onde seria necessário blindar os seus aviões para garantir o seu regresso com sucesso.
- Analisaram os pontos onde os aviões regressados tinham sido alvejados.
- Foi opinião unânime que os locais que precisavam de ser blindados eram as pontas das asas, o corpo central e os elevadores.
- Era aí que os aviões estavam todos a ser alvejados!



- Abraham Wald, matemático, discordou.
(mathshistory.st-andrews.ac.uk/Biographies/Wald)
- **Porquê?**





Métricas de Qualidade

- Porquê métricas de qualidade?
 - Para avaliar o desempenho do modelo.
- As métricas são usadas para monitorizar e medir o desempenho de um modelo:
 - Erro Médio Absoluto (*Mean Absolute Error* - MAE)
 - Erro Médio Quadrado (*Mean Squared Error* - MSE)
 - Precisão (*Precision*)
 - F1-Score,
 - entre outras...
- No entanto, depende do problema em mãos:
 - É um problema de classificação?
 - De regressão?
 - Séries temporais?





Métricas de Qualidade Modelos de Classificação

- Matrizes de Confusão

- Tabela utilizada para descrever o desempenho de um modelo de classificação.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN



Métricas de Qualidade Modelos de Classificação

- Matrizes de Confusão

- Tabela utilizada para descrever o desempenho de um modelo de classificação.

- *Accuracy*

- Quantidade de previsões corretas dividido pela quantidade total de observações:

- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN



Métricas de Qualidade Modelos de Classificação





■ Matrizes de Confusão

- Tabela utilizada para descrever o desempenho de um modelo de classificação.

■ Accuracy

- Quantidade de previsões corretas dividido pela quantidade total de observações:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

	Actually Pregnant	Actually NOT Pregnant
Predicted Pregnant	 True Positive (TP)	 False Positive (FP)
Predicted NOT Pregnant	 False Negative (FN)	 True Negative (TN)



Métricas de Qualidade Modelos de Classificação

■ Matrizes de Confusão

- Tabela utilizada para descrever o desempenho de um modelo de classificação.

■ Precisão (*Precision aka Sensitivity*)

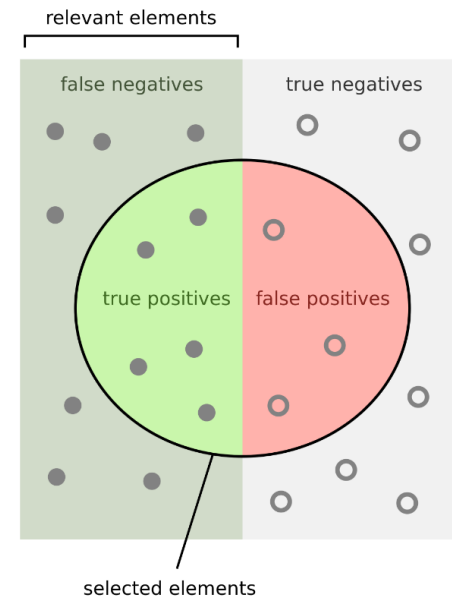
- É uma medida da exatidão;
- Determina a proporção de itens relevantes entre todos os itens:

$$\bullet \text{ Precision} = \frac{TP}{TP+FP}$$

■ Recall (*aka Specificity*)

- É uma medida de completude;
- Determina a proporção de itens relevantes obtidos:

$$\bullet \text{ Recall} = \frac{TP}{TP+FN}$$



How many selected
items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant
items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$



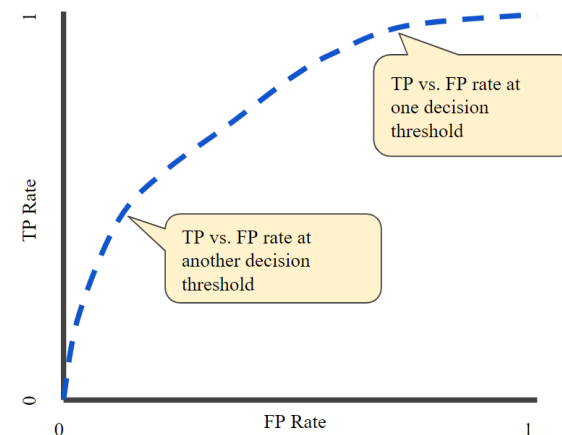
Métricas de Qualidade Modelos de Classificação

- Matrizes de Confusão

- Tabela utilizada para descrever o desempenho de um modelo de classificação.

- ROC curve:

- A curva *Receiver Operating Characteristics* (ROC) encontra o desempenho de um modelo de classificação em diferentes limites de classificação;
 - Reduzindo o patamar (*threshold*) de classificação, são classificados mais itens como positivos, aumentando os falsos positivos e os verdadeiros positivos.





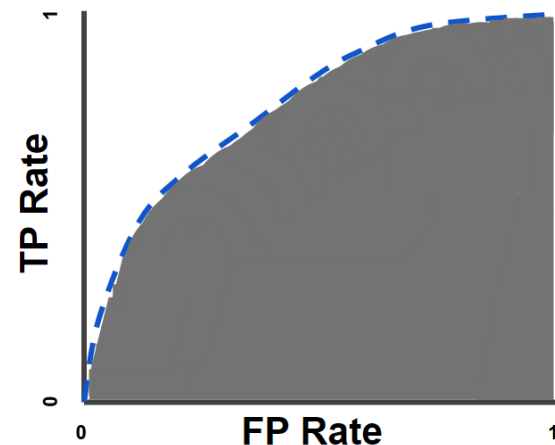
Métricas de Qualidade Modelos de Classificação

■ Matrizes de Confusão

- Tabela utilizada para descrever o desempenho de um modelo de classificação.

■ AUC curve:

- A Area Under the Curve (AUC) mede a área abaixo da curva ROC;
- Mede quão bem as previsões são classificadas, em vez de avaliar os seus valores absolutos (varia de 0 a 1);
- Um modelo cujas previsões estão 100% erradas tem uma AUC de 0; aquele cujas previsões estão 100% corretas tem uma AUC de 1.





Métricas de Qualidade Modelos de Regressão

- Erro Médio Absoluto (*Mean Absolute Error* - MAE)
 - Mede a magnitude média dos erros num conjunto de previsões (não considera a direção):

- $MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$

em que n é a quantidade de observações, y_j e \hat{y}_j são, respetivamente, a observação atual e o valor previsto.

$$MAE = \frac{1}{n} \sum \underbrace{|y - \hat{y}|}_{\text{The absolute value of the residual}}$$

Diagram annotations:

- Divide by the total number of data points (points to $\frac{1}{n}$)
- Actual output value (points to y)
- Predicted output value (points to \hat{y})
- Sum of (points to the summation symbol \sum)
- The absolute value of the residual (points to the absolute value bars)



Métricas de Qualidade Modelos de Regressão

■ Erro Médio Quadrado (*Mean Squared Error*- MSE)

- Consiste no cálculo da média das diferenças, ao quadrado, entre os erros num conjunto de previsões (não considera a direção):

- $MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$

em que n é a quantidade de observações, y_j e \hat{y}_j são, respetivamente, a observação atual e o valor previsto.

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2$$



Métricas de Qualidade Modelos de Regressão

- Raiz Quadrada do Erro Médio Quadrado (*Root Mean Squared Error* - RMSE)
 - Consiste no cálculo da média das diferenças, ao quadrado, entre os erros num conjunto de previsões (não considera a direção):

- $RMSE = \frac{1}{n} \sqrt{\sum_{j=1}^n (y_j - \hat{y}_j)^2}$

em que n é a quantidade de observações, y_j e \hat{y}_j são, respetivamente, a observação atual e o valor previsto.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$



Métricas de Qualidade Modelos de Regressão

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2 \quad RMSE = \frac{1}{n} \sqrt{\sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

- Três das métricas mais comuns usadas para medir a precisão de variáveis contínuas;
- Todas expressam o erro médio de previsão do modelo (valores mais baixos são melhores);
- Todos variam de 0 a ∞ e são indiferentes à direção dos erros;
- MAE e RMSE expressam o erro de previsão na mesma unidade da variável de interesse;
- MSE e RMSE, ao elevar o erro ao quadrado, dão um peso relativamente alto para erros grandes;
- MSE e RMSE são mais úteis quando grandes erros são especialmente indesejáveis.



Métricas de Qualidade Modelos de Regressão

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

$$MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$$

$$RMSE = \frac{1}{n} \sqrt{\sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

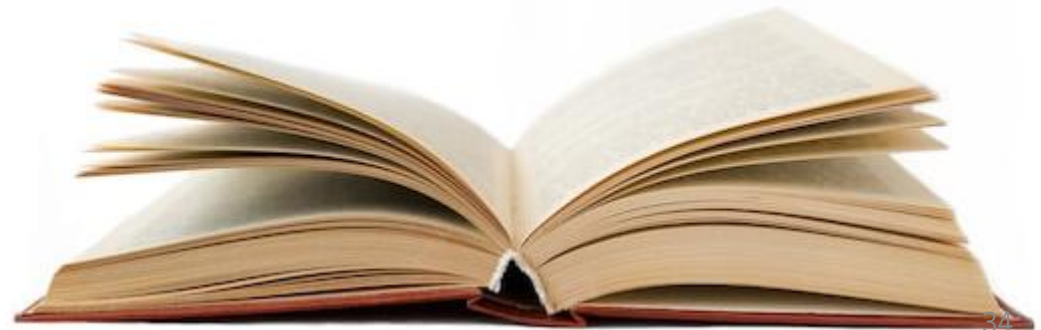
#	Error	Error	Error ²
1	1	1	1
2	-1	1	1
3	3	3	9
4	3	3	9
	MAE	MSE	RMSE
	2	5	2,24

#	Error	Error	Error ²
1	0	0	0
2	0	0	0
3	0	0	0
4	10	10	100
	MAE	MSE	RMSE
	2,5	25	2,5



Referências bibliográficas

- Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. Introduction to linear regression analysis. John Wiley & Sons, 2021
- Ranganathan, Priya, C. S. Pramesh, and Rakesh Aggarwal. “Common pitfalls in statistical analysis: logistic regression” Perspectives in clinical research 8.3, 2017
- Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984), “Classification and regression trees”, Monterey, CA
- Ross Quinlan (1993), “C4.5 Programs for Machine Learning”, Morgan Kaufmann





Universidade do Minho
Departamento de Informática

Avaliação de Modelos e Métricas de Qualidade

ADI³ - LEI/MiEI @ 2024/2025, 2º sem