

**Exercício 01** Num processo de diagnóstico médico, numa população de 2000 indivíduos, o método automático de classificação (diagnóstico positivo vs negativo) apresentou os seguintes resultados, expressos na matriz de confusão:

		Real	
	Tamanho da população	Condição Positiva	Condição Negativa
Predição	Predição Positiva	20	170
	Predição Negativa	10	1800

- Calcule a “accuracy” (exatidão/acurácia), precisão, sensibilidade (recall) e especificidade do classificador.
- Comente a qualidade do sistema de diagnóstico.

**Exercício 02** Está a resolver um problema de classificação binária com uma classe altamente desequilibrada, onde a classe maioritária é observada 99% das vezes nos dados de treino. O modelo, após o treino, apresenta 99% de “accuracy” no conjunto de teste.

- Qual ou quais das afirmações considera verdadeira e justifique:
  - A métrica de “accuracy” avalia corretamente o desempenho deste classificador.
  - A métrica de “accuracy” não deve ser usada, deve-se considerar a “precision”.
  - A métrica de “accuracy” não deve ser usada, as métricas de precisão e recall devem ser ambas avaliadas.

**Exercício 03** Pretende desenvolver um algoritmo para prever o número de visualizações de artigos num blog. A sua análise deve ser baseada em características tais como autor, número de artigos escritos pelo mesmo autor.

- Que outras “features” deve considerar?
- Como formaria o conjunto de treino e teste?

**IC 23/24**

***Exs - Métricas***

- c) Qual da(s) seguintes métrica(s) de avaliação escolheria para avaliar o classificador: MSE, Precisão ou F1 score? Justifique.

## Anexo - Métricas para Classificação

- **Matriz de Confusão:**

	<i>Tamanho da população</i>	<u>Real</u>	
		Condição Positiva	Condição Negativa
<u>Predição</u>	Predição Positiva	<b>TP</b>	<b>FP</b>
	Predição Negativa	<b>FN</b>	<b>TN</b>

Prevalência =  $\Sigma$  Condições positivas /  $\Sigma$  População Total

Taxa de Acerto (ACC-Accuracy) =  $(TP+TN)/\text{Tamanho da População}$

- **Métricas Lidas a partir da Matriz:**

Precisão ou PPV-“Positive Predicted Value” =  $TP/(TP+FP)$

False discovery rate (FDR) =  $FP/(TP+FP)$

False omission rate (FOR) =  $FN/(FN+TN)$

Negative predictive value (NPV) =  $TN/(FN+TN)$

Sensibilidade ou Recall (TPR-True Positive Rate) =  $TP/(TP+FN)$

False negative rate (FNR) =  $FN/(TP+FN)$

False positive rate (FPR) =  $FP/(FP+TN)$

Especificidade (TNR-True Negative Rate) =  $TN/FP+TN$

- **Métricas Combinadas:**

F-Measure =  $2 * (\text{Precison} * \text{Recall}) / (\text{Precision} + \text{Recall})$

F1 score =  $2 * TP / (2 * TP + FP + FN)$  – *media harmónica entre precisão e recall*

Positive likelihood ratio (LR+) =  $TPR/FPR$

Negative likelihood ratio (LR-) =  $FNR/TN$