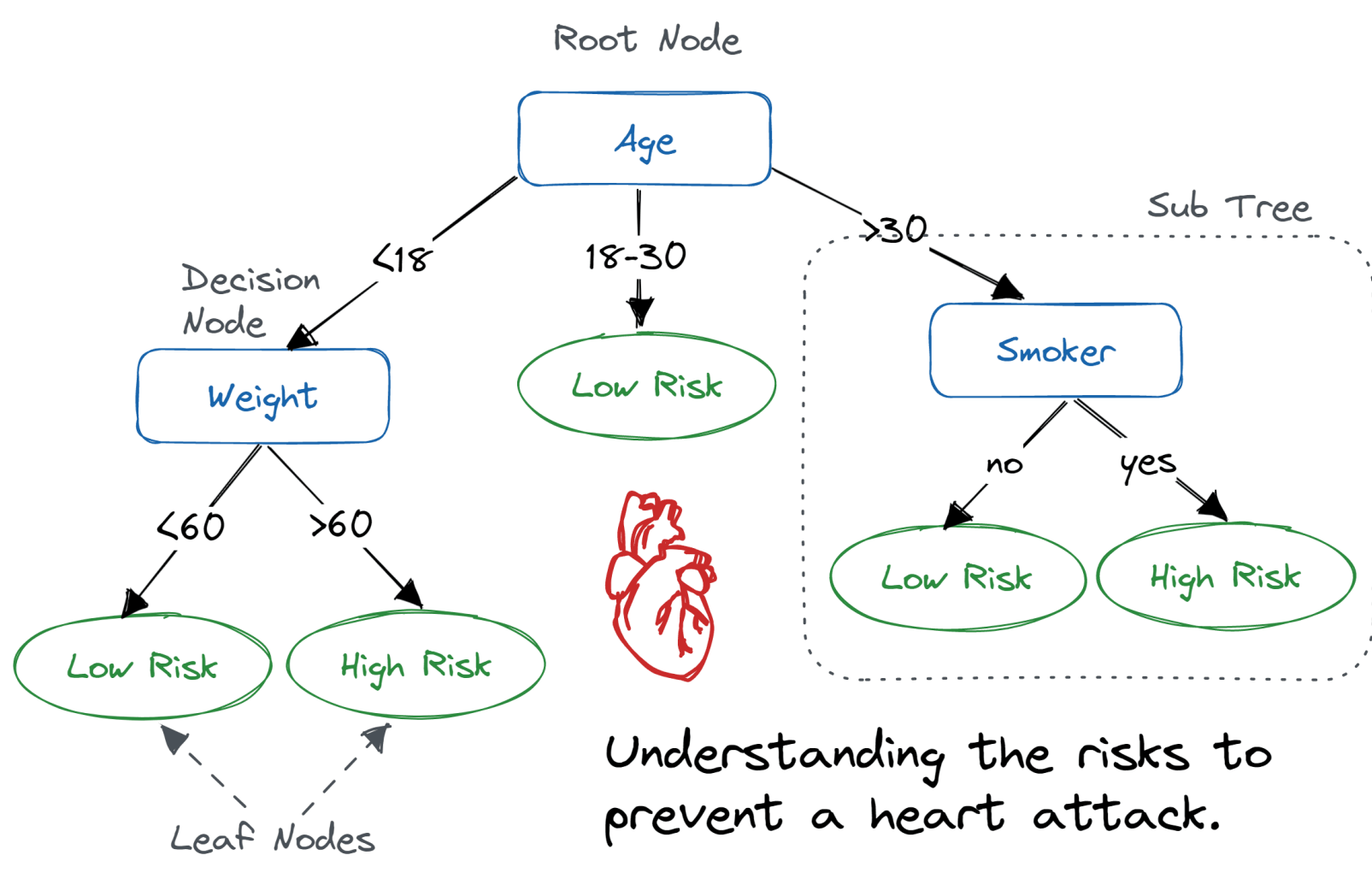


2 - DECISION TREES

Aprendizagem 2024/2025



CONCEITOS FUNDAMENTAIS

	y1	z
x1	A	B
x2	A	B
x3	B	C
x4	C	C

$$H(z) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$$

$$IG(y1) = H(z) - \left(\frac{2}{4}H(z|A) + \frac{1}{4}H(z|B) + \frac{1}{4}H(z|C)\right) = \dots$$

CONFUSION MATRIX

P/T	+	-
+	TP	FP
-	FN	TN

$$Recall = \frac{TP}{TP+FN}, \quad Precision = \frac{TP}{TP+FP}, \quad Specificity = \frac{TN}{TN+FP}$$

$$Accuracy = \frac{TP+TN}{\#All}$$

$$F - measure = \frac{1}{\alpha \frac{1}{P} + (1-\alpha) \frac{1}{R}}, \quad \beta^2 = \frac{1-\alpha}{\alpha} \quad (F1-measure \text{ com } \beta = 1)$$

UNDERFITTING E OVERFITTING

- **Underfitting:** modelo demasiado simples e incapaz de capturar os padrões nos dados (erro alto no conjunto de treino e de teste)
- **Overfitting:** modelo demasiado complexo e que captura noise para além dos padrões nos dados (erro baixo no conjunto de treino, erro alto no conjunto de teste)

SUMÁRIO

- Ficha 2