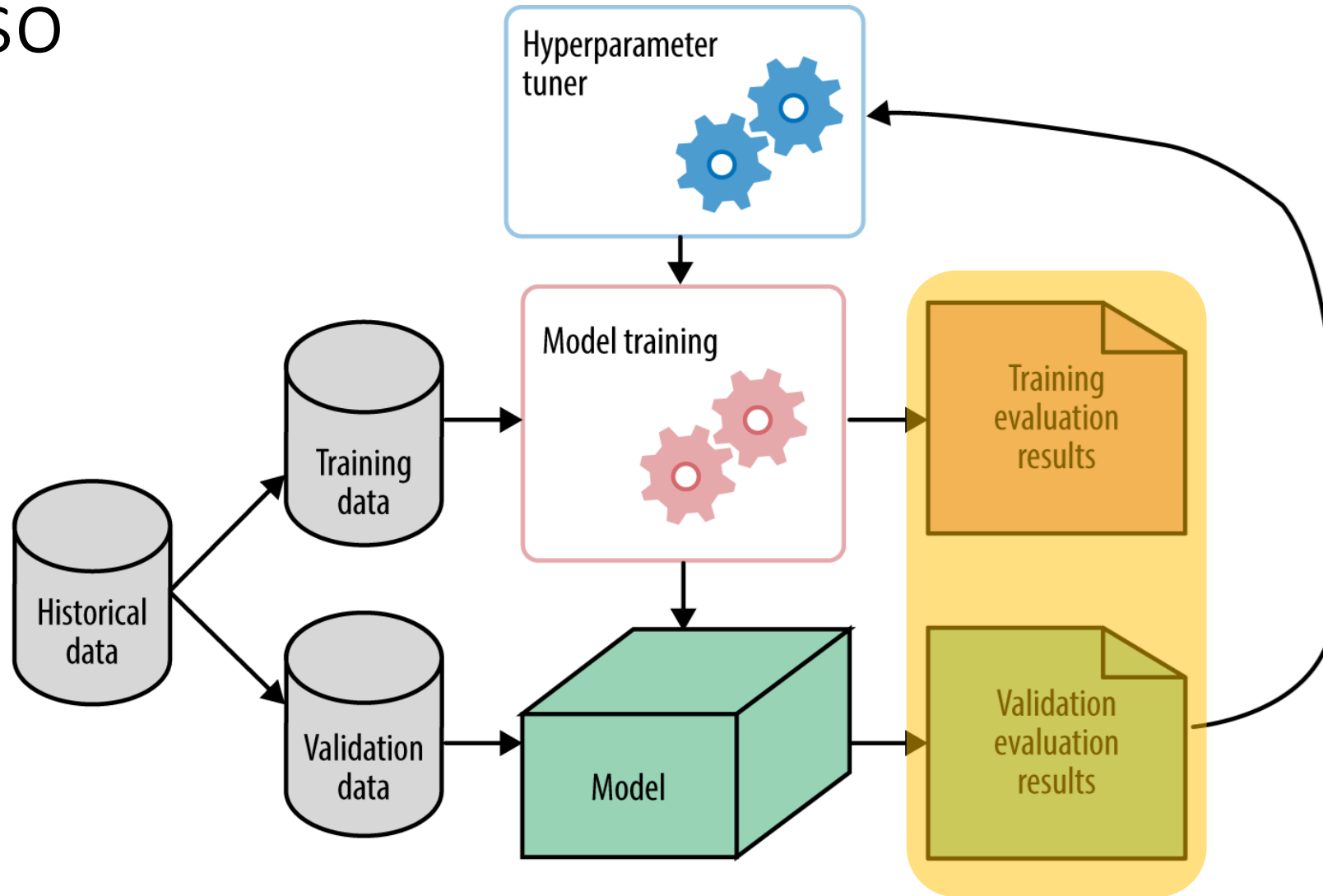


Aprendizado de Máquina e Deep Learning

Avaliação do modelo

Prof. Dr. Thiago Meirelles Ventura

Processo



Deep Learning Bible - A. End-to-End Pipelines

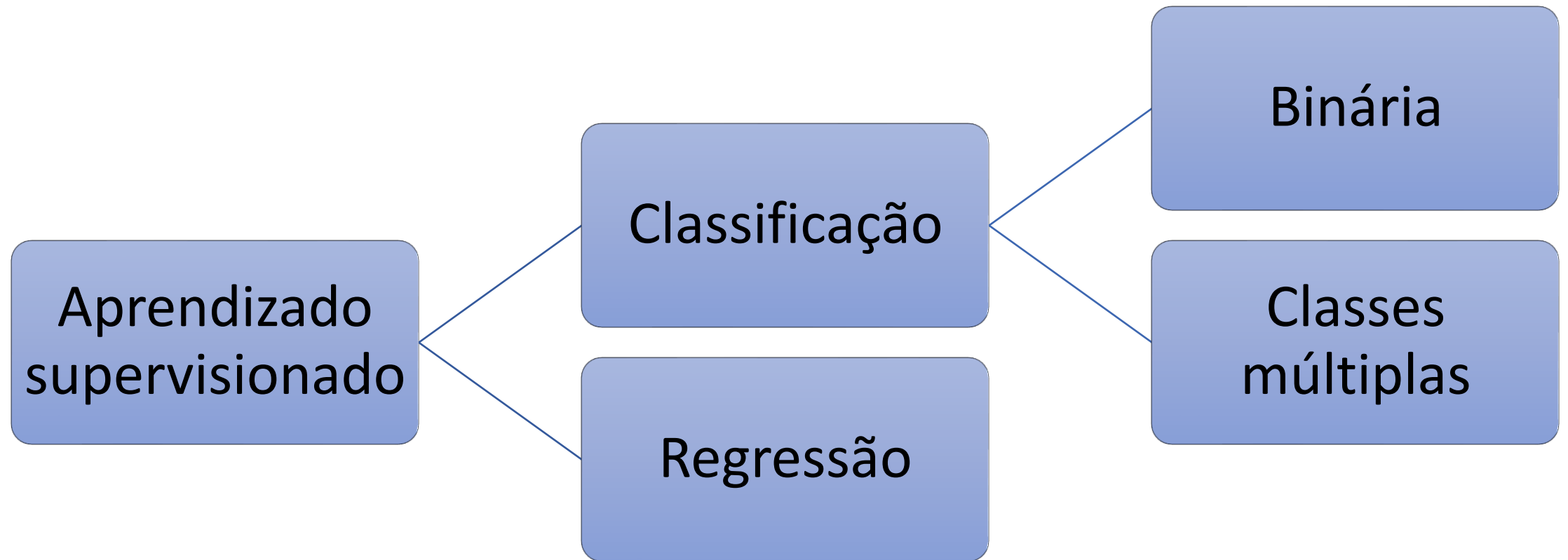
Avaliação

- Há a necessidade de avaliar o modelo para saber o quanto houve de aprendizado
- Isso acontece
 - durante o treinamento
 - após o treinamento
- Uma ou mais métricas devem ser escolhidas para a avaliação

Métricas

- Há diversas métricas
- Não há 1 melhor que todas
- A escolha da métrica depende
 - da característica dos dados
 - do problema que o modelo propõe resolver

Dependendo da característica dos dados



Dependendo do problema

- Qual foi o melhor modelo?

Modelo	Registros	Alvo	Detecção	Acerto
A	100	10	3	3
B	100	10	20	10

Dependendo do problema

- É mais importante acertar as ocorrências que foram detectadas ou não deixar que nenhuma ocorrência não seja detectada?
- O que deveria ser priorizado em um modelo que detecta câncer?
- O que deveria ser priorizado em um modelo que aplica multas automáticas?

Matriz de confusão

- Mostra quantos exemplos são
 - True positives (TP)
 - True negatives (TN)
 - False positives (FP)
 - False negatives (FN)

Matriz de confusão

- Mostra quantos exemplos são
 - True positives (TP): o modelo disse que era verdadeiro e acertou
 - True negatives (TN): o modelo disse que era falso e acertou
 - False positives (FP): o modelo disse que era verdadeiro e errou
 - False negatives (FN): o modelo disse que era falso e errou

Matriz de confusão

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

<https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>

Matriz de confusão

		True Class		
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

<https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>

Matriz de confusão

- Útil para visualizar o desempenho do modelo
- Possibilita observar com quais classes o modelo está se confundindo

*Os valores de TP, TN, FP e FN servirão para calcular diversas métricas

Métricas para classificação

- Acurácia
- Precisão
- Recall
- F1 score

Métricas para classificação

- Acurácia
 - Acerto das detecções em relação a todos os dados

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

Métricas para classificação

- Precisão
 - Acerto entre os valores positivos

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

Métricas para classificação

- Recall
 - Acerto entre os valores reais positivos

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

Métricas para classificação

- F1 score
 - Tenta fazer um balanceamento entre precision e recall

$$F1\ score = \frac{2 * precision * recall}{precision + recall}$$

Métricas para regressão

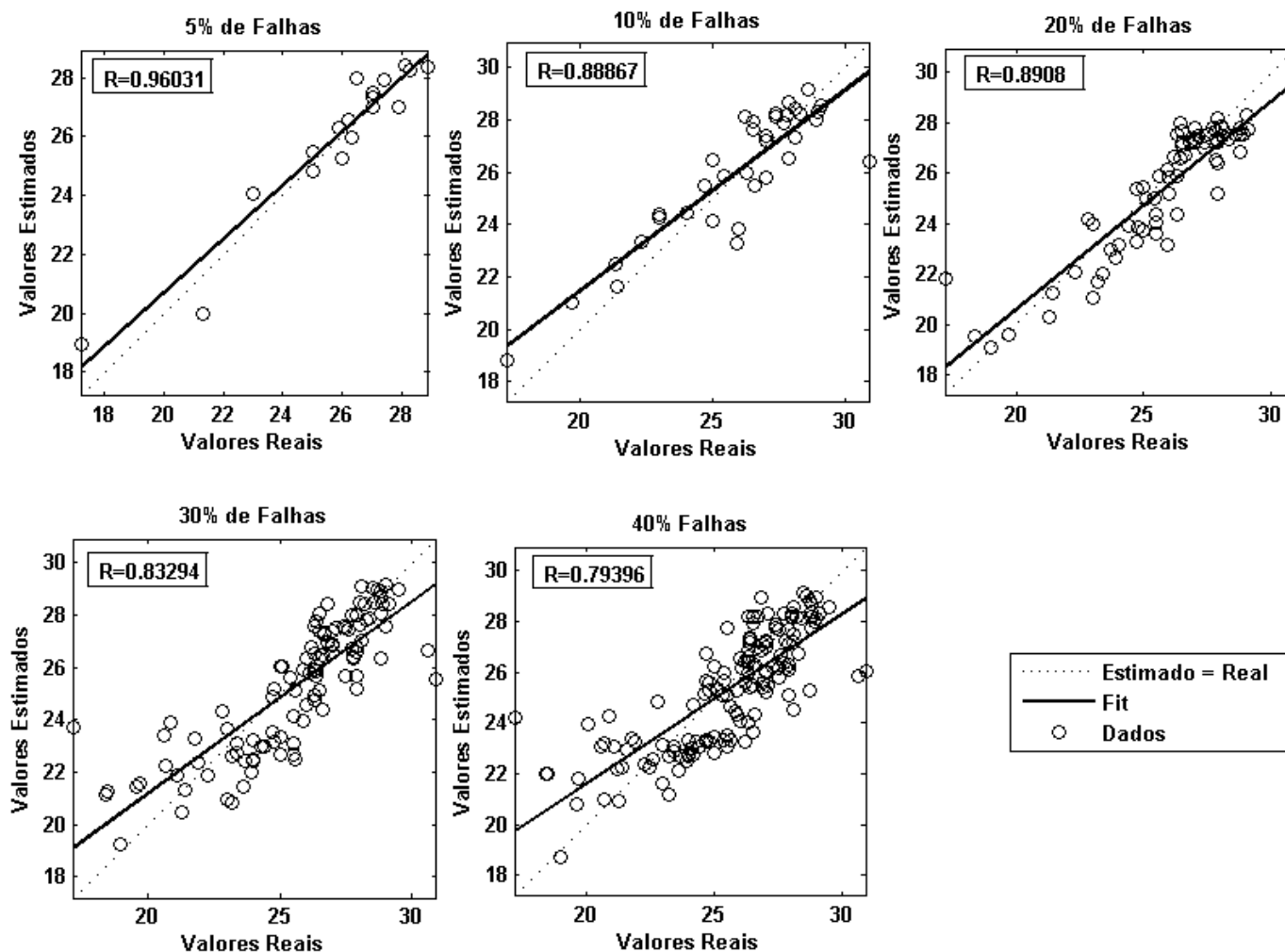
- r^2 (R square)
- MAE
- MSE
- RMSE

Métricas para regressão

- r^2
 - Coeficiente de determinação
 - Mede a variação entre os valores preditos com os valores reais
 - Valor entre 0 e 1
 - Quanto mais próximo de 1 melhor é a associação

Métricas para regressão

- r^2



Métricas para regressão

- MAE

- Mean absolute error
- Varia de 0 a infinito
- Quanto menor, melhor

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - d_i|$$

Métricas para regressão

- MSE

- Mean squared error

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - d_i)^2$$

- Maior penalização para erros maiores
 - Sensível a outliers

Métricas para regressão

- RMSE

- Root mean squared error

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

- Retorna à unidade do dado

script metricas.ipynb

Qual valor é bom o suficiente?

Qual valor é bom o suficiente?

- Não existe um limiar pré-definido para todos as aplicações
- Deve-se fazer uma comparação com alguma referência
 - Situação atual
 - Estado da arte
 - Teste com outros modelos

Discussão

- Já tiveram que debater no ambiente de trabalho sobre algum desempenho?