

GEX1082 - Tópicos Especiais em Computação XXXIII

Deep Learning

Métricas de Avaliação de Modelos



1100/1101 - CIÊNCIA DA COMPUTAÇÃO

Prof. Dr. Giancarlo D. Salton

Avaliação de Modelos

Hold-out Sampling

Método Padrão

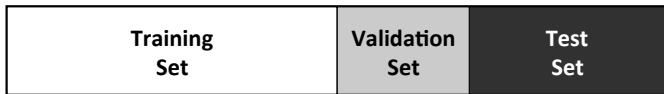
Métricas de Classificação

Métricas derivadas da matriz de confusão

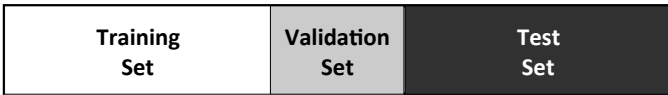
Medindo Lucro & Perda (*Profit & Loss*)

A parte mais importante ao desenvolvermos um experimento para avaliar um modelo preditivo é nos assegurarmos que os exemplos utilizados para o teste do modelo não sejam os mesmos utilizados durante o treino.

- São 3 os propósitos da avaliação:
 - ✓ determinar qual o melhor modelo
 - ✓ obter uma estimativa de como o modelo vai se comportar em produção
 - ✓ convencer os usuários de que ele funciona

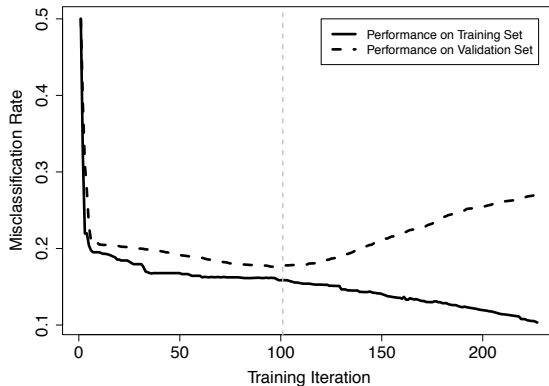


(a) A 50:20:30 split



(b) A 40:20:40 split

Hold-out sampling divide o *dataset* em subconjuntos de treino, teste e validação.



Pode-se utilizar o *dataset* de validação com algoritmos iterativos para reduzir o impacto do *overfitting*.

Método padrão: medir a taxa de erros em um *dataset* de teste

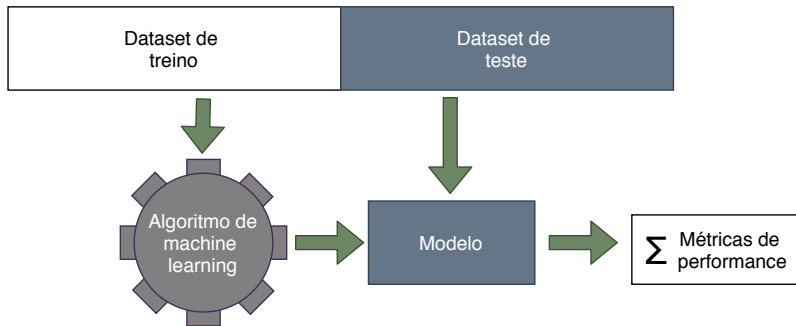


Imagem ilustrativa do processo de treinar e avaliar um modelo de *machine learning* utilizando um *dataset* de teste.

Tabela: Exemplo de um *dataset* de teste e as predições feitas pelo modelo.

ID	Target	Pred.	Resultado
1	spam	ham	FN
2	spam	ham	FN
3	ham	ham	TN
4	spam	spam	TP
5	ham	ham	TN
6	spam	spam	TP
7	ham	ham	TN
8	spam	spam	TP
9	spam	spam	TP
10	spam	spam	TP

ID	Target	Pred.	Resultado
11	ham	ham	TN
12	spam	ham	FN
13	ham	ham	TN
14	ham	ham	TN
15	ham	ham	TN
16	ham	ham	TN
17	ham	spam	FP
18	spam	spam	TP
19	ham	ham	TN
20	ham	spam	FP

$$\text{taxa de erros} = \frac{\text{previsões incorretas}}{\text{total de previsões}}$$

$$\text{taxa de acertos} = \frac{\text{previsões corretas}}{\text{total de previsões}}$$

$$\text{taxa de erros} = \frac{(2 + 3)}{(6 + 9 + 2 + 3)} = 0.25$$

$$\text{taxa de acertos} = \frac{(6 + 9)}{(6 + 9 + 2 + 3)} = 0.75$$

- Para classificações de *targets* binários, são 4 os possíveis resultados de uma previsão
 1. *True Positive* (TP) — verdadeiro positivo
 2. *True Negative* (TN) — verdadeiro negativo
 3. *False Positive* (FP) — falso positivo
 4. *False Negative* (FN) — falso negativo

Tabela: Estrutura de uma matriz de confusão.

		Predição	
		positive	negative
Target	positive	<i>TP</i>	<i>FN</i>
	negative	<i>FP</i>	<i>TN</i>

Tabela: Estrutura da uma matriz de confusão obtida para o *dataset* sobre classificação de spam.

		Predição	
		spam	ham
Target	spam	6	3
	ham	2	9

$$\text{taxa de erros} = \frac{(FP + FN)}{(TP + TN + FP + FN)} \quad (1)$$

$$\text{taxa de erros} = \frac{(2 + 3)}{(6 + 9 + 2 + 3)}$$

$$\text{taxa de erros} = 0.25$$

$$\begin{aligned}\text{taxa de acertos (acurácia)} &= \frac{(TP + TN)}{(TP + TN + FP + FN)} \\ \text{acurácia} &= \frac{(6 + 9)}{(6 + 9 + 2 + 3)} \\ \text{acurácia} &= 0.75\end{aligned}\tag{2}$$

$$TPR = \frac{TP}{(TP + FN)} \quad (3)$$

$$TNR = \frac{TN}{(TN + FP)} \quad (4)$$

$$FPR = \frac{FP}{(TN + FP)} \quad (5)$$

$$FNR = \frac{FN}{(TP + FN)} \quad (6)$$

$$\text{TPR} = \frac{6}{(6+3)} = 0.667$$

$$\text{TNR} = \frac{9}{(9+2)} = 0.818$$

$$\text{FPR} = \frac{2}{(9+2)} = 0.182$$

$$\text{FNR} = \frac{3}{(6+3)} = 0.333$$

Precision (precisão), Recall (especificidade) e F_1 Measure

$$\text{precision} = \frac{TP}{(TP + FP)} \quad (7)$$

$$\text{recall} = \frac{TP}{(TP + FN)} \quad (8)$$

Precision (precisão), Recall (especificidade) e F_1 Measure

$$precision = \frac{TP}{(TP + FP)} = \frac{6}{(6 + 2)} = 0.75$$

$$recall = \frac{TP}{(TP + FN)} = \frac{6}{(6 + 3)} = 0.667$$

Precision (precisão), Recall (especificidade) e F_1 Measure

$$F_1\text{-measure} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \quad (9)$$

Precision (precisão), Recall (especificidade) e F_1 Measure

$$F_1\text{-measure} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \quad (9)$$

$$\begin{aligned} F_1\text{-measure} &= 2 \times \frac{\left(\frac{6}{(6+2)} \times \frac{6}{(6+3)} \right)}{\left(\frac{6}{(6+2)} + \frac{6}{(6+3)} \right)} \\ &= 0.706 \end{aligned}$$

- Nem sempre é correto tratarmos todos as predições da mesma forma
- Nestes casos, é útil levar em consideração o “custo” de diferentes resultados ao avaliarmos os modelos

Tabela: Estrutura da “matriz de lucros” *profit matrix*.

		Previsão	
		positivo	negativo
Alvo	positivo	TP_{Profit}	FN_{Profit}
	negativo	FP_{Profit}	TN_{Profit}

Tabela: Um exemplo de *profit matrix* com os lucros e perdas de cada tipo de acerto e cada tipo de erro.

		Previsão	
		positivo	negativo
Alvo	positivo	140	−140
	negativo	−700	0

Tabela: Exemplo de matriz de confusão para um modelo k -NN com average class accuracy_{HM} = 83.824%

		Previsão	
		positivo	negativo
Alvo	positivo	57	3
	negativo	10	30

Tabela: Exemplo de matriz de confusão para uma árvore de decisão com average class accuracy_{HM} = 80.761%.

		Previsão	
		positivo	negativo
Alvo	positivo	43	17
	negativo	3	37

Tabela: Exemplo de cálculo de *profit & loss* para o modelo *k*-NN utilizando a *profit matrix* do slide anterior. Para o exemplo abaixo, *profit* = 560.

k-NN (Lucro = 560)

		Previsão	
		positivo	negativo
Alvo	positivo	7 980	−420
	negativo	−7 000	0

Tabela: Exemplo de cálculo de *profit & loss* para a árvore de decisão utilizando a *profit matrix* do slide anterior. Para o exemplo abaixo, *profit* = 1.540.

decision tree (Lucro = 1 540)

		Previsão	
		positivo	negativo
Alvo	positivo	6 020	−2 380
	negativo	−2 100	0