



APRENDIZAGEM DE MÁQUINA

PROF. JOSENALDE OLIVEIRA

josenalde.oliveira@ufrn.br

<https://github.com/josenalde/machinelearning>

ANÁLISE E DESENVOLVIMENTO DE SISTEMAS - UFRN

AVALIANDO MODELOS DE CLASSIFICAÇÃO

MATRIZ DE CONFUSÃO

	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	0	1	2	3	4	5	6	7	8	9
2	0	1	2	3	4	5	6	7	8	9
3	0	1	2	3	4	5	6	7	8	9
4	0	1	2	3	4	5	6	7	8	9
5	0	1	2	3	4	5	6	7	8	9
6	0	1	2	3	4	5	6	7	8	9
7	0	1	2	3	4	5	6	7	8	9
8	0	1	2	3	4	5	6	7	8	9
9	0	1	2	3	4	5	6	7	8	9

[1] Ferramenta comum para avaliar desempenho de classificador

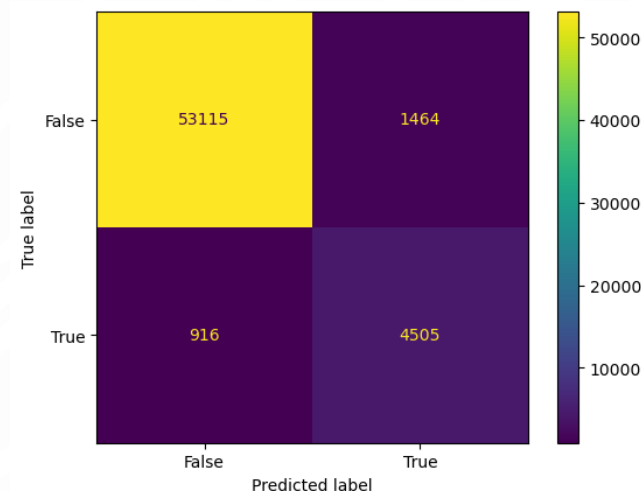
[2] Ideia geral: contar quantas vezes as instâncias de classe A são classificadas como classe B. Por exemplo, um classificador **multiclasse** (ou multinomial) como SGD, randomForest ou naiveBayes (que tratam multiclasse nativamente) teríamos uma matriz 10 x 10 para o dataset MNIST (0-9) e para saber quantas vezes imagens de 3s foram confundidas com imagens de 5s olharíamos na quarta linha e na sexta coluna

[3] Quando se tem a coluna target disponível no conjunto de teste final, pode-se prever as saídas com o modelo final escolhido e sintonizado e gerar as métricas a partir da matriz de confusão. Contudo, para comparar modelos, pode-se usar `cross_val_predict` para gerar previsões nos conjuntos de validação em cada fold

[4] Linhas: classes reais, Colunas: classes previstas (preditas)

- 53115 imagens não-5 (**classe negativa**) foram classificadas corretamente como não-5 (verdadeiro negativo: TN/VN) enquanto as 1464 restantes classificadas erroneamente como 5s (**classe positiva**) (falso positivo: FP)
- 916 imagens de 5s classificadas erroneamente como não-5 (falso negativo: FN) e as 4505 restantes corretamente classificadas como 5s (verdadeiro positivo).

Exemplo: [SGD@MNIST para dígito 5](#)



AVALIANDO MODELOS DE CLASSIFICAÇÃO

MATRIZ DE CONFUSÃO – métricas para além da acurácia

A escolha da(s) métrica(s) mais significativa(s) depende do problema, do balanceamento do conjunto em avaliação, do custo das avaliações incorretas

[1] **Acurácia:** o quanto o classificador acertou, seja da classe positiva, seja da classe negativa. No caso do

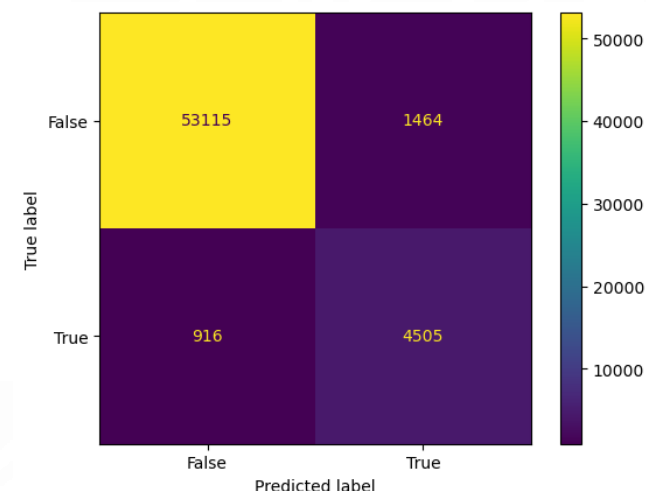
SPAM, quantos e-mails foram classificados corretamente

Um modelo perfeito teria 0FN e 0FP com acurácia 100%

É contudo métrica bem geral, com suposição de balanceamento

Quando o dataset é desbalanceado ou um dos erros de FN/FP são mais graves (doenças por exemplo), melhor outras métricas

		Valor Previsto	
		0	1
Valor Real	0	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	1	Falso Negativo (FN)	Verdadeiro Positivo (VP)



$$Acuracia(accuracy) = \frac{VP + VN}{VP + VN + FP + FN} = \frac{53115 + 4505}{60000} = 96,03\%$$

[2] **Acurácia das predições POSITIVAS (precisão/precision):** proporção de classificações positivas que são realmente positivas. No exemplo, proporção de imagens classificadas como 5 que realmente são imagens de 5

$$Precisao(precision) = \frac{VP}{VP + FP} = \frac{4505}{4505 + 1464} = 75,47\%$$

Fração de e-mails classificados como spam que eram realmente spam

“Quando um classificador é altamente preciso, ele tem maior confiabilidade ao rotular como positiva uma amostra. Quando ele diz que é de uma classe, é porque é!”

github@am@Profa.Laura

AVALIANDO MODELOS DE CLASSIFICAÇÃO

MATRIZ DE CONFUSÃO – métricas para além da acurácia

[3] **Revocação (recall, sensibilidade, taxa de verdadeiros positivos (TPR):**

Em geral usada junto com a precisão, pois é a proporção de instâncias positivas detectadas corretamente pelo classificador, pois

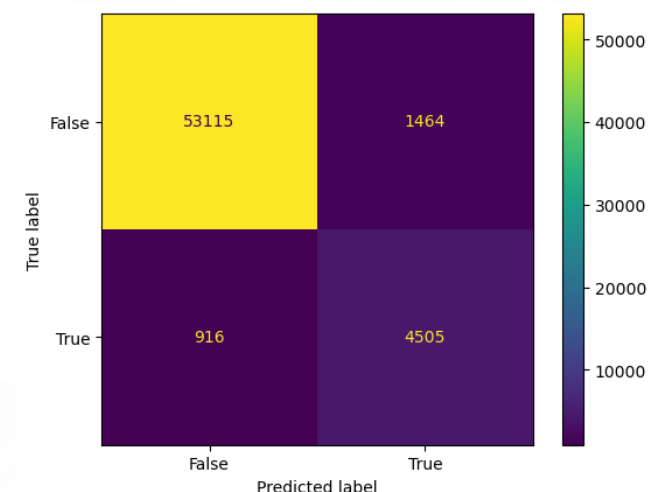
$$Revocacao(recall) = \frac{VP}{VP + FN} = \frac{4505}{4505 + 916} = 83,10\%$$

FN são positivos reais classificados incorretamente

Mede a fração de spams que foram corretamente classificados como spam

Também chamada **probabilidade de detecção**

		Valor Previsto	
		0	1
Valor Real	0	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	1	Falso Negativo (FN)	Verdadeiro Positivo (VP)



[4] **Taxa de falso positivo (TFP):** proporção de todos negativos reais que foram classificados incorretamente como positivos, conhecida como **probabilidade de alarme falso**

$$TFP = \frac{FP}{FP + VN} = \frac{1464}{1464 + 53115} = 2,68\%$$

Por quê?

AVALIANDO MODELOS DE CLASSIFICAÇÃO

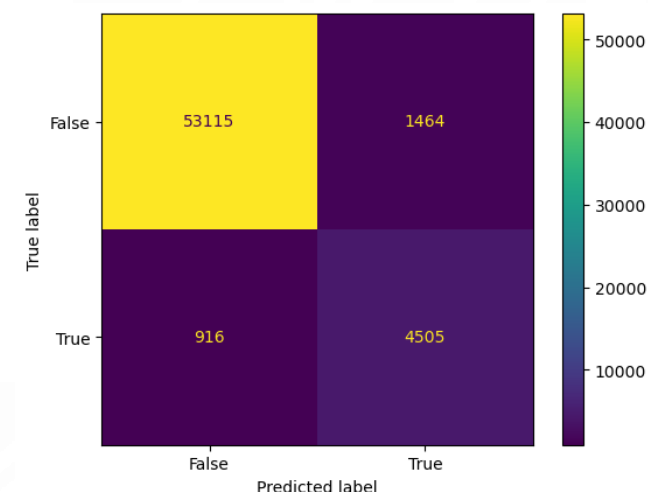
MATRIZ DE CONFUSÃO – métricas para além da acurácia

[5] Especificidade ou Taxa de Verdadeiros Negativos:

Avalia a capacidade do modelo detectar resultados negativos, de modo oposto à sensibilidade

$$Especificidade = \frac{VN}{VN + FP} = \frac{53115}{53115 + 1464} = 97,31\%$$

		Valor Previsto	
		0	1
Valor Real	0	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	1	Falso Negativo (FN)	Verdadeiro Positivo (VP)



[6] **F1-score:** média harmônica entre precisão e revocação, dá mais importância a valores mais baixos e só será alta se a revocação e precisão forem altas. Bom para comparar classificadores. Para pr e rec semelhante, f1 semelhante. Senão f1 se aproxima da pior métrica.

$$F_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = 2 \times \frac{precision \times recall}{precision + recall} = \frac{VP}{VP + \frac{FN + FP}{2}}$$

AVALIANDO MODELOS DE CLASSIFICAÇÃO

E como escolher?

Ocorre que há um trade-off entre precisão e revocação/sensibilidade. Aumentar um reduz o outro!

Exemplo 1: classificador para detectar vídeos seguros (classe positiva) para crianças

Objetivo: rejeitar muitos vídeos bons (baixa sensibilidade) e mantenha apenas os seguros (alta precisão)

OU uma revocação muito alta, mas permita que alguns vídeos ruins sejam exibidos em seu produto?

Exemplo 2: detectar ladrões de lojas em imagens de vigilância

Objetivo: pode ser bom que o classificador tenha somente 30% de precisão, desde que tenha 99% de recall (pode ter poucos alertas falsos)

Métrica	Orientação
Acurácia	Indicador bem geral sobre a performance do modelo em bases balanceadas
Precisão	Use quando é muito importante que predições POSITIVAS sejam acuradas
Revocação (Taxa de Verdadeiro Positivo)	Use quando o custo de FALSOS NEGATIVOS seja maior que de FALSOS POSITIVOS
Taxa de Falso Positivo	Use quando o custo de FALSOS POSITIVOS seja maior que de FALSOS NEGATIVOS

EXERCÍCIOS

- A) Seja uma matriz de confusão com 5VP, 6VN, 3FP e 2FN. Calcule a sensibilidade (0.455, 0.714, 0.625)
- B) Seja uma matriz de confusão com 3VP, 4VN, 2FP e 1FN. Calcule a precisão.
- C) Você está desenvolvendo um classificador binário que verifica se determinado inseto nocivo está presente em fotos de armadilhas inteligentes (smart traps). Se o modelo detecta o inseto, o entomologista é notificado. A detecção antecipada deste inseto é crítica para prevenir uma infestação. Um alarme falso (falso positivo) é fácil de tratar: o entomologista vê a foto que foi classificada incorretamente e a marca como tal. Presumindo um nível aceitável de acurácia, para qual métrica este modelo deve ser otimizado?
 1. Revocação
 2. Precisão
 3. Taxa de Falso Positivo

EXERCÍCIOS

1. Seja uma matriz de confusão com 5VP, 6VN, 3FP e 2FN. Calcule a sensibilidade (0.455, **0.714**, 0.625)

6	3
2	5

$$recall(s) = \frac{V_p}{V_p + F_n} = \frac{5}{5 + 2} = 0.714$$

2. Seja uma matriz de confusão com 3VP, 4VN, 2FP e 1FN. Calcule a precisão.

4	2
1	3

$$precision(p) = \frac{V_p}{V_p + F_p} = \frac{3}{3 + 2} = 0.625$$

3. Você está desenvolvendo um classificador binário que verifica se determinado inseto nocivo está presente em fotos de armadilhas inteligentes (smart traps). Se o modelo detecta o inseto, o entomologista é notificado. A detecção antecipada deste inseto é crítica para prevenir uma infestação. Um alarme falso (falso positivo) é fácil de tratar: o entomologista vê a foto que foi classificada incorretamente e a marca como tal. Presumindo um nível aceitável de acurácia, para qual métrica este modelo deve ser otimizado?

1. **Revocação**
2. Precisão
3. Taxa de Falso Positivo

Neste cenário, alarmes falsos (falsos positivos) tem custo baixo e falsos negativos tem custo muito alto (infestação). Faz sentido maximizar Revocação ou a probabilidade de detecção

4. Leia a documentação do `classification_report` do `sklearn.metrics`