

Métricas de Avaliação



Objetivos

- Conhecer as métricas mais comuns para avaliação de modelos.

Accuracy is telling the truth . . . Precision is telling the same story over and over again.

Yiding Wang

Tarefas de Classificação

Matriz de confusão

Esta matriz indica quantos exemplos existem em cada grupo: falso positivo (FP), falso negativo (FN), verdadeiro positivo (TP) e verdadeiro negativo (TN).

		P R E D I T O	
R E A L	POSITIVO	 POSITIVO  TP verdadeiro positivo	 NEGATIVO  FN falso negativo
	NEGATIVO	 NEGATIVO  FP falso positivo	 TN verdadeiro negativo

É interessante visualizar a contagem destes grupos tanto em números absolutos quanto em porcentagens da classe real, já que o número de exemplos em cada classe pode variar

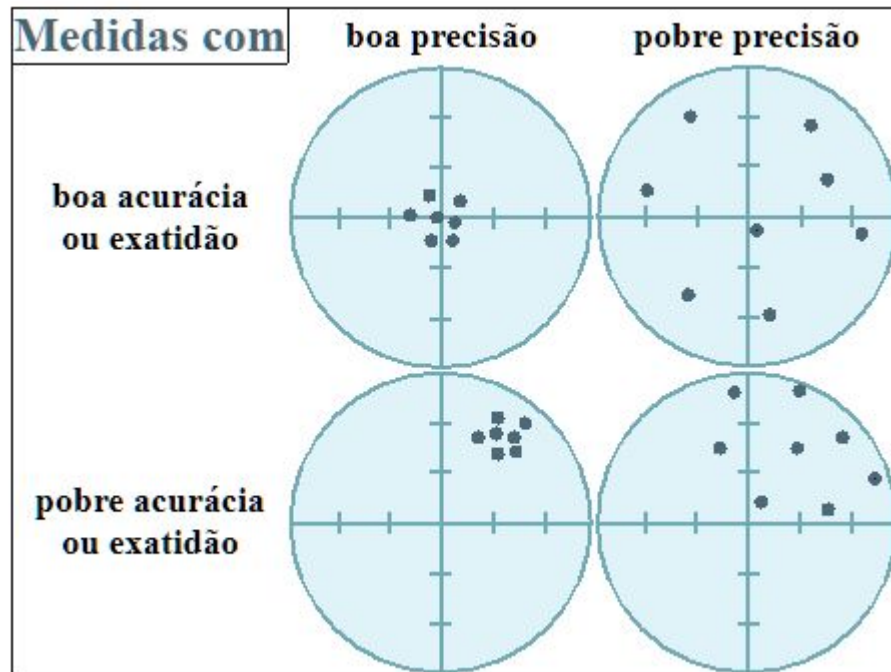


Acurácia (Precisão Geral)

$$Precisão\ Geral = \frac{P}{P + N}$$

- Número de acertos positivos pelo total de exemplos preditos;
- Sensível à distribuição dos dados no *dataset*;
- Suscetível a falsas conclusões para *datasets* desproporcionais;
- Mesma penalidades de acerto e erro para cada classe.





Precisão

$$Precisão = \frac{PV}{PV + FP}$$

Número de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe (positivos verdadeiros), dividido pela soma entre este número, e o número de exemplos classificados nesta classe, mas que pertencem a outras (falsos positivos).



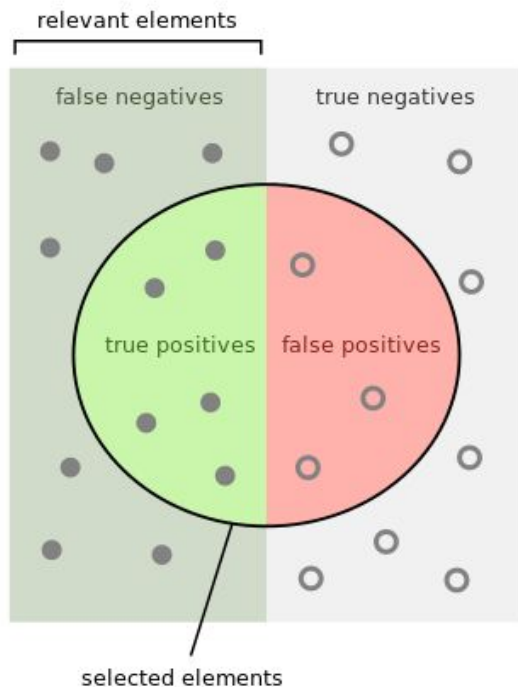
Recall

$$Recall = \frac{PV}{P}$$

Número de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe, dividido pela quantidade total de exemplos que pertencem a esta classe, mesmo que sejam classificados em outra. No caso binário, positivos verdadeiros divididos por total de positivos.



Precision vs Recall



How many selected items are relevant?

$$\text{Precision} = \frac{\text{PV}}{\text{PV} + \text{FP}}$$

$$\text{Precisão} = \frac{PV}{PV + FP}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{PV}}{P}$$

$$\text{Recall} = \frac{PV}{P}$$

F1 Score

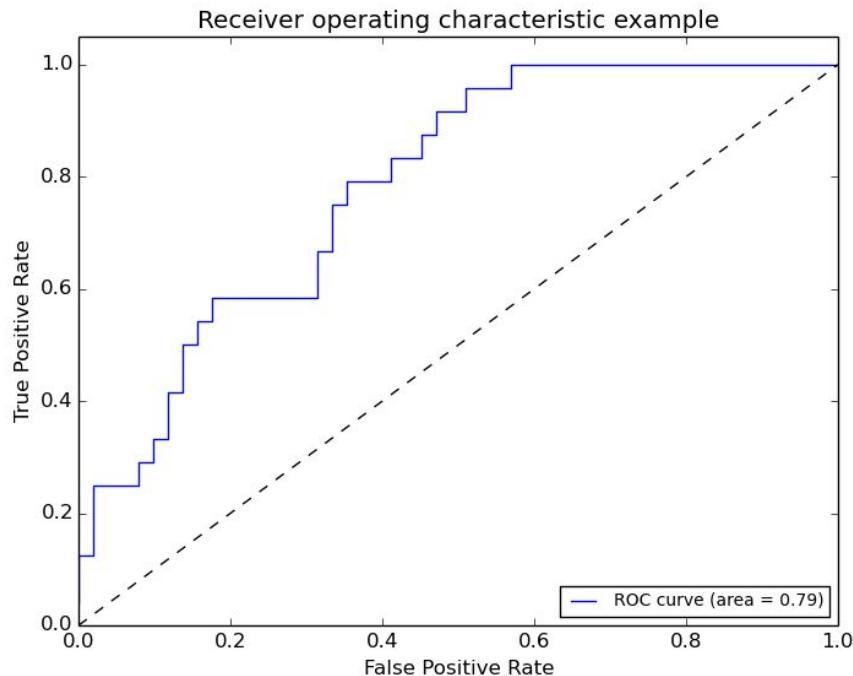
$$F1 = \frac{2 * precis\tilde{a}o * recall}{precis\tilde{a}o + recall}$$

- Medida harmônica entre Precision e Recall;
- Contorna o problema de classes desproporcionais no dataset;



AUC (Area Under the ROC Curve)

- Útil para classes desproporcionais;
- Mede o desempenho do modelo em vários “pontos de corte”;



Log Loss

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

- A fórmula do exemplo é para o caso binário, neste caso: **p** é a probabilidade do exemplo pertencer à classe 1, e **y** é o valor real da variável dependente.
- Esta função pune previsões incorretas muito confiantes. Por exemplo, prever uma classe com uma probabilidade de 95%, e na realidade a correta ser outra;
- Nas situações em que a **probabilidade** de um exemplo pertencer a uma classe for mais importante do que classificá-lo diretamente, esta função é preferível frente a precisão geral.



Tarefas de regressão



Mean Squared Error – MSE

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

$y_i = \text{valor real}$
 $\hat{y}_i = \text{valor previsto}$

- Diferenças menores têm menos importância, enquanto diferenças maiores recebem mais peso;
- Existe uma variação, que facilita a interpretação: o Root Mean Squared Error. Ele é simplesmente a raiz quadrada do primeiro.



Mean Absolute Error – MAE

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- Valor absoluto da diferença em vez do quadrado desta;
- Não define peso de acordo com a magnitude da diferença;
- Média vs Mediana



Mean Absolute Percentage Error – MAPE

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- Este erro calcula a média percentual do desvio absoluto entre as previsões e a realidade. É utilizado para avaliar sistemas de previsões de vendas e outros sistemas nos quais a diferença percentual seja mais interpretável, ou mais importante, do que os valores absolutos.

R^2 - Coeficiente de determinação

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$RSS = \sum_{i=1}^n (y_i - f(x_i))^2$$

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$$

A proporção da variância na variável dependente que é previsível a partir da variável independente. Valores maiores são desejáveis.



Métricas na prática

- **Keras**

- <https://keras.io/metrics>

- **Scikit-learn**

- https://scikit-learn.org/stable/modules/model_evaluation.html



Métricas na prática

- <https://keras.io/metrics>

```
model.compile(loss='mean_squared_error',  
              optimizer='sgd',  
              metrics=['mae', 'acc'])
```

```
from keras import metrics  
  
model.compile(loss='mean_squared_error',  
              optimizer='sgd',  
              metrics=[metrics.mae, metrics.categorical_accuracy])
```



Baseline

- **Classificação:** selecionar a classe que tem mais observações e usar essa classe como resultado para todas as previsões. Se o número de observações for igual para todas as classes em seu conjunto de dados de treinamento, você pode selecionar uma classe específica ou enumerar cada classe e ver qual dá o melhor resultado em seu equipamento de teste.
- **Regressão:** usar uma medida de tendência central como resultado para todas as previsões, como a média ou a mediana.
- **Otimização:** se você estiver trabalhando em um problema de otimização, poderá usar um número fixo de amostras aleatórias no domínio.



Exercício

- Apresente um gráfico com o comportamento das métricas Acurácia e Erro Quadrático Médio para o problema de classificação de diabetes;
- Apresente os resultados de r^2 para um problema de regressão.



Obrigado.