

Métricas de Avaliação

Mateus Fernandes de Souza

16 de março de 2022

1 Definições

1.1 Acurácia

É a razão entre as predições corretas pelo total.

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

1.2 Precision

É o número de verdadeiros positivos, dividido pelo número de positivos estimados pelo modelo

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

1.3 Recall

Dado que o estado verdadeiro é positivo, qual a proporção de verdadeiro positivo.

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

1.4 F1-Score

É uma média harmônica entre o Precision e o Recall.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

1.5 ROC-AUC

A curva Roc cria uma curva em um gráfico com 2 variáveis (figura 1): no eixo X é a taxa de falsos positivos e no eixo Y é a taxa de verdadeiros positivos. Já o AUC é a área de baixo, quanto mais próximo de 1, melhor será o seu modelo.

2 Quando Utilizar

2.1 Acurácia

Em geral é uma boa forma de se verificar a qualidade de um modelo, pois é uma métrica que avalia o corpo geral do resultado do seu código, porém ela considera os erros do tipo I e II com o mesmo peso, caso os tipos de erro tenham alguma importância diferente, sendo um mais impactante que o outro, não é o mais viável e recomendável

2.2 Precision

É uma boa opção quando o erro do tipo I é mais importante que o erro do tipo II. Define a quantidade que o modelo disse que são positivos, quantos realmente são positivos. Ele dá uma sensação de impacto.

2.3 Recall

É uma boa opção quando o erro do tipo II é mais importante que o erro do tipo I. Define a quantidade que são positivos, quantos o modelo acertou. Ele dá uma sensação de captura.

2.4 F1-Score

É uma boa forma para se verificar os erros em geral do modelo, vendo que ele faz uma média harmônica dos dois resultados anteriores. Quando os erros são mais importantes que os acertos na previsão, é uma métrica recomendável. OBS: Complementando com essa métrica, temos o FBeta Score, que simplesmente adiciona um peso para o Precision no cálculo gerando assim uma diferença de importância entre os erros do tipo I e II.

$$F\beta = (1 + \beta^2) \frac{Precision * Recall}{\beta^2 * Precision + Recall}$$

2.5 ROC-AUC

É uma boa escolha quando se procura um limiar específico entre a taxa de verdadeiros positivos e falsos positivos, pois, normalmente, a taxa de verdadeiros positivos e falsos positivos acaba sendo inversamente proporcional, quanto mais crescermos a taxa de um, a taxa do outro tende a crescer também.

3 O Recomendável Para Classes Desbalanceadas

A acurácia para dados desbalanceados não é provavelmente a melhor opção, visto que nosso modelo está enviesado pela diferença de classes, ele em sua maioria vai acertar grande parte dos casos, mesmo que os erros do tipo I e II fiquem mais altos pelo padrão a ser seguido, o seu resultado será alto.

Para um modelo desbalanceado o recomendável é se observar o F1 Score. Já que nosso modelo está inicialmente enviesado para a classe de maior quantidade, a taxa de erros provavelmente será muito alta, fazendo assim com que nosso modelo talvez fique com essa métrica baixíssima.

Quando aplicamos um método para igualar as classes, na acurácia não iremos, provavelmente, perceber uma mudança muito nítida, mas ao compararmos com o F1 Score, iremos verificar, provavelmente, uma mudança para melhor, visto que agora temos um modelo honesto e sem viés, assim provavelmente cometendo menos erros que anteriormente.

Mas, além de se observar isso, é muito necessário, também, observar o ROC-AUC, pois ela acaba sendo em muitas vezes mais precisa no que se procura e também justamente define o limiar para a diferenciação da melhor forma possível entre as duas classes.

4 Adendo

O erro do tipo I é a suposição que algo é verdadeiro, mas que na realidade é falso. O erro do tipo II é a suposição que algo é falso, mas que na realidade é verdadeiro.

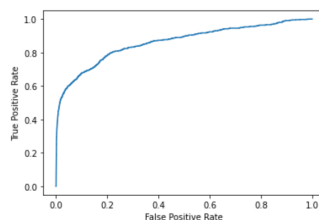


Figura 1: Curva ROC, retirada do código Bank Fraud Logistic Regression