



Data Science Academy

www.datascienceacademy.com.br

Big Data Analytics com R e Microsoft Azure
Machine Learning

Interpretando um Modelo de Classificação

Há duas subcategorias de problemas de classificação:

- Problemas com apenas duas classes (Classificação de duas classes ou binária)
- Problemas com mais de duas classes (Classificação multiclasse)

O Azure ML tem diferentes módulos para lidar com cada um desses tipos de classificação. Mas as maneiras de interpretar os resultados de previsão são muito semelhantes. Vamos estudar sobre os problemas de classificação de duas classes.

Avaliar um modelo de classificação binária

Em um cenário de classificação binária, a variável a ser prevista tem somente dois resultados possíveis, por exemplo: {0, 1}, {falso, verdadeiro}, {negativo, positivo}, {bom pagador, mau pagador}.

Depois de executar o experimento no Azure ML, você pode clicar na porta de saída do módulo Avaliar Modelo (Evaluate Model) e selecionar Visualizar para ver os resultados da avaliação. As métricas de avaliação disponíveis para modelos de classificação binária são:

- Exatidão (Accuracy)
- Precisão (Precision)
- Recuperação (Recall)
- Pontuação F1 (F-Score)
- AUC (Area Under the Curve)

Accuracy → mede a exatidão do modelo de classificação, como a proporção de resultados verdadeiros em relação ao total de casos analisados. Quanto maior, melhor!

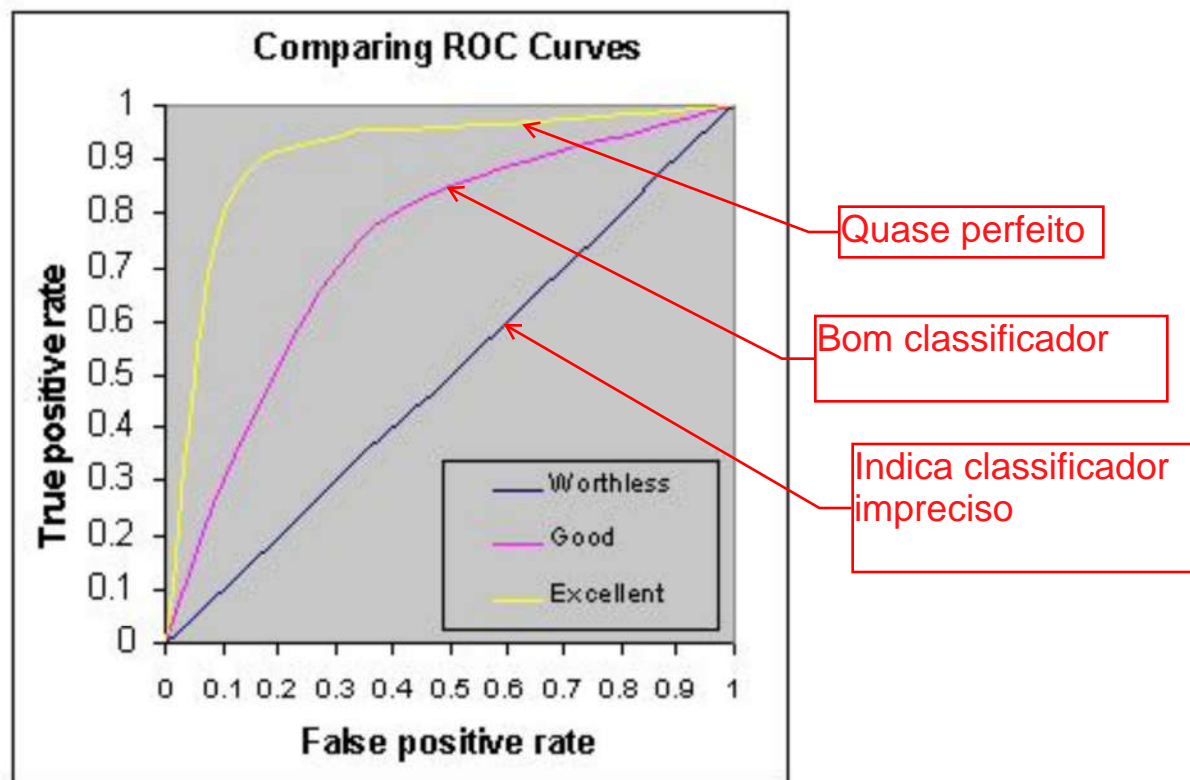
Precision → é a proporção de resultados verdadeiros sobre os resultados positivos. Quanto maior, melhor!

Recall → é a fração de resultados corretos retornados pelo modelo. Quanto maior, melhor!

F-Score → é a média ponderada entre a precisão e o recall. O valor ideal para o F-Score é igual a 1.

AUC → mede a área sob a curva, sendo os verdadeiros positivos no eixo y e os falsos positivos no eixo x. Esta métrica é útil, pois gera uma única medida que permite comparar modelos diferentes.

Além disso, o módulo gera uma matriz de confusão (Confusion Matrix) mostrando o número de verdadeiros positivos, de falsos negativos, de falsos positivos e de verdadeiros negativos, bem como as curvas ROC (Receiver Operating Characteristic).

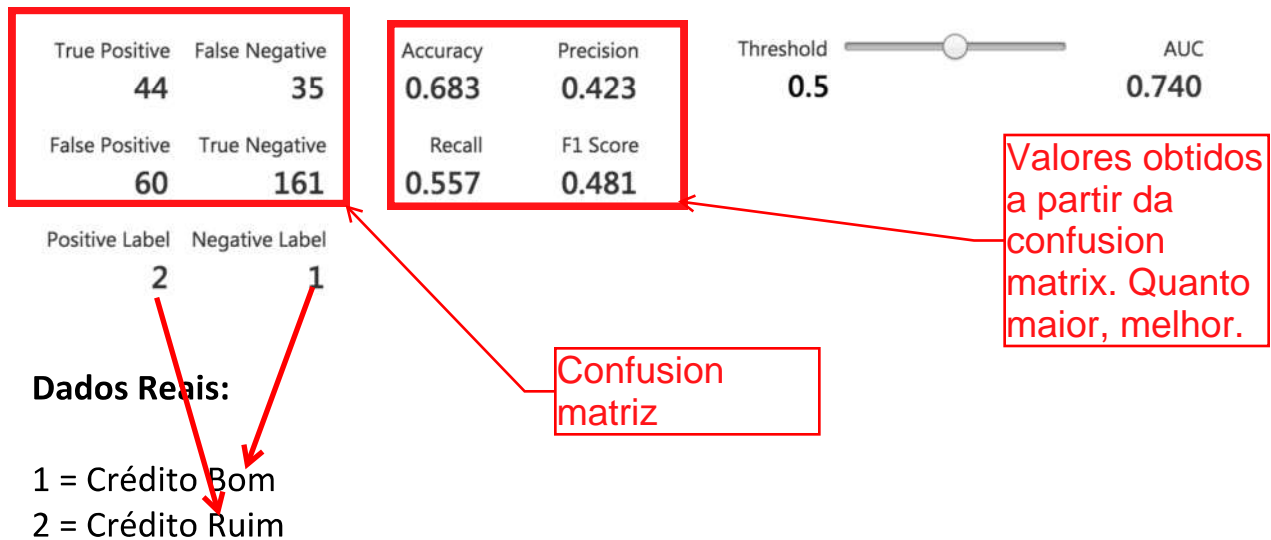


Quanto mais à esquerda a curva estiver, maior a precisão do seu modelo!

- .90-1 = excellent (A)
- .80-.90 = good (B)
- .70-.80 = fair (C)
- .60-.70 = poor (D)
- .50-.60 = fail (F)

A precisão é simplesmente a proporção das instâncias classificadas corretamente. Geralmente, ela é a primeira métrica a observar ao avaliar um classificador. No entanto, quando os dados de teste estão desbalanceados (em que a maioria das instâncias pertence a uma das classes) ou você está mais interessado no desempenho em qualquer uma das classes, a precisão pode não capturar a eficácia de um classificador.

Por esse motivo, é útil computar métricas adicionais que capturam aspectos mais específicos da avaliação. Antes de entrar em detalhes sobre essas métricas, é importante compreender a Confusion Matrix de uma avaliação de classificação binária. Os rótulos de classe no conjunto de treinamento podem assumir apenas 2 valores possíveis, o que normalmente chamamos como positivo ou negativo. As instâncias positivas e negativas que um classificador prevê corretamente são chamadas verdadeiros positivos (TP) e verdadeiros negativos (TN), respectivamente. Da mesma forma, as instâncias classificadas incorretamente são chamadas de falsos positivos (FP) e falsos negativos (FN). A Confusion Matrix é simplesmente uma tabela que mostra o número de instâncias que se enquadram em cada uma dessas 4 categorias. **O Azure ML decide automaticamente qual das duas classes no conjunto de dados é a classe positiva.** Se os rótulos de classe forem inteiros ou booleanos, as instâncias rotuladas como 'true' ou '1' serão atribuídas à classe positiva. Se os rótulos forem cadeias de caracteres, os rótulos serão classificados em ordem alfabética e o primeiro nível será escolhido para ser a classe negativa, enquanto o segundo nível será a classe positiva. Mas estas “rotulações” podem variar.



Azure ML:

Positive Label = 2 = Crédito Ruim

Negative Label = 1 = Crédito Bom

- TP - Crédito Ruim classificado como crédito Ruim
- FN - Crédito Ruim classificado como crédito Bom
- FP - Crédito Bom classificado como Crédito Ruim
- TN - Crédito Bom classificado como Crédito Bom

Você observará que há uma relação óbvia entre a precisão (Precision) e a recuperação (Recall). Por exemplo, dado um conjunto de dados relativamente equilibrado, uma classificação que prevê principalmente instâncias positivas, teria um Recall alto, mas um Precision bastante baixo com muitas instâncias negativas, resultando em um grande número de falsos positivos. Para ver um gráfico de como essas duas métricas variam, você pode clicar na curva 'PRECISION/RECALL' na página de saída do resultado de avaliação.

Outra métrica usada com frequência é a Pontuação F1, que calcula a relação entre Precision e Recall. Ela é a média ponderada dessas 2 métricas e é calculada como tal: $F1 = 2 (\text{precisão} \times \text{recuperação}) / (\text{precisão} + \text{recuperação})$. O score F1 é uma boa maneira de resumir a avaliação de um único número, mas é sempre uma boa prática examinar Precision e Recall para entender melhor como um classificador se comporta.

Além disso, é possível inspecionar a taxa de verdadeiros positivos versus a taxa de falsos positivos na curva ROC (Característica de Operação do Receptor) e o valor correspondente AUC (Área sob a Curva). **Quanto mais próxima essa curva estiver do canto superior esquerdo**, melhor será o desempenho do classificador (que é maximizar a taxa de positivos verdadeiros enquanto minimiza os falsos positivos).

Existem ainda 2 importantes métricas que podem ser consideradas:

Average log loss → é um único score usado para expressar as penalidades para resultados errados. Este score é calculado como a diferença entre 2 distribuições de probabilidade: a distribuição real dos dados e a distribuição gerada pelo modelo.

Training log loss → é um único score que representa a vantagem do classificador sobre predição randômica.