

Utilização de métricas para avaliar um modelo de Machine Learning

Ladies and Gentlemen, bem vindos à mais uma edição do Turing Talks! Hoje falaremos sobre Métricas para avaliação de modelos de Machine Learning.

Introdução

Primeiramente, o que é Machine Learning? Bom, Machine Learning (Aprendizado de Máquina) é um ramo da inteligência artificial que se concentra na criação de aplicativos que aprendem e melhoram de forma automática e periódica com a experiência, sem precisar de alguém o tempo todo dizendo ao computador como algo deve ser feito.

Um exemplo comum de aplicação é na recomendação de filmes e séries do catálogo de plataformas de streaming, à medida que você consome o conteúdo na plataforma, o aplicativo “aprende” quais são os seus gêneros preferidos, e a partir disso recomenda filmes e séries semelhantes com aqueles que você assistiu. O computador é capaz de fazer isso a partir de um modelo, que não precisa sofrer alterações cada vez que você assiste um filme diferente, isso daria muito trabalho! Esse modelo precisa passar por alguns testes e avaliações antes de chegar à sua versão final seguindo algumas métricas, e é justamente sobre essas métricas que falaremos nesse Turing Talks.

Acurácia

O modelo passa por uma série de testes para verificar a quantidade de acertos que ele foi capaz de fazer, após esses testes, pode-se utilizar esse método da acurácia para determinar os acertos em relação às tentativas totais, é um cálculo bem simples de se fazer:

$$\text{Acurácia} = \frac{\text{Número de predições corretas (acertos)}}{\text{Número total de previsões feitas}}$$

Por ser um cálculo simples, devem ser tomados alguns cuidados ao utilizá-lo, pois ela só funciona se houver um número igual de amostras pertencentes a cada classe. Por exemplo, considere que há 98% das amostras da classe A e 2% das amostras da classe B em nosso conjunto de treinamento. Se o nosso modelo acertar todas as amostras da classe A e errar todas da classe B, então será 98% de acerto, mesmo errando todas da classe B. Isso não acontece quando temos uma certa igualdade nas quantidades em cada classe.

Erro médio absoluto e Erro médio quadrático

Esses são dois tipos de métricas que também são muito utilizados. Ambas são bem semelhantes, elas basicamente calculam quão longe os dados apresentados pelo modelo estão do resultado ideal, e após o exemplo entenderemos qual é a diferença entre as duas e porque ela é importante.

Considere as notas de duas salas de aula diferentes, sala A e sala B, com cinco alunos cada uma. Por enquanto concentre-se apenas nas duas primeiras colunas da tabela:

Sala A	Sala B	ErroMA _A	ErroMA _B	ErroMQ _A	ErroMQ _B
7	9	1	1	1	1
9	10	1	2	1	4
6	10	2	2	4	4
8	1	0	7	0	49
10	10	2	2	4	4
Média: 8	Média: 8	ErroMA: 6	ErroMA: 14	ErroMQ: 10	ErroMQ: 62

Nas duas primeiras colunas temos a nota tirada por cada um dos alunos e a média da sala na última linha. A partir disso, o cálculo do Erro médio absoluto (ErroMA) é calculado pelo módulo da subtração de cada uma das notas pela média, por exemplo: para o primeiro aluno da sala A temos: $|7 - 8| = 1$.

Após feito o erro de cada aluno, é somado cada um deles, resultando em 6 para a turma A, e 14 para a turma B. Isso é o Erro médio absoluto. Já para o Erro médio quadrático, é feito apenas um passo a mais, que é elevar ao quadrado cada um dos erros da terceira e quarta coluna, e somente após isso realizar a soma.

Com o exemplo dado, vemos que o erro médio para a turma B é maior que o erro médio para a turma A, isso significa que as notas da turma B ficaram mais distantes da média, tanto para baixo quanto para cima. E agora você pode estar se perguntando: qual a utilidade do Erro médio quadrático se ambos retornam o mesmo resultado, apenas com uma diferença numérica?

Bom, o Erro médio quadrático destaca muito mais os valores que estão mais distantes da média. Na sala B, um dos alunos tirou nota 1, o que distancia muito da média 8, e isso influencia bastante quando calculamos o quadrado da diferença, portanto, essa métrica é muito importante para testar um modelo de machine learning que deve fazer

cálculos muito precisos, e que não tolera erros muito grandes, pois se esse tipo de erro ocorrer, ele será muito destacado.

Conclusão

Esses são apenas alguns exemplos de métricas para avaliar modelos de machine learning, e pudemos perceber que eles são muito importantes para testar se um determinado modelo cumpre bem a sua função. Portanto, escolher a métrica certa é crucial para a validação, pois nem todas as métricas servem para todos os tipos de modelos.

Referências

<https://www.g2.com/articles/machine-learning>

<https://www.flai.com.br/juscudilio/qual-a-melhor-metrica-para-avaliar-os-modelos-de-machine-learning/>

<https://www.mariofilho.com/as-metricas-mais-populares-para-avaliar-modelos-de-machine-learning/>

<https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/>

<https://medium.com/turing-talks/aprendizado-por-refor%C3%A7o-1-introdu%C3%A7%C3%A3o-7382ebb641ab>

<https://medium.com/turing-talks/como-avaliar-seu-modelo-de-classifica%C3%A7%C3%A3o-acd2a03690e>

<https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/>

<https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>

https://www.youtube.com/watch?v=_pJX2SFoFtY