

Aula 33 - Projeto Final (Parte 1)

Olá a todos, sejam muito bem-vindos à última parte do nosso curso de aprendizado de Máquina. Nestas últimas aulas, vamos abordar a criação de um projeto de aprendizado de máquina utilizando Python, e o objetivo principal é o reconhecimento de padrões.

Embora o aprendizado de máquina envolve muita matemática a gente não vai focar em como implementar esses algoritmos, afinal esse curso é apenas um curso introdutório e não para detalhar como esses conceitos são aplicados, mas sim a compreensão dos conceitos básicos de aprendizado de máquina, como os diferentes tipos, incluindo classificação e regressão, que exploramos ao longo do curso.

Além disso, graças às bibliotecas do python como o Pandas e o scikit-learn, a gente não precisa implementar esses algoritmos do zero, isso porque estas bibliotecas já fornecem todas as funcionalidades necessárias para treinar nossos algoritmos de aprendizado de máquina.

Para começar, é bom lembrar que a maneira como um computador aprende é semelhante à maneira como nós aprendemos, assim como eu disse no início deste curso. E como que uma criança que está nas suas fases iniciais de aprendizado aprende?

Uma criança aprende por meio da observação, experimentação e prática. Ela tenta algo novo, observa o resultado e aprende com os erros e acertos. Da mesma forma, um computador aprende por meio da análise de dados e padrões, treinamento em um conjunto de dados e ajuste de parâmetros até que os resultados sejam satisfatórios. É um processo iterativo e gradual, assim como o aprendizado humano.

Vamos considerar um exemplo clássico de classificação, frequentemente utilizado em cursos introdutórios de aprendizado de máquina. Nesse caso, nosso objetivo é classificar qual animal está sendo apresentado, se é um cachorro ou um gato.



Vamos considerar um exemplo clássico de classificação, frequentemente utilizado em cursos introdutórios de aprendizado de máquina. Nesse caso, nosso objetivo é classificar qual animal está sendo apresentado, se é um cachorro ou um gato.

Então vamos imaginar que a gente tenha uma tabela que representa algumas características que esses animais possuem, e a gente vai ter um valor para representar se esse animal possui ou não determinada característica.

Vamos usar 0 para representar a ausência de uma característica e 1 para representar a presença de uma característica.

Pelo longo?	Perna curta?	late?	gato ou cachorro?
0	1	0	1
0	1	1	1
0	1	0	1
0	0	1	0
1	1	1	0
1	0	1	0

Então, nessa tabela, existem diversos animais, essa tabela vai ser a nossa base de dados rotulados. Mas o que isso quer dizer? A gente pode fazer um paralelo com a nossa própria aprendizagem. Antes da gente saber o que era um cachorro, alguém nos disse o que representa um cachorro, que late, que tem um determinado tamanho, dentre outras coisas.

A partir daí a gente foi associando essas características, com outros cachorros que a gente vai encontrando, através do reconhecimento de padrões a gente é capaz de dizer se um outro animal é ou não um cachorro.

Os dados rotulados, são basicamente isso, é como a gente diz para o nosso algoritmo o que representa um cachorro, e a partir daí, ele vai ser capaz de fazer o mesmo que a gente, e dizer se um novo animal desconhecido é ou não um cachorro, com levando em consideração as características que a gente definiu como sendo de um cachorro. Isso é o que a gente chama de aprendizagem supervisionada. Que nesse caso os dados que a gente passa para o algoritmo foram rotulados por nós. Como fizemos na tabela.

É importante lembrar que isso é uma maneira muito simplificada de fazer um algoritmo desse, e que quanto maior a nossa base de dados, maior vai ser a precisão do nosso algoritmo.

Analisando a tabela, a gente vai ter o que a gente chama de *features*, que são as características do nosso animal, e serão armazenadas em uma tabela com três colunas, que chamaremos de X. Cada linha da tabela representa um animal e suas características.

Além disso, teremos uma coluna adicional, chamada Y, que indicará a qual categoria o animal pertence. Por exemplo, se estamos classificando animais, Y pode indicar se o animal é um cão, um gato ou um pássaro. Se estivéssemos classificando e-mails, Y poderia indicar se o e-mail é spam ou não.

Assim, o programa será treinado com base nessas características (X) e em suas respectivas categorias (Y) para aprender a identificar corretamente novos animais (ou e-mails, clientes, produtos, etc.) com base nas mesmas características.

Quando queremos ensinar um programa a classificar animais (ou outras coisas), é importante treinar o programa de forma que ele se torne eficiente na tarefa. Para medir a eficiência do programa, utilizamos uma medida chamada acurácia (ou accuracy, em inglês), que indica a porcentagem de classificações corretas em relação ao total de classificações feitas.

No entanto, é importante lembrar que os programas podem cometer erros e nem sempre classificam as coisas corretamente. Por isso, é fundamental avaliar a acurácia e outros parâmetros para identificar possíveis erros e melhorar a eficiência do programa.

Bom, e como eu já disse, a gente vai trabalhar com um algoritmo de classificação para classificar os nossos dados. E como eu disse também antes, a gente não precisa criar o código que implementa um algoritmo de classificação, a gente já tem isso pronto. Pra isso, a gente vai usar aqui, é uma biblioteca muito famosa para aprendizado de máquina, que é o Scykitlearn.

Além do mais, existem conceitos que a gente precisa saber para analisar o nosso algoritmo, saber se ele foi preciso ou não, entre outras coisas, e os conceitos que precisamos saber são os seguintes:

1.Features

2.Itens

3.Treino do estimador

4.Testes

5.Taxa de acerto

6.Otimização

7.Previsão

Mas estes conceitos a gente vai ver mais detalhadamente na próxima aula, muito obrigado por assistir até aqui e até a próxima.