

PROCESSO DE DESENVOLVIMENTO DOCUMENTAÇÃO DE PROJETO		
Nome do curso: <i>Introdução à Machine Learning</i>	Introdução à Classificação	Responsável: <i>Paulo Jarbas Camurça</i>

Introdução à Classificação

Olá, seja bem-vindo!

Uma das aplicações mais comuns de Machine Learning é a resolução de problemas, para que se possa tomar decisões, como definir se um cliente pode, ou não, receber um limite maior em uma análise de crédito de um banco; ou se um paciente tem, ou não, uma determinada doença, com base nos dados dos sintomas. Para essa categoria de problemas, utiliza-se uma técnica chamada de classificação.

Neste curso, você poderá conhecer melhor essa técnica, entendendo seus tipos e funcionamento. Antes de mais nada, é importante lembrar que regressão e classificação são métodos do aprendizado supervisionado, conteúdo que, provavelmente, é familiar para você, não é mesmo?

Pois bem, a classificação é uma técnica que tem o objetivo de classificar ou categorizar uma informação nova ou desconhecida, com base em um conjunto de informações previamente conhecidas. Por isso, para compreender melhor esse conceito, entenda os tipos de classificação, são eles: binária e multiclasse.

Na classificação binária, tem-se apenas dois tipos de classes como resposta, ou seja, possui apenas duas opções de escolha, como: 1 ou 0, sim ou não, e verdadeiro ou falso, por exemplo. Caso você crie um modelo, que faça a predição se um time vai ou não ganhar uma partida, esse modelo pode ser definido como classificação binária, pois também se aplica a um sistema que identifica se uma fruta está ou não estragada, com base em imagens de diversas frutas previamente rotuladas. Bem simples, não é?

Já a classificação multiclasse tenta classificar a resposta de interesse em mais de duas categorias ou classes. Acompanhe o exemplo, a seguir, para entender melhor esse tipo.

Imagine que o dono de uma fábrica de roupas deseja aumentar a velocidade e melhorar o fluxo de produção das peças. Para isso, ele pediu que os desenvolvedores da empresa implementassem um sistema que classifique, de forma automática, os tipos de roupas que são fabricadas diariamente, em três tipos diferentes: blusas, bermudas e calças.

Assim, com as roupas sendo classificadas automaticamente ao serem produzidas, o fluxo da produção seria otimizado, pois as peças seriam direcionadas, de modo muito rápido, de acordo com o seu tipo para o setor de entregas.

Para realizar essa tarefa, foram coletados vários dados sobre as roupas produzidas, como o tipo do tecido: se possui gola, mangas e pernas; nesse caso, o tamanho das pernas, dentre outras características. Em seguida, todos esses dados foram registrados em uma tabela, relacionando cada anotação com o verdadeiro tipo da roupa.

Com esses dados organizados, é possível utilizar um modelo de classificação para prever o tipo de uma nova roupa, logo que for fabricada. O modelo de classificação para esse problema é do tipo **multiclasse**, pois existem três categorias de roupas a serem classificadas. Agora ficou mais fácil de perceber a diferença entre classificação binária e de multiclasse, você não acha?

Bom, dando continuidade ao nosso tema desta aula, é importante que você entenda quais etapas são essenciais para construir um modelo de classificação, também chamado de classificador. Em resumo, é necessário realizar duas etapas: a de treino e a de teste do algoritmo; ou seja, um modelo, em Machine Learning, pode ser definido como um algoritmo que foi treinado e testado. Interessante, heim?

Conheça primeiro a etapa de treino do algoritmo: ela consiste em passar uma parte desses dados, em forma de tabela, como dados de entrada, e, em seguida, o algoritmo fará diversas iterações na busca de padrões de similaridade.

Finalizado o treino, será iniciada a etapa de testes. Nela, é submetido, para o algoritmo, os dados que não foram usados na etapa de treino, com isso é possível descobrir o quanto o modelo consegue acertar através de métodos de validação.

Depois de treinar e testar o algoritmo de classificação, o modelo está pronto para ser utilizado. Assim, no caso da fábrica de roupas, a cada nova roupa produzida, o algoritmo irá classificar automaticamente o seu tipo em uma das três classes pré-definidas. Muita coisa não é mesmo?

Até aqui, você pôde compreender o conceito de classificação, em quais problemas essa técnica pode ser utilizada, além de conhecer alguns tipos de classificação, como binária e a de multiclasse. Dominando essa técnica você será capaz de criar diversos modelos de Machine Learning.

Existem alguns exemplos de algoritmos muito usados em classificação, como K-Nearest Neighbor (Knn), Decision Trees ou Árvore de Decisão e Regressão logística. Ficou curioso sobre esses novos algoritmos? Bom, você pode continuar seu processo de aprendizagem, aprofundando

seus conhecimentos em Machine Learning para ir além do que foi apresentado até aqui. Portanto, não se esqueça de revisar o conteúdo abordado e resolver os exercícios propostos, está bem?

Bons estudos e até mais!