

Quando usar aprendizado de máquina

Então, acabamos de conversar sobre várias maneiras diferentes em que redações têm usado o aprendizado de máquina, e espero que tenha sido inspirador.

Mas deixe-me tentar dar algumas orientações sobre quando você deveria considerar usar ou não usar aprendizado de máquina para um projeto.

A primeira pergunta que você deve fazer é: Para o que quer que seja que você está tentando construir um modelo, você pode apenas fazer essa tarefa manualmente? Por exemplo, digamos que você está analisando crimes. Você tem 300 relatos de crimes diferentes que você tem que categorizar. Vai demorar muito mais tempo construir um modelo que possa categorizar esses exemplos e provavelmente precisar de mais dados para isso do que você, ou você e seus estagiários, categorizar esses dados manualmente. Se é concebível que você poderia fazê-lo de uma vez só, sempre prefira fazer sua tarefa manualmente.

A próxima pergunta é: você tem o conjunto de dados de que você precisa para fazer o que você quer? Há muitas tarefas gerais nas quais você pode usar um modelo pré-existente. Vamos falar sobre isso um pouco mais tarde na parte de ferramentas. Mas se você tem apenas uma foto e quer saber o que está na foto - É uma estátua? Uma cadeira? Um cão? - Você não precisa de um grande conjunto de dados de treinamento. Você provavelmente pode usar um modelo que já existe para fazer isso. Mas se você quer fazer algo muito específico, como identificar aviões espiões, você vai precisar de centenas ou milhares de exemplos de voos e seus rótulos. Assim, você precisa ter esses dados, saber onde encontrar esses dados, ou estar disposto a investir na criação desse tipo de conjunto de dados.

Próxima pergunta: Você precisa de explicabilidade? O que é explicabilidade? Muitas pessoas, acho eu, entendem que quando um modelo de aprendizado de máquina, especialmente uma rede neural, faz uma previsão, nós nem sempre entendemos como ele chegou a essa conclusão. Por exemplo, se um crime é classificado como sendo um crime violento, que palavras o modelo analisou para chegar a essa conclusão? Sabe, não é sempre claro nem mesmo para o criador do modelo. Às vezes, não importa se você entende por que um modelo chegou a uma decisão. Porque, por exemplo, no caso do LA Times, quando estávamos tentando ver se um crime foi ou não erroneamente classificado como sendo violento, os repórteres e os editores podiam analisar e dizer: "Ok, este modelo fez essa previsão, mas deixe-me conferir se isso está correto." Então o modelo está ajudando você de certa forma, mas o editor está sempre lá conferindo os resultados. Mas você tem que considerar se quando você está reportando uma história, o que o modelo fez, se isso importa, porque talvez você não consiga ter uma explicação satisfatória de por que o modelo tomou determinada decisão.

E, finalmente, e mais importante: você pode tolerar erros? Praticamente todos os modelos de aprendizado de máquina cometem erros. É inevitável. E a maneira correta de usar o aprendizado de máquina é dar conta desses erros, entender quando eles são realmente

sérios, entender quando eles são aceitáveis, e entender como lidar com eles. Então o que eu tenho enfatizado é que você realmente quer usar o aprendizado de máquina complementando-o com o trabalho de reportagem. E penso nisso como usar o aprendizado de máquina para encontrar agulhas em palheiros ou para encontrar pistas em meio a muitos e muitos dados, para que um modelo diga que você pode querer investigar este avião ou você pode querer investigar se este crime foi violento ou não. Então você trabalha com o modelo, mas você sabe que ele vai cometer erros. Então, simplesmente conte com eles em seu processo de reportagem.

Espero que tenha sido um bom resumo de quando você deve e não deve usar o aprendizado de máquina.