

Aprendizado de Máquina - Lista 1

Lucas Benjamim Cunha Bandeira - 388717

Fevereiro de 2019



Universidade Federal do Ceará
Campus Quixadá
Regis

Aprendizado de Máquina

1. O que é Aprendizado de Máquina e para que serve?

Aprendizado de Máquina é a ciência de programar computadores a aprenderem a partir de dados. O Aprendizado de máquina facilita o trabalho dos desenvolvedores de forma com o que eles não precisem criar um algoritmo altamente específico para determinada tarefa, tornando mais viável a criação dele, pois o Aprendizado de Máquina melhora sua performance a partir de uma experiência que ele obtém.

2. O que são dados rotulados (labels) e para que servem?

São dados que possuem informações na qual o algoritmo irá tirar base para prever informações sobre novos dados.

3. Quais os problemas mais comuns de aprendizado supervisionado?

Aprendizado Supervisionado requer uma quantidade de dados rotulados suficiente para o treinamento.

4. Quais os problemas mais comuns de aprendizado não-supervisionado?

Problemas com a falta de idéia do que os resultados devem aparentar.

5. Que tipo de algoritmo de Machine Learning (ML) você usaria para:

- Permitir um robô andar em diversos tipos de terreno?

Seria utilizado aprendizado supervisionado de classificação, para ele classificar o terreno e utilizar o seu determinado recurso para operar sobre o terreno.

- Segmentar clientes em múltiplos grupos?

Seria utilizado aprendizado não-supervisionado, para ele detectar padrões e determinar em quais grupos cada cliente deve se encaixar.


6. O que é um sistema de aprendizado online? Quais suas vantagens e desvantagens?

Em aprendizado Online o sistema é treinado incrementalmente, sendo alimentado por instâncias sequenciais.

Vantagens:

- O sistema aprende novos dados à medida que eles chegam.
- Cada um desses novos aprendizados são rápidos e baratos.
- Quando os dados são aprendidos, não são mais necessários, portanto podem ser excluídos para que não possa gerar um montante de dados.

- Pode usar o aprendizado out-of-core

Desvantagens: 

- Se dados ruins forem disponibilizados ao sistema, a performance do sistema irá cair gradualmente

7. O que significa aprendizado out-of-core?

Out-of-core é quando um sistema tenta aprender a partir de um conjunto de dados maior que sua memória, então o algoritmo aprende esses dados por partes até que ele tenha sido completamente aprendido.

8. Que tipo de aprendizado usa medidas de similaridade para fazer previsões?

É utilizado o instance-based learning. O sistema aprende os exemplos e então gera para novos casos usando a medida de similaridade.

9. Em um modelo de ML, qual a diferença entre parâmetros e hiper-parâmetros?

A diferença é que um hiper-parâmetro não é afetado pelo próprio algoritmo de aprendizado e deve ser definido antes do treinamento e permanecer constante.

10. O que são e qual a diferença entre Underfitting e Overfitting? O que fazer para solucionar cada um desses problemas?

- **Overfitting acontece quando o modelo é muito complexo em relação ao quantidade e distorção dos dados de treinamento. Uma forma de reduzir o Overfitting é simplificando o modelo, usando um com menos parâmetros, ter mais dados de treino e até mesmo reduzindo a distorção nos dados de treino.**
- **Underfitting é o contrário do Overfitting, ou seja, acontece quando o seu modelo é muito simples para aprender a estrutura dos dados. Uma forma de concertar isso é escolhendo um modelo mais poderoso, com mais parâmetros, é possível também disponibilizando melhores características para o algoritmo de aprendizagem e reduzindo as restrições no modelo.**

11. Para que servem: conjunto de treino, conjunto de validação e conjunto de teste?

Estes conjuntos servem para obter-se o resultado de como o seu algoritmo está performando. O conjunto de treino é a base de dados utilizada para treinar o algoritmo, já o conjunto de teste serve para testar o modelo criado a partir do treino. Logo, é normal ser criado um algoritmo que performe bem no seu conjunto

de teste, porém o algoritmo deverá ser utilizado para novos dados, e é possível que a performance dele não seja tão boa quanto a que ele estava obtendo com o conjunto de teste. Daí entra o conjunto de validação, que funciona como um conjunto de teste, só que antes de teste final, portanto é feito o algoritmo testando com o conjunto de validação e no final com o conjunto de testes para obter uma resposta mais fiel.

12. O que significa validação cruzada e qual a vantagem de usá-la ao invés de usar um conjunto de validação?

Validação Cruzada significa utilizar o conjunto de testes, separá-la em subconjuntos complementares e cada modelo é treinado com uma combinação diferente desses subconjuntos e validado com as restantes partes. A maior vantagem é que não há desperdício de dados para realização de validação



...

