

HOML - Chapter 1

O que é aprendizagem de máquina?

Aprendizagem de máquina é a ciência de programar de maneira que o computador aprenda através de dados.

Supervised learning

Supervised learning é quando temos dados com label e queremos prever o label de um novo dado com label desconhecido

Unsupervised learning

Unsupervised learning é quando tentamos detectar padrões, agrupamentos em dados que não contém label

Um exemplo dentro desta área é o anomaly detection, usado na detecção de fraude de cartão de crédito.

Association rule learning: identifica que um consumidor compra linguiça e ketchup junto, um padrão que antes poderia passar despercebido

Semisupervised learning

A mistura dos dois, alguns dados com label e outros sem.

Um exemplo é o google fotos, se tu identificar uma pessoa uma vez ele vai saber identificar essa pessoa em outras fotos.

Reinforcement learning

Aprende através de tentativa e erro; sistema de recompensa, punição e atualização de regras.

Batch/Online learning

Batch

Aprende de uma única vez, não é incremental. Se deseja aprimorar o sistema é necessário treinar ele do zero novamente

Online

Pode aprender aos poucos. Depois que utiliza os dados para aprender, esses dados são descartáveis, o tornando bastante útil para sistemas de pouco poder de armazenamento/processamento.

Instance-Based Versus Model-Based Learning

Instance

Classificação baseado em similaridades (KNN)

Model

A partir dos dados, um modelo é criado e esse modelo é usado para prever os novos dados

Questões

Questao 1

Aprendizagem de máquina é a ciência (ou arte) de programarmos de modo que nosso programa se aprimore com utilização de mais dados.

Uma outra definição seria a arte de desenvolver um programa sem programar ele de maneira explicita.

2. Can you name four types of problems where it shines?

Reconhecimento de voz, reconhecimento facial (ou reconhecimento de imagens, de uma maneira mais ampla), detecção de fraude (através de anomaly detection,

unsupervised learning) e processamento de linguagem natural, como por exemplo no reconhecimento de emoções em textos.

3. What is a labeled training set?

Um training set com label é um dataset que possui uma indicação de a que grupo aquele dado pertence, por exemplo, se é um cachorro, um gato, um papagaio.

4. What are the two most common supervised tasks?

Classificação e regressão

5. Can you name four common unsupervised tasks?

Anomaly detection, clustering, visualization and association rule learning

6. What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains?

Reinforcement learning

7. What type of algorithm would you use to segment your customers into multiple groups?

Clustering

8. Would you frame the problem of spam detection as a supervised learning problem or an unsupervised learning problem?

Unsupervised learning

9. What is an online learning system?

Um sistema de aprendizado online é aquele que permite a entrada incremental de dados no sistema de maneira a alterar a sua performance.

10. What is out-of-core learning?

É quando um sistema treina um modelo com dados que não cabem em sua memória. Os dados são carregados de maneira separada, uma parte de cada vez, pequena o suficiente para poder ser carregada na memória do computador.

11. What type of learning algorithm relies on a similarity measure to make predictions?

Instance based learning

12. What is the difference between a model parameter and a learning algorithm's hyperparameter?

Parametros são adquiridos através de dados, são internos ao modelo e são ajustados através de mais dados. Hiperparametros são externos ao modelo, não são adquiridos através de dados. Um exemplo de hiperparametro seria o K de KNN.

13. What do model-based learning algorithms search for? What is the most common strategy they use to succeed? How do they make predictions?

Algoritmos baseados em modelo buscam uma fórmula matemática que melhor solucione o problema que temos em mãos. A estratégia mais comum é a utilização de um grande conjunto de dados em conjunto com ajustes na fórmula de acordo com os dados utilizados.

14. Os principais desafios de aprendizagem de máquina são: dados insuficientes, dados de baixa qualidade, pouca representatividade e atributos irrelevantes

15. Ele está sofrendo de overfitting e uma possível solução seria a utilização de crossvalidation

16. Ao dividirmos em train e test, teremos uma maior chance de obtermos resultados similares em nossos testes e no cenário real, o test set nada mais é do

que uma parte dos nossos dados que é separado e não será utilizado nem no model training nem no hyperparameter tuning.

17. O propósito do validation set é otimizar o nosso modelo.

18.

19. Overfitting

What type of learning algorithm relies on a similarity measure to make predictions?

What type of learning algorithm relies on a similarity measure to make predictions?