
Aula 6 - Introdução ao Aprendizado de Máquina

AGENDA

- O que é aprendizado de máquina?
- Aprendizado Supervisionado
- Aprendizado Não Supervisionado
- Aprendizado Semi-supervisionado
- Classificação
- Regressão
- Agrupamento

O que é aprendizado de máquina?

- Segundo Arthur Samuel, 1959: “[Aprendizado de Máquina é o] campo de estudo que dá aos computadores a habilidade de aprender sem ser explicitamente programado.”
- Uma definição mais específica segundo Tom Mitchell, 1997: “Diz-se que um programa de computador aprende pela experiência E em relação a algum tipo de tarefa T e alguma medida de desempenho P se o seu desempenho em T , conforme medido por P , melhora com experiência E .”

O que é aprendizado de máquina?

- De maneira resumida o aprendizado de máquina se dá quando fornecemos a um dado algoritmo dados que irão permitir que o mesmo nos forneça uma resposta à um dado problema ou tome uma decisão.
- O filtro de spam ,por exemplo, consiste de um programa de AM, dado que são fornecidos e-mails **classificados** como spam de modo que quando o programa encontra um novo email ele consegue indicar se é ou não spam.
- Nessa abordagem não é necessário programar de forma explícita o que é um spam, o programa aprende com a experiência.

O que é aprendizado de máquina?

- Para o exemplo do filtro de spam, segundo a definição de Tom Mitchell:
 - Experiência E: Conjunto de email já classificados como spam
 - Tarefa T: Classificar um email como spam
 - Medida P: Acurácia, quantidade de acertos sobre o total de classificações

O que é aprendizado de máquina?

- O Aprendizado de máquina pode ser dividido em três subáreas:
 - Aprendizado Supervisionado
 - Aprendizado Semi Supervisionado
 - Aprendizado Não Supervisionado

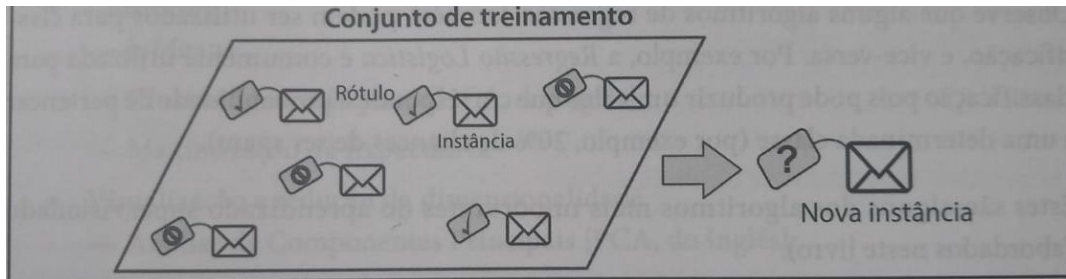
Aprendizado Supervisionado

Aprendizado Supervisionado

- No aprendizado supervisionado são fornecidos ao algoritmo dados para treinamento.
- Os dados de treino contém a resposta que se deseja obter, esse é chamado o atributo alvo.
- No exemplo do filtro de spam o conjunto de treino são os email já rotulados como sendo ou não spam. Esse rótulo é o atributo alvo.

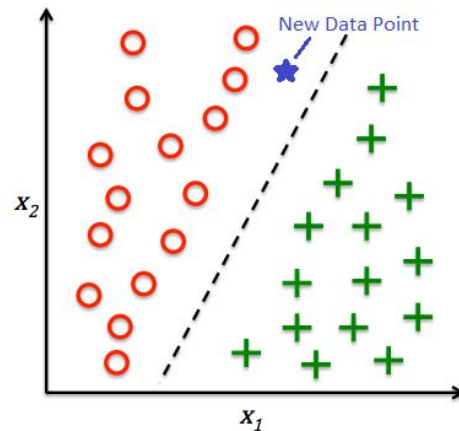
Aprendizado Supervisionado

- Dentro do conjunto de treinamento cada dado é chamado de instância e o valor do atributo alvo para aquela instância é conhecido.
- O atributo alvo pode ser numérico contínuo ou discreto:
 - Para o caso discreto tem-se um problema de Classificação
 - Para o caso contínuo tem-se um problema de Regressão



Classificação

- Quando se tem um conjunto de treino com o atributo alvo sendo um valor discreto tem-se um problema de **Classificação**.
- Com base nos atributos de cada instância o algoritmo aprende a prever a classe à qual ela pertence.
- Dessa forma quando um dado não rotulado é fornecido o conhecimento adquirido é utilizado para prever sua classe



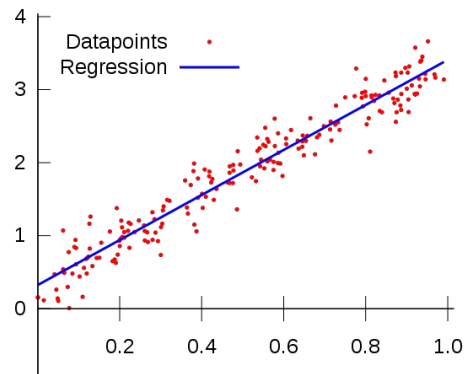
Fonte: [Supervised Learning: Basics of Classification and Main Algorithms | by Victor Roman | Towards Data Science](#)

Exemplos clássicos de Classificação

- Iris Dataset - Nesse caso com base no comprimento e a largura das sépalas e pétalas busca-se determinar a espécie da flor.
- Conjunto de imagens Dog vs Cat - de acordo com os atributos extraídos busca-se classificar o animal presente em uma foto como cachorro ou gato.
- Dataset MNIST - Conjunto de fotos de algarismos manuscritos onde busca-se determinar qual o número presente na imagem.

Regressão

- Quando se tem um conjunto de treino com o atributo alvo sendo um valor contínuo tem-se um problema de **Regressão**.
- Com base nos atributos de cada instância o algoritmo aprende a prever o valor do atributo alvo.
- Nesse caso o algoritmo busca reduzir o máximo possível o erro entre a previsão e o valor real.



Fonte: [Linear Regression — Intro To Machine Learning #6 | by David Fumo | Simple AI | Medium](#)

Exemplos clássicos de Regressão

- Houses Prices - Com base nas características de um imóvel busca-se prever seu valor de venda.
- Employees Salary - Com base nos atributos fornecidos busca-se estimar o valor do salário de um conjunto de trabalhadores.
- Fuel Consumption - Com base nas características técnicas busca-se estimar o consumo de combustível de um automóvel.

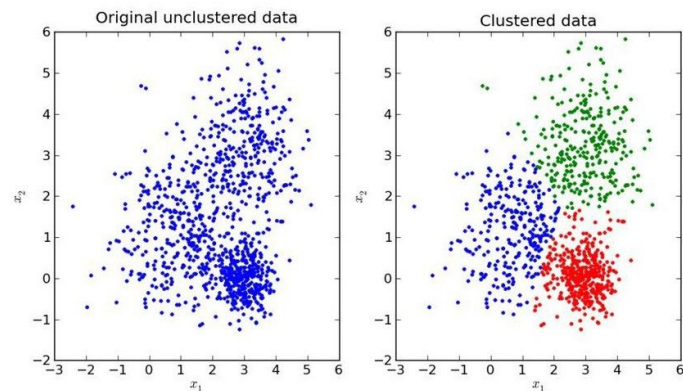
Aprendizado Não Supervisionado

Aprendizado Não Supervisionado

- Diferentemente do aprendizado supervisionado, agora o conjunto de treino não possui rótulo/atributo alvo.
- Em outras palavras não se conhece a “resposta”.
- Nessa abordagem o algoritmo busca aprender sem que haja uma experiência prévia.
- O exemplo mais famoso de aprendizado não supervisionado é o Agrupamento ou Clustering, como por exemplo o agrupamento dos visitantes de um site com base nos dados coletados.

Agrupamento

- No caso do agrupamento o objetivo é dividir as instâncias do conjunto de treino em grupos.
- Esses grupos não são previamente conhecidos. Eles são formados através do agrupamento de instâncias com características similares.
- Alguns exemplos de algoritmos: K-Means, Agrupamento hierárquico e Maximização de expectativa



Fonte: [Workflow of a Machine Learning project | by Ayush Pant | Towards Data Science](#)

Aprendizado Semi-supervisionado

- Nesta abordagem tem-se o uso de dados para treino parcialmente rotulados.
- Consiste basicamente em uma junção do aprendizado supervisionado e do não supervisionado.
- Um exemplo comum apresentado em [1], o reconhecimento de pessoas em fotos.
- O algoritmo não supervisionado consegue detectar a mesma pessoas em diferentes fotos.
- Em seguida basta que seja fornecido o nome da pessoa em uma foto para que as demais sejam rotuladas.

Referências bibliográficas

[1]Mãos À Obra: Aprendizado De Máquina Com Scikit-Learn & Tensorflow - Aurélien Géron

[2]Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina

GRATIDÃO!

