Aula 1 - O *Dataset*

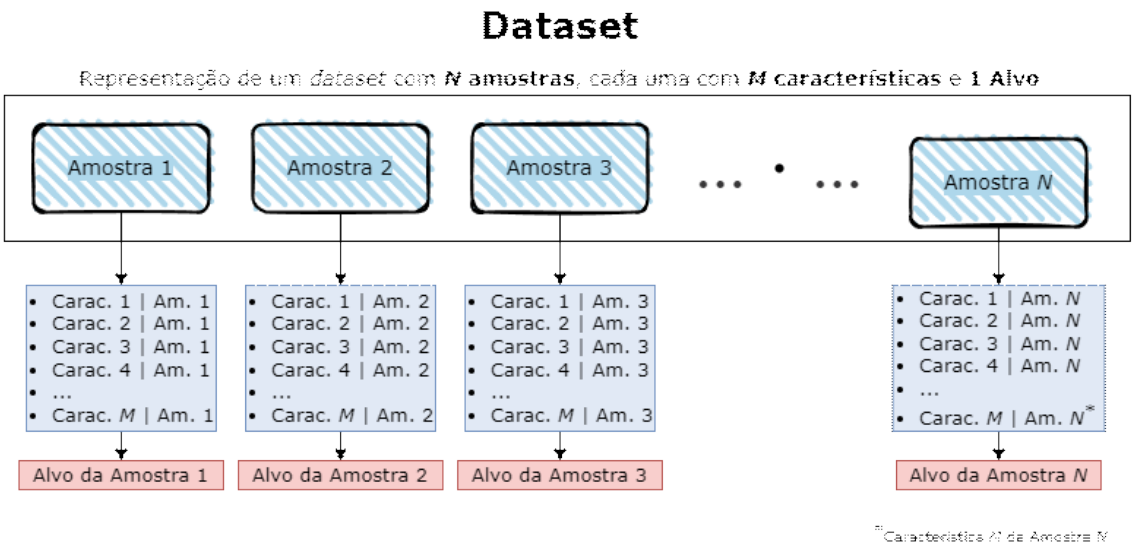
Saturday, November 21, 2020 7:52 PM

Referências
- Notas de aula do professor Andrew Ng
- Notebook (ipynb)

Tópicos Abordados
1. Como representar o conjunto de dados
2. Aprendizado Supervisionado e Não-Supervisionadp

Como representar o conjunto de dados

Estabeleceremos algumas notações matemáticas para os conjuntos de dados afim de melhorar a compreensão teórica de técnicas e algoritmos. Suponha um conjunto de dados genérico com *N* amostras onde cada amostra possui *M* características e uma variável alvo, como na figura abaixo:



Basicamente, a presença ou ausência da variável alvo no problema determinará se este problema é:
1. Supervisionado: presença de variável alvo.
2. Não-Supervisionado: ausência de variável alvo.

> *Mais adiante iremos entender o significado da variável alvo.*

Por enqunato iremos abordar aprendizado supervisionado. Deste modo, um conjunto de dados supervisionado (tabular, por hora) pode ser descrito da seguinte meneira:

	Caract. da amostra	Alvo
Amostra 1:	$X^{(1)}$	y_1
Amostra 1	$X^{(2)}$	y_2
...	...	...
Amostra *N*	$X^{(N)}$	y_N

Expandindo a tabela acima,

	Caract. 1	Caract. 2	...	Caract. *M*	Alvo
Amostra 1:	$x_{1,1}$	$x_{1,2}$	...	$x_{1,M}$	y_1
Amostra 1	$x_{2,1}$	$x_{2,2}$	...	$x_{2,M}$	y_2
...	...	...	...	...	...
Amostra *N*	$x_{N,1}$	$x_{N,2}$	...	$x_{N,M}$	y_N

- $X^{(i)}$: conjunto de *inputs* (**features** ou característica) da i-ésima amostra do conjunto de dados.
- $y^{(i)}$: *output* (**target** ou alvo) da i-ésima amostra do conjunto de dados.
- $\left(X^{(i)}, y^{(i)}\right)$: exemplo de treinamento ou amostras de treinamento (**training example** ou **training sample**), composto por variáveis aleatórias de *input* e *output*.
- $\left\{\left(X^{(i)}, y^{(i)}\right) \forall\, i \in [1, n]\right\}$: conjunto de dados (**dataset**) composto pelas amostras de treinamento.

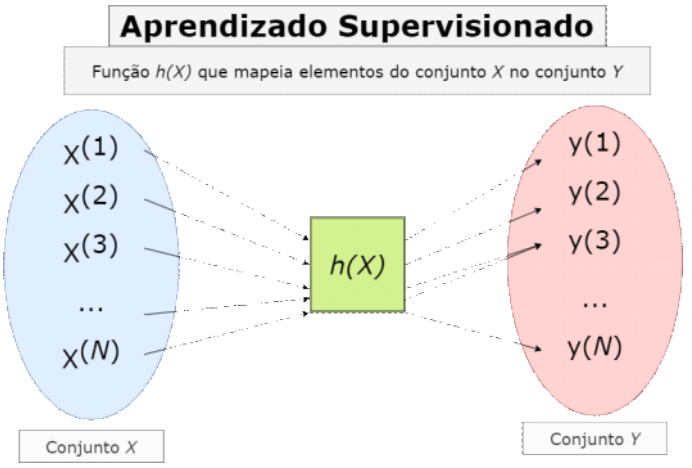
Aula 1 - Aprendizado Supervisionado

Wednesday, January 18, 2023 5:19 PM

Aprendizado Supervisionado

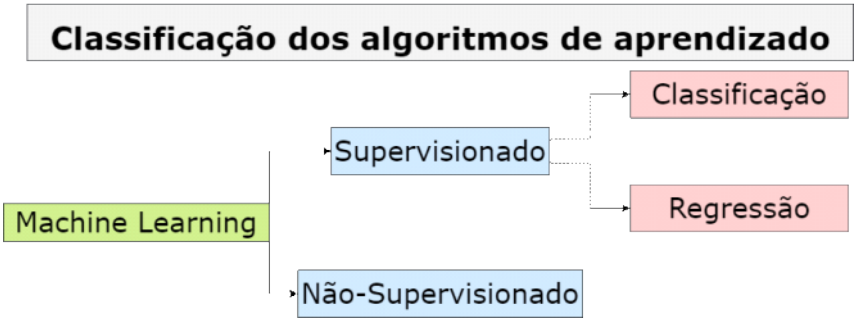
Agora que já vimos como definir o dataset, vamos formalizar o que é aprendizado supervisionado:

Queremos encontrar uma função $h : X \rightarrow Y$ *que mapeia o espaço das variáveis de input no espaço das variáveis de output.*



A função *h(X)* é chamada de hipótese, além disso o *aprendizado supervisionado* é subdividido em:

1. Classificação: quando a variável alvo é qualitativa
2. Regressão : quando a variável alvo é quantitativas contínua



Alguns exemplos

Com base na descrição seguir, é possível identificar qual o escopo de *Machine Learning* do problema?

1. *"Você trabalha em uma empresa de vendas e deseja agrupar todos os clientes com o mesmo perfil de compra."*
2. *"Você é um pesquisador que deseja identificar possíveis pessoas com uma doença específica e para isso você conta com um banco de dados com diversas características de pacientes com e sem a doença."*
3. *"Você está num projeto cuja demanda é tentar prever o valor de ação de mercado de uma companhia com base nos valores passados"*