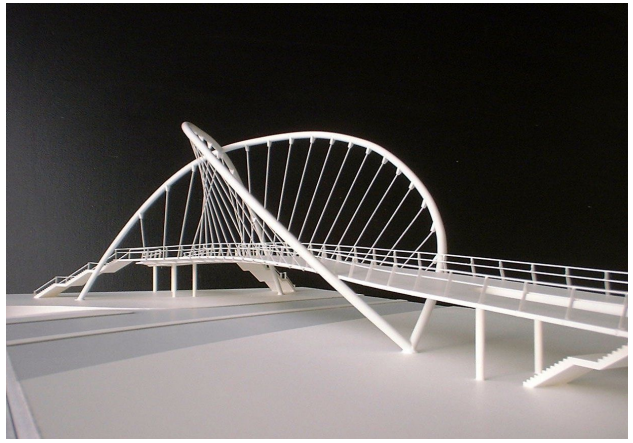
Preditiva.ai

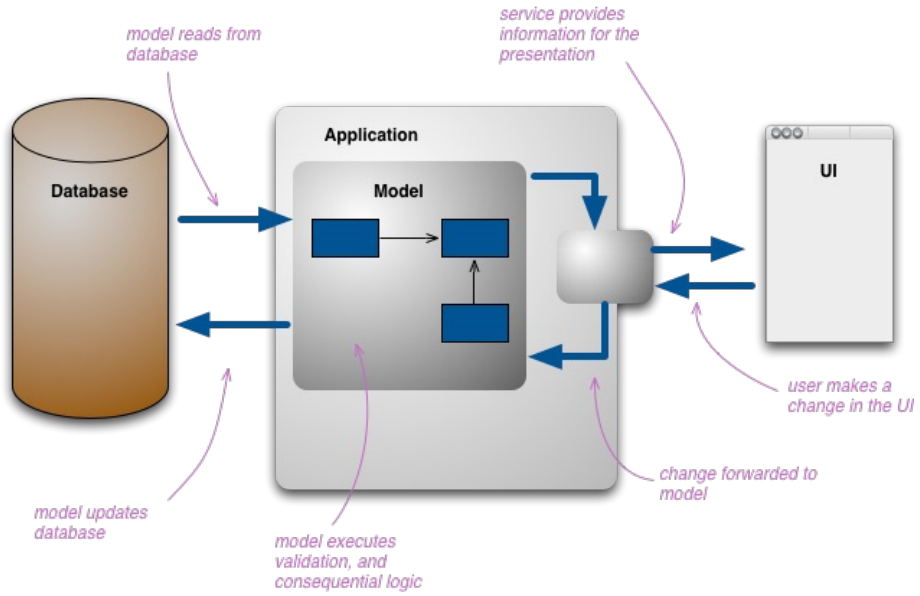
Introdução à Data Science
O que são modelos?

O que são modelos?

Modelos estão presentes em todos os lugares...

Preditiva.ai





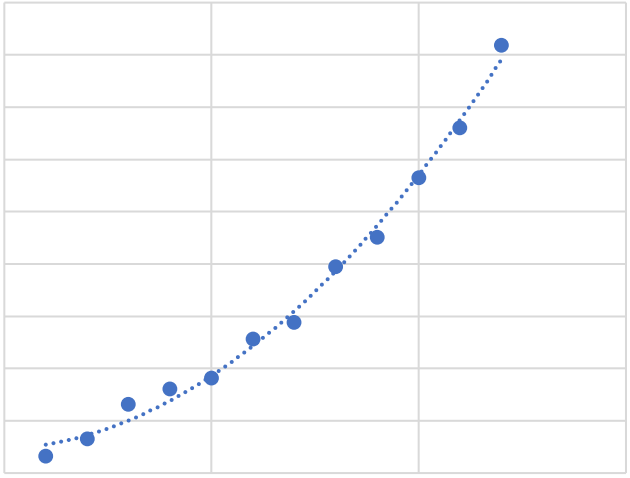
O que são modelos?

Mas existem diferentes tipos de modelos

Preditiva.ai

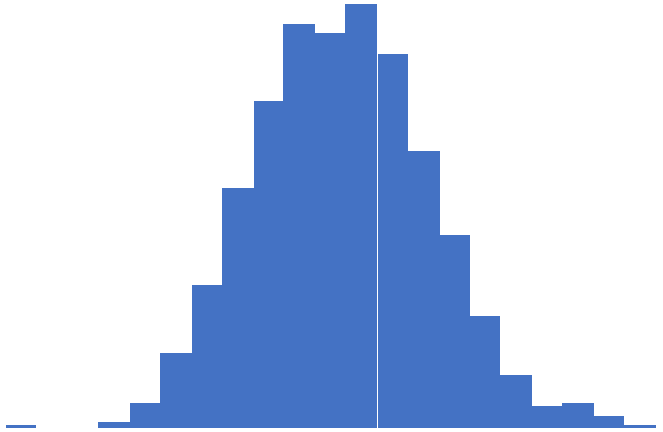
Modelos Matemáticos

Representação ou interpretação simplificada de um fenômeno considerando **conceitos matemáticos**.



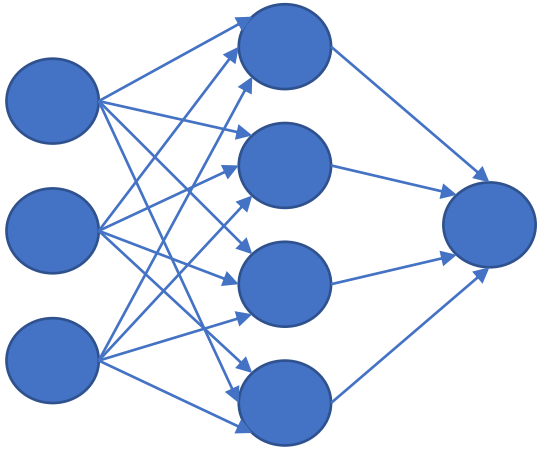
Modelos Estatísticos

Representação ou interpretação simplificada de um fenômeno considerando a **distribuição dos dados de origem**.



Modelos *Machine Learning*

Representação ou interpretação simplificada de um fenômeno considerando **exemplos de dados para seu treinamento**.



Modelos de Dados

Representação ou interpretação simplificada de um fenômeno.

https://plato.stanford.edu/entries/models-science

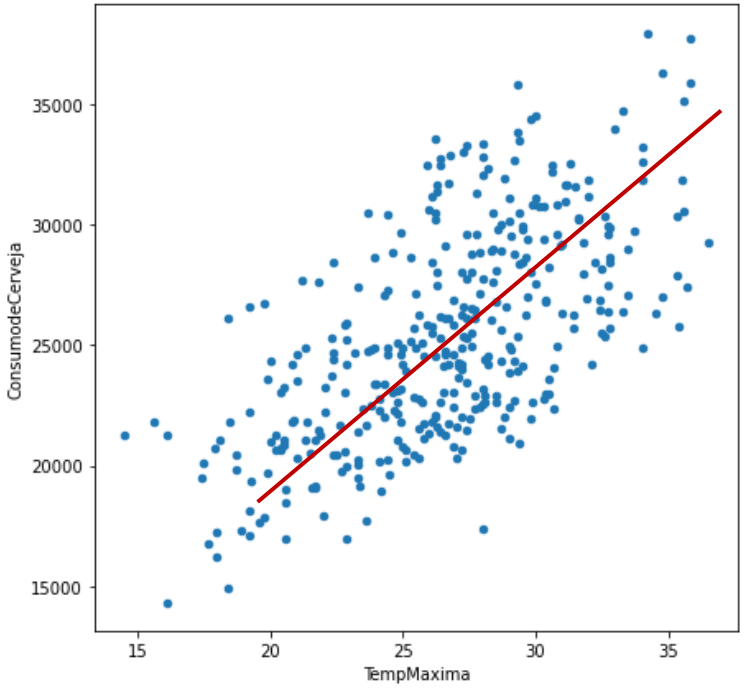
O que são modelos?

Exemplo: Modelo de Regressão

Preditiva.ai

Modelos de Regressão são utilizados quando a variável resposta é quantitativa, e busca-se identificar quais fatores a influenciam.

Neste exemplo temos um estudo que visa **estimar o consumo de cerveja** (em litros) a partir da **temperatura máxima** registrada no dia.



Ajustando um modelo de **regressão linear simples**, obtemos os seguintes **coeficientes / parâmetros**:

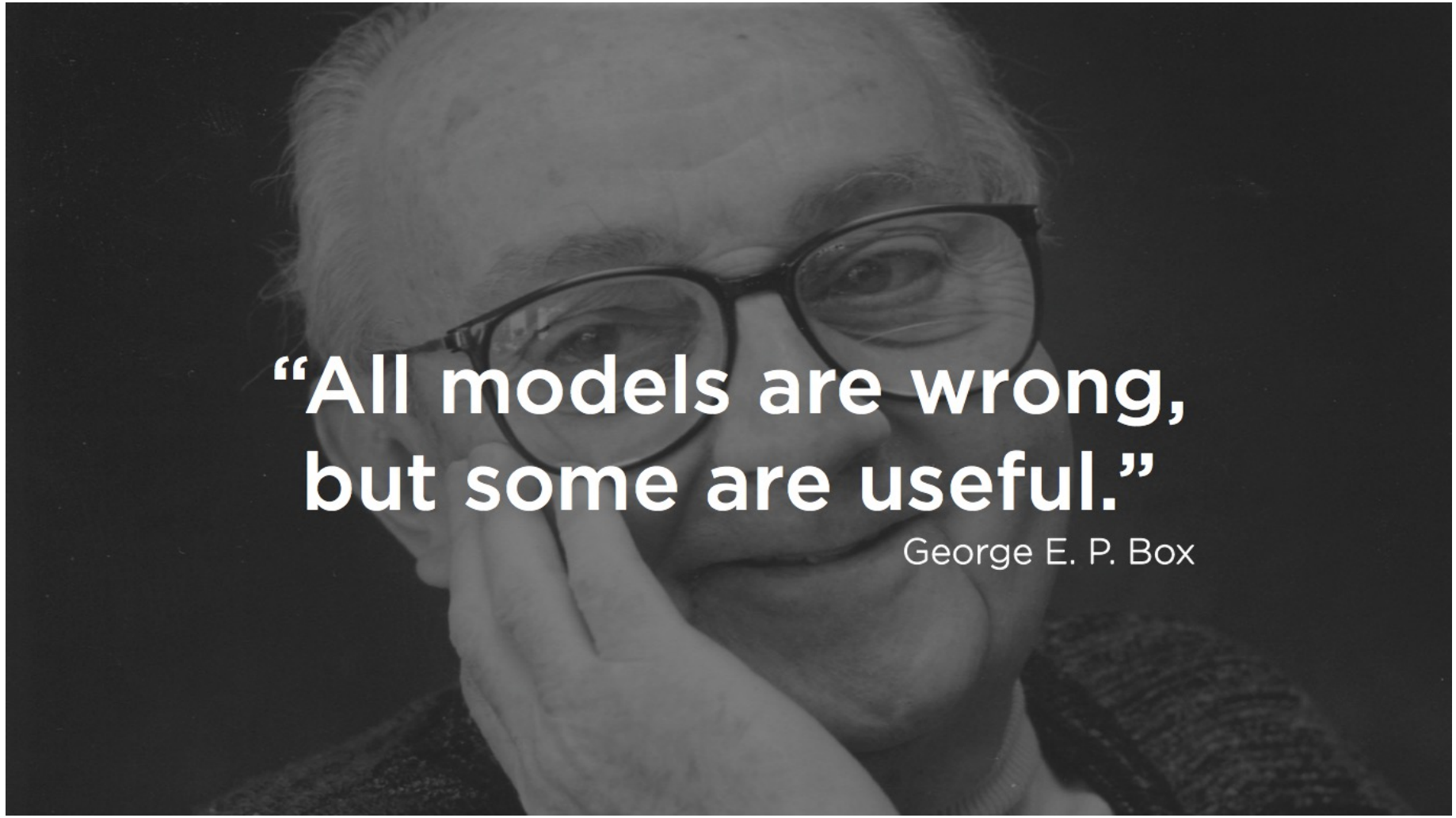
$$\text{Consumo Cerveja} = 7975 + 655 \cdot \text{Temperatura Máxima}$$

Ou seja, a cada **incremento de 1 grau Celsius** o consumo médio de cerveja **aumenta 655 litros**.

Para que servem os modelos?

Todos os modelos estão errados, mas alguns são úteis.

Preditiva.ai



Preditiva.ai

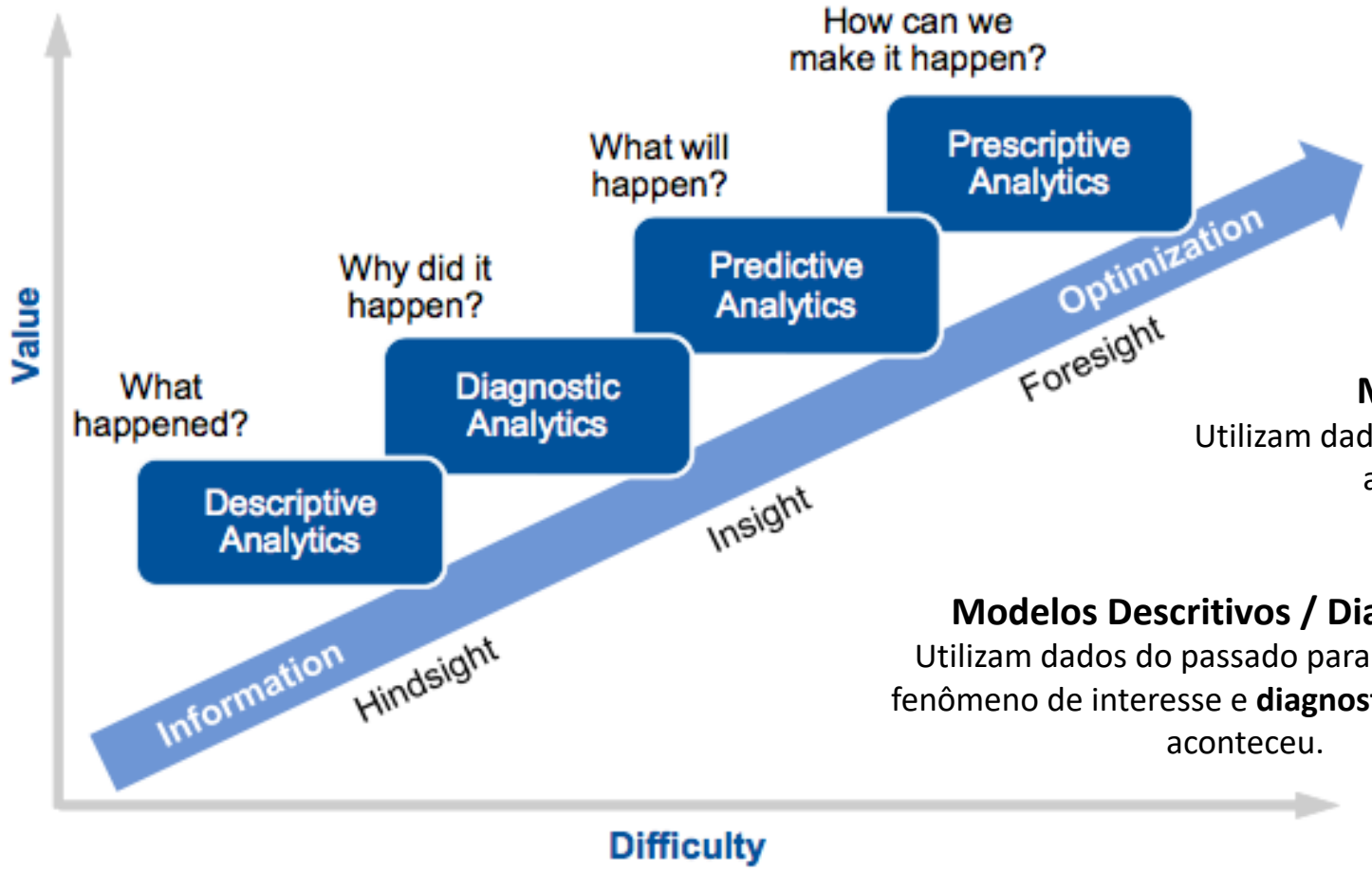
Introdução à Data Science

Para que servem os modelos?

Para que servem os modelos?

Existem modelos específicos para cada finalidade

Preditiva.ai



Modelos Prescritivos

Utilizam dados do passado e presente em conjunto com outras informações de negócios para **sugerir a melhor decisão**.

Modelos Preditivos

Utilizam dados do passado para **prever** o que acontecerá no futuro.

Modelos Descritivos / Diagnósticos

Utilizam dados do passado para **descrever** um fenômeno de interesse e **diagnosticar** porque ele aconteceu.

Source: Gartner (March 2012)

Para que servem os modelos?

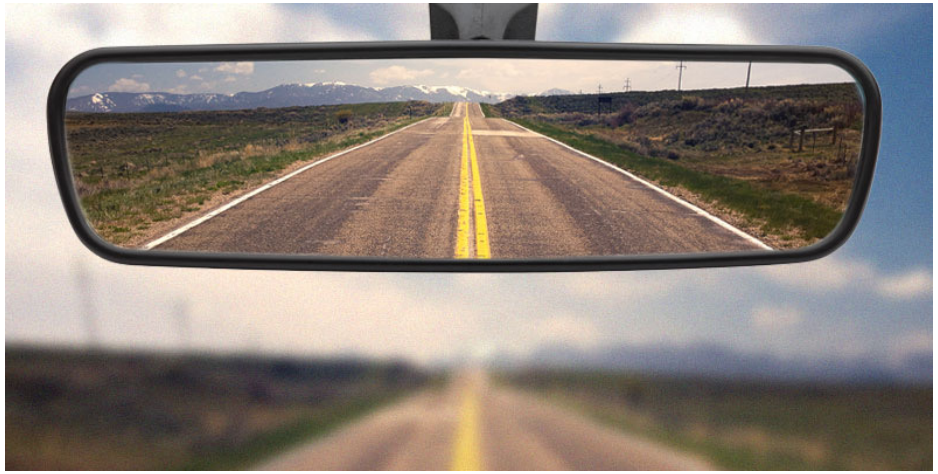
Modelos Descritivos

Preditiva.ai

Modelos Descritivos são utilizados para **compreender** como os diversos **fatores** (variáveis explicativas) **influenciam** o comportamento de um **fenômeno** (variável resposta) e são baseados exclusivamente em **dados do passado**. Por esse motivo são utilizados para responder perguntas do tipo: "**O que aconteceu?**".

Alguns exemplos de aplicações de **Modelos Descritivos**:

- Quais fatores **estiveram** mais presentes no turnover dos funcionários?
- Quais perfis de clientes mais **compraram** no ano passado?
- Quais regiões **foram** mais favoráveis aos clientes em ações cíveis revisionais?
- Quais combinações de produtos e lojas **tiveram** maior rentabilidade?



Fonte: https://www.rottenbytes.info/post/adventures-in-hindsight/

Perceba que todos os verbos acima estão no **passado**. Na prática é como se estivéssemos olhando para o **retrovisor** de um carro e vendo tudo o que **já aconteceu**.

Para que servem os modelos?

Modelos Preditivos

Preditiva.ai

Modelos Preditivos são utilizados para **prever** o que acontecerá com o **fenômeno** em estudo (variável resposta) se os **fatores** exercerem determinada **influência** (variáveis explicativas). São utilizados para responder perguntas do tipo: "**O que acontecerá?**".

Alguns exemplos de aplicações de **Modelos Preditivos**:

- Quais fatores **influenciam** o turnover dos funcionários?
- Quais perfis de clientes **possuem** maior volume de compras?
- Quais regiões **favorecem** os clientes em ações cíveis revisionais?
- Quais combinações de produtos e lojas **geram** maior rentabilidade?

Neste caso todos os verbos acima estão no **presente**. Ou seja, a partir de um modelo descritivo, analisa-se as relações de influência dos fatores no fenômeno em estudo, buscando associações que possam ajudar a prever **resultados futuros**.

Fonte: https://www.techrepublic.com/article/analytics-prediction-confidence-its-all-about-semantics/

Para que servem os modelos?

Modelos Prescritivos

Preditiva.ai

Modelos Prescritivos são um dos mais avançados níveis de utilização de dados para tomada de decisão. Além dos **dados históricos** e das **análises de associação** para previsão, os **Modelos Prescritivos** utilizam dados do presente para responder a pergunta: "**Como faço isso acontecer?**".

Alguns exemplos de aplicações de **Modelos Prescritivos**:

- Como **reduziremos** o turnover dos funcionários?
- Como **estimularemos** os diferentes perfis de clientes para aumentar o volume de compras?
- Como **aumentaremos** o êxito em ações cíveis revisionais?
- Como **escolheremos** a melhor combinação de produtos e lojas para maior rentabilidade?

Agora todos os verbos acima estão no **futuro**. Após compreender o fenômeno e estudar as associações entre ele e os fatores de influência, chega o momento de obter recomendações para os possíveis resultados desejados.

Fonte: https://www.lucidchart.com/blog/webinar-critical-elements-for-better-decision-making

Preditiva.ai

Introdução à Data Science
Categorias de Modelos

Categorias de Modelos

Método Supervisionado

Preditiva.ai

No **método supervisionado** temos sempre as **variáveis explicativas** (*features*) e a **variável resposta** (*target*).

Por esse motivo, os tipos de problemas mais comuns nesse método são:

- **Classificação**: variável resposta qualitativa.
- **Regressão**: variável resposta quantitativa.

Nas duas situações dizemos que os **dados estão rotulados**, pois além das características de cada observação, temos o que normalmente é o objetivo do modelo: a **categoria** em problemas de **classificação** e os **valores** em problemas de **regressão**.

Iris Versicolor **Iris Setosa** **Iris Virginica**

Fonte: http://www.lac.inpe.br/~rafael.santos/Docs/CAP394/WholeStory-Iris.html

Largura Pétala	Comprimento Pétala	Largura Sépala	Comprimento Sépala	Espécie
0,2	1,4	3,5	5,1	Setosa
1,6	4,5	3,0	5,4	Versicolor
1,8	6,0	3,2	7,2	Virgínica
...	...	...	...	...

Target

Categorias de Modelos

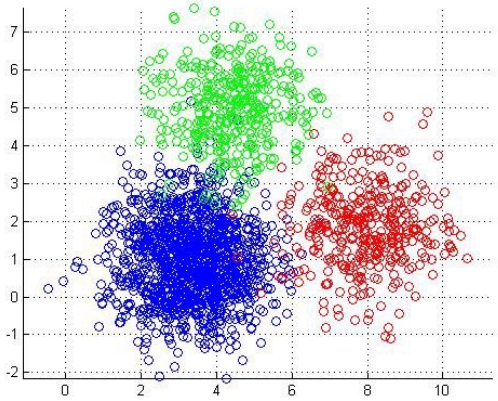
Método Não Supervisionado

Preditiva.ai

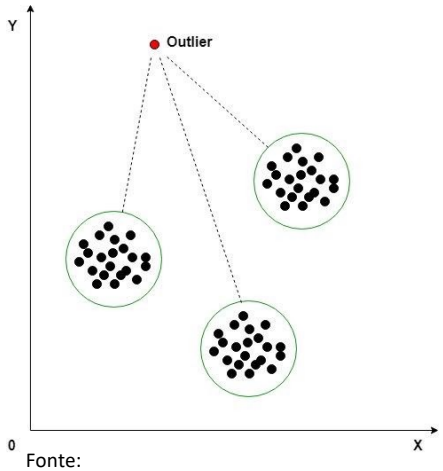
No **método não supervisionado** é muito utilizado quando **não temos** os **dados rotulados**, ou seja, temos apenas as **variáveis explicativas** (*features*).

Dessa forma, os tipos de problemas para os métodos **não supervisionados** são:

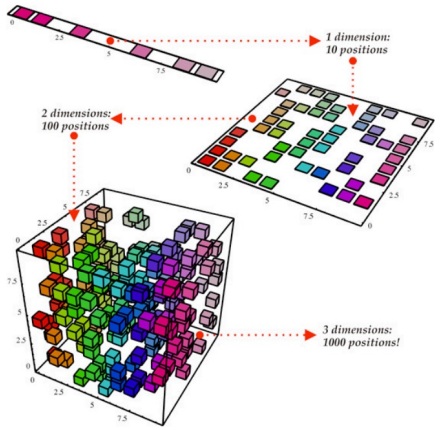
- **Clusterização**: agrupamento de observações com variáveis explicativas semelhantes.
- **Detecção de anomalias**: análise de outliers.
- **Redução de dimensionalidade**: transformação dos dados para espaço com menor dimensão.



Fonte: https://www.kdnuggets.com/2019/09/hierarchical-clustering.html



Fonte: https://medium.com/datadriveninvestor/how-machine-learning-can-enable-anomaly-detection-eed9286c5306



Fonte: https://medium.com/@snehasathishdeva/introduction-to-dimensionality-reduction-2970a7d2a918

Categorias de Modelos
Método Semi-Supervisionado

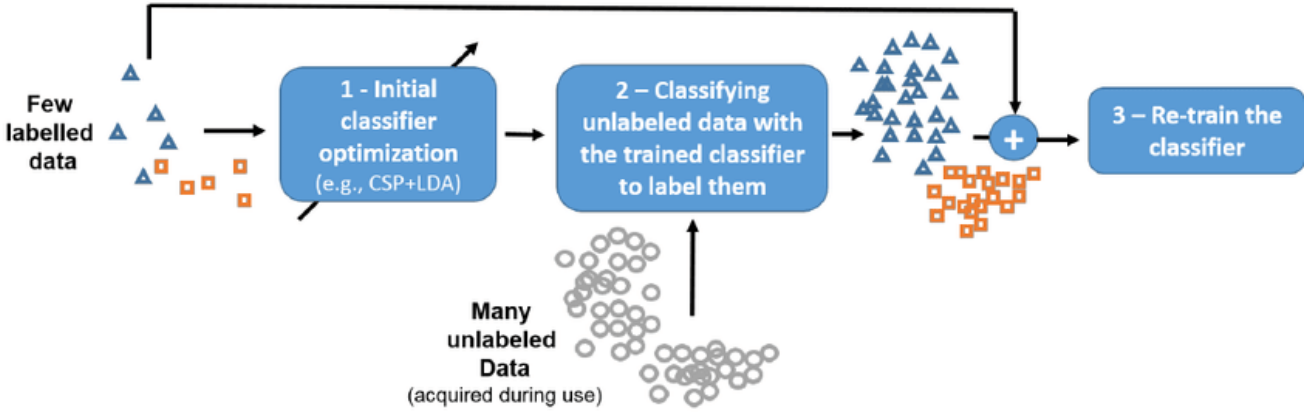
Preditiva.ai

Rotular os dados normalmente é uma **tarefa muito custosa** em termos de tempo e dinheiro. Por esse motivo, surgiu o **método semi-supervisionado**.

No **método semi-supervisionado** os problemas também são de **classificação** e **regressão**, porém são utilizados **dados rotulados** em pequena quantidade e uma grande quantidade de **dados não rotulados**.

Existem algumas estratégias para utilização do **método semi-supervisionado**.

Uma delas consiste em **treinar** um estimador com os **dados rotulados** e aplicar esse estimador nos **dados não rotulados**, e em seguida **re-treinar** o estimador.



Fonte: https://www.researchgate.net/figure/Principle-of-semi-supervised-learning-1-a-model-eg-CSP-LDA-classifier-is-first_fig4_277605013

Preditiva.ai

Introdução à Data Science

Modelos Estatísticos vs. *Machine Learning*

Processo de Ajuste vs. Aprendizado de Máquina

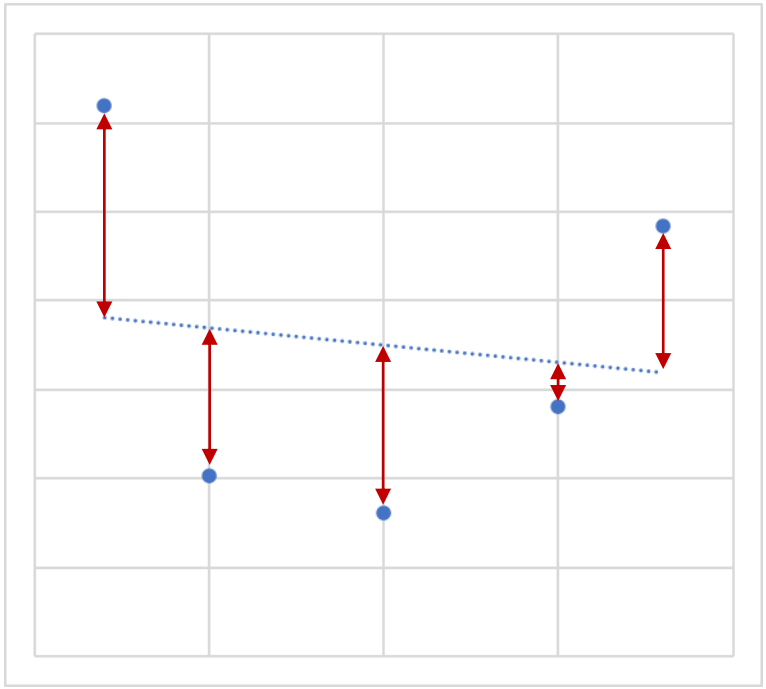
Diferentes estratégicas para transformar dados em informação

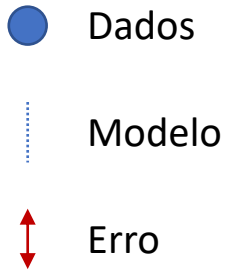
Preditiva.ai

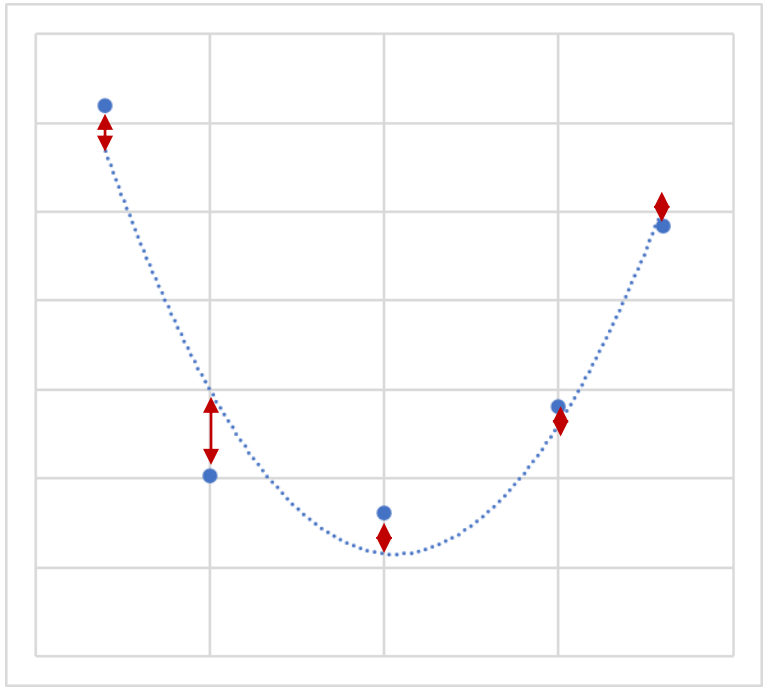
A terminologia depende do tipo de **Modelos de Dados** considerado:

- Modelos Matemáticos / Estatísticos: **Ajuste**
- Modelos *Machine Learning*: **Aprendizado ou Treinamento**

O **ajuste** e o **aprendizado** têm o **mesmo objetivo**: **minimizar** uma **função de erro** pré-definida. O que muda é a estratégia para se obter os menores erros.







Processo de Ajuste vs. Aprendizado de Máquina

Diferentes estratégicas para transformar dados e informação

Preditiva.ai

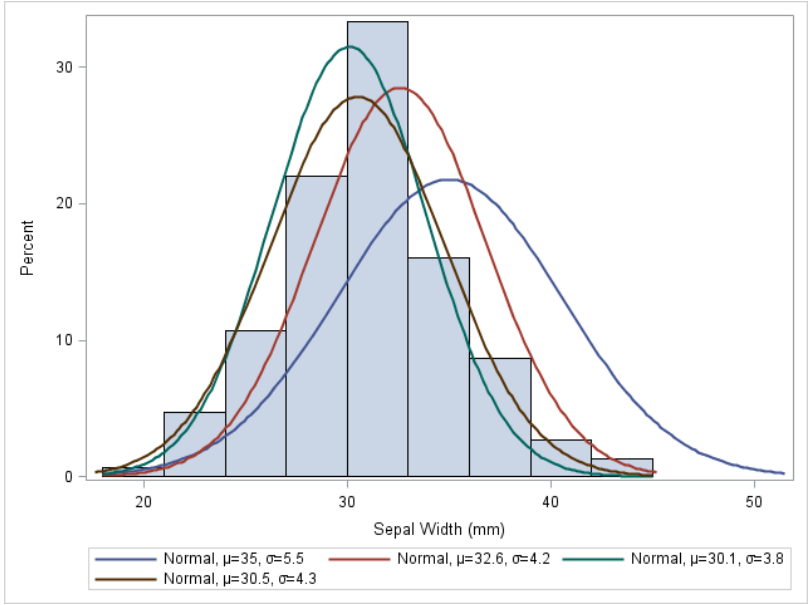
Em **Modelos Estatísticos**, um método bastante comum para encontrar os parâmetros que minimizam o erro é o **Método da Máxima Verossimilhança**.

Esse método consiste em 2 principais etapas:

1. Definir a **função de verossimilhança**: depende da **distribuição de probabilidades** utilizada.
2. Encontrar os valores dos **parâmetros** que **maximizam** a **função de verossimilhança**: quando essa função for diferenciável é possível obter os parâmetros de forma explícita. Caso contrário, é necessário utilizar métodos numéricos.

Por considerar a distribuição de **probabilidades dos dados**, os **Modelos Estatísticos** possuem premissas que devem ser respeitadas.

Em geral, nas bibliotecas e plataformas de modelagem estatística, já estão incluídos alguns testes de hipóteses para avaliar se os **valores encontrados** são **estatisticamente significantes**.



Fonte: https://blogs.sas.com/content/iml/2011/10/12/maximum-likelihood-estimation-in-sasiml.html

Processo de Ajuste vs. Aprendizado de Máquina

Diferentes estratégicas para transformar dados e informação

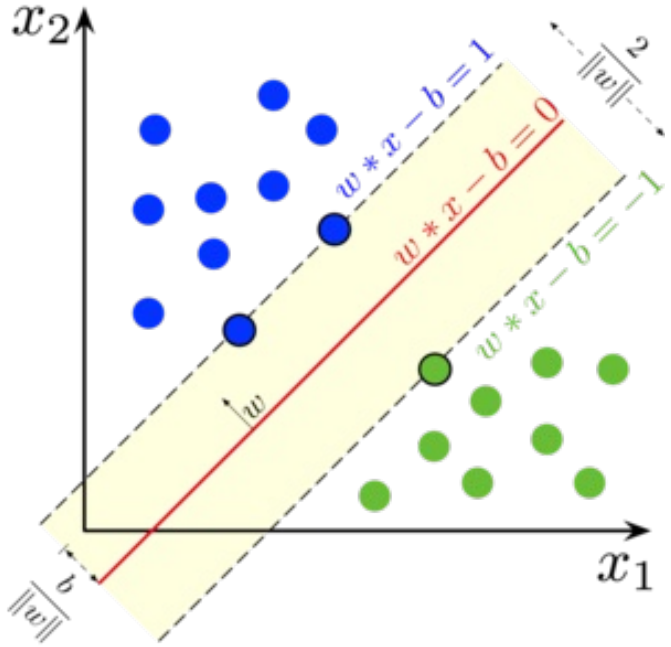
Preditiva.ai

Em **Modelos *Machine Learning*** não existe o conceito de uma distribuição de probabilidades associada aos dados. Por esse motivo, nesses modelos não existem as premissas dos **Modelos Estatísticos**.

Com a ausência dessas premissas, a **estrutura do modelo é mais flexível** e o custo dessa flexibilidade é a necessidade de um **maior volume de dados** para que os modelos sejam treinados adequadamente.

Devido a diversidade de técnicas utilizadas em **Modelos *Machine Learning***, também existem diversos métodos de treinamento desses modelos:

1. **Árvores de Decisão**: Ganho de informação e Entropia.
2. **Redes Neurais Artificiais**: Gradiente Descendente e outros.
3. **Suport Vector Machines**: Construção de hiperplanos.



Fonte: https://en.wikipedia.org/wiki/Support_vector_machine

Processo de Ajuste vs. Aprendizado de Máquina

Diferentes estratégicas para transformar dados e informação

Preditiva.ai

Principais diferenças entre ajuste e/ou aprendizado em **Modelos de Dados**:

Características	Modelo Estatístico	Modelo *Machine Learning*
Suposições sobre a distribuição de probabilidade dos dados	Sim	Não
Quantidade de dados necessários	Média	Alta
Processo de ajuste / aprendizado	Explícito ou Numérico Iterativo	Numérico Iterativo
Recursos computacionais necessários para ajuste / aprendizado	Médio	Alto
Tempo do Cientista de Dados no desenvolvimento do modelo	Alto	Médio

Preditiva.ai