

Mineração de Dados

Aula 1: Introdução

Rafael Izbicki

Por que você está aqui?

- ▶ Porque você gosta do assunto
- ▶ Porque é um requerimento
- ▶ Porque Mineração de Dados é A matéria que vai te fazer ganhar \$\$\$
- ▶ ...

Por que você está aqui?

- ▶ Porque você gosta do assunto
- ▶ Porque é um requerimento
- ▶ Porque Mineração de Dados é A matéria que vai te fazer ganhar \$\$\$
- ▶ ...

Por que você está aqui?

- ▶ Porque você gosta do assunto
- ▶ Porque é um requerimento
- ▶ Porque Mineração de Dados é A matéria que vai te fazer ganhar \$\$\$
- ▶ ...

Por que você está aqui?

- ▶ Porque você gosta do assunto
- ▶ Porque é um requerimento
- ▶ Porque Mineração de Dados é A matéria que vai te fazer ganhar \$\$\$
- ▶ ...



Por que você está aqui?

- ▶ Porque você gosta do assunto
- ▶ Porque é um requerimento
- ▶ Porque Mineração de Dados é A matéria que vai te fazer ganhar \$\$\$
- ▶ ...



Avaliação do Conhecimento Prévio

Estrutura do Curso

- ▶ Lembre-se de verificar o site da disciplina!
- ▶ 48 horas para responder os mensagens!

Estrutura do Curso

- ▶ Lembre-se de verificar o site da disciplina!
- ▶ 48 horas para responder os mensagens!

Estrutura do Curso: Material

- ▶ Slides em Aula

- ▶ Livro texto:

Hastie T., Tibshirani R. And Friedman J. *The Elements of Statistical Learning*, Springer, 2009.

James, G., Witten, D., Hastie, T. e Tibshirani, R. *An Introduction to Statistical Learning, with Applications in R*, Springer 2013.

Disponível gratuitamente em

<http://statweb.stanford.edu/~tibs/ElemStatLearn/>

<http://www-bcf.usc.edu/~gareth/ISL/>

Estrutura do Curso: Material

- ▶ Slides em Aula
- ▶ Livro texto:

Hastie T., Tibshirani R. And Friedman J. *The Elements of Statistical Learning*, Springer, 2009.

James, G., Witten, D., Hastie, T. e Tibshirani, R. *An Introduction to Statistical Learning, with Applications in R*, Springer 2013.

Disponível gratuitamente em

<http://statweb.stanford.edu/~tibs/ElemStatLearn/>

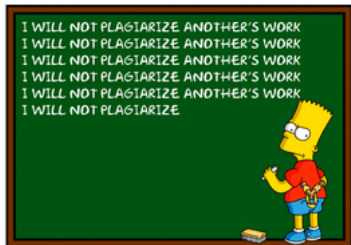
<http://www-bcf.usc.edu/~gareth/ISL/>

Estrutura do Curso: Notas

⇒ Listas de Exercício (15% da nota): não deixe para a última hora! Não copie!

Para cada dia de atraso, pontos serão descontados

Duplas sorteadas dentro e fora da sala



Estrutura do Curso: Notas

⇒ Listas de Exercício (15% da nota): não deixe para a última hora! Não copie!

Para cada dia de atraso, pontos serão descontados

Duplas sorteadas dentro e fora da sala



Estrutura do Curso: Notas

⇒ Prova 1 (25% da nota)

⇒ Prova 2 (25% da nota)

⇒ Seminário (15% da nota)

⇒ Trabalho Prático (20% da nota)

Mais detalhes em breve

Estrutura do Curso: Notas

⇒ Prova 1 (25% da nota)

⇒ Prova 2 (25% da nota)

⇒ Seminário (15% da nota)

⇒ Trabalho Prático (20% da nota)

Mais detalhes em breve

Estrutura do Curso: Notas

⇒ Prova 1 (25% da nota)

⇒ Prova 2 (25% da nota)

⇒ Seminário (15% da nota)

⇒ Trabalho Prático (20% da nota)

Mais detalhes em breve

Estrutura do Curso: Notas

⇒ Prova 1 (25% da nota)

⇒ Prova 2 (25% da nota)

⇒ Seminário (15% da nota)

⇒ Trabalho Prático (20% da nota)

Mais detalhes em breve

Estrutura do Curso: Notas

⇒ Prova 1 (25% da nota)

⇒ Prova 2 (25% da nota)

⇒ Seminário (15% da nota)

⇒ Trabalho Prático (20% da nota)

Mais detalhes em breve

O que é o Aprendizado de Máquina?

Também conhecida como:

- ▶ Data Mining (Mineração de Dados)
- ▶ Big Data
- ▶ Inteligência Artificial

Aprendizado de Máquina é a ciência de **descobrir estruturas** e **fazer previsões** em conjuntos de dados (grandes).

- ▶ Aprendizado **supervisionado**: Dadas medições $(X_1, Y_1), \dots, (X_n, Y_n)$, aprender um modelo para **prever** Y_i baseado em X_i
- ▶ Aprendizado **não supervisionado**: Dadas medições X_1, \dots, X_n , descobrir alguma **estrutura** baseado em similaridade

O que é o Aprendizado de Máquina?

Também conhecida como:

- ▶ Data Mining (Mineração de Dados)
- ▶ Big Data
- ▶ Inteligência Artificial

Aprendizado de Máquina é a ciência de **descobrir estruturas** e **fazer previsões** em conjuntos de dados (grandes).

- ▶ Aprendizado **supervisionado**: Dadas medições $(X_1, Y_1), \dots, (X_n, Y_n)$, aprender um modelo para **prever** Y_i baseado em X_i
- ▶ Aprendizado **não supervisionado**: Dadas medições X_1, \dots, X_n , descobrir alguma **estrutura** baseado em similaridade

O que é o Aprendizado de Máquina?

Também conhecida como:

- ▶ Data Mining (Mineração de Dados)
- ▶ Big Data
- ▶ Inteligência Artificial

Aprendizado de Máquina é a ciência de **descobrir estruturas** e **fazer previsões** em conjuntos de dados (grandes).

- ▶ Aprendizado **supervisionado**: Dadas medições $(X_1, Y_1), \dots, (X_n, Y_n)$, aprender um modelo para **prever** Y_i baseado em X_i
- ▶ Aprendizado **não supervisionado**: Dadas medições X_1, \dots, X_n , descobrir alguma **estrutura** baseado em similaridade

O que é o Aprendizado de Máquina?

Também conhecida como:

- ▶ Data Mining (Mineração de Dados)
- ▶ Big Data
- ▶ Inteligência Artificial

Aprendizado de Máquina é a ciência de **descobrir estruturas** e **fazer previsões** em conjuntos de dados (grandes).

- ▶ Aprendizado **supervisionado**: Dadas medições $(X_1, Y_1), \dots, (X_n, Y_n)$, aprender um modelo para **prever** Y_i baseado em X_i
- ▶ Aprendizado **não supervisionado**: Dadas medições X_1, \dots, X_n , descobrir alguma **estrutura** baseado em similaridade

O que é o Aprendizado de Máquina?

Também conhecida como:

- ▶ Data Mining (Mineração de Dados)
- ▶ Big Data
- ▶ Inteligência Artificial

Aprendizado de Máquina é a ciência de **descobrir estruturas** e **fazer previsões** em conjuntos de dados (grandes).

- ▶ Aprendizado **supervisionado**: Dadas medições $(X_1, Y_1), \dots, (X_n, Y_n)$, aprender um modelo para **prever** Y_i baseado em X_i
- ▶ Aprendizado **não supervisionado**: Dadas medições X_1, \dots, X_n , descobrir alguma **estrutura** baseado em similaridade

O que é o Aprendizado de Máquina?

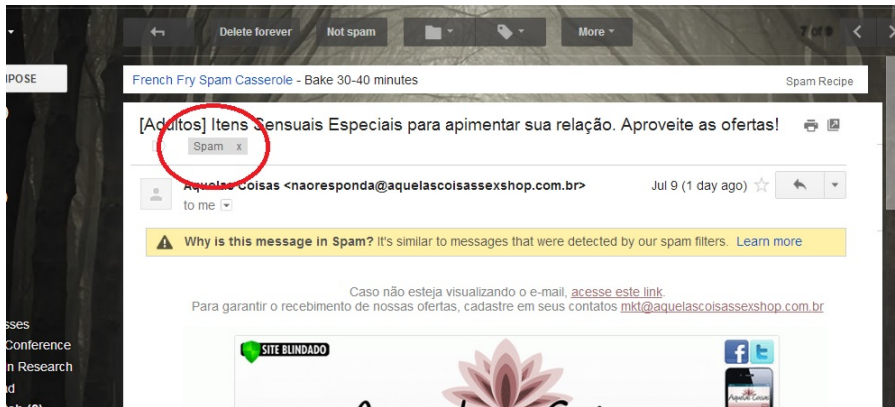
Também conhecida como:

- ▶ Data Mining (Mineração de Dados)
- ▶ Big Data
- ▶ Inteligência Artificial

Aprendizado de Máquina é a ciência de **descobrir estruturas** e **fazer previsões** em conjuntos de dados (grandes).

- ▶ Aprendizado **supervisionado**: Dadas medições $(X_1, Y_1), \dots, (X_n, Y_n)$, aprender um modelo para **prever** Y_i baseado em X_i
- ▶ Aprendizado **não supervisionado**: Dadas medições X_1, \dots, X_n , descobrir alguma **estrutura** baseado em similaridade

Exemplo: Detecção de Spams



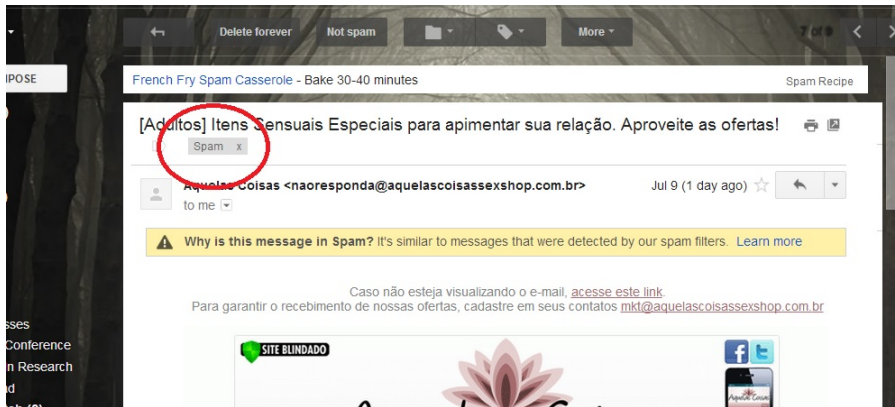
X_i → email

Y_i → spam/não spam

Objetivo: prever Y_i com base em X_i

Conjunto de dados grande!

Exemplo: Detecção de Spams



$X_i \rightarrow$ email

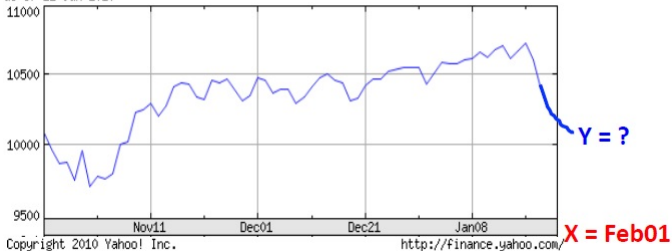
$Y_i \rightarrow$ spam/não spam

Objetivo: prever Y_i com base em X_i

Conjunto de dados grande!

Exemplo: Predição da Bolsa

DJ INDU AVERAGE (DOW JONES & CO
as of 22-Jan-2010



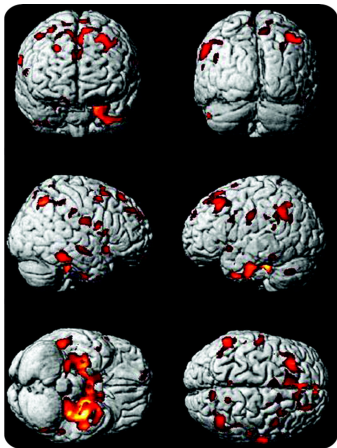
Exemplo: Reconhecimento de Dígitos

7 2 1 0 4 1 4 9 5 9
0 6 9 0 1 5 9 7 3 4
9 6 6 5 4 0 7 4 0 1
3 1 3 4 7 2 7 1 2 1
1 7 4 2 3 5 1 2 4 4

X_i → imagem de um dígito

Y_i → dígito correspondente

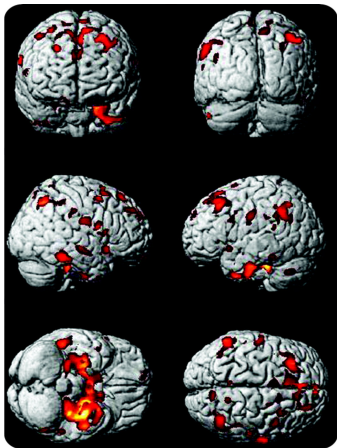
Exemplo: Predição de Alzheimer



X_i → imagem da ressonância magnética

Y_i → Paciente com/sem Alzheimer

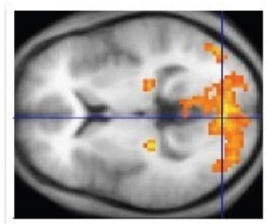
Exemplo: Predição de Alzheimer



X_i → imagem da ressonância magnética

Y_i → Paciente com/sem Alzheimer

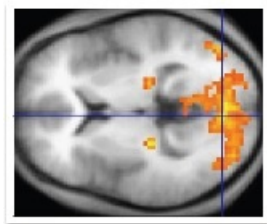
Exemplo: Leitura de Pensamentos



X_i → imagem da ressonância magnética

Y_i → “Pensamento”






Exemplo: Leitura de Pensamentos



X_i → imagem da ressonância magnética


Y_i → “Pensamento”

Exemplo: Busca por Imagens Semelhantes

Google  x descreva a imagem aqui    


Web **Imagens** Notícias Shopping Mapas Mais ▾ Ferramentas de pesquisa

Aproximadamente 3 resultados (0,58 segundos)

 Tamanho da imagem:
958 × 1280

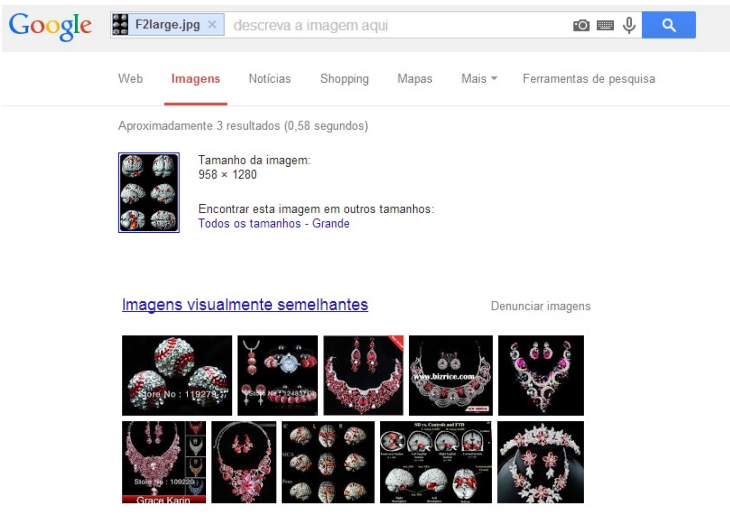
Encontrar esta imagem em outros tamanhos:
[Todos os tamanhos - Grande](#)

[Imagens visualmente semelhantes](#) [Denunciar imagens](#)



X_i → imagens na internet
Busca por estrutura

Exemplo: Busca por Imagens Semelhantes



The screenshot shows a Google search interface. The search bar contains the text "F2large.jpg" and "descreva a imagem aqui". Below the search bar, the "Imagens" tab is selected. The search results show "Aproximadamente 3 resultados (0,58 segundos)". The first result is a 3x3 grid of images showing jewelry. To the right of the grid, the text reads "Tamanho da imagem: 958 x 1280" and "Encontrar esta imagem em outros tamanhos: Todos os tamanhos - Grande". Below the grid, there are two links: "Imagens visualmente semelhantes" and "Denunciar imagens". The "Imagens visualmente semelhantes" link leads to a grid of 12 images showing various pieces of jewelry, including necklaces, earrings, and bangles, with some text overlays like "Store No : 11927837" and "www.birrice.com".

$X_i \rightarrow$ imagens na internet
Busca por estrutura

Exemplo: Recomendação de Amizades

The screenshot shows a Facebook interface in Google Chrome. The browser tabs include 'Inbox (2) - rafaelizbici', 'Facebook', 'www.cs.cmu.edu/~ab...', 'www.stat.cmu.edu/~...', and 'My Drive - Google Drive'. The address bar shows 'https://www.facebook.com/friends/requests/?fcref=ff'. The Facebook search bar contains the text 'Search for people, places and things'. The user's name 'Rafael' and 'Home' are visible in the top right navigation bar.

The main content area is titled 'People You May Know' and lists several user recommendations:

- Lingxue Zhu**: Carnegie Mellon University. Johnson Jining Qin and 7 other mutual friends. [Add Friend](#)
- Karina Pellegrinelli**: Instituto de Psiquiatria do HCFMUSP. Lee Fu-H and 3 other mutual friends. [Add Friend](#)
- Gustavo Cabral Duarte**: (UFS) Universidade Federal de Sergipe. Ricardo Bueno is a mutual friend. [Add Friend](#)
- Eunice De Fatima França**: São Paulo, Brazil. Elder Veras and 13 other mutual friends. [Add Friend](#)
- Marcia Benite França**: São Paulo, Brazil. Moises Oliveira and 4 other mutual friends. [Add Friend](#)
- Eduardo Rocha Cesar**: Técnico em Reabilitação Física at Fundação Faculdade de Medicina. [Add Friend](#)

In the bottom right corner, there is a 'Chat (Off)' button.

$X_i \rightarrow$ existe um link entre dois usuários

Exemplo: Recomendação de Propagandas

The screenshot shows a Google Chrome browser window with the search query "apartamentos sao carlos". The search results are displayed in a list format. The first result is "Apartamentos Novos Lopes - Lopes.com.br", which includes a link to "www.lopes.com.br/Apartamentos" and a description: "A Lopes tem mais de 18 mil opções de aptos p/ você escolher. Confira. Lopes Consultoria de Imóveis tem 85.103 seguidores no Google+ Jardim das Perdizes SP - Invista em um Ibis Hotel - Imóveis Até 24% Desconto". The second result is "Apartamento em São Carlos - MRV.com.br", with a link to "www.mrv.com.br/saocarlos" and the text "Apartamentos MRV em até 360 meses Corretor online 24h. Visite o site!". The third result is "Avenida Imóveis S. Carlos - avenidaimob.com.br", with a link to "www.avenidaimob.com.br/" and the text "Compra - Venda - Administra Ofertas Especias para Estudantes". Below these are two more results: "Imobiliária Cardinali" with a link to "www.imobiliariacardinali.com.br/" and "Venda - Casas e Apartamentos São Carlos" with a link to "saocarlos-saopaulo.olx.com.br" and the text "Casa Linda Jardim Ipanema Muito Bem localizada!! Venda | Casas e Apartamentos | São Carlos. R\$ 340.000. Topa negociar. Ontem., 13:07 ...".

apartamentos sao carlos - Pesquisa Google - Google Chrome

Inbox (2) - rafaelzbici x Facebook x www.cs.cmu.edu/~abi x www.stat.cmu.edu/~ x apartamentos sao carl x

https://www.google.com/search?q=machine+learning&oq=machine+learning&aqs=chrome..69i57j0j69i65j69i59j0l2.2696j0j7&sourceid=chrome&es_sm=122

Google apartamentos sao carlos

+Rafael

Web Mapas Imagens Shopping Vídeos Mais Ferramentas de pesquisa

Aproximadamente 266.000 resultados (0,42 segundos)

Apartamentos Novos Lopes - Lopes.com.br
Anúncio www.lopes.com.br/Apartamentos

A Lopes tem mais de 18 mil opções de aptos p/ você escolher. Confira.
Lopes Consultoria de Imóveis tem 85.103 seguidores no Google+
Jardim das Perdizes SP - Invista em um Ibis Hotel - Imóveis Até 24% Desconto

Apartamento em São Carlos - MRV.com.br
Anúncio www.mrv.com.br/saocarlos

Apartamentos MRV em até 360 meses Corretor online 24h. Visite o site!

Avenida Imóveis S. Carlos - avenidaimob.com.br
Anúncio www.avenidaimob.com.br/

Compra - Venda - Administra Ofertas Especias para Estudantes

Imobiliária Cardinali
www.imobiliariacardinali.com.br/

Imobiliária Cardinali - Imóveis, apartamentos, condomínios, casas, loteamentos, empreendimentos e lançamentos nos melhores bairros de São Carlos. Confira!
Locação - Vendas - Imóveis Próximos - Lançamento Ana Beatriz

Venda - Casas e Apartamentos São Carlos
saocarlos-saopaulo.olx.com.br - São Paulo

Casa Linda Jardim Ipanema Muito Bem localizada!! Venda | Casas e Apartamentos | São Carlos. R\$ 340.000. Topa negociar. Ontem., 13:07 ...

Exemplo: Sistemas de Recomendação

Imagine que temos um banco de dados em que cada linha representa a ida de uma pessoa a um supermercado, e cada coluna representa se ela comprou ou não determinado produto.

Objetivo: descobrir regras do tipo

“Quem compra **leite** em geral também compra **pão**”,

“Quem compra **cerveja e refrigerante** em geral também compra **carne**”.

“Quem compra **fralda** em geral também compra **cerveja**”.

Exemplo: Sistemas de Recomendação

Imagine que temos um banco de dados em que cada linha representa a ida de uma pessoa a um supermercado, e cada coluna representa se ela comprou ou não determinado produto.

Objetivo: descobrir regras do tipo

“Quem compra **leite** em geral também compra **pão**” ,

“Quem compra **cerveja e refrigerante** em geral também compra **carne**” .

“Quem compra **fralda** em geral também compra **cerveja**” .

Exemplo: Sistemas de Recomendação

Imagine que temos um banco de dados em que cada linha representa a ida de uma pessoa a um supermercado, e cada coluna representa se ela comprou ou não determinado produto.

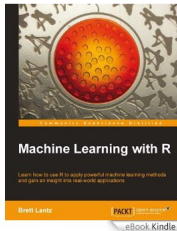
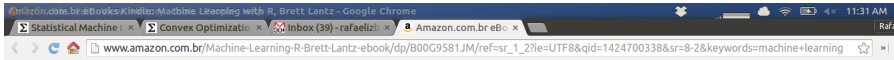
Objetivo: descobrir regras do tipo

“Quem compra **leite** em geral também compra **pão**”,

“Quem compra **cerveja e refrigerante** em geral também compra **carne**”.

“Quem compra **fralda** em geral também compra **cerveja**”.

Exemplo: Sistemas de Recomendação



Clique para abrir visualização expandida



Machine Learning with R [EBook Kindle]

Brett Lantz (Autor)

★★★★★ (2 avaliações de clientes)

Preço digital sugerido: ~~R\$ 62,00~~ O que é isso?

Preço Kindle: **R\$ 58,65** inclui envio sem fio internacional gratuito via **Amazon Whispernet**

Você está economizando: **R\$ 4,34 (7%)**

- Número de páginas: 398 páginas (Contém números de páginas reais)
- Idioma: Inglês
- Ainda não possui um Kindle? [Compre seu Kindle aqui](#) ou baixe um de nossos [Aplicativos de Leitura Kindle GRATUITOS](#).

Formatos	Preço da Amazon
eBook Kindle	R\$ 58,65
Capa comum	R\$ 206,30

[Compre agora com 1-Clique](#)

Entregue no seu Kindle ou em outro dispositivo

Disponível em seu computador

[Insira aqui o cupom de desconto](#)

[Adicione à Lista de Desejos](#)

Amostra grátis

Leia agora o início deste eBook gratuitamente

[Envie amostra agora](#)

Entregue no seu Kindle ou em outro dispositivo

Como as amostras funcionam

Clientes que compraram este item também compraram

Página 1 de 2



R Cookbook (O'Reilly Cookbooks)
Paul Teator
★★★★★ 1
eBook Kindle



Predictive Analytics: The Power to Predict Who...
Eric Siegel
★★★★★ 1
eBook Kindle



R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data...)
Jared P. Lander
eBook Kindle



Practical Machine Learning: Innovations in Recommendation
Ted Dunning
eBook Kindle



Artificial Intelligence for Humans, Volume 1: Fundamental Algorithms...
Jeff Heaton
eBook Kindle



Leia nossos eBooks mesmo sem ter um dispositivo Kindle: basta baixar um de nossos [Aplicativos de Leitura Kindle GRATUITOS](#) para smartphones, tablets e computadores.

[Compartilhar](#) [Facebook](#) [Twitter](#) [Pinterest](#)

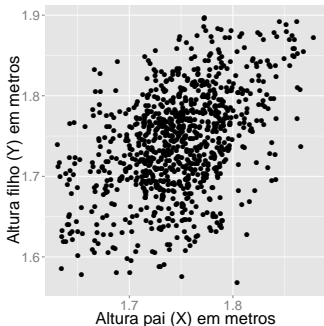
Aprendizado de Máquina é portanto uma ferramenta extremamente útil hoje para fazer **predições** (*aprendizado supervisionado*) e **aprender estruturas** (*aprendizado não supervisionado*) em conjuntos de dados **grandes**

Vamos entrar mais fundo no problema de **predição**.

Aprendizado de Máquina é portanto uma ferramenta extremamente útil hoje para fazer **predições** (*aprendizado supervisionado*) e **aprender estruturas** (*aprendizado não supervisionado*) em conjuntos de dados **grandes**

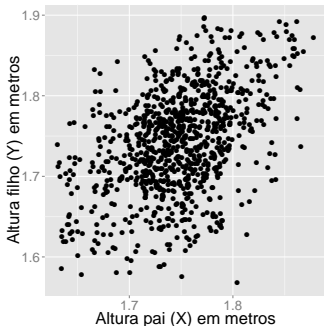
Vamos entrar mais fundo no problema de **predição**.

Vamos assumir que deseja-se prever a altura de um filho (Y) com base na altura de seu pai (X), e que coletamos uma amostra $(X_1, Y_1), \dots, (X_n, Y_n)$ com observações sobre n indivíduos.



Com base nesses dados, podemos criar uma **função de predição** $g(x)$: dado que a altura do pai de um indivíduo é $X = x$, $g(x)$ representa nossa predição sobre a altura do filho Y .

Vamos assumir que deseja-se prever a altura de um filho (Y) com base na altura de seu pai (X), e que coletamos uma amostra $(X_1, Y_1), \dots, (X_n, Y_n)$ com observações sobre n indivíduos.



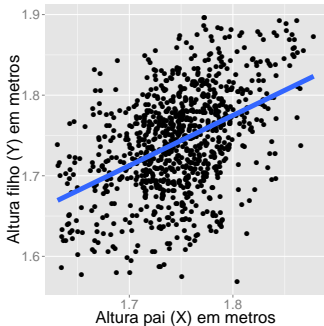
Com base nesses dados, podemos criar uma **função de predição** $g(x)$: dado que a altura do pai de um indivíduo é $X = x$, $g(x)$ representa nossa predição sobre a altura do filho Y .

Os principais objetivos de métodos de predição são (i) **construir** g de modo a se obter **boas predições**, e (ii) saber **quantificar** o **quão boa** uma (função de) predição é.

Uma maneira simples de se criar uma função de predição g

Os principais objetivos de métodos de predição são (i) **construir** g de modo a se obter **boas predições**, e (ii) saber **quantificar** o **quão boa** uma (função de) predição é.

Uma maneira simples de se criar uma função de predição g neste caso é através de uma **regressão linear** (vamos rever isto mais tarde).



Digamos que queremos prever a altura de um filho cujo pai tem $x = 1,80$ m. Nossa previsão neste caso é de $g(1,80) = 1,77$. Se observamos uma altura de $1,76$ m, como quantificar quão errados estamos? (Exemplo no R)

Uma maneira de se fazer isso é usando-se o **erro quadrático**: neste caso, diríamos que o erro incorrido por g para prever $x = 1,80$ foi de

$$(g(x) - y)^2 = (1,76 - 1,77)^2 = 0.0001.$$

Quanto maior o valor de $(g(x) - y)^2$, pior é nossa previsão.

Digamos que queremos prever a altura de um filho cujo pai tem $x = 1,80$ m. Nossa previsão neste caso é de $g(1,80) = 1,77$. Se observamos uma altura de $1,76$ m, como quantificar quão errados estamos? (Exemplo no R)

Uma maneira de se fazer isso é usando-se o **erro quadrático**: neste caso, diríamos que o erro incorrido por g para prever $x = 1,80$ foi de

$$(g(x) - y)^2 = (1,76 - 1,77)^2 = 0.0001.$$

Quanto maior o valor de $(g(x) - y)^2$, pior é nossa previsão.

Digamos que queremos prever a altura de um filho cujo pai tem $x = 1,80$ m. Nossa previsão neste caso é de $g(1,80) = 1,77$. Se observamos uma altura de $1,76$ m, como quantificar quão errados estamos? (Exemplo no R)

Uma maneira de se fazer isso é usando-se o **erro quadrático**: neste caso, diríamos que o erro incorrido por g para prever $x = 1,80$ foi de

$$(g(x) - y)^2 = (1,76 - 1,77)^2 = 0.0001.$$

Quanto maior o valor de $(g(x) - y)^2$, pior é nossa previsão.

Note que desta maneira quantificamos quão boa g é apenas para um dado par (x, y) . Uma maneira de se generalizar isso (i.e., atribuir um único número quantificando quão boa g é) é através da função de risco $R(g)$:

$$R(g) = \mathbb{E} [(Y - g(X))^2].$$

Note que desta maneira quantificamos quão boa g é apenas para um dado par (x, y) . Uma maneira de se generalizar isso (i.e., atribuir um único número quantificando quão boa g é) é através da função de risco $R(g)$:

$$R(g) = \mathbb{E} [(Y - g(X))^2].$$

$R(g)$ que avalia o erro preditivo de g . Por quê? Digamos que observamos um novo conjunto $(X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$.

Pela lei dos grandes números sabemos que, se m é grande,

$$\frac{1}{m} \sum_{i=1}^m (Y_{n+i} - g(X_{n+i}))^2 \approx R(g) := \mathbb{E} [(Y - g(X))^2]$$

Em outras palavras, $R(g)$ é aproximadamente uma média do erro quadrático para cada uma das novas observações.

Assim, se conseguimos g tal que $R(g)$ seja pequeno, teremos boas predições em novas observações.

$R(g)$ que avalia o **erro preditivo** de g . Por quê? Digamos que observamos um novo conjunto $(X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$.

Pela **lei dos grandes números** sabemos que, se m é grande,

$$\frac{1}{m} \sum_{i=1}^m (Y_{n+i} - g(X_{n+i}))^2 \approx R(g) := \mathbb{E} [(Y - g(X))^2]$$

Em outras palavras, $R(g)$ é aproximadamente uma **média do erro quadrático** para cada uma das novas observações.

Assim, se conseguimos g tal que $R(g)$ seja pequeno, teremos boas **predições em novas observações**.

$R(g)$ que avalia o **erro preditivo** de g . Por quê? Digamos que observamos um novo conjunto $(X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$.

Pela **lei dos grandes números** sabemos que, se m é grande,

$$\frac{1}{m} \sum_{i=1}^m (Y_{n+i} - g(X_{n+i}))^2 \approx R(g) := \mathbb{E} [(Y - g(X))^2]$$

Em outras palavras, $R(g)$ é aproximadamente uma **média do erro quadrático** para cada uma das novas observações.

Assim, se conseguimos g tal que $R(g)$ seja pequeno, teremos boas **previsões em novas observações**.

$R(g)$ que avalia o erro preditivo de g . Por quê? Digamos que observamos um novo conjunto $(X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$.

Pela lei dos grandes números sabemos que, se m é grande,

$$\frac{1}{m} \sum_{i=1}^m (Y_{n+i} - g(X_{n+i}))^2 \approx R(g) := \mathbb{E} [(Y - g(X))^2]$$

Em outras palavras, $R(g)$ é aproximadamente uma média do erro quadrático para cada uma das novas observações.

Assim, se conseguimos g tal que $R(g)$ seja pequeno, teremos boas predições em novas observações.

$R(g)$ que avalia o erro preditivo de g . Por quê? Digamos que observamos um novo conjunto $(X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$.

Pela lei dos grandes números sabemos que, se m é grande,

$$\frac{1}{m} \sum_{i=1}^m (Y_{n+i} - g(X_{n+i}))^2 \approx R(g) := \mathbb{E} [(Y - g(X))^2]$$

Em outras palavras, $R(g)$ é aproximadamente uma média do erro quadrático para cada uma das novas observações.

Assim, se conseguimos g tal que $R(g)$ seja pequeno, teremos boas predições em novas observações.

Resumindo até agora

Em um problema de predição:

- ▶ Observamos um conjunto de treinamento $(X_1, Y_1), \dots, (X_n, Y_n)$. X são chamados de **preditores**, ou **variáveis explicativas**, ou **variáveis independentes**, ou **features**. Y é chamado de **resposta**, ou **variável dependente**, ou **labels**
- ▶ Desejamos criar uma **função de predição** $g(x)$ para prever novas observações X_{n+1}, \dots, X_{n+m} bem
- ▶ Prever **novas observações bem** significa criar g tal que o risco quadrático $R(g) = \mathbb{E} [(Y - g(X))^2]$ seja baixo (dependendo da situação, outras funções de risco podem ser mais adequadas; veremos mais sobre isso depois)

Vamos ver ao longo do curso como, em diferentes contextos, criar g tal que $R(g)$ seja baixo, e como comparar diferentes g 's.

Resumindo até agora

Em um problema de predição:

- ▶ Observamos um conjunto de treinamento $(X_1, Y_1), \dots, (X_n, Y_n)$. X são chamados de **preditores**, ou **variáveis explicativas**, ou **variáveis independentes**, ou **features**. Y é chamado de **resposta**, ou **variável dependente**, ou **labels**
- ▶ Desejamos criar uma **função de predição** $g(x)$ para prever **novas observações** X_{n+1}, \dots, X_{n+m} bem
- ▶ Prever **novas observações bem** significa criar g tal que o risco quadrático $R(g) = \mathbb{E} [(Y - g(X))^2]$ seja **baixo** (dependendo da situação, outras funções de risco podem ser mais adequadas; veremos mais sobre isso depois)

Vamos ver ao longo do curso como, em diferentes contextos, criar g tal que $R(g)$ seja baixo, e como comparar diferentes g 's.

Resumindo até agora

Em um problema de predição:

- ▶ Observamos um conjunto de treinamento $(X_1, Y_1), \dots, (X_n, Y_n)$. X são chamados de **preditores**, ou **variáveis explicativas**, ou **variáveis independentes**, ou **features**. Y é chamado de **resposta**, ou **variável dependente**, ou **labels**
- ▶ Desejamos criar uma **função de predição** $g(x)$ para prever **novas observações** X_{n+1}, \dots, X_{n+m} bem
- ▶ Prever **novas observações bem** significa criar g tal que o risco quadrático $R(g) = \mathbb{E} [(Y - g(X))^2]$ seja **baixo** (dependendo da situação, outras funções de risco podem ser mais adequadas; veremos mais sobre isso depois)

Vamos ver ao longo do curso como, em diferentes contextos, criar g tal que $R(g)$ seja baixo, e como comparar diferentes g 's.

Resumindo até agora

Em um problema de predição:

- ▶ Observamos um conjunto de treinamento $(X_1, Y_1), \dots, (X_n, Y_n)$. X são chamados de **preditores**, ou **variáveis explicativas**, ou **variáveis independentes**, ou **features**. Y é chamado de **resposta**, ou **variável dependente**, ou **labels**
- ▶ Desejamos criar uma **função de predição** $g(x)$ para prever **novas observações** X_{n+1}, \dots, X_{n+m} bem
- ▶ Prever **novas observações bem** significa criar g tal que o risco quadrático $R(g) = \mathbb{E} [(Y - g(X))^2]$ seja **baixo** (dependendo da situação, outras funções de risco podem ser mais adequadas; veremos mais sobre isso depois)

Vamos ver ao longo do curso como, em diferentes contextos, criar g tal que $R(g)$ seja baixo, e como comparar diferentes g 's.

Qual a melhor função $g(x)$?

O problema de encontrar g que minimize $R(g) = \mathbb{E} [(Y - g(X))^2]$ possui solução analítica!!

Seja $r(x) = \mathbb{E}[Y|X = x]$ a **função de regressão** (note que não estamos assumindo que ela é linear).

Então $R(r) \leq R(g)$ para toda função $g(x)$. Este é um exercício da lista.

O problema então está resolvido? Não!! $r(x) = \mathbb{E}[Y|X = x]$ depende de quantidades desconhecidas!

Contudo, este resultado indica que **uma estimativa boa da função de regressão r nos fornece um bom preditor para y com base em x .**

Qual a melhor função $g(x)$?

O problema de encontrar g que minimize $R(g) = \mathbb{E} [(Y - g(X))^2]$ possui solução analítica!!

Seja $r(x) = \mathbb{E}[Y|X = x]$ a **função de regressão** (note que não estamos assumindo que ela é linear).

Então $R(r) \leq R(g)$ para toda função $g(x)$. Este é um exercício da lista.

O problema então está resolvido? Não!! $r(x) = \mathbb{E}[Y|X = x]$ depende de quantidades desconhecidas!

Contudo, este resultado indica que **uma estimativa boa da função de regressão r nos fornece um bom preditor para y com base em x .**

Qual a melhor função $g(x)$?

O problema de encontrar g que minimize $R(g) = \mathbb{E} [(Y - g(X))^2]$ possui solução analítica!!

Seja $r(x) = \mathbb{E}[Y|X = x]$ a **função de regressão** (note que não estamos assumindo que ela é linear).

Então $R(r) \leq R(g)$ para toda função $g(x)$. Este é um exercício da lista.

O problema então está resolvido? Não!! $r(x) = \mathbb{E}[Y|X = x]$ depende de quantidades desconhecidas!

Contudo, este resultado indica que **uma estimativa boa da função de regressão r nos fornece um bom preditor para y com base em x .**

Qual a melhor função $g(x)$?

O problema de encontrar g que minimize $R(g) = \mathbb{E} [(Y - g(X))^2]$ possui solução analítica!!

Seja $r(x) = \mathbb{E}[Y|X = x]$ a **função de regressão** (note que não estamos assumindo que ela é linear).

Então $R(r) \leq R(g)$ para toda função $g(x)$. Este é um exercício da lista.

O problema então está resolvido? Não!! $r(x) = \mathbb{E}[Y|X = x]$ depende de quantidades desconhecidas!

Contudo, este resultado indica que **uma estimativa boa da função de regressão r nos fornece um bom preditor para y com base em x .**

Qual a melhor função $g(x)$?

O problema de encontrar g que minimize $R(g) = \mathbb{E} [(Y - g(X))^2]$ possui solução analítica!!

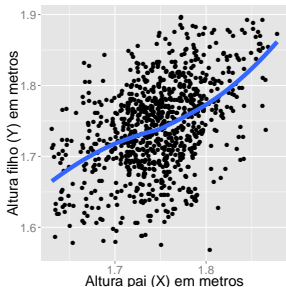
Seja $r(x) = \mathbb{E}[Y|X = x]$ a **função de regressão** (note que não estamos assumindo que ela é linear).

Então $R(r) \leq R(g)$ para toda função $g(x)$. Este é um exercício da lista.

O problema então está resolvido? Não!! $r(x) = \mathbb{E}[Y|X = x]$ depende de quantidades desconhecidas!

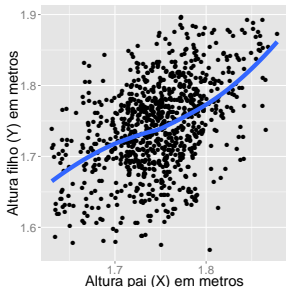
Contudo, este resultado indica que **uma estimativa boa da função de regressão r nos fornece um bom preditor para y com base em x .**

Uma forma de se criar/estimar g é através de uma **regressão linear**. Ao longo das próximas aulas veremos algumas metodologias mais **sofisticadas** para isso; inclusive métodos que podem ser aplicados quando x consiste em **imagens**, **textos** etc. Veja por exemplo um método mais sofisticado (não linear) para os dados das idades:



Antes disso, precisamos entender mais a fundo os elementos de um problema de predição.

Uma forma de se criar/estimar g é através de uma **regressão linear**. Ao longo das próximas aulas veremos algumas metodologias mais **sofisticadas** para isso; inclusive métodos que podem ser aplicados quando x consiste em **imagens**, **textos** etc. Veja por exemplo um método mais sofisticado (não linear) para os dados das idades:



Antes disso, precisamos entender mais a fundo os elementos de um problema de predição.

Regressão linear sob um ponto de vista preditivo

Regressão linear é (ou ao menos deveria ser) um velho conhecido seu.

É um estimador de $r(x) = \mathbb{E}[Y|X = x]$ quando ele assume uma forma linear, i.e.,

$$r(x) = \beta_0 + \beta_1 x,$$

ou, no caso em que $x = (x_1, \dots, x_p)$ é um vetor,

$$r(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Em notação matricial,

$$r(x) = \beta^t x,$$

com $\beta = (\beta_0, \dots, \beta_p)$ e $x = (x_1, \dots, x_p)$

Regressão linear sob um ponto de vista preditivo

Regressão linear é (ou ao menos deveria ser) um velho conhecido seu.

É um estimador de $r(x) = \mathbb{E}[Y|X = x]$ quando ele assume uma forma linear, i.e.,

$$r(x) = \beta_0 + \beta_1 x,$$

ou, no caso em que $x = (x_1, \dots, x_p)$ é um vetor,

$$r(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Em notação matricial,

$$r(x) = \beta^t x,$$

com $\beta = (\beta_0, \dots, \beta_p)$ e $x = (x_1, \dots, x_n)$

Regressão linear sob um ponto de vista preditivo

Regressão linear é (ou ao menos deveria ser) um velho conhecido seu.

É um estimador de $r(x) = \mathbb{E}[Y|X = x]$ quando ele assume uma forma linear, i.e.,

$$r(x) = \beta_0 + \beta_1 x,$$

ou, no caso em que $x = (x_1, \dots, x_p)$ é um vetor,

$$r(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Em notação matricial,

$$r(x) = \beta^t x,$$

com $\beta = (\beta_0, \dots, \beta_p)$ e $x = (x_1, \dots, x_n)$

Regressão linear sob um ponto de vista preditivo

Regressão linear é (ou ao menos deveria ser) um velho conhecido seu.

É um estimador de $r(x) = \mathbb{E}[Y|X = x]$ quando ele assume uma forma linear, i.e.,

$$r(x) = \beta_0 + \beta_1 x,$$

ou, no caso em que $x = (x_1, \dots, x_p)$ é um vetor,

$$r(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Em notação matricial,

$$r(x) = \beta^t x,$$

com $\beta = (\beta_0, \dots, \beta_p)$ e $x = (x_1, \dots, x_n)$

Regressão linear sob um ponto de vista preditivo

Relembrando: o estimador para os parâmetros β usual é o **estimador de mínimos quadrados** que em sua **forma matricial** é dado por

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p) = (X^t X)^{-1} X^t Y.$$

Este é o estimador que minimiza

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p})^2$$

Assim, **uma estimativa** para $r(x) = \mathbb{E}[Y|X = x]$ é dada por

$$\hat{r}(x) = \hat{\beta}^t x.$$

Regressão linear sob um ponto de vista preditivo

Relembrando: o estimador para os parâmetros β usual é o **estimador de mínimos quadrados** que em sua **forma matricial** é dado por

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p) = (X^t X)^{-1} X^t Y.$$

Este é o estimador que minimiza

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p})^2$$

Assim, **uma estimativa** para $r(x) = \mathbb{E}[Y|X = x]$ é dada por

$$\hat{r}(x) = \hat{\beta}^t x.$$

Regressão linear sob um ponto de vista preditivo

Relembrando: o estimador para os parâmetros β usual é o **estimador de mínimos quadrados** que em sua **forma matricial** é dado por

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p) = (X^t X)^{-1} X^t Y.$$

Este é o estimador que minimiza

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p})^2$$

Assim, **uma estimativa** para $r(x) = \mathbb{E}[Y|X = x]$ é dada por

$$\hat{r}(x) = \hat{\beta}^t x.$$

Regressão linear sob um ponto de vista preditivo

No R, podemos usar a função `lm`:

```
ajuste=lm("y ~ x",data=dados);
```

também é possível usar a forma matricial

Predição versus Inferência:

Em **inferência** em geral **assume-se que o modelo é correto**. Isso ocorre pois o principal objetivo está na interpretação dos parâmetros β . Ex: quais parâmetros são significantes? Qual o efeito do aumento da dose do remédio no medicamento? etc.

Já em **predição**, nosso objetivo maior é simplesmente criar $g(x)$ que tenha **bom poder preditivo**. **Não estamos assumindo que a verdadeira regressão é linear!!** Isto não quer dizer que não possamos interpretar nossos resultados, mas este em geral não é o foco das análises.

Predição versus Inferência:

Em **inferência** em geral **assume-se que o modelo é correto**. Isso ocorre pois o principal objetivo está na interpretação dos parâmetros β . Ex: quais parâmetros são significantes? Qual o efeito do aumento da dose do remédio no medicamento? etc.

Já em **predição**, nosso objetivo maior é simplesmente criar $g(x)$ que tenha **bom poder preditivo**. **Não estamos assumindo que a verdadeira regressão é linear!!** Isto não quer dizer que não possamos interpretar nossos resultados, mas este em geral não é o foco das análises.

As duas culturas

L. Breiman: Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231, 2001.

Duas culturas no uso de modelos estatísticos:

- ▶ Data Modeling Culture:

- ▶ Algorithmic Modeling Culture:

As duas culturas

L. Breiman: Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231, 2001.

Duas culturas no uso de modelos estatísticos:

- ▶ Data Modeling Culture:

- ▶ Algorithmic Modeling Culture:

As duas culturas

L. Breiman: Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231, 2001.

Duas culturas no uso de modelos estatísticos:

- ▶ **Data Modeling Culture:** Domina a comunidade estatística. Nela se assume que o modelo para $r(\vec{x})$ é correto. Testar suposições é fundamental. Foco em inferência.
- ▶ **Algorithmic Modeling Culture:**

As duas culturas

L. Breiman: Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231, 2001.

Duas culturas no uso de modelos estatísticos:

- ▶ **Data Modeling Culture:** Domina a comunidade estatística. Nela se assume que o modelo para $r(\vec{x})$ é correto. Testar suposições é fundamental. Foco em inferência.
- ▶ **Algorithmic Modeling Culture:** Domina a comunidade de machine learning. Nela **não** se assume que o modelo utilizado para $r(\vec{x})$ é correto; o modelo é utilizado apenas para criar bons algoritmos preditivos.

“Oddly, we are in a period where there has never been such a wealth of new statistical problems and sources of data. The danger is that if we define the boundaries of our field in terms of familiar tools and familiar problems, we will fail to grasp the new opportunities”. (Breiman, 2001)

Resumo da Aula: Mineração de Dados é uma ferramenta extremamente útil hoje para fazer **predições** (*aprendizado supervisionado*) e **aprender estruturas** (*aprendizado não supervisionado*) em conjuntos de dados **grandes**

Próxima Aula: Vamos entrar mais fundo no problema de **predição**

Resumo da Aula: Mineração de Dados é uma ferramenta extremamente útil hoje para fazer **predições** (*aprendizado supervisionado*) e **aprender estruturas** (*aprendizado não supervisionado*) em conjuntos de dados **grandes**

Próxima Aula: Vamos entrar mais fundo no problema de **predição**