



Introdução

Prof.: Eduardo Vargas Ferreira

O que é Machine Learning?



Estatística



Data Mining



O que é Machine Learning?



Estatística



Machine Learning



Data Mining



O que é Machine Learning?



Estatística

Machine Learning

Data Mining



Métodos

O que é Machine Learning?



Estatística

Machine Learning

Data Mining



Métodos

problemas

O que é Machine Learning?



Estatística

Machine Learning

Data Mining

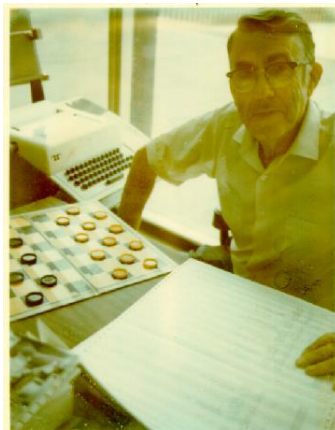


Métodos

aplicados a

problemas

- Área de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados (Arthur Samuel, 1959).

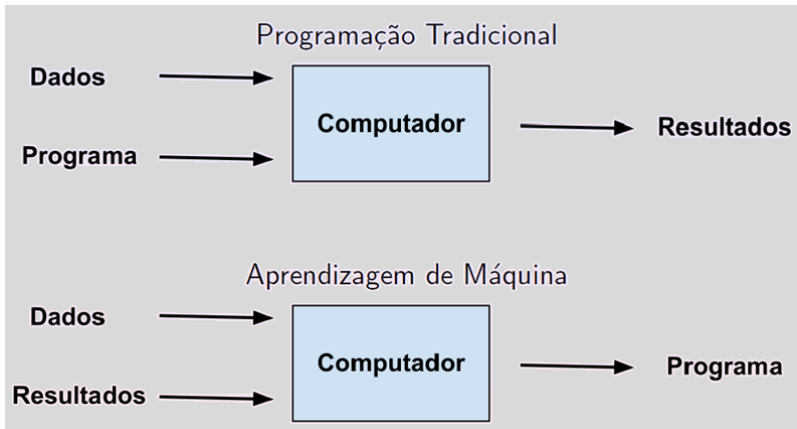


- Um programa de computador é dito aprender a partir de uma experiência **E** com respeito a alguma classe de tarefas **T** e medida de desempenho **P**, se seu desempenho em tarefas **T**, medido por **P**, melhora com a experiência **E** (Tom Mitchell, 1998).

Exemplo

Suponha que seu Gmail observe/registre quais e-mails você marca ou não como spam. E, baseado nisso, aprende como melhor filtrar mensagens de spam de sua caixa de e-mails.

- **Tarefa T:** Classificar e-mails como spam ou não.
- **Experiência E:** observação de quais e-mails você marca como spam.
- **Desempenho/performance P:** número (ou proporção) de e-mails corretamente classificados como spam/não spam.



- Machine Learning é a ciência de **descobrir estruturas** e **fazer previsões** em conjunto de dados;
- O aprendizado é efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado;
- Ele pode ser dividido em:
 - ★ Aprendizado **supervisionado**;
 - ★ Aprendizado **não-supervisionado**.

- Machine Learning é a ciência de **descobrir estruturas** e **fazer previsões** em conjunto de dados;
- O aprendizado é efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado;
- Ele pode ser dividido em:
 - ★ Aprendizado **supervisionado**;
 - ★ Aprendizado **não-supervisionado**.

- Machine Learning é a ciência de **descobrir estruturas** e **fazer previsões** em conjunto de dados;
- O aprendizado é efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado;
- Ele pode ser dividido em:
 - ★ Aprendizado **supervisionado**;
 - ★ Aprendizado **não-supervisionado**.

- Machine Learning é a ciência de **descobrir estruturas** e **fazer previsões** em conjunto de dados;
- O aprendizado é efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado;
- Ele pode ser dividido em:
 - ★ Aprendizado **supervisionado**;
 - ★ Aprendizado **não-supervisionado**.

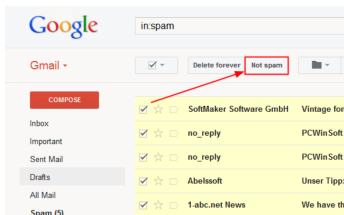
- Aprendizado **supervisionado**: é fornecido ao algoritmo de aprendizado, ou *indutor*, um conjunto de exemplos de treinamento para os quais o **rótulo da classe associada é conhecido**;
 - ★ Dadas as medições $(x_1, y_1), \dots, (x_n, y_n)$, aprender um modelo para **prever** y baseado em x ;
 - ★ Ou seja, o objetivo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados;
 - ★ Para rótulos de classe discretos, esse problema é conhecido como **classificação** e para valores contínuos como **regressão**.

- Aprendizado **supervisionado**: é fornecido ao algoritmo de aprendizado, ou *indutor*, um conjunto de exemplos de treinamento para os quais o **rótulo da classe associada é conhecido**;
 - ★ Dadas as medições $(x_1, y_1), \dots, (x_n, y_n)$, aprender um modelo para **prever** y baseado em x ;
 - ★ Ou seja, o objetivo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados;
 - ★ Para rótulos de classe discretos, esse problema é conhecido como **classificação** e para valores contínuos como **regressão**.

- Aprendizado **supervisionado**: é fornecido ao algoritmo de aprendizado, ou *indutor*, um conjunto de exemplos de treinamento para os quais o **rótulo da classe associada é conhecido**;
 - ★ Dadas as medições $(x_1, y_1), \dots, (x_n, y_n)$, aprender um modelo para **prever** y baseado em x ;
 - ★ Ou seja, o objetivo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados;
 - ★ Para rótulos de classe discretos, esse problema é conhecido como **classificação** e para valores contínuos como **regressão**.

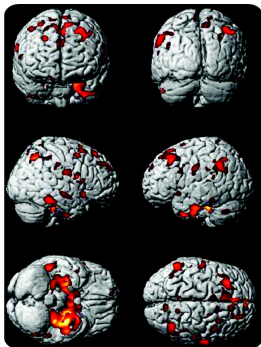
- Aprendizado **supervisionado**: é fornecido ao algoritmo de aprendizado, ou *indutor*, um conjunto de exemplos de treinamento para os quais o **rótulo da classe associada é conhecido**;
 - ★ Dadas as medições $(x_1, y_1), \dots, (x_n, y_n)$, aprender um modelo para **prever** y baseado em x ;
 - ★ Ou seja, o objetivo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados;
 - ★ Para rótulos de classe discretos, esse problema é conhecido como **classificação** e para valores contínuos como **regressão**.

Detecção de Spams



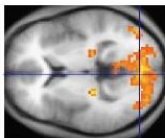
- x_i : e-mail;
- y_i : spam/não spam;
- **Objetivo:** prever y_j com base em x_j .

Predição de Alzheimer



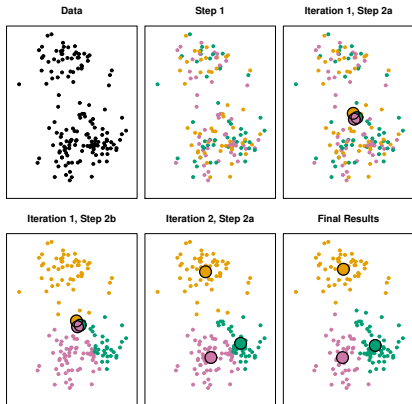
- x_i : imagem da ressonância magnética;
- y_i : Paciente com/sem Alzheimer;
- **Objetivo:** prever y_j com base em x_j .

Leitura de pensamentos



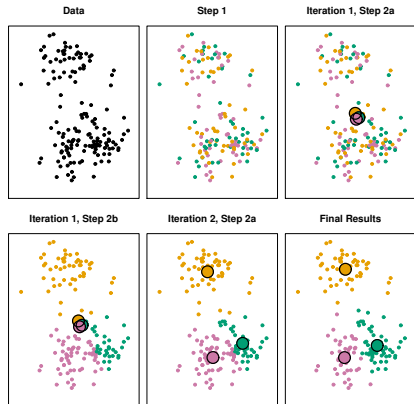
- x_i : imagem da ressonância magnética;
- y_i : "Pensamento";
- **Objetivo:** prever y_j com base em x_j .

- Aprendizado **não-supervisionado**: o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira.



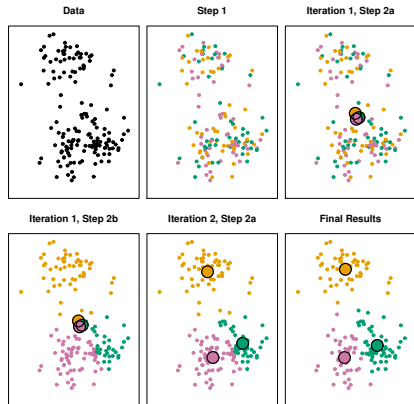
- ★ Dadas as medições x_1, \dots, x_n , descobrir **alguma estrutura** baseada na similaridade;
- ★ Após a determinação dos agrupamentos, normalmente, é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema.

- Aprendizado **não-supervisionado**: o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira.



- ★ Dadas as medições x_1, \dots, x_n , descobrir **alguma estrutura** baseada na similaridade;
- ★ Após a determinação dos agrupamentos, normalmente, é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema.

- Aprendizado **não-supervisionado**: o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira.



- ★ Dadas as medições x_1, \dots, x_n , descobrir **alguma estrutura** baseada na similaridade;
- ★ Após a determinação dos agrupamentos, normalmente, é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema.

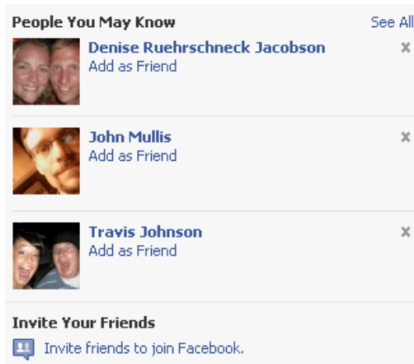
Exemplo de aprendizado não-supervisionado






- Busca por imagens semelhantes
 - ★ x_i : imagens na internet;
 - ★ **Objetivo:** Busca por estrutura semelhante.

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 results section shows 'Aproximadamente 3 resultados (0,58 segundos)'. A thumbnail of the image is shown with the text 'Tamanho da imagem: 958 x 1280' and 'Encontrar esta imagem em outros tamanhos: Todos os tamanhos - Grande'. Below this, there is a link for 'Imagens visualmente semelhantes' and a 'Denunciar imagens' button. The main content area displays a grid of 12 image thumbnails, all showing various styles of jewelry, including necklaces and earrings, which are visually similar to the original image.

- Recomendação de amizades
 - ★ x_i : existe um link entre dois usuários;



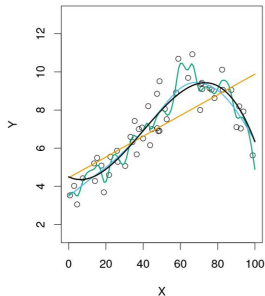
The screenshot shows the 'People You May Know' section of a Facebook profile. It features three recommended users, each with a profile picture, name, and an 'Add as Friend' button. To the right of each user's name is a small 'x' icon. At the top right of the section is a 'See All' link. Below the recommendations is a section titled 'Invite Your Friends' with a button that says 'Invite friends to join Facebook.'

Profile Picture	Name	Action
	Denise Rueherschneck Jacobson Add as Friend	x
	John Mullis Add as Friend	x
	Travis Johnson Add as Friend	x

Como avaliar a precisão dos modelos?



- Suponha que estamos interessados em estudar a relação entre X e Y

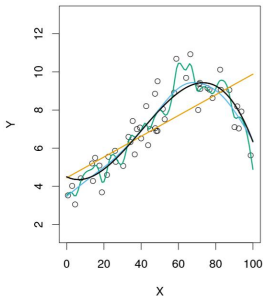


- A verdadeira relação está representada pela curva em preto;
- Podemos definir varias funções $h(x)$, mas qual fornece a melhor predição?
- **Resposta:** a que apresentar menos custo (ou risco).

Como avaliar a precisão dos modelos?



- Suponha que estamos interessados em estudar a relação entre X e Y

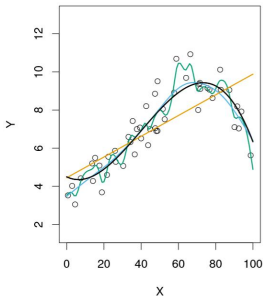


- A verdadeira relação está representada pela curva em preto;
- Podemos definir varias funções $h(x)$, mas qual fornece a melhor predição?
- **Resposta:** a que apresentar menos custo (ou risco).

Como avaliar a precisão dos modelos?



- Suponha que estamos interessados em estudar a relação entre X e Y



- A verdadeira relação está representada pela curva em preto;
- Podemos definir varias funções $h(x)$, mas qual fornece a melhor predição?
- **Resposta:** a que apresentar menos custo (ou risco).

Função custo

- Descreve o quão bom a superfície de resposta $h(\mathbf{x})$ ajustou-se aos dados;
- Valores pequenos da função custo corresponde à melhores ajustes;
- Notação: $J(y_i, h(\mathbf{x}))$;

- No gráfico da direita, a linha cinza representa os dados de treinamento e a vermelha de teste (veremos isso mais adiante).

Função custo

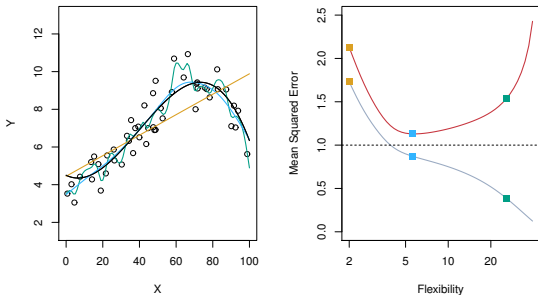
- Descreve o quão bom a superfície de resposta $h(\mathbf{x})$ ajustou-se aos dados;
- Valores pequenos da função custo corresponde à melhores ajustes;
- Notação: $J(y_i, h(\mathbf{x}))$;

- No gráfico da direita, a linha cinza representa os dados de treinamento e a vermelha de teste (veremos isso mais adiante).

Função custo



- Descreve o quão bom a superfície de resposta $h(\mathbf{x})$ ajustou-se aos dados;
- Valores pequenos da função custo corresponde à melhores ajustes;
- Notação: $J(y_i, h(\mathbf{x}))$;



- No gráfico da direita, a linha cinza representa os dados de treinamento e a vermelha de teste (veremos isso mais adiante).

Tipos de função custo

Soma de quadrado dos desvios (SQD)

- Definido como

$$J(y_i, h(x)) = \frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2$$

- Note que

$$f(y|x) \sim e^{-\left(\frac{y-h(x)}{\sigma}\right)^2}.$$

- Assim, se quisermos minimizar o $-\log$ da verossimilhança (equivalente a maximizar o \log da verossimilhança), temos (desprezando as constantes)

$$\begin{aligned} \min\{-(\log\text{-verossimilhança})\} &\sim \min\left\{-\log\left(\prod_{i=1}^n e^{-\left(\frac{y_i-h(x_i)}{\sigma}\right)^2}\right)\right\} \\ &\sim \min\left\{\frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2\right\}. \end{aligned}$$

Soma de quadrado dos desvios (SQD)

- Definido como

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2$$

- Note que

$$f(y|x) \sim e^{-\left(\frac{y-h(x)}{\sigma}\right)^2}.$$

- Assim, se quisermos minimizar o $-\log$ da verossimilhança (equivalente a maximizar o \log da verossimilhança), temos (desprezando as constantes)

$$\begin{aligned} \min\{-(\log\text{-verossimilhança})\} &\sim \min\left\{-\log\left(\prod_{i=1}^n e^{-\left(\frac{y_i-h(x_i)}{\sigma}\right)^2}\right)\right\} \\ &\sim \min\left\{\frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2\right\}. \end{aligned}$$

Soma de quadrado dos desvios (SQD)

- Definido como

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2$$

- Note que

$$f(y|x) \sim e^{-\left(\frac{y-h(x)}{\sigma}\right)^2}.$$

- Assim, se quisermos minimizar o $-\log$ da verossimilhança (equivalente a maximizar o \log da verossimilhança), temos (desprezando as constantes)

$$\begin{aligned} \min\{-(\log\text{-verossimilhança})\} &\sim \min\left\{-\log\left(\prod_{i=1}^n e^{-\left(\frac{y_i-h(x_i)}{\sigma}\right)^2}\right)\right\} \\ &\sim \min\left\{\frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2\right\}. \end{aligned}$$

Tipos de função custo

Soma de quadrado dos desvios (SQD)

- Definido como

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2$$

- Note que

$$f(y|x) \sim e^{-\left(\frac{y-h(x)}{\sigma}\right)^2}.$$

- Assim, se quisermos minimizar o $-\log$ da verossimilhança (equivalente a maximizar o \log da verossimilhança), temos (desprezando as constantes)

$$\begin{aligned} \min\{-(\log\text{-verossimilhança})\} &\sim \min\left\{-\log\left(\prod_{i=1}^n e^{-\left(\frac{y_i-h(x_i)}{\sigma}\right)^2}\right)\right\} \\ &\sim \min\left\{\frac{1}{n} \sum_{i=1}^n [y_i - h(x_i)]^2\right\}. \end{aligned}$$

Tipos de função custo

Soma dos desvios absolutos (SDA)

- Definido como

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n |y_i - h(\mathbf{x}_i)|$$

- Temos

$$f(y|x) \sim e^{-|y-h(x)|} \text{ (Distribuição Laplace).}$$

- Mais robusta na presença de outliers.

Tipos de função custo

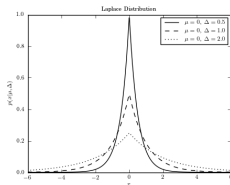
Soma dos desvios absolutos (SDA)

- Definido como

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n |y_i - h(x_i)|$$

- Temos

$$f(y|x) \sim e^{-|y-h(x)|} \text{ (Distribuição Laplace).}$$



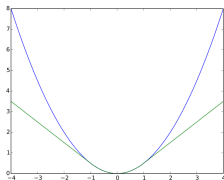
- Mais robusta na presença de outliers.

Huber-M cost

- Definido como

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n \begin{cases} \frac{1}{2} [y_i - h(\mathbf{x}_i)]^2, & \text{para } |y - h(\mathbf{x}_i)| \leq \delta, \\ \delta |y_i - h(\mathbf{x}_i)| - \frac{1}{2} \delta^2, & \text{caso contrário.} \end{cases}$$

- Abaixo, a curva azul representa a **SQD** e verde a **função Huber**.



- Ela combina as qualidades da **SQD** e **SQA**.
- O parâmetro δ é obtido automaticamente para um específico percentil dos erros absolutos.

- Fase de treino:
 - ★ Entra com os **dados** e os **verdadeiros níveis**;

- Fase de teste (ou aplicação do modelo):

- Fase de treino:
 - ★ Entra com os **dados** e os **verdadeiros níveis**;
 - ★ Obtém um “classificador”.
- Fase de teste (ou aplicação do modelo):

- Fase de treino:
 - ★ Entra com os **dados** e os **verdadeiros níveis**;
 - ★ Obtém um “classificador”.
- Fase de teste (ou aplicação do modelo):
 - ★ Entra com os **dados**;

- Fase de treino:
 - ★ Entra com os **dados** e os **verdadeiros níveis**;
 - ★ Obtém um “classificador”.

- Fase de teste (ou aplicação do modelo):
 - ★ Entra com os **dados**;
 - ★ Obtém seu nível (de acordo com o “classificador”).

- ATCCGTATAGTCGATCAGTCAGCTACTATGCGTAT **CANCER**
- TGCATGCATGCAGATCGATCGATCGCCAACGTAC **NO CANCER**
- ATTATATTCTGCGATCGAAGCTATGCGATCGTCGA **CANCER**
- TATGCGCGCGAGTTTTATGAGGCGATCGATGCTA **CANCER**
- ATCGCATCGACGTACGATGCTGATTATTATAGCCG **NO CANCER**
- GATCATGCTGCGAGAGGAGATTTTATGCGATAGA **CANCER**

...

ATCGTCTGATGCAGCGAGCTATGCGTACGTAGCA ?????

- Em inferência em geral **assume-se que o modelo é correto**.
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- P. 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 $h(x)$ que tenha **bom poder preditivo**.
- Não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas este, em geral, não é o foco das análises.

- Em inferência em geral **assume-se que o modelo é correto.**
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- P. 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 $h(x)$ que tenha **bom poder preditivo.**
- Não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas este, em geral, não é o foco das análises.

- Em inferência em geral **assume-se que o modelo é correto**.
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- *P. 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 $h(x)$ que tenha **bom poder preditivo**.
- Não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas este, em geral, não é o foco das análises.

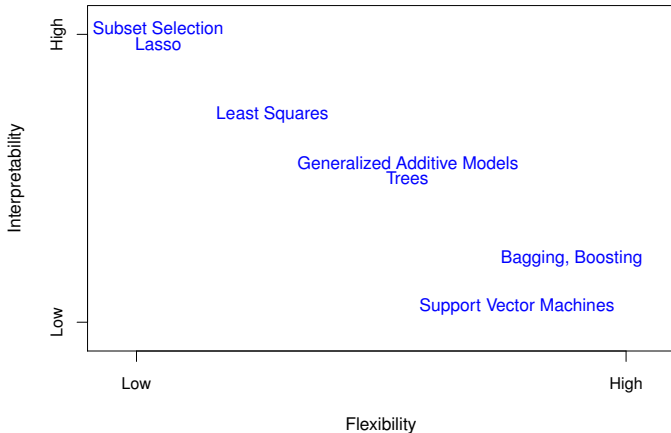
- Em inferência em geral **assume-se que o modelo é correto**.
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- P. ex., *quais parâmetros são significantes? Qual o efeito do aumento da dose do remédio no medicamento?* etc.
- Já em predição, nosso objetivo maior é simplesmente criar $h(x)$ que tenha **bom poder preditivo**.
- Não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas este, em geral, não é o foco das análises.

- Em inferência em geral **assume-se que o modelo é correto**.
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- P. 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 $h(\mathbf{x})$ que tenha **bom poder preditivo**.
- Não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas este, em geral, não é o foco das análises.

- Em inferência em geral **assume-se que o modelo é correto**.
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- P. 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 $h(\mathbf{x})$ que tenha **bom poder preditivo**.
- Não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas este, em geral, não é o foco das análises.

- Em inferência em geral **assume-se que o modelo é correto**.
- Isso ocorre pois o principal objetivo está na interpretação dos parâmetros.
- P. 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 $h(\mathbf{x})$ que tenha **bom poder preditivo**.
- Não estamos assumindo que a verdadeira regressão é linear!!
- Podemos continuar interpretando os resultados, mas este, em geral, não é o foco das análises.

Predição versus Inferência



- Duas culturas no uso de modelos estatísticos:
 - ★ **Data Modeling Culture:** Domina a comunidade estatística. Testar suposições é fundamental. Foco em inferência;
 - ★ **Algorithmic Modeling Culture:** Domina a comunidade de machine learning. O modelo é utilizado apenas para criar bons algoritmos preditivos.

L. Breiman: **Statistical modeling: The two cultures**. *Statistical Science*, **16(3):199-231, 2001** (disponível no material complementar).

- Duas culturas no uso de modelos estatísticos:
 - ★ **Data Modeling Culture:** Domina a comunidade estatística. Testar suposições é fundamental. Foco em inferência;
 - ★ **Algorithmic Modeling Culture:** Domina a comunidade de machine learning. O modelo é utilizado apenas para criar bons algoritmos preditivos.

Breve revisão de Álgebra Linear

- Um **Espaço Vetorial** consiste do seguinte:
 - 1 Um conjunto não vazio V de objetos, denominados vetores;
 - 2 Um corpo F (\mathbb{R} ou \mathbb{C}) de escalares.
 - 3 Uma **operação de adição de vetores**, que associa a cada par de elementos $u, v \in V$ um elemento $u + v \in V$, isto é, V é **fechado** com relação à operação de adição (relembre as 4 propriedades).
 - 4 Uma **operação de multiplicação por escalar**, que associa a cada elemento $u \in V$ e cada escalar $\alpha \in F$ um elemento $\alpha u \in V$, isto é, V é **fechado** com relação à operação de multiplicação por escalar.

Exemplo: O conjunto $\mathbb{R}_n = \{u = (x_1, \dots, x_n) / x_i \in \mathbb{R}\}$, conjunto de todas as n -uplas reais, com a operação de adição de elementos definida por:

$$u + v = (x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$$

e a operação de multiplicação por escalar definida por:

$$\lambda u = (\lambda x_1, \dots, \lambda x_n)$$

é um espaço vetorial real.

- Um **subespaço vetorial** de V é um subconjunto U de V que é ele mesmo um espaço vetorial sobre o corpo F com as operações de adição de vetores e multiplicação por escalar definidas em V .

Exemplo: O subconjunto S do \mathbb{R}^3 definido da forma:

$$S = \{w \in \mathbb{R}_3 / w = a(1, -1, 1) + b(2, 1, -1); a, b \in \mathbb{R}\},$$

é um subespaço vetorial de \mathbb{R}^3 .

- Sejam V um espaço vetorial sobre o corpo F e $v_1, \dots, v_n \in V$.
- Dizemos que o conjunto $S = \{v_1, \dots, v_n\} \subset V$ é **Linearmente Independente** (LI) sse toda combinação linear nula

$$\alpha_1 v_1 + \dots + \alpha_n v_n = 0_V; \quad \alpha_i \in F$$

implicar que $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$.

Exemplo: O conjunto $S = \{1, x, x^2\}$ é linearmente independente no espaço vetorial $P_3(\mathbb{R})$.

- Sejam V um espaço vetorial sobre o corpo F e $v_1, \dots, v_n \in V$.
- Dizemos que o conjunto $S = \{v_1, \dots, v_n\} \subset V$ é **Linearmente Dependente** (LD) sse é possível uma combinação linear nula

$$\alpha_1 v_1 + \dots + \alpha_n v_n = 0_V; \quad \alpha_i \in F$$

sem que os escalares $\alpha_1, \alpha_2, \dots, \alpha_n$ sejam todos nulos.

Exemplo: O conjunto $S = \{1, x, x^2, 2 - 3x + 2x^2\}$ é linearmente dependente no espaço vetorial $P_3(\mathbb{R})$. Por simplicidade, vamos denotar

$$p_1(x) = 1, p_2(x) = x, p_3(x) = x^2 \quad \text{e} \quad p_4(x) = 2 - 3x + 2x^2.$$

Podemos verificar facilmente que $p_4(x) = 2p_1(x) - 3p_2(x) + 2p_3(x)$.

- Seja V um espaço vetorial sobre o corpo F . Uma base de V é um conjunto linearmente independente de elementos de V que gera V .

Exemplo: Considere o espaço vetorial real \mathbb{R}^3 . O conjunto

$$\beta = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$$

é linearmente independente em \mathbb{R}^3 e gera o espaço \mathbb{R}^3 . Logo, β é uma base para \mathbb{R}^3 , denominada base canônica.

Exemplo: Considere o espaço vetorial real \mathbb{R}^2 . O conjunto

$$\Gamma = \{(1, 1), (-1, 1)\}$$

é linearmente independente em \mathbb{R}^2 e gera o espaço \mathbb{R}^2 . Logo, Γ é uma base para \mathbb{R}^2 .

- A dimensão de V é definida como sendo o número de elementos de uma base de V .

- Em geometria Euclidiana as propriedades que nos possibilitam expressar o comprimento de vetor e o ângulo entre dois vetores são denominadas de propriedades métricas;
- No estudo do \mathbb{R}^n , definimos comprimento de vetores e ângulo entre vetores através do produto escalar

$$x \cdot y = \sum_{i=1}^n x_i y_i \quad \text{para } x, y \in \mathbb{R}^n.$$

- Denotamos o **produto interno** entre dois elementos u e v de um espaço vetorial da seguinte forma: $\langle u, v \rangle$;
- Utilizando as propriedades de simetria, distributividade e homogeneidade tem-se
 - ★ $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ para todos $u, v, w \in V$;
 - ★ $\langle u, \lambda v \rangle = \lambda \langle u, v \rangle$ para todos $u, v \in V$ e $\lambda \in \mathbb{R}$.

- Seja V um espaço vetorial sobre o corpo F ;
- Uma norma, ou comprimento, em V é uma aplicação $\|\cdot\|$;
- Para cada elemento $u \in V$ associa um número real $\|u\|$, que possui as seguintes propriedades:
 - ★ **Positividade:** $\|u\| > 0$, para $u \neq 0_v$, com $\|u\| = 0 \iff u = 0_v$;
 - ★ **Homogeneidade:** $\|\lambda u\| = |\lambda| \|u\|$ para todo $u \in V, \lambda \in F$;
 - ★ **Desigualdade triangular:** $\|u + v\| \leq \|u\| + \|v\|, \forall u, v \in V$.
- Exemplos de normas ℓ^p :
 - ★ $\|u\|_1 = \sum_{i=1}^n |u_i|$;
 - ★ $\|u\|_2 = \left(\sum_{i=1}^n u_i^2 \right)^{1/2}$;
 - ★ $\|u\|_\infty = \max\{u_i ; 1 \leq i \leq n\}$.

- Seja V um espaço vetorial real munido do produto interno $\langle \cdot, \cdot \rangle$;
- O ângulo entre dois elementos não nulos $u, v \in V$ é definido como o valor $\theta \in [0, \pi]$ que satisfaz a equação

$$\cos(\theta) = \frac{\langle u, v \rangle}{\|u\|_2 \|v\|_2}.$$

- Assim, dizemos que os elementos $u, v \in V$ são ortogonais se, e somente se, $\langle u, v \rangle = 0$, e denotamos por $u \perp v$.

(Teorema de Pitágoras):

- Sejam V um espaço vetorial real munido do produto interno $\langle \cdot, \cdot \rangle$ e $\|\cdot\|_2$ a norma proveniente do produto interno;
- Então, os elementos $u, v \in V$ são ortogonais se, e somente se,

$$\|u + v\|_2^2 = \|u\|_2^2 + \|v\|_2^2.$$