

RECONHECIMENTO DE PADRÕES | RP

Prof. André Backes | @progdescomplicada

O que é?

- Ramo do aprendizado de máquina
- Consiste em atribuir um rótulo (ou classe) para uma certa amostra ou valor de entrada
- Exemplo: classificação
 - Podemos tentar determinar se um e-mail é "spam" ou "não-spam"
 - 2 classes

O que é?

- Só classificação?
 - Reconhecimento de padrões é algo mais geral do que apenas classificar amostras
- O que mais?
 - Regressão
 - *Sequence labeling*: usado para análise de fala
 - Análise da estrutura sintática
 - Etc

Para que serve?

- Em geral, visa fornecer uma resposta razoável para todas as entradas possíveis e realizar a correspondência "mais provável" das entradas, tendo em conta a sua variação estatística.
- É estudado em muitas áreas
 - psicologia, psiquiatria, fluxo de tráfego, ciência da computação, etc

O que é ser inteligente?

- Seria resolver um problema específico com exatidão?
 - Ter um mestre do xadrez ou médico especialista
- Resolver problemas genéricos de modo aproximado?
 - Determinar a vaga adequada no estacionamento
- Ter conhecimento Enciclopédico?
 - Saber tudo
- Tocar um instrumento? Falar outras línguas? Jogar bola bem?

RP e a inteligência

- Seres vivos são bastante habilidosos em reconhecer padrões
 - Comportamentais
 - Sonoros
 - Táteis
 - Visuais
 - Olfativos
 - Lógico Matemáticos

RP e a inteligência

- Reconhecer padrões equivale a classificar determinado objeto físico ou situação como pertencente ou não a um certo número de categorias previamente estabelecidas.
 - Dada uma amostra, a qual das classes conhecidas ela pertence?

Usando a intuição

- A qual grupo pertence este objeto?



Usando a intuição

- E agora?
- A qual grupo pertence este objeto?



Como o computador reconhece padrões?

- Seres humanos
 - A decisão de a qual grupo pertence é tomada com base no grau de similaridade entre a fruta desconhecida e os grupos de frutas conhecidas.
- Como o cérebro humano realiza esta tarefa?
 - comparação entre o objeto novo e objetos armazenados/conhecidos?

Como o computador reconhece padrões?

- Podemos “replicar” este processo em uma máquina. Para isso, precisamos
 - Representar os atributos físicos das frutas
 - Aprender o conceito laranja/maçã.
 - Armazenar as frutas aprendidas
 - De uma regra de decisão para classificar a nova fruta

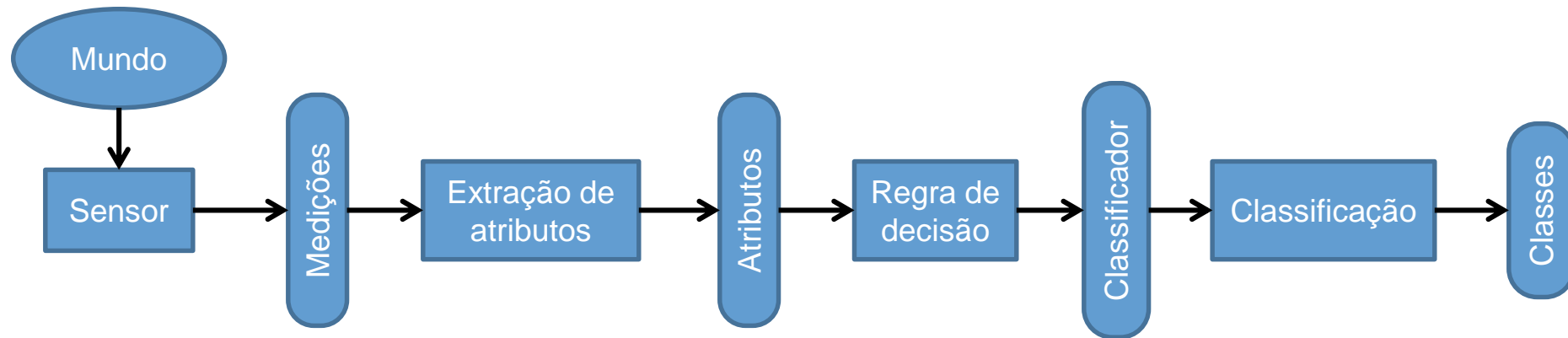
Como o computador reconhece padrões?

- Jornadas nas Estrelas Voyager
 - O médico da nave (IA holograma) tem que lidar com as questões éticas durante o processo de tomada de decisão.
 - Ele se torna obcecado com o fato de talvez ter tomado a decisão "errada".



<https://www.youtube.com/watch?v=FqvmoyLozzY>

Modelo básico de um sistema RP



Definição Formal de RP

- Para definir um problema de RP precisamos
 - Um número finito de K classes: C_1, C_2, \dots, C_K
 - Um número finito de N_i objetos por classe C_i
 - Um número finito de p atributos (*features*) para representar numericamente cada objeto físico.
 - Um mecanismo de memória e/ou aprendizado.
 - Uma regra de decisão para classificar novos objetos.
 - Um critério de avaliação do classificador.

Definições básicas

- Classe
 - Trata-se de um conjunto de objetos que compartilham um mesmo conjunto de características ou atributos comuns a todos.
- Objeto ou amostra
 - Trata-se de um único exemplar de uma classe conhecida ou não. É representado por um conjunto de atributos.

Definições básicas

- Atributo (*feature*)
 - É cada uma das características ou propriedades mais representativas de um determinado objeto
- Vetor de atributos ou *padrão* (X)
 - Em geral, mais de um atributo é necessário para descrever um objeto. Assim, é interessante considerar um conjunto de atributos organizados na forma de um vetor

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{bmatrix}$$

Definições básicas | atributo

- Tipos de atributos possíveis
 - Nominal
 - cor, identificação, profissão, ...
 - Ordinal
 - gosto (ruim, médio, bom), dias da semana , ...
 - Intervalar
 - temperatura em Celsius, ...
 - Racional
 - peso, tamanho, idade, temperatura em Kelvin, ...

Definições básicas | atributo

**Numérico
(Quantitativo)**

**Categórico
(Qualitativo)**

Tipo de Atributo	Descrição	Exemplos
Nominal	Valores são simplesmente nomes (símbolos) diferentes, i.e., atributos nominais provêm apenas informação suficiente para distinguir uma instância de outra: ($=$, \neq)	Sexo, Estado Civil, CEP, ...
Ordinal	Os valores de atributos ordinais provêm informação suficiente para distinguir e ordenar instâncias, i.e.: ($=$, \neq) e ($<$, $>$)	Grau de Educação, Números de Endereço, ...
Intervalo	Atributos para os quais a diferença entre valores faz sentido, i.e., existe uma unidade de medida com referência (zero) arbitrário. Suporta as operações anteriores e ainda (+, -)	Datas, Temperatura em Fahrenheit, ...
Razão	Atributos para os quais não apenas a diferença entre valores faz sentido, mas também a razão entre valores (zero é absoluto). Suporta as ops. anteriores e ainda (*, /)	Contagens, Massa, Largura, Corrente Elétrica, Quantidades Monetárias, ...

Definições básicas | atributo

- Pose-se estabelecer uma taxonomia independente para atributos pelo seu número de valores
 - Atributos Discretos
 - Atributos Contínuos

Definições básicas | atributo

- Atributos Discretos
 - assumem um número contável (enumerável) de valores
 - estações do ano, cores elementares, ...
 - n^0 de filhos, n^0 estrelas no universo, n^0 de anos
 - Caso especial: Atributos Binários
 - 0 ou 1
 - V ou F
 - etc

Definições básicas | atributo

- Atributos Contínuos
 - assumem uma quantidade incontável de valores
 - valores que são números reais
 - temperatura
 - peso
 - distância

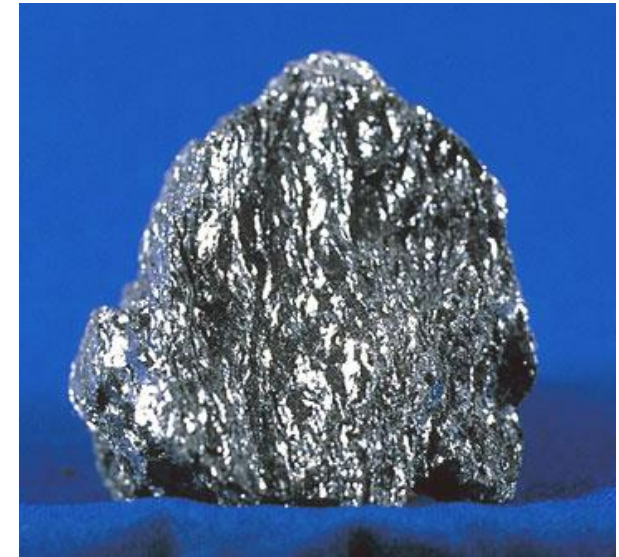
Definições básicas | atributo

- Quais atributos podemos usar para descrever a fruta ao lado?
 - Forma?
 - Cor?
 - Textura?
 - Cheiro?
 - Sabor?
 - Etc



Definições básicas | atributo

- Quais atributos podemos usar para descrever uma amostra de minério de ferro?
 - Teor de ferro?
 - Teor de umidade?
 - Granulometria?
 - Etc
- Quais atributos são mais importantes?
 - Difícil responder...



Definições básicas | atributo

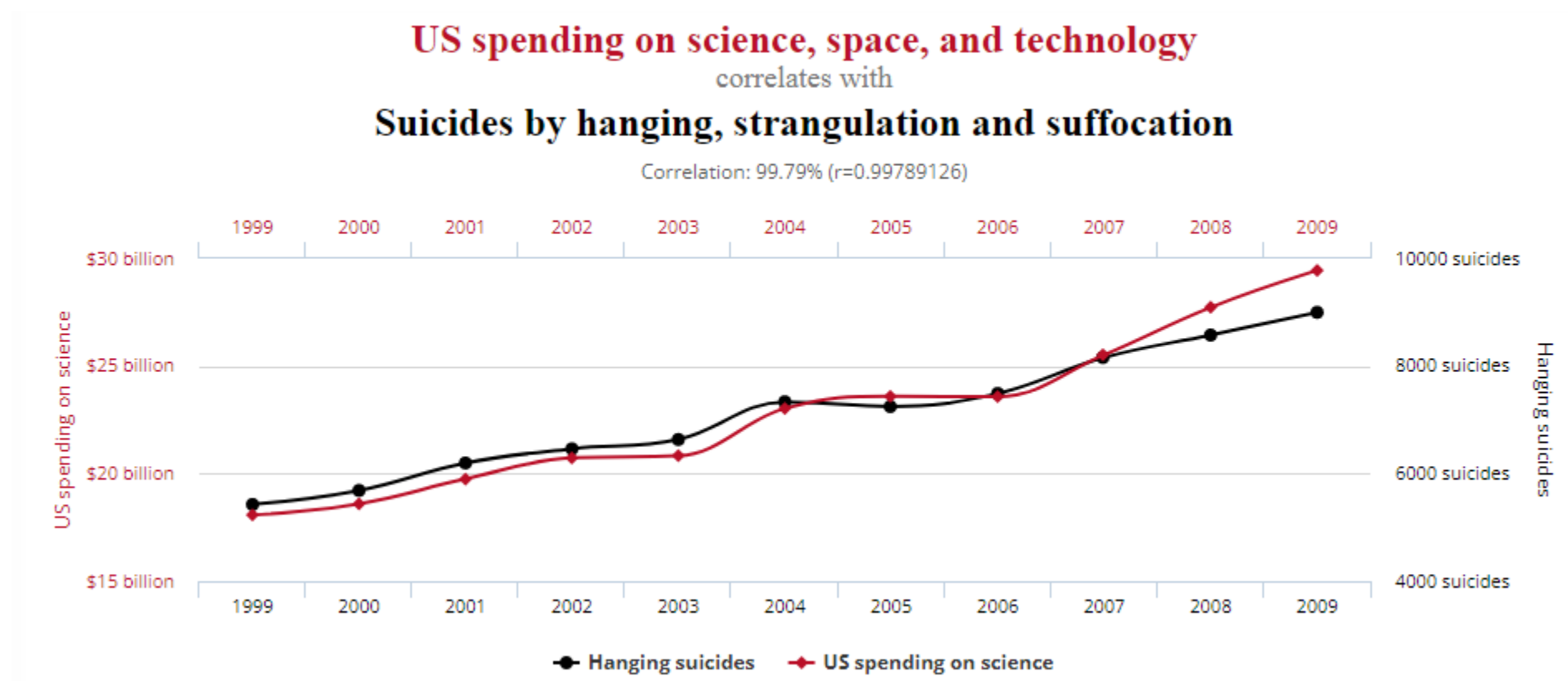
- Atributos inadequados
 - São atributos que não contém nenhuma informação relevante para a separação das classes, não importa o classificador usado.
 - Solução: definir novos atributos

Definições básicas | atributo

- Atributos correlacionados
 - Pode acontecer de 2 atributos distintos serem influenciados por um mecanismo comum de modo que variem juntos.
 - Isso pode degradar o desempenho do classificador
 - Exemplos:
 - raio e comprimento da circunferência
 - o preço de um produto e a quantidade de imposto pago por ele

Definições básicas | atributo

- Correlação não implica causalidade!



<https://www.tylervigen.com/spurious-correlations>

Definições básicas | atributo

- O computador entende apenas números!
 - Cada atributo deve ser representado por um valor numérico
 - Teor de ferro: 0,5 (50%)
 - Teor de umidade: 0,1 (10%)
 - Granulometria: 1,5 mm
- Vetor de atributos da amostra



$$X = \begin{bmatrix} 0,5 \\ 0,1 \\ 1,5 \end{bmatrix}$$

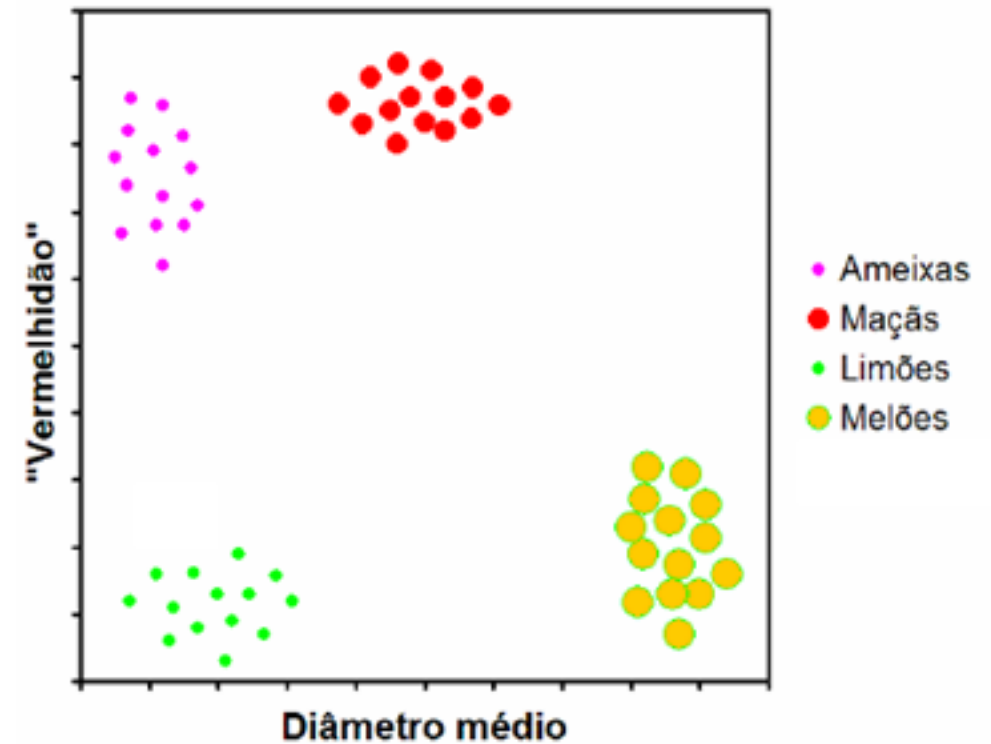
Classificação

- Consiste em tentar discriminar em diferentes classes um conjunto de objetos com características mensuráveis
 - Exemplo: classificação de frutas
 - Forma, cor, sabor, etc



Classificação

- Essas características, ou atributos, do objeto formam um espaço multidimensional
 - Espaço de características
- Cada objeto é então representado como sendo um ponto nesse espaço



Tipos de aprendizado

- Principais paradigmas de treinamento
 - Supervisionado
 - Semi-supervisionado
 - Não supervisionado
 - Reforço

Treinamento Supervisionado

- Consiste em apresentar um padrão a ser reconhecido juntamente com a resposta que o sistema deve fornecer ao deparar-se novamente com o esse padrão

Treinamento Supervisionado

- Guiado por um “professor” externo que possui conhecimento sobre o ambiente
 - Parâmetros do sistema são ajustados por apresentações sucessivas do padrão de modo a reproduzir comportamento do “professor”



Treinamento Supervisionado

- Métodos Paramétricos
 - Assumem que a distribuição dos dados é conhecida
 - distribuição normal, por exemplo
 - Em muitos casos não se tem conhecimento da distribuição
- Métodos Não-Paramétricos
 - Não consideram essa hipótese
 - Um exemplo é o **k-NN** (*k Nearest Neighbor*)

Treinamento por Reforço

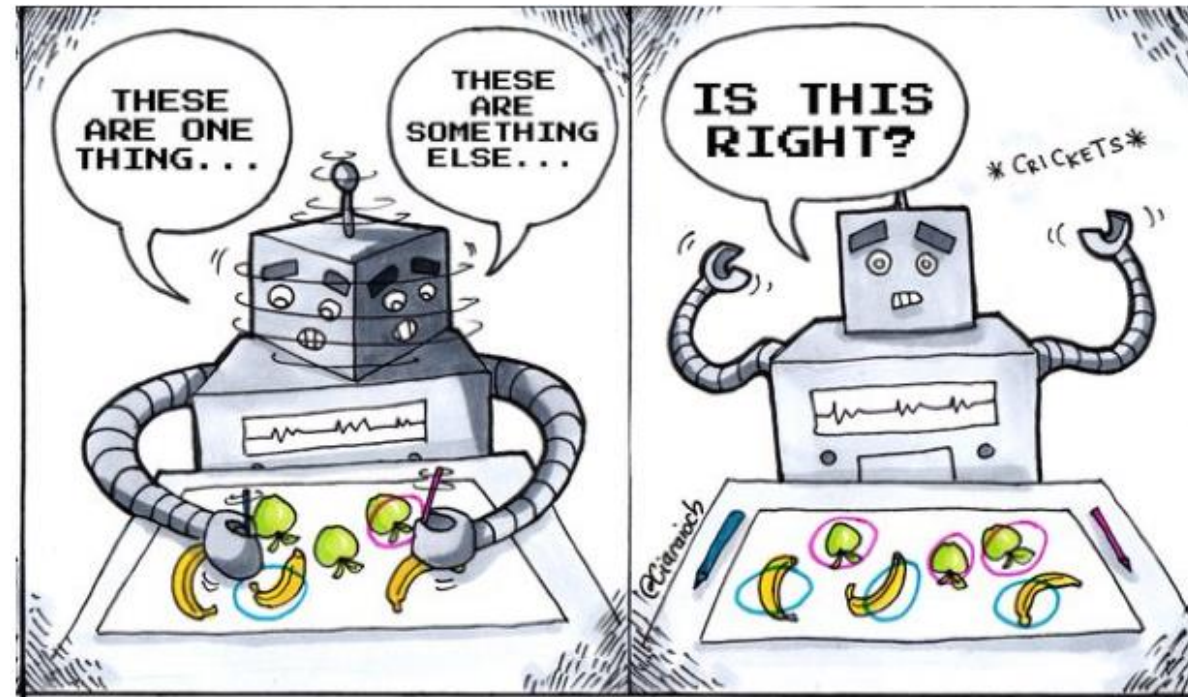
- Tipo de treinamento intermediário entre o supervisionado e não supervisionado
 - Processo de tentativa e erro que procura maximizar sinal de reforço
 - Exemplo: agente jogador de damas, onde o sistema é reforçado de acordo com o número de peças capturadas ou perdidas

Treinamento por Reforço

- Guiado por um “crítico” externo
 - Se ação tomada por sistema é seguida por estado satisfatório, sistema é fortalecido, caso contrário, sistema é enfraquecido (lei de Thorndike)
 - Tipos de reforço:
 - Positivo (recompensa)
 - Negativo (punição)
 - Nulo

Treinamento Não Supervisionado

- Não tem “crítico” ou “professor” externo, apenas os dados de entrada
- Tem-se um conjunto de exemplos mas não se conhece as categorias envolvidas
- Busca extrair as propriedades estatisticamente relevantes



Treinamento Não Supervisionado

- Exemplos: *Clustering*
 - Organização dos objetos similares (em algum aspecto) em grupos
 - Descobre categorias automaticamente
- Quantização: atribui valores discretos para um atributo que aceita infinitos valores

Treinamento Semi-Supervisionado

- Combina uma pequena quantidade de amostras classificadas com um grande número de amostras não classificadas para produzir melhores classificadores
- Tem um “professor” externo apenas para parte dos exemplos de treinamento
 - Exemplo: busca por páginas de internet similares.

Função Alvo

- Trata-se da função objetivo.
 - Ela estabelece qual conhecimento será aprendido
- Permite também verificar quão bem o conhecimento foi aprendido
 - Função discriminante entre classes
 - Função de similaridade intra grupos
 - etc

Teste diagnóstico

- Utilizado para a verificação da ocorrência ou não ocorrência de uma doença
 - Compara o resultado de um teste com um padrão de referência (padrão ouro)

True negative



False positive



False negative



True positive



Teste diagnóstico

- Indicadores que podem ser calculados a partir dos resultados da tabela
 - Classificação correta (acurácia):
 - $(VP+VN)/(VP+FP+FN+VN)$
 - Classificação incorreta:
 - $(FP+FN)/(VP+FP+FN+VN)$

Teste diagnóstico

- Indicadores que podem ser calculados a partir dos resultados da tabela
 - Sensibilidade: capacidade de detectar corretamente os doentes
 - $VP/(VP+FN)$
 - Especificidade: capacidade de detectar corretamente os indivíduos sadios
 - $VN/(VN+FP)$

Revisão de estatística

- Nos próximos slides são apresentadas algumas definições estatísticas muito comuns no trato com dados multivariados.
- Essas estatísticas se aplicam, de modo geral, a cada atributo do nosso vetor de atributos.

Revisão de estatística

- Amplitude Total

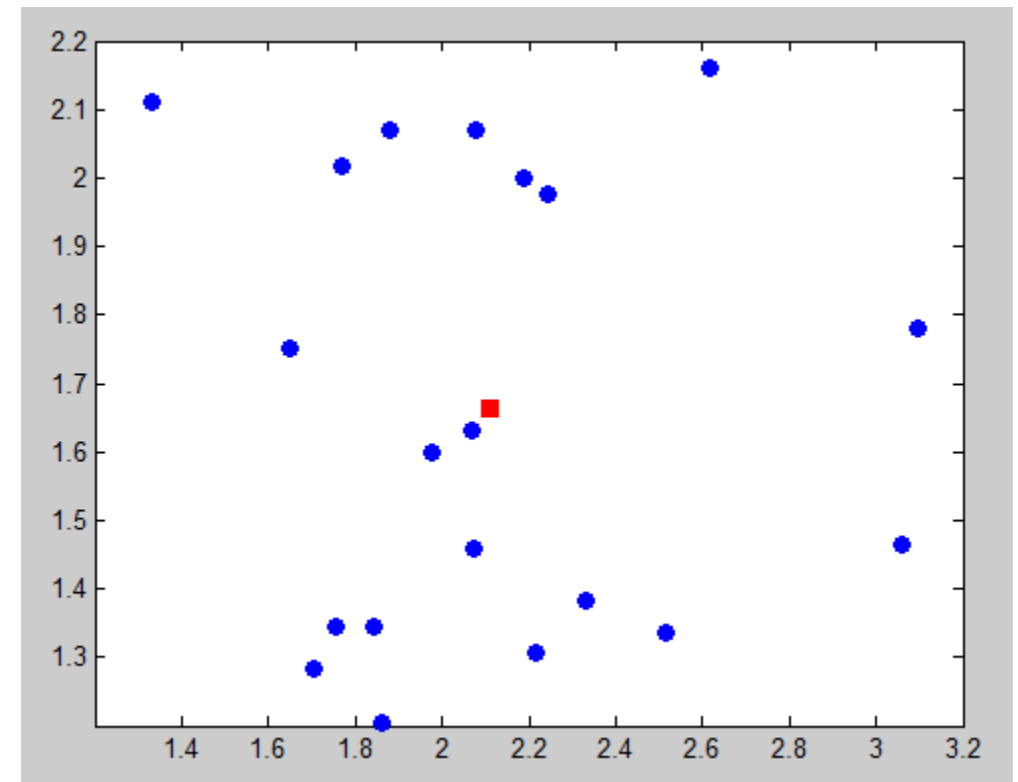
- Trata-se da dispersão entre o maior e o menor valor de um determinado atributo.

$$R = \max_j X_i(j) - \min_j X_i(j)$$

- Exemplo: para um atributo “idade” temos o seguinte conjunto de valores:
 - 20, 25, 27, 28, 40, 30, 31 e 19
 - $R = 40 - 19 = 21$

Revisão de estatística

- Média ou esperança
 - É o valor que aponta para onde mais se concentram os dados de uma distribuição
 - Pode também ser chamado de centróide



Revisão de estatística

- Média ou esperança
 - A média aritmética é a forma mais simples de calcular uma média

$$\mu = E[X] = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \dots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_p \end{bmatrix}$$

$$\mu_i = \frac{1}{N} \sum_{j=1}^N X_i(j)$$

Revisão de estatística

- Muito cuidado ao calcular a média, pois o resultado obtido pode não refletir a realidade
 - Exemplo: uma pesquisa feita em uma população de 8 domicílios em duas regiões (A e B)

Local	B	A	B	B	B	A	A	B
Renda	13	17	6	5	10	12	19	6

Média	11
Média A	16
Média B	8

Revisão de estatística

- Mediana
 - Dado um conjunto de dados organizados em ordem crescente, a mediana é o valor que ocupa a posição central do conjunto.
 - Dado o conjunto {2, 2, 3, 5, 5, 6, 7, 7, 9, 9, 10}
 - A mediana \tilde{x} será igual a 6

Revisão de estatística

- Mediana

- Se a quantidade de valores é ímpar, a mediana, será simplesmente o valor central.
- Se a quantidade de valores é par, a mediana será a média dos dois valores centrais
 - Dado o conjunto $\{0, 1, 1, 2, 3, 4, 5, 5, 6, 6, 7, 8\}$, a mediana \tilde{x} será igual a $(4+5)/2 = 4,5$.

Revisão de estatística

- Mediana é muito mais resistente a dados espúrios
 - Para o conjunto {2, 2, 3, 5, 5, 6, 7, 7, 9, 9, 10}
 - Média: 5.91
 - Mediana: 6
 - Para o conjunto {2, 2, 3, 5, 5, 6, 7, 7, 9, 9, 10, 1000}
 - Média: 88.75
 - Mediana: 6,50

Revisão de estatística

- Moda
 - Dado um conjunto de dados, a moda é o valor com maior frequência individual, ou seja, aquele que mais se repete dentro do conjunto de dados
 - Dado o conjunto $\{0, 1, 2, 2, 2, 3, 4, 5, 5, 6, 6, 7, 8\}$, a moda será igual a 2

Revisão de estatística

- Variância
 - Trata-se de uma medida da dispersão estatística de um atributo
 - É uma medida unidimensional. Não leva em consideração as outras dimensões
 - Indica quão longe, em geral, os valores se encontram da média.
- Para o seu cálculo deve-se ter em mente a natureza dos dados estudados

$$\sigma^2 = \text{var}(X) = E[(X - \mu)^2]$$

Revisão de estatística

- Os dados estudados podem constituir uma *população* ou uma *amostra*
 - *Variância da População*: nosso conjunto de dados observados representa todos os elementos existentes na população
 - *Variância da Amostra*: nosso conjunto de dados observados representa **apenas** uma amostra de todos os elementos existentes na população

Revisão de estatística

- Variância da População

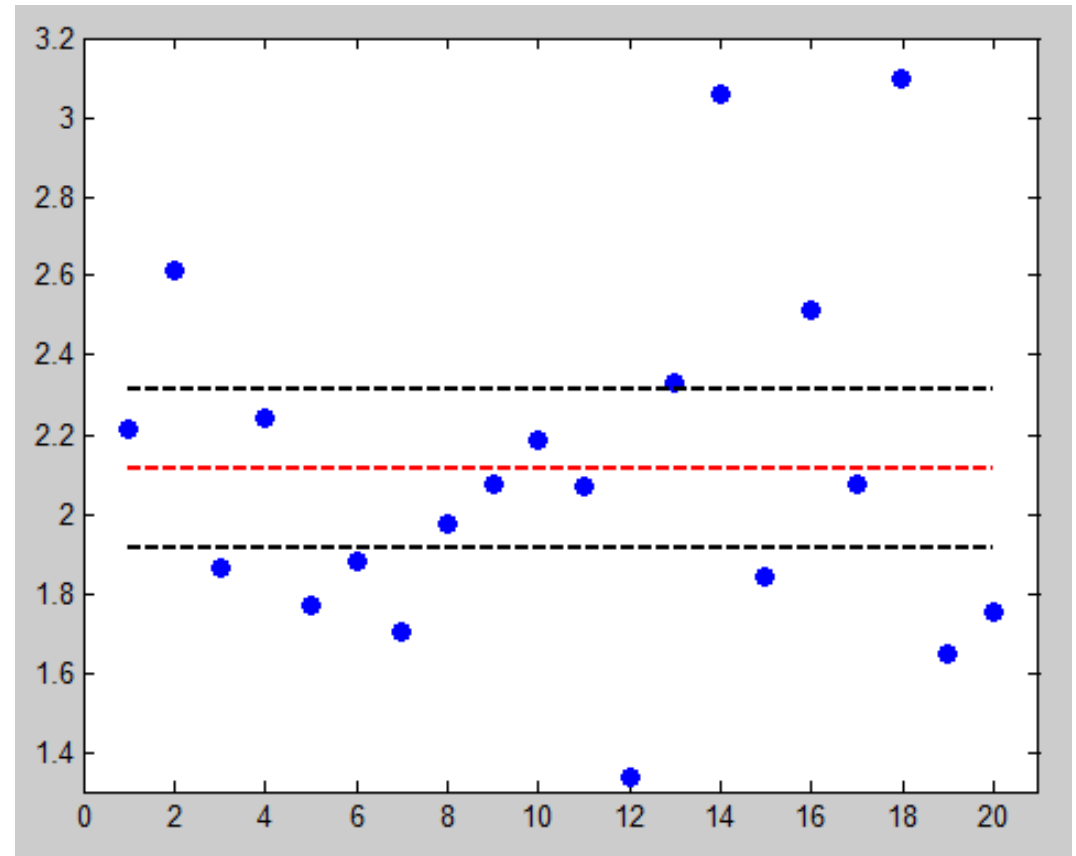
$$\sigma_i^2 = \frac{1}{N} \sum_{j=1}^N (X_i(j) - \mu_i)^2$$

- Variância da Amostra

$$\sigma_i^2 = \frac{1}{N-1} \sum_{j=1}^N (X_i(j) - \mu_i)^2$$

Revisão de estatística

- Média e variância



Revisão de estatística

- Desvio Padrão
 - É dado pela raiz quadrada da variância

$$\sigma = \sqrt{\text{var}(X)} = \sqrt{E[(X - \mu)^2]}$$

$$\sigma_i = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \mu_i)^2}$$

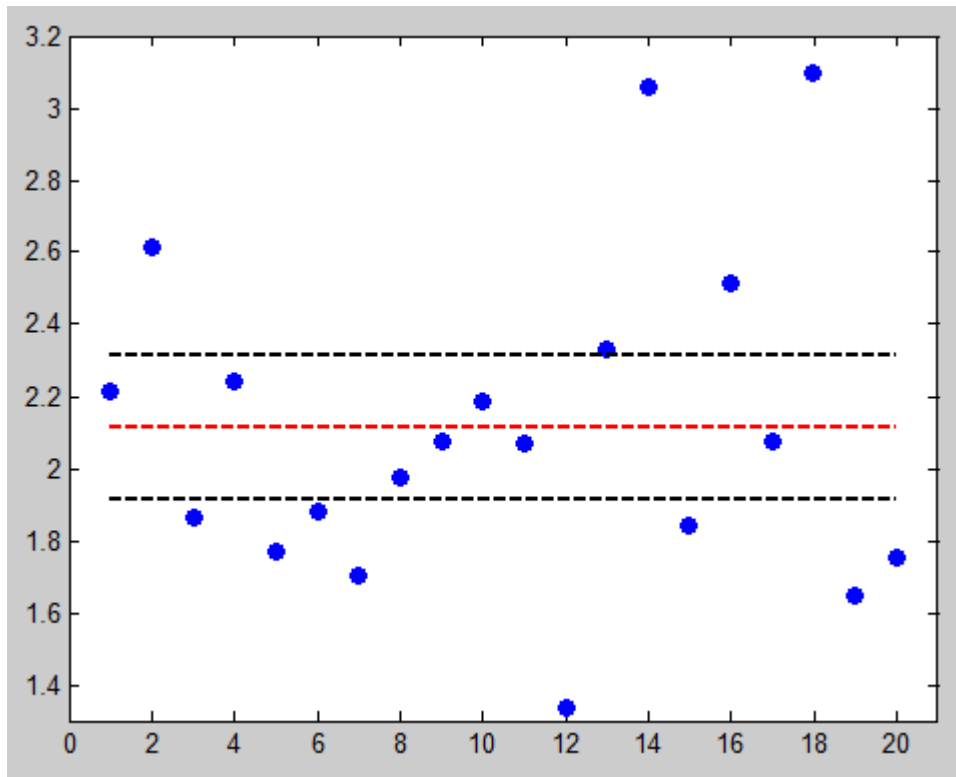
- Como no cálculo da variância, deve-se ter em mente a natureza dos dados estudados
 - População ou amostra

Revisão de estatística

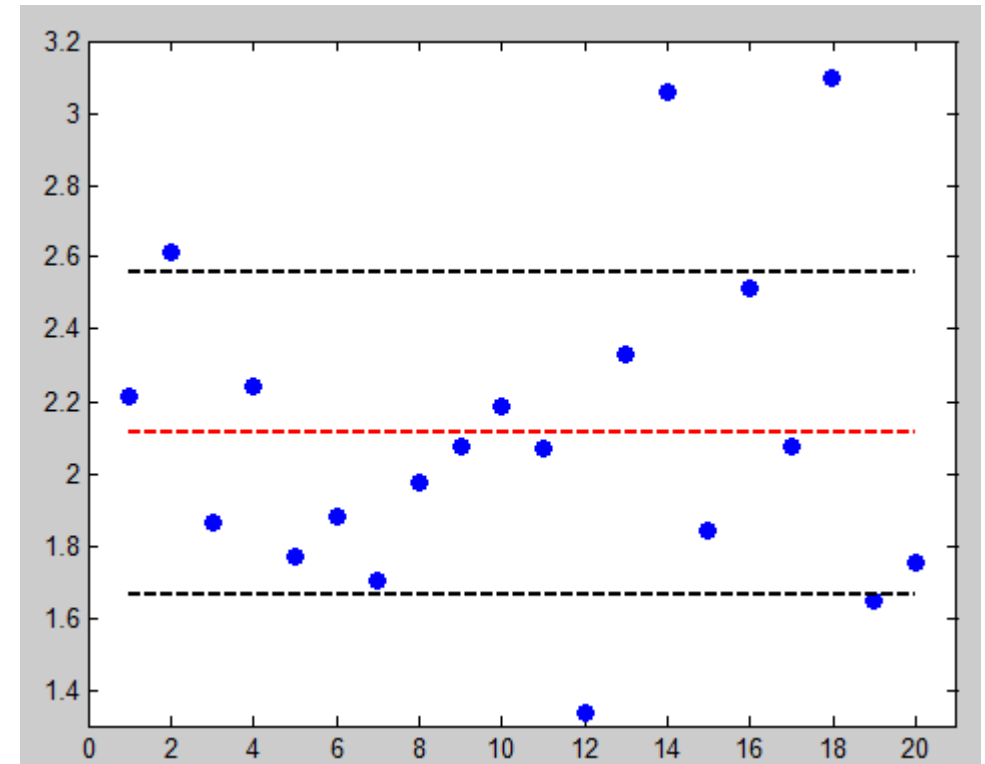
- Qual a diferença entre Variância e Desvio Padrão?
 - O desvio padrão é, em geral, mais útil para descrever a variabilidade dos dados e tem a conveniência de ser expresso nas mesmas unidades do dado original
 - O desvio padrão é utilizado para normalização antes dos testes estatísticos
 - A variância é geralmente mais útil matematicamente

Revisão de estatística

Variância



Desvio Padrão



Revisão de estatística

- Coeficiente de Variação
 - É uma medida de dispersão empregada para estimar a precisão de experimentos
 - É definido como a razão entre o desvio padrão e a média

$$CV = \frac{\sigma}{\mu}$$

Revisão de estatística

- O coeficiente de variação permite representar o desvio-padrão como uma porcentagem da média.
 - Capacidade de comparar distribuições que apresentem diferentes médias e desvios

Revisão de estatística

- Covariância

- Trata-se de uma medida da dispersão estatística de entre dois atributos
 - É uma medida bidimensional.
 - Verifica a dispersão de duas variáveis aleatórias
- Permite a nós medir o grau de relacionamento linear entre duas variáveis aleatórias (atributos) X_i e X_j

$$\sigma_{ij} = cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] =$$

$$\frac{1}{N-1} \sum_{k=1}^N (X_i(k) - \mu_i)(X_j(k) - \mu_j)$$

Revisão de estatística

- Matriz de Covariâncias
 - Trata-se de uma matriz simétrica que sumariza a covariância entre p atributos.

$$\text{cov}(X) = \Sigma_{p \times p} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

- Exemplo:

$$\Sigma_{2 \times 2} = \begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix}$$

Revisão de estatística

- Matriz de Covariâncias
 - Note que temos as variâncias dos atributos ao longo da diagonal principal e das covariâncias entre cada par de variáveis nas outras posições da matriz

$$\sigma_{ii} = \sigma_i^2$$

$$\sigma_{ij} = \sigma_{ji}$$

Revisão de estatística

- Calcular a matriz de covariâncias tem um custo alto
 - Devemos buscar alternativas para ter uma boa estimativa utilizando menos dados

Revisão de estatística

- Podemos assumir que os atributos são estatisticamente independentes

$$\Sigma_{p \times p} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_p^2 \end{bmatrix}$$

- Assumir que a matriz de covariâncias é a mesma para todas as classes
 - Abordagem utilizada pelo classificador *Naive-Bayes*

Revisão de estatística

- Variância total
 - É representada pela soma de todos os atributos envolvidos no vetor X .
 - Altos valores de variância indicam uma maior dispersão dos atributos
 - O traço da matriz de covariâncias é uma forma sintetizada da variância total

$$\text{traço}(\Sigma_{p \times p}) = \text{tr}(\Sigma_{p \times p}) = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}$$

Revisão de estatística

- Variância generalizada
 - É representada pelo determinante da matriz de covariâncias
 - Como na variância, a raiz da variância generalizada é o desvio-padrão generalizada

Revisão de estatística

- Coeficiente de Correlação
 - Também chamada de correlação, indica a força e a direção do relacionamento linear entre dois atributos
 - Trata-se de uma medida da relação entre dois atributos, embora correlação não implique causalidade
 - Duas variáveis podem estar altamente correlacionadas e não existir relação de causa e efeito entre elas

Revisão de estatística

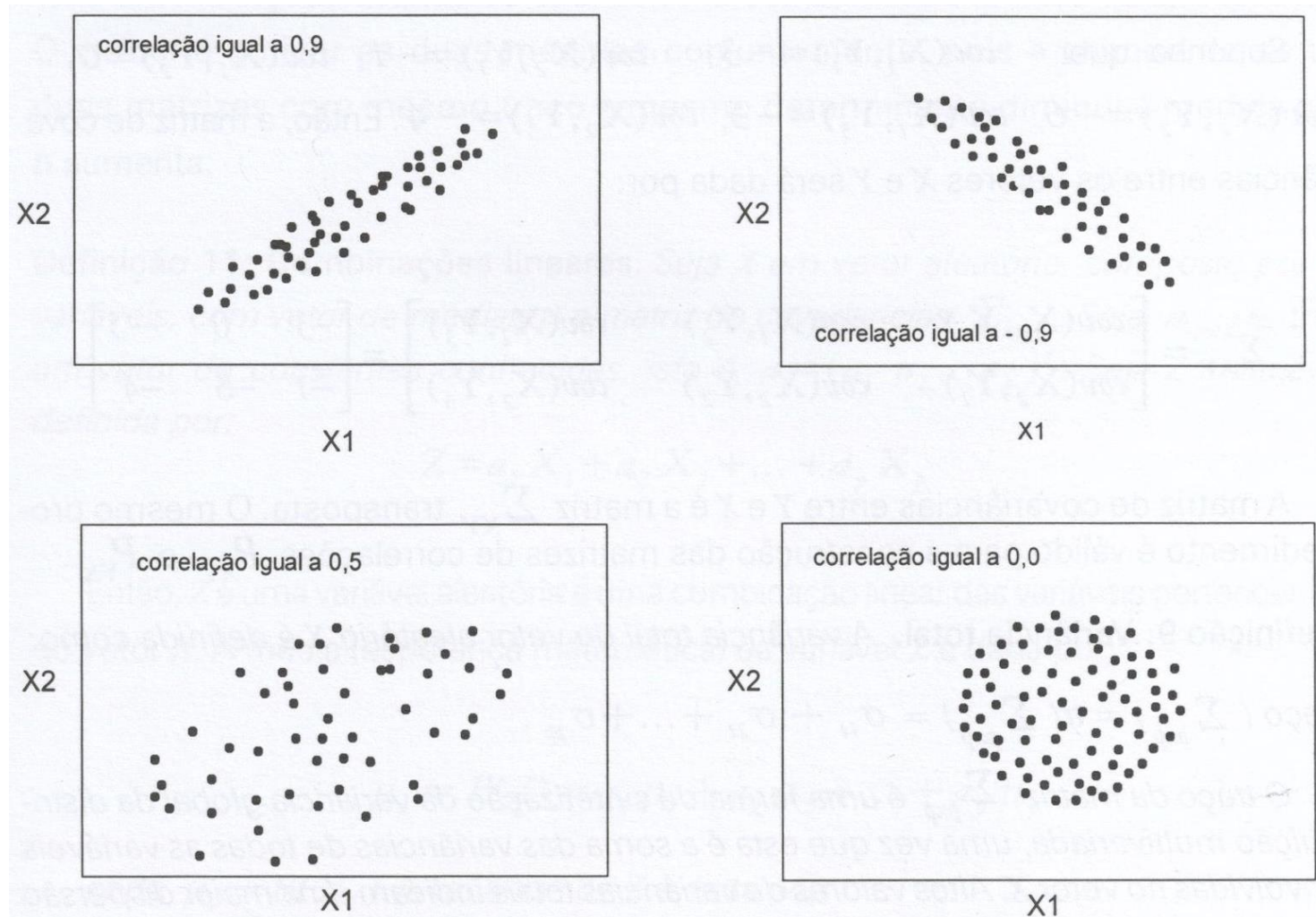
- Coeficiente de Correlação

- Existem vários coeficientes medindo o grau de correlação, adaptados à natureza dos dados.
- O coeficiente de correlação de Pearson ou "coeficiente de correlação produto-momento" é um dos mais utilizados

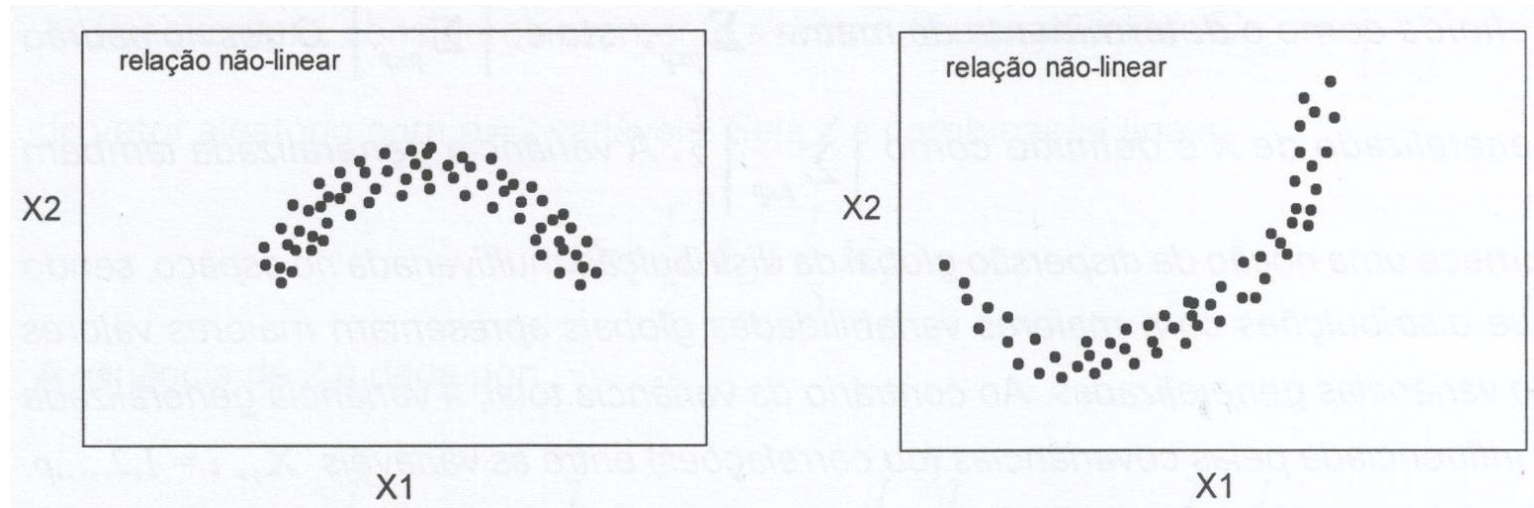
$$p_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} = \frac{\sigma_{ij}}{\sigma_i\sigma_j}$$

- O estudo de correlação pressupõe que os dois atributos tenham distribuição normal

Revisão de estatística

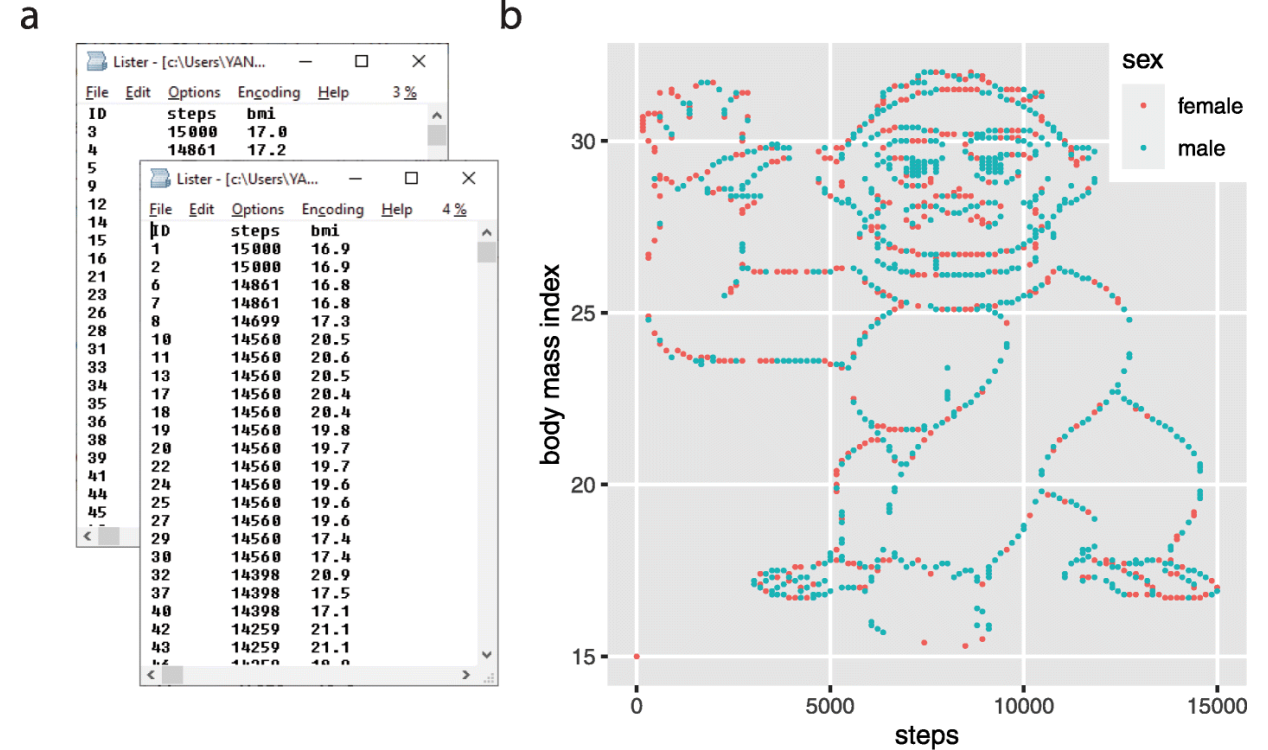


Revisão de estatística



Revisão de estatística

- O experimento do gorila
 - Sofremos de atenção seletiva
 - O foco numa hipótese pode nos cegar para algumas descobertas



c

	Gorilla <u>not</u> discovered	Gorilla discovered
Hypothesis-focused	14	5
Hypothesis-free	5	9

Revisão de estatística

- Matriz de Correlação

- Trata-se de uma matriz simétrica que sumariza a covariância entre p atributos.

$$P_{p \times p} = \begin{bmatrix} 1 & p_{12} & \dots & p_{1p} \\ p_{21} & 1 & \dots & p_{2p} \\ \dots & \dots & \dots & \dots \\ p_{p1} & p_{p2} & \dots & 1 \end{bmatrix}$$

- Exemplo:

$$P_{2 \times 2} = \begin{bmatrix} 1 & -0,3162 \\ -0,3162 & 1 \end{bmatrix}$$

Revisão de estatística

- Combinação linear
 - Trata-se de uma ferramenta importante na análise de dados multivariados, pois permite sumarizar as informações dos p -atributos originais
 - Um atributo Z é combinação linear de outros p atributos se existe um conjunto de escalares a_1, a_2, \dots, a_p , tal que

$$Z = a_1X_1 + a_2X_2 + \dots + a_pX_p$$

Agradecimentos

- Agradeço ao professor Guilherme de Alencar Barreto da Universidade Federal do Ceará (UFC) pelo material disponibilizado