

RECONHECIMENTO DE PADRÕES - RP

Definições Básicas

O que é?

2

- 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?

3

- 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?

4

- 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?

5

- 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

6

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

RP e a inteligência

7

- 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

8

- A qual grupo pertence este objeto?



Usando a intuição

9

- E agora? A qual grupo pertence este objeto?



Como o computador reconhece padrões?

10

- 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?

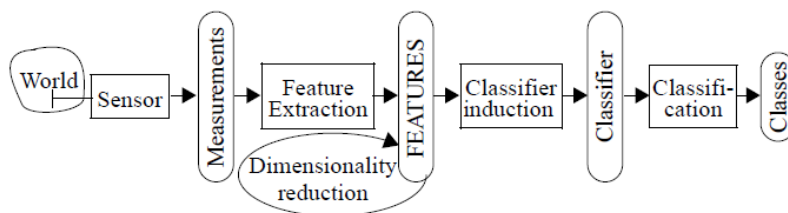
Como o computador reconhece padrões?

11

- 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

Modelo básico de um sistema de reconhecimento de padrões?

12



Definição Formal de RP

13

- 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

14

- 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

15

- 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

16

- 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

17

**Catagórico
(Qualitativo)**

**Numérico
(Quantitativo)**

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

18

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

Definições básicas - atributo

19

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

Definições básicas - atributo

20

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

Definições básicas - atributo

21

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



Definições básicas - atributo

22

- 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

23

- 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

24

- 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

25

- 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

26

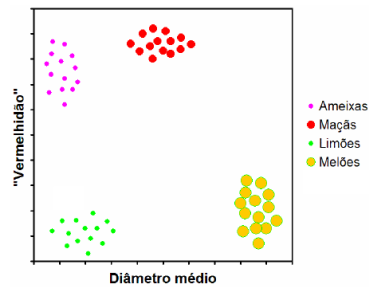
- Consiste em tentar discriminar em diferentes classes um conjunto de objetos com características mensuráveis
 - ▣ Exemplo: classificação de frutas
 - Forma, cor, sabor, etc



Classificação

27

- 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

28

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

Tipos de aprendizado

29

- 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
 - ▣ 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”

Tipos de aprendizado

30

- 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*)

Tipos de aprendizado

31

- 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

Tipos de aprendizado

32

- 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) e Nulo

Tipos de aprendizado

33

- 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

Tipos de aprendizado

34

- 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

Tipos de aprendizado

35

- 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

36

- 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

Revisão de estatística

37

- 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

38

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

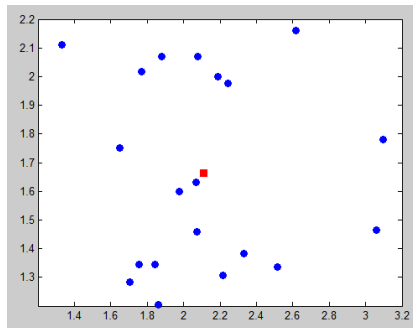
$$R = \max_j X_i(j) - \min_j X_i(j)$$

- ▣ 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

39

- 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

40

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

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

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

Revisão de estatística

41

□ 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

42

□ 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

43

□ 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

44

□ 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.

$$\sigma^2 = \text{var}(X) = E[(X - \mu)^2]$$
- ▣ Para o seu cálculo deve-se ter em mente a natureza dos dados estudados

Revisão de estatística

45

- 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

46

- Variância da População

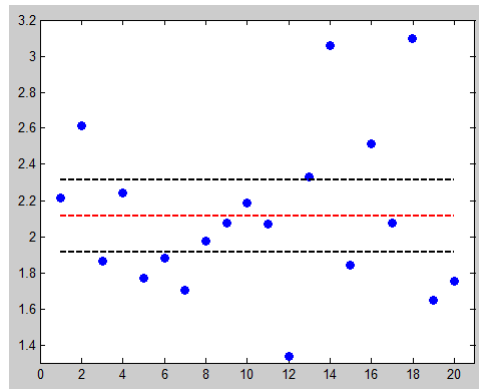
$$\sigma_i^2 = \frac{1}{N} \sum_{j=1}^N (X_i(j) - \mu_i)^2$$

- Variância da Amostra

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

Revisão de estatística

47



Revisão de estatística

48

- 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

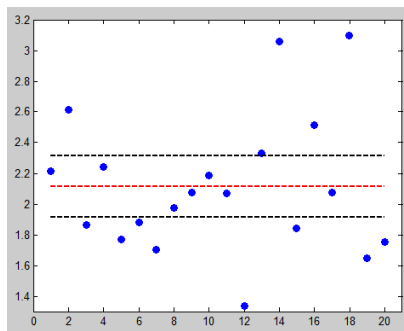
49

- 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

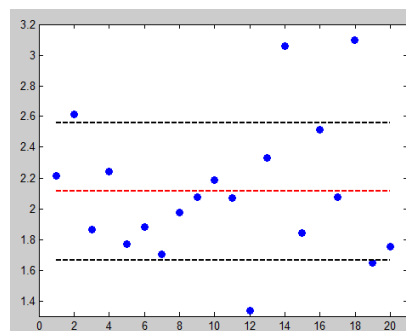
Revisão de estatística

50

Variância



Desvio Padrão



Revisão de estatística

51

- 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

52

- 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

53

□ 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

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

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

Revisão de estatística

54

□ 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

55

□ 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

56

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

Revisão de estatística

57

- 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

58

- 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

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

Revisão de estatística

59

- 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

60

- 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

61

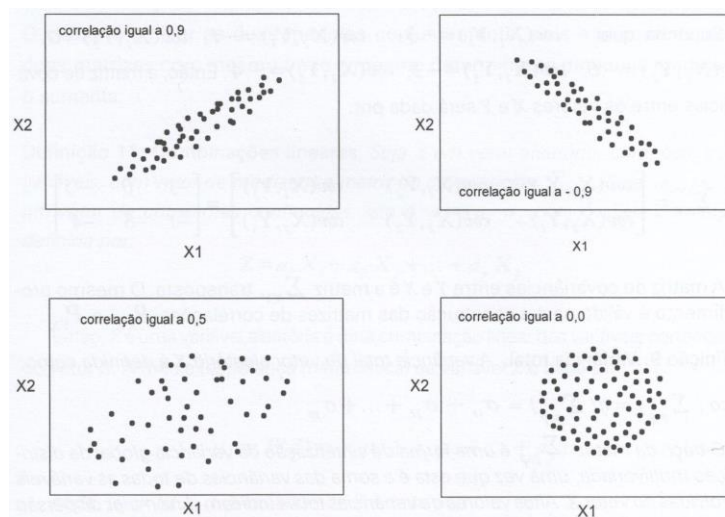
- Coeficiente de Correlação
 - ▣ Existem vários coeficientes medindo o grau de correlação, adaptados à natureza dos dados.
 - ▣ O coeficiente de correlação de Pearson ou "coeficiente de correlação produto-momento" é um dos mais utilizados

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

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

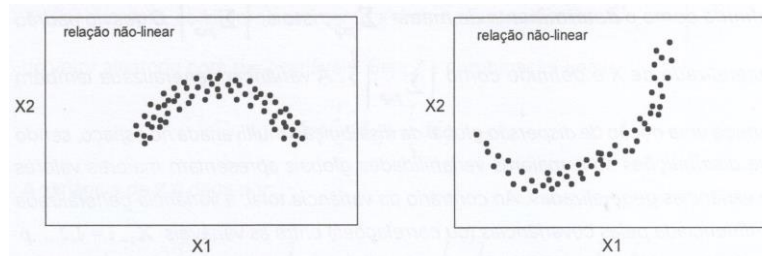
Revisão de estatística

62



Revisão de estatística

63



Revisão de estatística

64

- 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

65

- 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

66

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