

MINERAÇÃO MASSIVA DE DADOS

Parte 3 – Introdução à Álgebra Linear e Estatística

Thelmo de Araújo

thelmo@larces.uece.br

Mestrado Acadêmico em Ciência da Computação (MACC)

Universidade Estadual do Ceará (UECE)

Laboratório de Matemática Computacional (LAMAC)

Mineração Massiva de Dados

Aula 1

Thelmo de Araujo

LABORATÓRIO DE MATEMÁTICA COMPUTACIONAL

Semestre 2018-2

Uma *imagem digital monocromática* pode ser descrita por uma matriz $M \times N$ cujos elementos $f(x, y)$ assumem valores inteiros no intervalo $[0, L - 1]$.

Em geral, $L = 2^b$, para algum inteiro positivo b , chamado de *profundidade* da imagem.

Por exemplo, se $b = 8$, então $L = 256$ e a imagem é dita de profundidade 8.

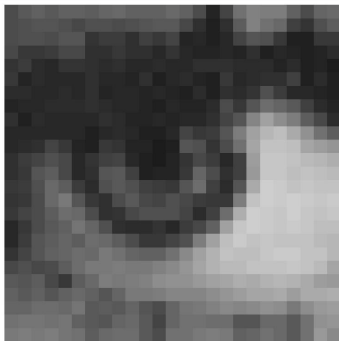


Figura: Imagem original 512×512 e ampliação de uma região de 25×25 pixels.

78	89	96	91	98	99	93	100	107	97	98	98	104	96	75	37	70	112	129	127	124	96	77	103	105
76	81	85	87	79	74	79	81	78	80	84	89	88	73	48	34	50	92	128	114	105	64	52	82	94
77	80	82	78	84	83	54	59	61	49	66	58	67	36	41	31	41	69	85	80	58	34	33	57	55
71	58	56	59	72	72	50	50	56	50	48	41	54	39	38	36	39	38	50	32	42	35	33	36	44
51	61	44	42	51	53	49	45	48	42	39	43	42	40	49	47	47	46	46	43	37	44	33	38	30
56	38	45	39	42	40	48	41	41	37	48	33	39	46	35	31	55	41	44	50	49	64	31	45	33
44	45	42	42	46	41	47	38	40	41	35	39	37	34	43	31	50	45	54	80	60	40	42	38	32
35	42	42	43	45	44	48	44	41	39	37	32	39	38	38	43	75	57	86	109	100	88	64	46	35
48	34	43	42	39	42	42	48	43	35	36	41	37	39	42	48	69	86	132	163	142	127	95	72	44
42	40	39	42	40	41	31	41	43	42	34	34	41	68	55	60	84	124	161	184	192	170	144	113	93
40	44	48	67	55	35	33	57	52	38	37	30	32	85	82	54	76	114	167	193	194	189	166	156	148
50	59	70	80	73	39	53	66	68	41	34	29	38	51	72	62	42	64	158	198	199	193	186	181	175
48	57	83	92	71	41	48	87	91	66	34	36	35	58	103	75	47	64	152	200	197	189	195	192	186
55	61	79	104	93	62	45	81	90	79	63	62	57	83	108	64	46	84	171	206	201	195	200	196	193
53	57	78	104	120	84	57	50	76	89	73	68	79	73	61	55	58	124	194	206	202	205	204	198	201
57	67	77	92	110	105	78	49	53	72	81	81	79	69	46	63	108	173	205	201	199	206	204	194	198
40	55	77	90	99	118	103	82	50	45	44	51	42	62	74	96	160	199	206	202	199	203	198	192	191
39	59	86	95	100	95	108	104	92	75	56	58	67	82	126	160	189	201	194	198	193	193	194	187	181
53	68	87	96	100	108	105	121	114	100	98	98	112	137	160	179	195	191	192	199	202	203	194	194	184
74	79	72	77	95	111	120	123	126	123	121	130	137	149	167	179	187	188	194	193	194	197	195	185	165
90	85	95	71	55	92	104	125	120	125	130	143	146	145	154	161	165	174	171	177	184	177	171	156	152
96	89	97	87	99	104	98	83	101	121	122	126	135	135	135	141	149	158	164	167	169	157	170	169	174
102	101	110	111	106	118	90	93	88	101	86	76	103	118	120	129	138	135	138	145	143	123	142	155	157
112	117	117	115	116	99	88	87	103	107	90	68	101	114	112	111	108	85	86	105	105	88	112	146	150
120	127	121	126	117	113	92	107	108	109	91	76	112	114	122	125	123	99	100	114	105	90	114	154	139

Figura: Níveis de intensidade da região de 25×25 pixels da imagem original.

Representação Vetorial de uma Imagem

Além do formato matricial, podemos tratar uma imagem $M \times N$ como um vetor com MN componentes, concatenando suas colunas.

Assim, as primeiras N componentes do vetor correspondem à primeira coluna da imagem, as próximas N componentes correspondem à segunda coluna da imagem, e assim sucessivamente.

Esse processo nos permite considerar imagens como vetores em um espaço vetorial, que sabemos manipular por meio das técnicas de álgebra linear.

Entretanto, isso nos leva a espaços de dimensões muito elevadas, por exemplo, uma imagem 512×512 é um vetor no \mathbb{R}^{262144} .

DEFINIÇÃO

Um *espaço vetorial* consiste de

- 1 um conjunto não vazio V , cujos elementos são chamados de vetores;
- 2 um corpo \mathbb{F} de escalares;
- 3 uma operação, chamada adição vetorial, que associa dois vetores \mathbf{u} e \mathbf{v} em V ao vetor $\mathbf{u} + \mathbf{v}$ em V ;
- 4 e uma operação, chamada multiplicação por escalar, que associa um escalar $\alpha \in \mathbb{F}$ e um vetor $\mathbf{u} \in V$ ao vetor $\alpha\mathbf{u}$ em V ,

que obedecem às seguintes propriedades, $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ e

$\forall \alpha, \beta \in \mathbb{F}$:

DEFINIÇÃO (CONTINUAÇÃO)

- 1 $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (comutativa);
- 2 $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (associativa);
- 3 existe um (único) vetor $\mathbf{0}$ em V , chamado de vetor nulo, tal que $\mathbf{u} + \mathbf{0} = \mathbf{u}$ (elemento neutro);
- 4 existe um (único) vetor $-\mathbf{u}$ em V tal que $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (simétrico);
- 5 $(\alpha\beta)\mathbf{u} = \alpha(\beta\mathbf{u})$ (associativa);
- 6 $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ (distributiva);
- 7 $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$ (distributiva);
- 8 $1\mathbf{u} = \mathbf{u}$ (unidade).

São exemplos de espaços vetoriais:

- Os vetores do \mathbb{R}^2 sobre o corpo dos reais, com as operações (usuais) de adição vetorial:

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix},$$

e multiplicação por escalar:

$$\alpha \mathbf{u} = \alpha \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \alpha u_1 \\ \alpha u_2 \end{bmatrix}.$$

- Os vetores do \mathbb{R}^n sobre o corpo dos reais, com as operações (usuais) de adição vetorial:

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix},$$

e multiplicação por escalar:

$$\alpha \mathbf{u} = \alpha \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} \alpha u_1 \\ \alpha u_2 \\ \vdots \\ \alpha u_n \end{bmatrix}.$$

Espaços Vetoriais – Exemplos

- Os vetores do \mathbb{C}^n sobre o corpo dos números complexos, com as operações usuais de adição vetorial e multiplicação por escalar.
- As matrizes reais $m \times n$ sobre o corpo dos reais, com as operações usuais de adição de matrizes e multiplicação por escalar.
- As funções reais contínuas no intervalo $[a, b]$ sobre o corpo dos reais, com as operações usuais de adição de funções e multiplicação por escalar.

Neste minicurso, consideraremos somente o espaço vetorial dos vetores do \mathbb{R}^n sobre o corpo dos reais, com as as operações usuais de adição vetorial e multiplicação por escalar.

DEFINIÇÃO

Um subconjunto não vazio U do espaço vetorial V é dito um *subespaço vetorial* de V se, para quaisquer \mathbf{u} e \mathbf{v} em U e qualquer escalar α ,

① $\mathbf{u} + \mathbf{v} \in U$ e

② $\alpha\mathbf{u} \in U$,

com a adição vetorial e a multiplicação por escalar herdadas de V .

Subespaços Vetoriais – Exemplos

Assim, o plano xy é um subespaço vetorial do \mathbb{R}^3 .

A reta $y = 2x$ é um subespaço vetorial do \mathbb{R}^2 ,

mas a reta r dada pela equação $y = 2x + 1$ **não** é um subespaço vetorial do \mathbb{R}^2 , pois, $\mathbf{u} \in r$, $\mathbf{v} \in r$, mas:

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \notin r.$$

Se consideramos imagens como vetores do \mathbb{R}^n , podemos dizer que duas imagens são próximas, se a diferença entre os vetores que as representam é “pequena”.

Assim, precisamos definir uma medida de “comprimento” de um vetor no \mathbb{R}^n .

DEFINIÇÃO

Uma *norma* em um espaço vetorial V é uma função que associa a cada vetor $\mathbf{v} \in V$ um número real não negativo $\|\mathbf{v}\|$, satisfazendo, para quaisquer $\mathbf{u}, \mathbf{v} \in V$ e qualquer escalar α , as propriedades:

- 1 $\|\mathbf{u}\| \geq 0$, com $\|\mathbf{u}\| = 0$ se, e somente se, $\mathbf{u} = \mathbf{0}$;
- 2 $\|\alpha\mathbf{u}\| = |\alpha|\|\mathbf{u}\|$;
- 3 $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Neste minicurso, só usaremos a *norma euclidiana* para vetores no \mathbb{R}^n , definida por:

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

Então, por exemplo, a norma de $\mathbf{u} = [1 \ 2 \ -1]^T$ é:

$$\|\mathbf{u}\| = \sqrt{(1)^2 + (2)^2 + (-1)^2} = \sqrt{6}.$$

A distância euclidiana entre dois vetores \mathbf{u} e \mathbf{v} em \mathbb{R}^n é definida por:

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}.$$

Uma outra noção geométrica do \mathbb{R}^2 e do \mathbb{R}^3 que podemos estender ao \mathbb{R}^n é a de ângulo entre vetores.

Para isso, precisamos definir o análogo do produto escalar:

DEFINIÇÃO

Um *produto interno* é uma função que associa a cada par de vetores \mathbf{u} e \mathbf{v} de um espaço vetorial V um número (real ou complexo) $\langle \mathbf{u}, \mathbf{v} \rangle$, tal que, para quaisquer \mathbf{u} , \mathbf{v} e \mathbf{w} em V e qualquer α escalar:

- 1 $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$, sendo que a barra significa o complexo conjugado;
- 2 $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$;
- 3 $\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \bar{\alpha} \langle \mathbf{u}, \mathbf{v} \rangle$;
- 4 $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, com $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ se, e somente se, $\mathbf{v} = \mathbf{0}$.

Produto Interno Euclidiano

Novamente, neste minicurso, usaremos somente o produto interno euclidiano (real), definido por:

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Outra maneira de escrever o produto interno euclidiano (real) é:

$$\mathbf{u}^T \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Notemos que a norma euclidiana provém do produto interno euclidiano:

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Produto Interno e Ângulo entre Vetores

Utilizando a Desigualdade de Cauchy-Schwarz

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\| ,$$

podemos definir o *ângulo* θ entre os vetores não nulos \mathbf{u} e \mathbf{v} por:

$$\cos \theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} .$$

O único ângulo que estamos interessado é o de $\pi/2$ radianos (90°). Assim, definimos:

DEFINIÇÃO

Dois vetores \mathbf{u} e \mathbf{v} de um espaço vetorial real com produto interno são *ortogonais* se $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Nesse caso, escrevemos $\mathbf{u} \perp \mathbf{v}$.

Vetores Ortogonais – Exemplos

São exemplos de vetores ortogonais no \mathbb{R}^2 :

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ e } \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \quad \text{e também } \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \text{ e } \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}.$$

São exemplos de vetores ortogonais dois a dois no \mathbb{R}^3 :

$$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1/\sqrt{6} \\ -1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}.$$

Vetores Ortonormais e Matriz Ortogonal

Os vetores dos exemplos anteriores são ditos *ortonormais*, pois:

DEFINIÇÃO

Os vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ são ditos *ortonormais* se forem ortogonais dois a dois e se todos possuírem norma unitária. Ou seja, se $\mathbf{v}_i \perp \mathbf{v}_j$, para $i \neq j$, e $\|\mathbf{v}_i\| = 1$, para todo i , com $i, j \in \{1, 2, \dots, n\}$.

Podemos formar uma matriz Q com colunas que sejam vetores ortonormais.

Uma matriz **quadrada** com colunas ortonormais é chamada de *matriz ortogonal*.

PROPOSIÇÃO

Uma matriz Q será ortogonal se, e somente se, $Q^{-1} = Q^T$.

Uma matriz retangular com colunas ortonormais não possui inversa, mas, mesmo assim, tem a seguinte propriedade:

$$Q^T Q = I.$$

No entanto, $QQ^T \neq I$.

Matriz Ortogonal

$$\begin{aligned} Q^T Q &= \begin{bmatrix} - & \mathbf{q}_{:,1}^T & - \\ - & \mathbf{q}_{:,2}^T & - \\ \vdots & & \\ - & \mathbf{q}_{:,n}^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{q}_{:,1} & \mathbf{q}_{:,2} & \cdots & \mathbf{q}_{:,n} \\ | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{q}_{:,1}^T \mathbf{q}_{:,1} & \mathbf{q}_{:,1}^T \mathbf{q}_{:,2} & \cdots & \mathbf{q}_{:,1}^T \mathbf{q}_{:,n} \\ \mathbf{q}_{:,2}^T \mathbf{q}_{:,1} & \mathbf{q}_{:,2}^T \mathbf{q}_{:,2} & \cdots & \mathbf{q}_{:,2}^T \mathbf{q}_{:,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_{:,n}^T \mathbf{q}_{:,1} & \mathbf{q}_{:,n}^T \mathbf{q}_{:,2} & \cdots & \mathbf{q}_{:,n}^T \mathbf{q}_{:,n} \end{bmatrix} \\ &= \begin{bmatrix} \langle \mathbf{q}_{:,1}, \mathbf{q}_{:,1} \rangle & \langle \mathbf{q}_{:,1}, \mathbf{q}_{:,2} \rangle & \cdots & \langle \mathbf{q}_{:,1}, \mathbf{q}_{:,n} \rangle \\ \langle \mathbf{q}_{:,2}, \mathbf{q}_{:,1} \rangle & \langle \mathbf{q}_{:,2}, \mathbf{q}_{:,2} \rangle & \cdots & \langle \mathbf{q}_{:,2}, \mathbf{q}_{:,n} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{q}_{:,n}, \mathbf{q}_{:,1} \rangle & \langle \mathbf{q}_{:,n}, \mathbf{q}_{:,2} \rangle & \cdots & \langle \mathbf{q}_{:,n}, \mathbf{q}_{:,n} \rangle \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I. \end{aligned}$$

Veamos alguns exemplos no GNU Octave.



LABORATÓRIO DE MATEMÁTICA COMPUTACIONAL

Toda matriz A possui uma decomposição QR, *i.e.*:

$$A = QR,$$

sendo Q uma matriz ortogonal e R uma matriz triangular superior.

Aqui, trataremos somente de matrizes com colunas linearmente independentes, assim, Q será uma matriz com colunas ortonormais (com a mesma dimensão de A) e R será uma matriz quadrada triangular superior invertível.

Decomposição QR – GNU Octave `qr()`

No GNU Octave, temos a função:

$$[Q R] = \text{qr}(A).$$

No entanto, nenhum computador digital trabalha com aritmética exata e isso produz erros.

Um experimento no GNU Octave:

```
[Q X] = qr(randn(1000));
```

```
R = triu(randn(1000));
```

```
A = Q*R;
```

```
[Q1 R1] = qr(A);
```

```
norm(Q - Q1)
```

```
max(max(Q - Q1))
```

```
norm(R - R1)
```

```
max(max(R - R1))
```

```
norm(A - Q1*R1)
```

```
max(max(A - Q1*R1))
```

DEFINIÇÃO

Seja V um espaço vetorial. Uma *combinação linear* dos vetores $\mathbf{u}_1, \dots, \mathbf{u}_n$ de V é um vetor da forma

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \dots + \alpha_n \mathbf{u}_n,$$

sendo $\alpha_1, \dots, \alpha_n$ escalares do corpo de V , com as operações de adição vetorial e multiplicação por escalar definidas em V .

Por exemplo, o vetor \mathbf{w} pode ser escrito como uma combinação linear dos vetores \mathbf{u} e \mathbf{v} :

$$\mathbf{w} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 2\mathbf{u} + 3\mathbf{v}.$$

DEFINIÇÃO

Seja S um subconjunto de vetores de um espaço vetorial V . O *subespaço vetorial gerado* pelos vetores de S (ou por S) é o conjunto de todas as combinações lineares

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_n \mathbf{u}_n$$

de vetores de S . Alternativamente, dizemos que S *gera* o subespaço U se todo vetor \mathbf{u} em U puder ser escrito como uma combinação linear de vetores em S .

Por exemplo, os vetores

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

geram o plano xy , que é um subespaço do \mathbb{R}^3 .

DEFINIÇÃO

Dizemos que os vetores $\mathbf{u}_1, \dots, \mathbf{u}_n$ são *linearmente independentes* (abreviadamente, L.I.) se

$$\alpha_1 \mathbf{u}_1 + \dots + \alpha_n \mathbf{u}_n = \mathbf{0}$$

somente quando $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$. Caso contrário, eles são *linearmente dependentes* (abreviadamente, L.D.).

Por exemplo, tomando os vetores $\mathbf{u} = [1 \ 0 \ 0]^T$, $\mathbf{v} = [0 \ 1 \ 0]^T$ e $\mathbf{w} = [2 \ 3 \ 0]^T$, observamos que

$$\alpha_1 \mathbf{u} + \alpha_2 \mathbf{v} = \mathbf{0}$$

somente se $\alpha_1 = \alpha_2 = 0$.

Independência Linear

De fato, o sistema

$$\begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

tem solução única $\alpha_1 = \alpha_2 = 0$. Ou seja \mathbf{u} e \mathbf{v} são vetores L.I.

De forma análoga, podemos concluir que \mathbf{u} e \mathbf{w} são L.I. e \mathbf{v} e \mathbf{w} são L.I.

Mas \mathbf{u} , \mathbf{v} e \mathbf{w} não são L.I., mas linearmente dependentes, pois

$$2\mathbf{u} + 3\mathbf{v} - 1\mathbf{w} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 1 \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}.$$

Vale notar que \mathbf{w} pode ser escrito como uma combinação linear de \mathbf{u} e \mathbf{v} , ou, dito de outra maneira, \mathbf{w} depende linearmente de \mathbf{u} e \mathbf{v} .

DEFINIÇÃO

Um conjunto não vazio S de vetores (de V) será chamado de *base* do espaço vetorial V se

- 1 os vetores de S forem linearmente independentes e
- 2 os vetores de S gerarem V .

O conjunto vazio é a base do subespaço trivial $\text{span}\{\mathbf{0}\}$.

Exemplo: A base mais simples do \mathbb{R}^n é a *base canônica*, formada pelos n vetores da forma

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

O vetor $\mathbf{v} = [2 \quad -3]^T$ do plano, por exemplo, é a seguinte combinação linear dos vetores da base canônica do \mathbb{R}^2 :

$$\mathbf{v} = \begin{bmatrix} 2 \\ -3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

DEFINIÇÃO

Diz-se que uma base \mathcal{B} de um espaço vetorial de dimensão finita é uma *base ordenada*, se os elementos de \mathcal{B} são considerados em uma ordem fixa (isto é, primeiro elemento, segundo elemento etc.).

Podemos considerar a base canônica $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ como uma base ordenada do \mathbb{R}^3 , bem como os vetores

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}.$$

DEFINIÇÃO

Seja $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ uma base ordenada de um espaço vetorial V de dimensão finita. Se

$$\mathbf{v} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \dots + \alpha_n \mathbf{u}_n,$$

então os escalares $\alpha_1, \alpha_2, \dots, \alpha_n$ são as *coordenadas* do vetor $\mathbf{v} \in V$ em relação à base ordenada \mathcal{B} .

Então, 2, 3 e -1 são as coordenadas do vetor $\mathbf{w} = [2 \ 3 \ -1]^T$ em relação à base canônica \mathcal{C} do \mathbb{R}^3 , pois:

$$\begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}_{\mathcal{C}} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\mathcal{C}} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{C}} - 1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_{\mathcal{C}}.$$

Assim como -9 , 6 e 5 são as coordenadas desse mesmo vetor em relação à base ordenada:

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_c, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_c, \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}_c \right\}.$$

De fato,

$$\begin{bmatrix} -9 \\ 6 \\ 5 \end{bmatrix}_B = -9 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_c + 6 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_c + 5 \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}_c = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}_c.$$

Base Ordenada de um Subespaço Vetorial

No caso do vetor \mathbf{v} pertencente ao subespaço do \mathbb{R}^3 com base ordenada

$$\mathcal{B} = \left\{ \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}_c, \begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix}_c \right\}$$

e dado por

$$\mathbf{v} = 1 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix},$$

suas coordenadas são 1, 1 e 0, pois:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{B}} = 1 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix} + 0 \begin{bmatrix} * \\ * \\ * \end{bmatrix},$$

Base Ordenada de um Subespaço Vetorial

não importando o vetor escolhido para completar uma base do \mathbb{R}^3 .

Então, se conhecermos uma base de um dado subespaço vetorial, podemos representar seus vetores utilizando somente as coordenadas associadas aos vetores da base do subespaço.

No exemplo anterior, podemos representar \mathbf{v} por:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{B}} = 1 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix}.$$

Isso é um tipo de compressão de dados, pois “economizamos” a quantidade de informação (coordenadas) a ser processada.



Math is Fun – Correlation.

<http://www.mathsisfun.com/data/correlation.html>.

Accessado em: 2014-09-05.



James E. Gentle. *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer, New York, 2007.



William W. Hines, Douglas C. Montgomery, David M. Goldsman, and Connie M. Borror. *Probabilidade e Estatística na Engenharia*. LTC, Rio de Janeiro, 4ª edition, 2006. Trad. Vera Regina Lima de Farias e Flores.



Ian T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 2nd edition, 2010.

-  Heysem Kaya, Pinar Tüfekci, and Sadık Fikret Gürgen. Local and global learning methods for predicting power of a combined gas & steam turbine. In *Proceedings of the International Conference on Emerging Trends in Computer and Electronics Engineering ICETCEE 2012*, pages 13 – 18, 2012.
-  Gilbert Strang. *Linear Algebra and Its Applications*. Harcourt, Orlando, FL, 3 edition, 1998.
-  Lloyd N. Trefethen and David Bau, III. *Numerical Linear Algebra*. SIAM, Philadelphia, 1997.
-  Pinar Tüfekci. Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, 60(0):126 – 140, 2014.



FIM

Estas notas estão disponíveis na página

<http://sites.google.com/site/thelmodearaujo/>
cursos/2014-2/pca

Mineração Massiva de Dados

Aula 2

Thelmo de Araujo

LABORATÓRIO DE MATEMÁTICA COMPUTACIONAL

Semestre 2018-2

Vamos estudar um pouco de estatística.

Como primeiro exemplo, utilizaremos os seguintes dados, que representam a temperatura média diária e o valor total diário de vendas de sorvetes em uma sorveteria durante 12 dias.

Na Teoria das Probabilidades, dizemos que a temperatura e o total de vendas são *variáveis aleatórias*, e denominamo-las de X_1 e X_2 , respectivamente.

Os dados serão importados no formato do GNU Octave do arquivo `ice_cream_data.m`.

Temperatura ($^{\circ}\text{C}$) – X_1	Total da Venda (\$) – X_2
14.2	215
16.4	325
11.9	185
15.2	332
18.5	406
22.1	522
19.4	412
25.1	614
23.4	544
18.1	421
22.6	445
17.2	408

As primeiras informações que extrairemos das variáveis aleatórias serão suas *médias* (amostrais), definidas por:

$$\mu = \frac{1}{m} \sum_{k=1}^m x_k ,$$

sendo m a quantidade de amostras e x_k o k -ésimo valor da variável aleatória X .

Assim, no exemplo do sorvete:

$$\mu_1 = 18.675 \quad \text{e} \quad \mu_2 = 402.417 .$$

Um Pouco de Estatística – Variância

Mas a média não nos dá nenhuma informação sobre a dispersão dos dados.

De fato, se uma pessoa comer um frango e outra passar fome, na média cada uma comeu meio frango, o que é muito diferente de cada uma ter comido, de fato, meio frango. Pergunte àquele que passou fome...

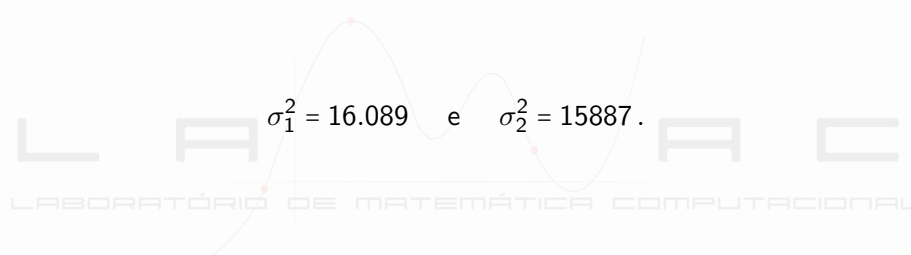
Para medir a dispersão dos dados, utilizamos a *variância amostral*, definida por:

$$\sigma^2 = \frac{1}{m-1} \sum_{k=1}^m (x_k - \mu)^2.$$

A raiz quadrada da variância é chamada de *desvio padrão* (σ).

Um Pouco de Estatística – Variância

No exemplo do sorvete, temos:


$$\sigma_1^2 = 16.089 \quad \text{e} \quad \sigma_2^2 = 15887.$$

Observe que as variâncias diferem em três ordens de grandeza.
Voltaremos a este problema em breve.

Um Pouco de Estatística – Covariância

Se pusermos num gráfico $X_1 \times X_2$, obteremos:

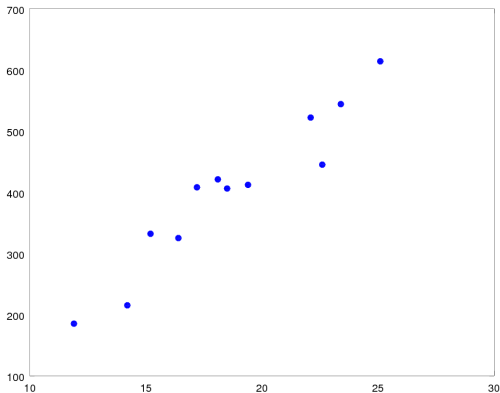


Figura: Nuvem de dados.

Podemos perceber, na figura anterior, que, à medida em que a temperatura (X_1) aumenta, o total de vendas de sorvete (X_2) também aumenta.

Essa relação entre variáveis aleatórias X_1 e X_2 é medida pela *covariância*, que é definida por:

$$\sigma_{12} = \frac{1}{m-1} \sum_{k=1}^m (x_{k1} - \mu_1)(x_{k2} - \mu_2).$$

A covariância σ_{12} é uma medida da variação conjunta das variáveis X_1 e X_2 , isto é, como X_1 se comporta quando X_2 aumenta ou diminui.

Um Pouco de Estatística – Covariância

Para normalizar a covariância, obtendo números que vão de -1 a 1 , definimos a *correlação* entre as variáveis aleatórias X_1 e X_2 como:

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2} .$$

Assim, a correlação entre uma variável aleatória e ela mesma é máxima e igual a 1 .

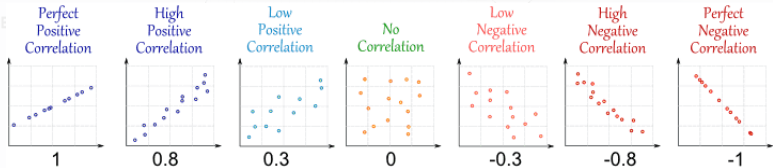


Figura: Exemplos de correlações.

Para visualizarmos o quão duas variáveis aleatórias são correlacionadas, devemos normalizar os dados.

Isto é, devemos subtrair a média de cada coluna e dividir os resultados pelo desvio padrão dessa coluna:

$$z_k = \frac{x_k - \mu}{\sigma},$$

chamando de Z a normalização da variável aleatória X .

Assim, obteremos para o exemplo do sorvete:

Um Pouco de Estatística – Covariância

Temperatura – Z_1	Total da Venda – Z_2
-1.115640	-1.486928
-0.567169	-0.614209
-1.689042	-1.724942
-0.866335	-0.558672
-0.043628	0.028429
0.853870	0.948751
0.180746	0.076032
1.601785	1.678661
1.177967	1.123295
-0.143350	0.147437
0.978523	0.337848
-0.367725	0.044297

Um Pouco de Estatística – Covariância

No gráfico:

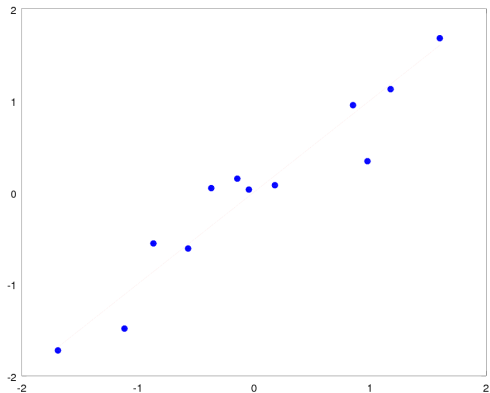


Figura: Nuvem de dados normalizados.

Um Pouco de estatística – Exemplo

A estatística do exemplo do sorvete é:

Médias:

$$\mu_1 = 18.675 \quad \text{e} \quad \mu_2 = 402.417.$$

Variâncias:

$$\sigma_1^2 = 16.089 \quad \text{e} \quad \sigma_2^2 = 15887.$$

Covariância e correlação:

$$\sigma_{12} = 484.09 \quad \text{e} \quad \rho_{12} = 0.95751.$$

O que nos mostra que a temperatura média diária e o total de vendas diárias de sorvete, nos dados coletados, são fortemente positivamente correlacionados (ρ próximo de 1).

Ou seja, quanto mais quente, mais sorvetes são vendidos.

E se tivermos três variáveis aleatórias?

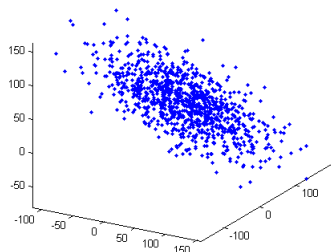


Figura: Nuvem de dados tridimensional.

Um Pouco de Estatística Multivariada

A maneira de tratarmos mais de duas variáveis aleatórias (e.g., n variáveis) é utilizar uma *matriz de covariância*:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}.$$

Notemos que Σ é simétrica (i.e., $\Sigma^T = \Sigma$), já que:

$$\sigma_{ij} = \frac{1}{m-1} \sum_{k=1}^m (x_{ki} - \mu_i)(x_{kj} - \mu_j) = \frac{1}{m-1} \sum_{k=1}^m (x_{kj} - \mu_j)(x_{ki} - \mu_i) = \sigma_{ji}.$$

Voltemos ao problema original, no qual tratamos de várias imagens formando uma matriz X $m \times n$, sendo m o tamanho de cada imagem (em forma de vetor-coluna) e n o número de imagens.

Nesse caso, é comum que $n \ll m$ (já sabemos que m tende a ser muito grande), mas mesmo assim as dimensões envolvidas são intratáveis numericamente.

O que fazer, então?

Uma ideia é a redução de dimensionalidade por meio de projeções.

Qual das três vistas fornece mais informação?



Figura: Celular visto sob três vistas diferentes.

Vejamos um exemplo de dados em três dimensões, adaptado do *Combined Cycle Power Plant Data Set* [8, 5].

Extraímos 60 amostras de três características: temperatura, umidade relativa e consumo líquido horário de energia elétrica.

Temperatura (°C) – X_1	Umidade Relativa (%) – X_2	Consumo Elétrico (MW) – X_3
08.34	90.01	480.48
23.64	74.20	445.75
29.74	41.91	438.76
⋮	⋮	⋮
29.11	74.01	432.43

Num gráfico tridimensional:

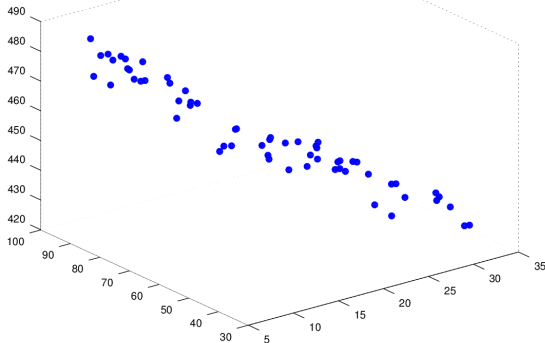


Figura: Dados de consumo de energia, temperatura e umidade relativa.

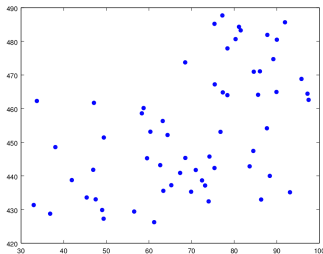
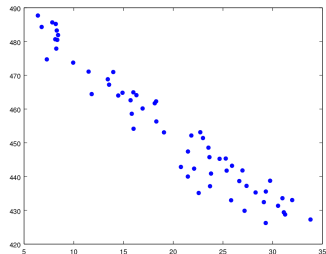
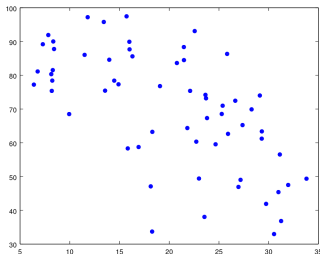
Tomemos os dados em:
`uece/cursos/pca/matlab/`

carregando os dados com
`load power_plant_data.m;`

e fazemos:
`scatter3(X(:,1),X(:,2),X(:,3),'filled')`

Variabilidade e Informação – Um Exemplo 3D

Ou, isolando pares de variáveis:



Calculando as estatísticas de X , obtemos:

Médias:

$$\mu_1 = 19.886, \quad \mu_2 = 70.124, \quad \mu_3 = 453.492.$$

Variâncias:

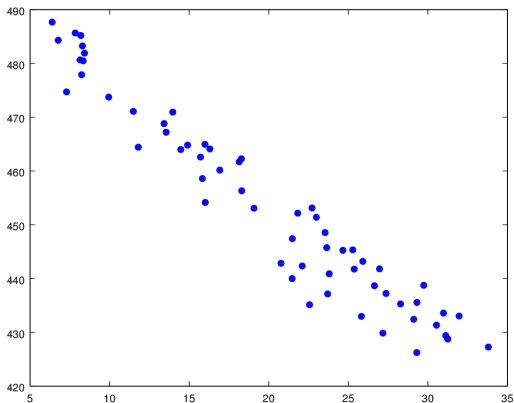
$$\sigma_1^2 = 62.445, \quad \sigma_2^2 = 295.968, \quad \sigma_3^2 = 326.219.$$

Matriz de correlação:

$$P = \begin{bmatrix} 1.00000 & -0.62092 & -0.96692 \\ -0.62092 & 1.00000 & 0.49953 \\ -0.96692 & 0.49953 & 1.00000 \end{bmatrix}.$$

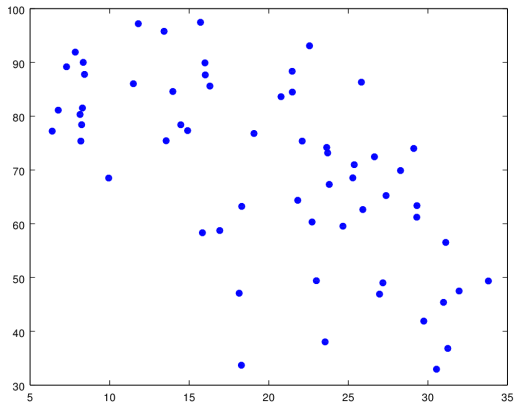
Variabilidade e Informação – Um Exemplo 3D

Assim, a maior correlação (negativa) é entre as variáveis X_1 e X_3 :
 $\rho_{13} = -0.96692$, como vemos na figura abaixo.



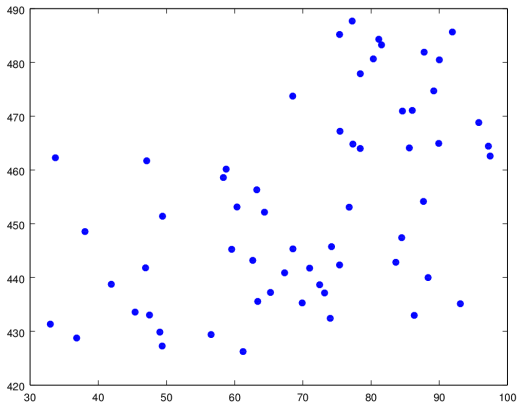
Variabilidade e Informação – Um Exemplo 3D

A segunda maior correlação (negativa) é entre as variáveis X_1 e X_2 : $\rho_{12} = -0.62092$, como vemos na figura abaixo.



Variabilidade e Informação – Um Exemplo 3D

A menor correlação (positiva) é entre as variáveis X_2 e X_3 :
 $\rho_{13} = 0.49953$, como vemos na figura abaixo.



Se considerarmos somente as variáveis X_2 e X_3 , o que corresponde a projetar os dados no plano X_2X_3 , manteríamos a maior variabilidade, mas perderíamos as duas outras variações.

O que devemos fazer para desconsiderar somente a menor variabilidade?

Para isso, é necessário projetarmos os dados num determinado plano.

Executar: `gerar_melhor_plano`.

Estudaremos como encontrar esse plano nas próximas aulas.

Projeções

Antes, porém, de encontrarmos o que seria o “melhor” plano de projeção, devemos saber o que é uma projeção e como aplicá-la. Estamos interessados somente em projeções ortogonais:

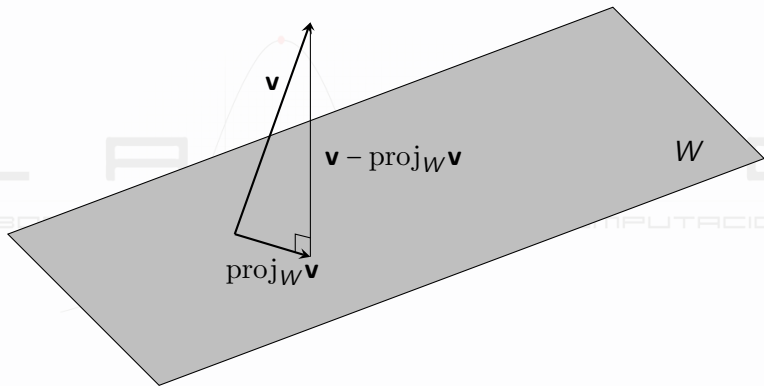


Figura: Projeção ortogonal de \mathbf{v} sobre W .

Na figura anterior, projetamos \mathbf{v} ortogonalmente sobre o subespaço vetorial W .

Lembrando que podemos falar de ortogonalidade porque definimos um produto interno $\langle \cdot, \cdot \rangle$.

Vejam, então, como projetar ortogonalmente um vetor $\mathbf{v} \in \mathbb{R}^3$ sobre um subespaço $W \subset \mathbb{R}^3$ de dimensão 2, definido por uma base ortonormal $\{\mathbf{w}_1, \mathbf{w}_2\}$.

Chamando a projeção de $\text{proj}_W \mathbf{v}$, temos:

$$\mathbf{v} = \text{proj}_W \mathbf{v} + (\mathbf{v} - \text{proj}_W \mathbf{v})$$

$$\text{proj}_W \mathbf{v} = \alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2,$$

pois $\text{proj}_W \mathbf{v} \in W$.

Assim,

$$\begin{aligned}\langle \mathbf{w}_1, \mathbf{v} \rangle &= \langle \mathbf{w}_1, \text{proj}_W \mathbf{v} + (\mathbf{v} - \text{proj}_W \mathbf{v}) \rangle \\ &= \langle \mathbf{w}_1, \text{proj}_W \mathbf{v} \rangle + \langle \mathbf{w}_1, \mathbf{v} - \text{proj}_W \mathbf{v} \rangle \\ &= \langle \mathbf{w}_1, \alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2 \rangle + 0 \\ &= \alpha_1 \langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \alpha_2 \langle \mathbf{w}_1, \mathbf{w}_2 \rangle \\ &= \alpha_1 \cdot 1 + \alpha_2 \cdot 0 \\ &= \alpha_1,\end{aligned}$$

e, analogamente,

$$\langle \mathbf{w}_2, \mathbf{v} \rangle = \alpha_2.$$

Ou seja,

$$\text{proj}_W \mathbf{v} = \langle \mathbf{w}_1, \mathbf{v} \rangle \mathbf{w}_1 + \langle \mathbf{w}_2, \mathbf{v} \rangle \mathbf{w}_2.$$

Podemos estender o último resultado para projeções ortogonais sobre um subespaço W com base ortonormal $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$:

$$\text{proj}_W \mathbf{v} = \langle \mathbf{w}_1, \mathbf{v} \rangle \mathbf{w}_1 + \langle \mathbf{w}_2, \mathbf{v} \rangle \mathbf{w}_2 + \dots + \langle \mathbf{w}_r, \mathbf{v} \rangle \mathbf{w}_r.$$

Uma projeção (ortogonal ou não) pode ser representada por uma matriz P , chamada de *projetor*.

No caso da projeção:

$$\text{proj}_W \mathbf{v} = \langle \mathbf{w}_1, \mathbf{v} \rangle \mathbf{w}_1 + \langle \mathbf{w}_2, \mathbf{v} \rangle \mathbf{w}_2,$$

podemos escrever:

$$\begin{aligned} P\mathbf{v} &= \langle \mathbf{w}_1, \mathbf{v} \rangle \mathbf{w}_1 + \langle \mathbf{w}_2, \mathbf{v} \rangle \mathbf{w}_2 \\ &= \mathbf{w}_1 \langle \mathbf{w}_1, \mathbf{v} \rangle + \mathbf{w}_2 \langle \mathbf{w}_2, \mathbf{v} \rangle \\ &= \mathbf{w}_1 (\mathbf{w}_1^T \mathbf{v}) + \mathbf{w}_2 (\mathbf{w}_2^T \mathbf{v}) \\ &= (\mathbf{w}_1 \mathbf{w}_1^T) \mathbf{v} + (\mathbf{w}_2 \mathbf{w}_2^T) \mathbf{v} \\ &= (\mathbf{w}_1 \mathbf{w}_1^T + \mathbf{w}_2 \mathbf{w}_2^T) \mathbf{v}. \end{aligned}$$

Ou seja,

$$P = \mathbf{w}_1 \mathbf{w}_1^T + \mathbf{w}_2 \mathbf{w}_2^T .$$

Um *projetor ortogonal* é uma matriz que realiza uma projeção ortogonal. Mais formalmente:

DEFINIÇÃO

Um *projetor* P é uma matriz idempotente, isto é, uma matriz que obedece à relação

$$P^2 = P .$$

PROPOSIÇÃO

A projeção realizada pelo projetor P será ortogonal se, e somente se,

$$P = P^T .$$

É fácil verificar que $P = \mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_2\mathbf{w}_2^T$ é um projetor ortogonal.
De fato:

$$\begin{aligned}P^2 &= (\mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_2\mathbf{w}_2^T)(\mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_2\mathbf{w}_2^T) \\&= \mathbf{w}_1\mathbf{w}_1^T\mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_1\mathbf{w}_1^T\mathbf{w}_2\mathbf{w}_2^T + \\&\quad + \mathbf{w}_2\mathbf{w}_2^T\mathbf{w}_2\mathbf{w}_2^T + \mathbf{w}_2\mathbf{w}_2^T\mathbf{w}_1\mathbf{w}_1^T \\&= \mathbf{w}_1\mathbf{1}\mathbf{w}_1^T + \mathbf{w}_1\mathbf{0}\mathbf{w}_2^T + \mathbf{w}_2\mathbf{1}\mathbf{w}_2^T + \mathbf{w}_2\mathbf{0}\mathbf{w}_1^T \\&= \mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_2\mathbf{w}_2^T \\&= P,\end{aligned}$$

e

$$\begin{aligned}P^T &= (\mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_2\mathbf{w}_2^T)^T = (\mathbf{w}_1\mathbf{w}_1^T)^T + (\mathbf{w}_2\mathbf{w}_2^T)^T = \\&= \mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_2\mathbf{w}_2^T = P.\end{aligned}$$

Vejamos como ficam os dados projetados no exemplo do consumo de energia.

```
./matlab/gerar_projecoes
```

A base ortonormal do plano é dada pelos vetores:

$$\mathbf{w}_1 = \begin{bmatrix} 0.328127 \\ -0.614479 \\ -0.717459 \end{bmatrix} \quad \text{e} \quad \mathbf{w}_2 = \begin{bmatrix} -0.164868 \\ -0.785106 \\ 0.597015 \end{bmatrix}.$$

E o projetor ortogonal é:

$$P = \begin{bmatrix} 0.134849 & -0.072189 & -0.333846 \\ -0.072189 & 0.993977 & -0.027856 \\ -0.333846 & -0.027856 & 0.871175 \end{bmatrix}.$$

Agora que sabemos projetar vetores, precisamos descobrir o plano sobre o qual serão feitas as projeções.

Esse será o assunto de nossa próxima aula.



Math is Fun – Correlation.

<http://www.mathsisfun.com/data/correlation.html>.

Accessado em: 2014-09-05.







James E. Gentle. *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer, New York, 2007.



William W. Hines, Douglas C. Montgomery, David M. Goldsman, and Connie M. Borror. *Probabilidade e Estatística na Engenharia*. LTC, Rio de Janeiro, 4ª edition, 2006. Trad. Vera Regina Lima de Farias e Flores.



Ian T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 2nd edition, 2010.

-  Heysem Kaya, Pinar Tüfekci, and Sadık Fikret Gürgen. Local and global learning methods for predicting power of a combined gas & steam turbine. In *Proceedings of the International Conference on Emerging Trends in Computer and Electronics Engineering ICETCEE 2012*, pages 13 – 18, 2012.
-  Gilbert Strang. *Linear Algebra and Its Applications*. Harcourt, Orlando, FL, 3 edition, 1998.
-  Lloyd N. Trefethen and David Bau, III. *Numerical Linear Algebra*. SIAM, Philadelphia, 1997.
-  Pinar Tüfekci. Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, 60(0):126 – 140, 2014.



FIM

Estas notas estão disponíveis na página

[http://sites.google.com/site/thelmodearaujo/
cursos/2014-2/pca](http://sites.google.com/site/thelmodearaujo/cursos/2014-2/pca)