Revisão

1 Álgebra linear matricial

Um vetor $x \in \mathbb{R}^n$ será representado por um vetor coluna

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

enquanto o transposto de x corresponde a um vetor linha

$$x^\top = [x_1 \;\; x_2 \;\; \ldots x_n] \, .$$

1.1 Independência linear

Se existem escalares não-nulos $\lambda_1, \lambda_2, \ldots, \lambda_m$ tais que

$$\sum_{i=1}^{m} \lambda_i v_i = 0,$$

dizemos que os vetores $v_1, v_2, \ldots, v_m$ são *linearmente dependentes*. Caso contrário o conjunto de vetores $\{v_1, v_2, \ldots, v_m\}$ é dito *linearmente independente*.

O $\mathbb{R}^n$ é gerado por *combinações lineares* de qualquer conjunto de n vetores linearmente independentes. Neste caso, tal conjunto é chamado de *base* para o $\mathbb{R}^n$.

1.2 Combinações linear, afim, cônica e convexa

- Combinação linear

$$x = \sum_{i=1}^{m} \lambda_i v_i.$$

- Combinação afim

$$x = \sum_{i=1}^{m} \lambda_i v_i, \quad \text{com} \; \sum_{i=1}^{m} \lambda_i = 1.$$

- Combinação cônica

$$x = \sum_{i=1}^{m} \lambda_i v_i, \quad \text{com } \lambda_i \geq 0, \text{ para } i = 1, 2, \ldots m.$$

- Combinação convexa

$$x = \sum_{i=1}^{m} \lambda_i v_i, \quad \text{com} \; \sum_{i=1}^{m} \lambda_i = 1 \text{ e } \; \lambda_i \geq 0, \text{ para } i = 1, 2, \ldots m.$$

1.3 Normas e produto interno

O produto interno em $\mathbb{R}^n$ é definido por

$$x^\top y = \sum_{i=1}^{n} x_i y_j.$$

Dois vetores x e y são ditos *ortogonais* quando $x^\top y = 0$.

Uma aplicação $\|.\| : \mathbb{R}^n \to \mathbb{R}$ é chamada de norma vetorial se satisfaz as seguintes propriedades:

- $\|x\| \geq 0, \quad \forall x \in \mathbb{R}^n$;
- $\|\alpha x\| = |\alpha|\|x\|$, para qualquer vetor x e escalar α;
- $\|x + y\| \leq \|x\| + \|y\|, \quad \forall x, y \in \mathbb{R}^n$; (*desigualdade triangular*)
- $\|x\| = 0 \Rightarrow x = 0$.

Abaixo temos alguns exemplos de normais vetoriais:

$$\|x\|_2 = \sqrt{x^\top x},$$

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|,$$
$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Lembre que quando x e y são ortogonais, i.e. $x^\top y = 0$, vale o teorema de Pitágoras:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

A *desigualdade de Cauchy-Schwarz*:

$$|x^\top y| \leq \|x\|_2 \|y\|_2,$$

relaciona o produto interno com a norma Euclidiana $\|.\|_2$.

Uma propriedade importante da norma Euclidiana $\|.\|_2$ é a *invariância por transformações ortogonais*, i.e., se Q é ortogonal, então:

$$\|Qx\|_2 = \|x\|_2.$$

Outro fato importante é que em dimensão finita todas as normas são equivalentes, i.e., dadas duas normas vetoriais $\|.\|_a$ e $\|.\|_b$, existem $c_2 > c_1 > 0$ tais que

$$c_1 \|x\|_b \leq \|x\|_a \leq c_2 \|x\|_b.$$

1.4 Subspaços vetoriais

Dizemos que S é um subspaço de $\mathbb{R}^n$ se S é fechado em relação a soma e produto por escalar, i.e., se $x, y \in S$ então $\lambda x + \mu y \in S$ para quaisquer escalares λ, μ.

Seja M um subspaço vetorial de $\mathbb{R}^n$. Denotamos por $M^\perp$ o conjunto de todos os vetores de $\mathbb{R}^n$ ortogonais a todos os vetores de M:

$$M^\perp = \{y \in \mathbb{R}^n \mid x^\top y = 0\}.$$

O conjunto $M^\perp$ é chamado *complemento ortogonal* de M. Note que

$$\mathbb{R}^n = M \oplus M^\perp.$$

1.5 Matrizes

Denotamos por $A \in \mathbb{R}^{m\times n}$, uma matriz com entradas a_{ij} reais, de m linhas por n colunas:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Operações

Sejam $\alpha \in \mathbb{R}$, $A \in \mathbb{R}^{m\times n}$ e $B \in \mathbb{R}^{p\times q}$.

- Adição e substração: $C = A \pm B$

  Se $m = p$ e $n = q$ então $c_{ij} = a_{ij} \pm b_{ij}$, com $C \in \mathbb{R}^{m\times n}$.

- Multiplicação por escalar: $C = \alpha A$

  $c_{ij} = \alpha a_{ij}$, com $C \in \mathbb{R}^{m\times n}$.

- Multiplicação: $C = AB$

  Se $n = p$ então $c_{ij} = \displaystyle\sum_{k=1}^{n} a_{ik} b_{kj}$, e $C \in \mathbb{R}^{m\times q}$.

  Atenção: Em geral $AB \neq BA$.

- Matriz transposta: $C = A^\top$

  $c_{ij} = a_{ji}$, e $C \in \mathbb{R}^{n\times m}$.

- Matriz Inversa: $C = A^{-1}$

  Se $m = n$ então $CA = AC = I$, com $C \in \mathbb{R}^{n\times n}$, e onde I denota a matriz identidade de ordem n.

Operações por bloco

Seja $A \in \mathbb{R}^{m \times n}$:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1q} \\ A_{21} & A_{22} & \dots & A_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \dots & A_{pq} \end{bmatrix},$$

uma matriz decomposta em blocos $A_{ij} \in \mathbb{R}^{n_i \times n_j}$, para $i = 1, \dots, p$ e $j = 1, \dots, q$, tais que $\sum_{i=1}^{p} n_i = m$ e $\sum_{j=1}^{q} n_j = n$.

Tomando os devidos cuidados, as operações da seção anterior podem ser realizadas por blocos, desde que as dimensões dos blocos sejam compatíveis.

1.6 Matrizes especiais

- Linha: $m = 1$
- Coluna: $n = 1$.
- Nula: $a_{ij} = 0$.
- Quadrada: $m = n$.
- Diagonal: $a_{ij} = 0, \forall i \neq j$.
- Identidade: $a_{ij} = 0, \forall i \neq j$ e $a_{ii} = 1$ se $i = j$.
- Triangular superior: $a_{ij} = 0$ para $i > j$.
- Triangular inferior: $a_{ij} = 0$ para $j > i$.
- Simétrica: $m = n$ e $a_{ij} = a_{ji}$, ou $A = A^\top$.
- Anti-simétrica: $m = n$ e $a_{ij} = -a_{ji}$, ou $A = -A^\top$.
- Nilpotente: $m = n$ e $A^k = 0$, para algum k inteiro positivo.
- Idempotente: $m = n$ e $A^2 = A$.
- Auto-reflexiva: $m = n$ e $A^2 = I$.
- Ortogonal: $m = n$ e $A^{-1} = A^\top$.
- Normal: $m = n$ e $A^\top A = AA^\top$.

1.7 Posto

Seja $A \in \mathbb{R}^{m \times n}$. O posto coluna(linha) de A é o número de colunas(linhas) linearmente independentes. É possível mostrar que posto linha é igual ao posto coluna e, então, diremos simplesmente posto de A e denotaremos por posto(A).

Dizemos que uma matriz A tem *posto completo* se posto$(A) = \min\{m, n\}$.

Uma matriz quadrada é não-singular se seu posto é completo.

1.8 Subspaços fundamentais

Seja $A \in \mathbb{R}^{m \times n}$.

- Núcleo de A: $\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$.
- Imagem de A: $\mathcal{R}(A) = \{y \in \mathbb{R}^m \mid y = Ax \text{ para algum } x \in \mathbb{R}^n\}$
- Espaço linha de A: $\mathcal{R}(A^\top) = \{u \in \mathbb{R}^n \mid u = A^\top y \text{ para algum } y \in \mathbb{R}^m\}$.
- Núcleo a esquerda de A: $\mathcal{N}(A^\top) = \{y \in \mathbb{R}^m \mid A^\top y = 0\}$.

Teorema fundamental da álgebra linear matricial:

$$\mathcal{N}(A) \oplus \mathcal{R}(A^\top) = \mathbb{R}^n,$$

$$\mathcal{N}(A^\top) \oplus \mathcal{R}(A) = \mathbb{R}^m.$$

1.9 Determinantes

Seja $A \in \mathbb{R}^{n \times n}$. O determinante de A é definido por

$$\det(A) = \sum_{p \in P} \sigma(p) a_{1p_1} a_{2p_2} \dots a_{np_n},$$

onde p é uma permutação do vetor $p_0 = (1, 2, 3, \dots, n)$ e P o conjunto de todas as permutações. A função $\sigma(p)$ é igual a 1, se p pode ser obtido através de um número par de permutações de p_0, e igual a -1 se o número de permutações for ímpar.

Cada parcela da soma acima é chamada de "produto elementar com sinal". Assim, dizemos que o determinante é a soma de todos os produtos elementares com sinal.

Propriedades

Algumas propriedades do determinante:

- $\det(A) = \det(A^\top)$
- $\det(AB) = \det(A) \det(B)$
- Se A é não-singular: $\det(A^{-1}) = 1/\det(A)$.

1.10 Autovalores e autovetores

Seja $A \in \mathbb{R}^{n\times n}$. Dizemos que $\lambda \in \mathbb{C}$ é um *autovalor* de A, associado ao *autovetor* $v \in \mathbb{C}^n \setminus \{0\}$ quando

$$Av = \lambda v.$$

Decorre desta definição que λ é autovalor de A, se e somente se, λ é raiz do polinômio característico:

$$p(\lambda) = \det(A - \lambda I).$$

1.11 Matrizes simétricas, positivas semidefinidas

Uma matriz $A \in \mathbb{R}^{n\times n}$ é dita *positiva semidefinida* se

$$x^\top Ax \geq 0, \quad \forall x \in \mathbb{R}^n,$$

e positiva definida quando

$$x^\top Ax > 0, \quad \forall x \in \mathbb{R}^n \setminus \{0\}.$$

Quando A é simétrica, i.e. $A = A^\top$, todos seus autovalores são reais e seus autovetores são ortogonais. Neste caso faz sentido dizer que A é positiva semidefinida se e somente se todos seus autovalores são não-negativos.

1.12 Normas matriciais

Podemos definir normas matriciais induzidas por normas vetoriais:

$$\|A\| = \max_{\|x\|=1} \|Ax\|,$$

onde $\|.\|$ no lado direito denota uma norma vetorial. Uma consequência dessa definição é a relação de *consistência*:

$$\|Ax\| \leq \|A\|\|x\|,$$

válida para toda norma matricial induzida.

Se A é não singular também temos que $\|Ax\| \geq \|x\|/\|A^{-1}\|$. Neste caso

$$\frac{\|x\|}{\|A^{-1}\|} \leq \|Ax\| \leq \|A\|\|x\|.$$

Há outras normas matriciais que não são induzidas mas são consistentes, como a norma de Frobenius:

$$\|A\|_F = \left(\sum_i \sum_j a_{ij}^2\right)^{1/2}.$$

Seja $A \in \mathbb{R}^{n\times n}$ uma matriz não-singular. Definimos o *número de condição* de A por

$$\kappa(A) = \|A\|\|A^{-1}\|,$$

onde $\|.\|$ é uma norma matricial induzida.

1.13 Decomposição espectral e SVD

Seja $A \in \mathbb{R}^{n\times n}$ simétrica. Então

$$A = Q\Lambda Q^\top,$$

onde Q é uma matriz ortogonal e $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ é uma matriz diagonal com os autovalores de A, e sem perda de generalidade assuma que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. A decomposição acima é conhecida como decomposição espectral de A. Note que a decomposição espectral também pode ser escrita na forma: $A = \sum_{i=1}^n \lambda_i q_i q_i^\top$, onde o autovetor unitário q_i corresponde a i-ésima coluna de Q. Note ainda que as colunas de Q formam uma base ortonormal para $\mathbb{R}^n$.

Seja $A \in \mathbb{R}^{m\times n}$, com $m \geq n$. Então

$$A = U \begin{bmatrix} S \\ 0 \end{bmatrix} V^\top,$$

onde U é uma matriz ortogonal de ordem m, V é uma matriz ortogonal de ordem n e $S = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$, com $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$. Cada σ_i é chamado de valor singular, e decomposição acima é conhecida como decomposição em valores singulares ou SVD, do inglês singular value decomposition.

De modo análogo, para o caso $m \leq n$ temos que

$$A = U \begin{bmatrix} S & 0 \end{bmatrix} V^\top,$$

mas agora $S \in \mathbb{R}^{m\times m}$. Se posto$(A) = r$, é comum escrever a SVD de A na forma

$$A = \sum_{i=1}^r \sigma_i u_i v_i^\top.$$

Note que se A é simétrica e positiva definida então a decomposição espectral coincide com a decomposição SVD, com $Q = U = V$ e $\Lambda = S$.

Para uma matriz A simétrica, positiva definida, temos que

$$\|A\|_2 = \sigma_1(A), \qquad \|A^{-1}\|_2 = 1/\sigma_n(A).$$

Ainda

$$\sigma_n(A)\|x\|_2^2 \leq x^\top Ax \leq \sigma_1(A)\|x\|_2^2.$$

Se A é apenas simétrica, então

$$\lambda_n(A)\|x\|_2^2 \leq x^\top Ax \leq \lambda_1(A)\|x\|_2^2.$$

2 Elementos de análise e cálculo

2.1 Sequências

Uma sequência de vetores $x_0, x_1, \dots, x_k, \dots,$ é denotada por $\{x_k\}_{k=0}^\infty$ ou simplesmente $\{x_k\}$ se o conjunto de índices estiver claro. Dizemos que $\{x_k\}$ converge a x_* se $\|x_k - x_*\| \to 0$

quando $k \to \infty$, e escrevemos $x_k \to x_*$ ou ainda $\lim_{k\to\infty} x_k = x_*$.

Um ponto $\hat{x}$ é dito *ponto limite* de uma sequência $\{x_k\}$, se existe um subconjunto de índices $\mathbb{K} \subset \mathbb{N}$, tal que a subsequência $\{x_k\}_{k\in\mathbb{K}}$ converge para $\hat{x}$.

2.2 Topologia do $\mathbb{R}^n$

Chamamos de *bola* de centro em x_* e raio $\epsilon > 0$ o seguinte conjunto de pontos do $\mathbb{R}^n$:

$$B(x_*, \epsilon) = \{x \in \mathbb{R}^n \mid \|x - x_*\| < \epsilon\}.$$

Por vezes $B(x_*, \epsilon)$ também é chamada de ϵ-vizinhança de x_*.

Um subconjunto S de $\mathbb{R}^n$ é dito *aberto* se para todo ponto x de S existe um $\epsilon > 0$ tal que $B(x, \epsilon) \subset S$.

Um conjunto C é dito *fechado* se $x_k \to x$, com $x_k \in C$ implica que $x \in C$.

Dizemos que um conjunto $S \subset \mathbb{R}^n$ é *limitado* se existe M finito tal que $\|x\| \leq M$, para todo $x \in S$.

O *interior* de um conjunto $S \subset \mathbb{R}^n$ é o maior conjunto aberto contido em S. O *fecho* de S, denotado por $\bar{S}$, é o menor conjunto fechado que contem S.

Um subconjunto S de um espaço métrico é dito *compacto* se toda sequência em S possui uma subsequência convergente(em S). Em dimensão finita, como em $\mathbb{R}^n$, um subconjunto S é compacto se e somente se é fechado e limitado.

2.3 Continuidade

Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é dita *contínua* se $x_k \to x$ implica que $f(x_k) \to f(x)$.

Um resultado importante sobre funções contínuas é que se f é contínua e S é compacto, então $f(S)$ é compacto. (Prove!). Este fato implica no teorema de Bolzano-Weierstrass: *uma função contínua f definida em um compacto S possui um minimizador(global) em S*, i.e., existe $x_* \in S$ tal que $f(x_*) \leq f(x), \forall x \in S$.

Uma função $f : D \subset \mathbb{R}^n \to \mathbb{R}$ é dita *Lipschitz* em D, se existe $L > 0$ tal que

$$\|f(x) - f(y)\| \leq L\|x - y\|, \quad \forall x, y \in D.$$

Perceba que toda função Lipschitz é contínua, mas a recíproca não é verdadeira (dê um contra-exemplo!).

2.4 Derivadas

Seja $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$, i.e., uma função contínua com derivadas parciais(até segunda ordem) contínuas.

Denotamos o gradiente de f em x por um vetor coluna $\nabla f(x)$, cujas componentes são $\nabla f(x)_i = \dfrac{\partial f}{\partial x_i}(x)$. Já a Hessiana de f é uma matriz simétrica $n \times n$, denotada por $\nabla^2 f(x)$, cujas entradas são dadas por $[\nabla^2 f(x)]_{ij} = \dfrac{\partial^2 f}{\partial x_j \partial x_i}(x)$.

A *derivada direcional* de f em x, na direção d, é dada por

$$\nabla f(x)^\top d = \lim_{t \to 0} \frac{f(x+td) - f(x)}{t}.$$

Se $F : \mathbb{R}^n \to \mathbb{R}^m$ é uma função vetorial com m componentes, i.e.:

$$F(x) = \begin{bmatrix} F_1(x) \\ \vdots \\ F_m(x) \end{bmatrix},$$

onde cada $F_i(x)$ é uma função continuamente diferenciável, denotamos o Jacobiano de F por:

$$J(x) = F'(x) = \begin{bmatrix} \nabla F_1(x)^\top \\ \vdots \\ \nabla F_m(x)^\top \end{bmatrix}.$$

2.5 Notação $o(.)$

Seja g uma função de uma variável real. A notação $g(x) = o(x)$ significa que $g(x)$ vai para zero mais rápido que x, i.e.

$$\lim_{x \to 0} \left| \frac{g(x)}{x} \right| = 0.$$

2.6 Teoremas do Valor Médio e de Taylor

Relembremos o teorema do valor médio para uma função de uma variável. Se $\phi : \mathbb{R} \to \mathbb{R}$ é continuamente diferenciável então dados dois valores reais $t_1 > t_0$ temos que

$$\phi(t_1) = \phi(t_0) + \phi'(\zeta)(t_1 - t_0),$$

onde $\zeta \in (t_0, t_1)$.

Uma versão em $\mathbb{R}^n$ deste teorema, envolvendo a derivada direcional de uma função $f \in \mathcal{C}^1$ é dada por

$$f(x+d) = f(x) + \nabla f(x + \alpha d)^\top d, \quad \text{para algum } \alpha \in (0,1),$$

ou ainda

$$f(y) = f(x) + \nabla f(\xi)^\top (y - x),$$

onde $\xi = \alpha x + (1-\alpha)y$ e $\alpha \in (0,1)$.

Quando $f \in \mathcal{C}^2$, uma versão refinada deste resultado é

$$f(x+d) = f(x) + \nabla f(x)^\top d + \frac{1}{2} d^\top \nabla^2 f(\xi) d.$$

Em verdade a expressão acima é uma das formas do teorema de Taylor. Outra forma útil é

$$f(x+d) = f(x) + \nabla f(x)^\top d + \frac{1}{2} d^\top \nabla^2 f(x) d + o(\|d\|^2).$$

No caso de uma função vetorial $F : \mathbb{R}^n \to \mathbb{R}^m$, o teorema do valor médio se apresenta na forma

$$F(x+p) - F(x) = \int_0^1 J(x+tp)p\,dt.$$

2.7 Função Lipschitz

Dizemos que uma função $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ é Lipschitz de constante $L > 0$ (em Ω) quando

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x \in \Omega.$$

Para uma função $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$ tal que o gradiente ∇f é Lipschitz em $\mathbb{R}^n$, temos uma cota superior para o *erro de linearização* dada por

$$|f(y) - f(x) - \nabla f(x)^\top (y-x)| \leq \frac{L}{2}\|y - x\|^2.$$

2.8 Função Convexa

Dizemos que um subconjunto não vazio $\Omega \subset \mathbb{R}^n$ é um conjunto *convexo* quando para quaisquer $x, y \in \Omega$ tem-se que

$$\alpha x + (1-\alpha)y \in \Omega, \quad \forall \alpha \in [0,1].$$

Uma função $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, com Ω convexo, é dita *convexa* se para quaisquer $x, y \in \Omega$, tem-se que

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \quad \forall \alpha \in [0,1].$$

Geometricamente isto significa que o gráfico de f está sempre abaixo da reta secante aos pontos $(x, f(x))$ e $(y, f(y))$.

Para uma f convexa e diferenciável temos também que o gráfico de f está sempre acima da reta tangente a $(x, f(x))$:

$$f(y) \geq f(x) + \nabla f(x)^\top (y-x), \quad \forall x, y \in \Omega.$$

2.9 Teorema da Função Implícita

Seja $h : \mathbb{R}^n \to \mathbb{R}^m$ e considere o seguinte sistema de equações não-lineares:

$$h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix} = 0.$$

onde cada $h_i : \mathbb{R}^n \to \mathbb{R}$ é continuamente diferenciável e $m \leq n$.

Particione o vetor de variáveis $x = (x_B, x_N)$, onde x_B corresponde as primeiras m e x_N as últimas $n - m$ variáveis. Se $x^* = (x_B^*, x_N^*)$ é tal que $h(x^*) = 0$ e a matriz

$$J_B = \frac{\partial h}{\partial x_B}(x^*) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x^*) & \dots & \frac{\partial h_1}{\partial x_m}(x^*) \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1}(x^*) & \dots & \frac{\partial h_m}{\partial x_m}(x^*) \end{bmatrix}$$

é não-singular, então existe uma vizinhança de x_N^* tal que para x_N nessa vizinhança, existe $\phi : \mathbb{R}^{n-m} \to \mathbb{R}^m$ que satisfaz

1. $x_B^* = \phi(x_N^*)$;
2. $h(\phi(x_N), x_N) = 0$;
3. $\phi_i \in \mathcal{C}^1$, para $i = 1, 2, \dots, m$;
4. $\phi'(x_N^*) = -\left(\frac{\partial h}{\partial x_B}(x^*)\right)^{-1} \frac{\partial h}{\partial x_N}(x^*)$.