Matrizes - 2

Normas Matriciais

Existem multiplas normas distintas que podem ser calculadas de uma matriz. Normas matriciais são de alguma forma similares a normas vetoriais em que cada norma fornece um numero que caracteriza a matriz. A norma é indicada usando duas linhas verticais, por exemplo, a norma da matriz A é indicada por $\|A\|$.

As normas matriciais podem ser amplamente divididas em duas familias: normas elementares (ou de entrada) e normas induzidas. As normas elementares são calculadas baseadas nos elementos individuais da matriz e portanto essas normas podem ser interpretadas para refletir as magnitudes dos elementos na matriz.
As normas induzidas pode ser interpretada da seguinte maneira: uma das funções de uma matriz é codificar a transformação de um vetor; a norma induzida de uma matriz é a medida de quanto essa transformação escalou (alongando ou contraindo) esse vetor.

Este capitulo ira abordar as normas elementares. A norma Euclidiana também chamada de norma Frobenius, é uma extensão direta da norma vetorial. Ela é calculada como a raiz quadrada da soma de todos os elementos da matriz ao quadrado

A norma Frobenius

$$\|A\|_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij}^2}$$

Os indices i e j correspondem as M linhas e N colunas. E a anotação subescrita $_F$ indica a norma Frobenius.

Essa norma também é chamada de norma ℓ_2. Esse nome tem origem da fórmula geral para normas-p elementares:

$$\|A\|_p = \left(\sum_{i=1}^{M} \sum_{j=1}^{N} |a_{ij}|^p\right)^{1/p}$$

Perceba que você obtem a norma frobenius quando $p=2$

Uma das aplicações mais importantes das normas matriciais na analise estatistica e Machine Learning é sua aplicação na regularização, cujo objetivo visa melhorar o ajuste do modelo e aumentar a generalização dos modelos aos dados novos. A ideia basica da regularização é adicionar uma norma matricial como uma função de custo a um algoritmo de minimização. Essa norma ajudará a evitar que os parâmetros do modelo se tornem muito grandes (a regularização l_2 também é conhecida por regressão Ridge) ou incentivara soluções esparsas (a regularização l_1, tambem conhecida como regressão lasso). As arquiteturas modernas de deep learning dependem das normas matriciais para alcancar desempenhos tão impressionantes na resolução de problemas de visão computacional.

Outra aplicação da norma de Frobenius é medir a "distância da Matriz". A distancia matricial de uma matriz com ela mesmo é 0, e a distância entre duas matrizes distintas aumenta a medida que os valores numericos dessas matrizes se tornam mais diferentes.

Essa distancia pode ser usada como um criterio de otimização em machine learning, por exemplo, para reduzir o tamanho de armazenamento de dados de uma imagem enquanto minimiza a distância de Frobenius entre as matrizes reduzida e a original.

Traço de Matriz e Norma de Frobenius

O traço de uma matriz é a soma de seus elementos diagonais, indicado por $tr(A)$, e so existe para matrizes quadradas. As duas matrizes a seguir tem o mesmo traço (14):

$$\begin{bmatrix} 4 & 5 & 6 \\ 0 & 1 & 4 \\ 9 & 9 & 9 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 8 & 0 \\ 1 & 2 & 6 \end{bmatrix}$$

O traço tem algumas propriedades interessantes. Por exemplo, o traço de uma matriz é igual a soma de seus autovalores e portanto é uma medida de volume de seu autoespaço. Outra propriedade interessante é:

$$\|A\|_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij}^2} = \sqrt{tr(A^T A)}$$

A norma de Frobenius pode ser calculada como a raiz quadrada do traço de uma matriz vezes sua transposta. Isso se deve ao caso que cada elemento diagonal da matriz A^TA é definido pelo produto escalar de cada linha com ela mesmo

Espaços Matriciais (colunas, linhas e nulos)

Espaços matriciais são conceitos simples e são essencialmente combinações lineares ponderadas de diferentes características de uma matriz.

Espaço colunar

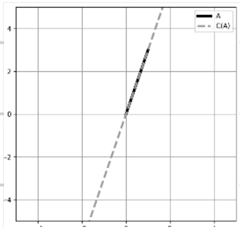
Primeiro nos conceitualizamos uma matriz como um conjunto de vetores coluna. Depois, consideramos um infinito de valores reais escalares em vez de trabalhar com um conjunto específico de escalares. Isso resultara em infinitas formas de combinar um conjunto de vetores. Esse resultado infinito de conjunto de vetores é chamado de espaço colunar de uma matriz.

$$C\left(\begin{bmatrix}1\\3\end{bmatrix}\right)=\lambda\begin{bmatrix}1\\3\end{bmatrix},\quad \lambda\in\mathbb{R}$$

O $C(A)$ indica o espaço colunar da matriz A.

Considerando alguns casos, o vetor $[-2\ \ -6]$ está no espaço colunar pois pode ser expressado com $\lambda=-2$. Porem $[1\ \ 4]$ não é um vetor do espaço colunar da matriz por que não há escalar que possa multiplicar a matriz e produzir esse vetor.

Para uma matriz com uma coluna, o espaco colunar e uma linha que atravesse a origem na direção do espaço vetorial e se alonga ao infinito em ambas as direções.



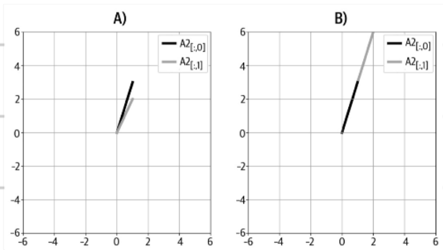
Considerando uma matriz com mais colunas:

$$C\left(\begin{bmatrix}1&1\\3&2\end{bmatrix}\right)=\lambda_1\begin{bmatrix}1\\3\end{bmatrix}+\lambda_2\begin{bmatrix}1\\2\end{bmatrix},\quad \lambda\in\mathbb{R}$$

Com duas colunas temos dois λs diferentes.

O conjunto de todos os vetores que podem ser alcançados por alguma combinação linear desses dois vetores de coluna são todos os vetores $\mathbb{R}^2$

Por exemplo, o vetor $[-4 \ \ 3]$ pode ser obtido escalonando as duas colunas por 11 e -15, respectivamente (esse escalares foram obtidos pelo metodo de projeção dos minimos escalares)



Outro exemplo é:

$$C\left(\begin{bmatrix}1 & 2\\3 & 6\end{bmatrix}\right) = \lambda_1\begin{bmatrix}1\\3\end{bmatrix} + \lambda_2\begin{bmatrix}2\\6\end{bmatrix}, \ \lambda\in\mathbb{R}$$

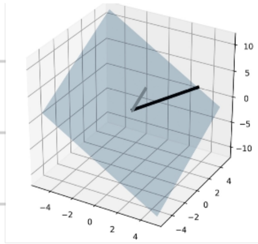
Essas duas colunas são colineares por que uma Jo é a versão reduzida da outra. O que significa que o espaco colunar dessa matriz 2x2 e apenas uma linha - Um subespaço 1D.

A dimensionalidade do espaço colunar é igual ao numero de colunas apenas se as colunas forem um conjunto linear independente.

Um ultimo exemplo:

$$C\left(\begin{bmatrix}3 & 0\\5 & 2\\1 & 2\end{bmatrix}\right) = \lambda_1\begin{bmatrix}3\\5\\1\end{bmatrix} + \lambda_2\begin{bmatrix}0\\2\\2\end{bmatrix}$$

Agora há 2 colunas em $\mathbb{R}^3$. Elas são linearmente independentes. Então o espaço colunar dessa matriz esta em 2D, mas é um plano 2D que esta embutido em $\mathbb{R}^3$



Há muitos vetores no plano. Mas há muito mai vetores fora dele.

Espaço linear

O espaco linear de uma matriz tem o mesmo conceito do colunar, mas consideramos todas as possiveis combinações ponderadas das linhas ao inves das colunas.

$$R(A) = C(A^T)$$

O espaço linear é invariante as operações de redução de linha

Como $R(A) = C(A^T)$, esses dois espaços matriciais serão idênticos para matrizes simetricas.

Matrizes - 2

Espaços Nulos

O espaço colunar pode ser resumido como:

$$Ax = b$$

Enquanto o espaço nulo, em contraste:

$$Ay = 0$$

"Podemos achar um conjunto de pesos — que não são todos 0 — e produzir o vetor zero?"

Qualquer vetor y que satisfaça essa equação esta no espaço nulo de A, escrito por $N(A) = \{\}$

Há uma estreita relação entre o dimensionalidade do espaco nulo e a independencia linear das colunas em uma matriz. O espaço nulo é vazio quando as colunas da matriz formam um conjunto linearmente independente

```
A = np.array([[1,-1],[-2,2]])
B = np.array([[1,-1],[-2,3]])
print(scipy.linalg.null_space(A))
print(scipy.linalg.null_space(B))
```

Dada os infinitos vetores possiveis, o Python ira retorna um vetor unidade.

Os vetores unidades são convenientes de trabalhar e tem otimas propriedades como a estabilidade numérica.

Rank (classificação)

Rank é um numero associado a uma matriz. Ele está relacionado a dimensionalidade dos subespaços da matriz e tem importantes implicações para as operações de matrizes, incluindo a inversão de matrizes e a determinação do número de soluções para um sistema de equações.

Algumas propriedades do rank:

- É un numero inteiro não-negativo.
- Toda matriz tem um unico rank. (rank é uma caracteristica da matriz, não de linhas ou colunas)
- É indicado por $r(A)$. ou $rank(A)$
- O rank maximo possivel é o menor quantidade entre linhas e colunas.
- Uma matriz com o maior rank possivel é chamada de "full-rank". Se $r < \{M,N\}$ então é chamada de "rank-reduzido".
- Multiplicações escalares não afetam o rank da matriz (com exceção do 0)

Há muitas interpretações e definições para o rank.

Matrizes - 2

• O maior numero de colunas (ou linhas) que formam um conjunto linearmente independente.

• A dimensionalidade do espaco colunar (que é a mesma dimensionalidade do espaço linear)

• O numero de dimensões que contem informações na matriz.

• A quantidade de valores singulares diferente de zero na matriz.

Ranks de Matrizes especiais

• Vetores: Todos os vetores tem rank 1. Isso porque vetores so possuem uma coluna (ou linha) de informação. A unica excessão é o vetor zero.

• Matriz zero: uma matriz zero de qualquer tamanho (incluindo o vetor zero) tem o rank 0.

• Matrizes Identidade: O rank da matriz identidade é igual ao numero de linhas/coluna. Ou seja, $r(I_N) = N$.

• Matrizes Diagonais: O rank de uma matriz diagonal é igual a quantidade de elementos diferentes de zero na diagonal. Isso porque cada linha contem no maximo um elemento diferente de zero, e é impossivel criar um numero diferente de zero por meio de combinações ponderadas de zero.

• Matrizes Triângulares: Uma matriz triangular tem sua classificação completa apenas se houver valores diferentes de zero em todos os elementos da diagonal. Se houver pelo menos um zero na diagona a matriz será de rank reduzido. (O rank exato irá depender dos valores numericos da matriz).

• Matrizes Aleatorias: A classificação de uma matriz aleatoria é impossivel de ser determinada a priori, porque ela depende da distribuição de numeros do qual os elemento da matriz foram extraidos e da probabilidade de extrair cada numero. Matrizes criadas com np.random.randn() terão o maximo de classificacao possivel.

Rank de Matrizes adicionadas e multiplicadas

A classificação de duas matrizes individuais fornecem limites para a maxima classificação possivel:

$$\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$$

$$\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$$

Matrizes - 2

Memorize:

- Você não consegue saber a classificação exata de uma matriz somada ou multiplicada se baseando apenas na classificação das matrizes individuais.

As matrizes individuais fornecem limites superiores a classificação das matrizes somadas ou multiplicadas.

- A classificação de uma matriz somada pode ser maior que a classificação das matrizes individuais
- A classificação de uma matriz multiplicada não pode ser maior do que a maior classificação das matrizes de multiplicação.

Rank de Matrizes Deslocadas

Simplificando: matrizes deslocadas tem a classificação completa. Um dos objetivos principais de deslocamento de uma matriz quadrada é aumentar seu rank de $r < M$ para $r = M$.

Aplicações de Rank

No espaço colunar?

Antes de explicar o algoritmo para determinar se $v \in C(A)$, é necessario explicar o procedimento chamado aumento de uma matriz.

Aumentar uma matriz significa adicionar colunas extras ao lado direito da matriz

$$\begin{bmatrix} 4 & 5 & 6 \\ 0 & 1 & 2 \\ 9 & 9 & 4 \end{bmatrix} \sqcup \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 & 1 \\ 0 & 1 & 2 & 2 \\ 9 & 9 & 4 & 3 \end{bmatrix}$$

O algoritmo para determinar se um vetor esta no espaço colunar da matriz e:

1. Aumente a matriz com o vetor.

Vamos chamar a matriz original de A e a matriz aumentada de $\tilde{A}$.

2. Calcule o rank das duas matrizes
3. Compare os dois ranks.

 a. $rank(A) = rank(\tilde{A}) \rightarrow$ vetor v está no espaço colunar da matriz A.

 b. $rank(A) < rank(\tilde{A}) \rightarrow$ vetor v não está no espaço colunar da matriz A.

Independencia linear de um conjunto vetor

O algoritmo é direto: coloque os vetores em uma matriz, calcule o rank da matriz e então compare o rank ao rank maximo possivel da matriz.

Matrizes - 2

· $r = M$: O conjunto vetor é linearmente independente

· $r < M$: O conjunto vetor é linearmente dependente

Determinante

A determinante é um numero associado a matriz quadrada. Na algebra linear abstrata, ela é a fundamental para muitas operações, incluindo a matriz inversa. Mas calcular a determinante na pratica pode ser numericamente instavel para grandes matrizes, devido a problemas de subfluxo e excesso de valores numericos.

As duas propriedades mais importantes das determinantes são:

1. Só é valida para matrizes quadradas
2. E zero para matrizes de rank reduzido

A determinante é anotada como $\det(A)$ ou $|A|$. A letra Grega maiuscula delta Δ é usada quando você não precisa fazer referecia a uma matriz especifica.

A determinante tem uma interpretação geometrica, que é relacionada a quanto a matriz alonga vetores durante a multiplicação matriz-vetor. Uma determinante negativa significa que um eixo e rotacionado durante a transformação.

Calcular a Determinante

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

A determinante não é limitada a valores inteiros ou positivos.

Calcular a determinante e simplesmente o produto da diagonal menos o produto da diagonal inversa.

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{vmatrix} = \begin{matrix} afkp - aflo - agjp + agln + ahjo - ahkn - bekp + belo \\ + bgip - bglm - bhio + bhkm + cejp - celn - cfip + cflm \\ + chin - chjm - dejo + dekn + dfio - dfkm - dgin + dgjm \end{matrix}$$

No python:

```
np.linalg.det() # ou
scipy.linalg.det()
```

Determinantes com Dependencias Lineares

Como foi dito anteriormente, a determinante para qualquer matriz reduzida e zero.

$$\begin{vmatrix} a & \lambda a \\ c & \lambda c \end{vmatrix} = ac\lambda - a\lambda c = 0$$

Esse conceito generaliza para matrizes

maiores.

A característica polinomial

A combinação da matriz deslocada com o determinante é chamada de característica polinomial da matriz

$$\det(A-\lambda I) = \Delta$$

$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2-(a+d)\lambda+(ad-bc) = 0$$

$$\begin{vmatrix} a-\lambda & b & c \\ d & e-\lambda & f \\ g & h & i-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \begin{gathered} -\lambda^3+(a+e+i)\lambda^2 \\ +(-ae+bd-ai+cg-ei+fh)\lambda \\ +aei-afh-bdi+bfg+cdh-ceg = 0 \end{gathered}$$

$$\det\left(\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} - \lambda I\right) = 0$$

$$\begin{vmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{vmatrix} = 0 \rightarrow (1-\lambda)^2 - 9 = 0 \rightarrow 1-\lambda = \pm 3 \rightarrow$$

$$\rightarrow \lambda_1 = -2, \quad \lambda_2 = 4$$

$$\lambda = -2 \rightarrow \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$$

$$\lambda = 4 \rightarrow \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix}$$

As duas matrizes tem rank 1. Além disso, ambas possuem espacos nulos não triviais; o que significa que você pode encontrar alguns vetores y diferentes de zero que satisfaça $(A-\lambda I)y=0$.

Resumo

- Há muitos tipos de normas matriciais, que podem ser amplamente classificadas em elementares e induzidas. O primeiro reflete a magnitude dos elementos na matriz enquanto o ultimo reflete o efeito geometrico-transformativo da matriz no vetor.
- A norma elementar mais comumente usada é a norma de Frobenius (conhecida por norma Euclidiana ou norma l_2) e é calculada como a raiz quadrada da soma dos elementos ao quadrado.
- O traço da matriz é a soma dos elementos diagonais
- Há quatro espaços matriciais (colunar, linear, nulo e nulo-esquerdo), e eles são definidos como o conjunto de combinações lineares ponderadas de diferentes características da matriz.
- O espaco coluna da matriz compreende toda a combinação linear ponderada das colunas na matriz e é escrito como $C(A)$.
- O espaco linha da matriz compreende toda a combinação linear ponderada das linhas na matriz e é escrito como $R(A)$ ou $C(A^T)$

Matrizes - 2

- O espaco nulo da matriz é o conjunto de vetores que combina linearmente as colunas para produzir o vetor zero, em outras palavras, qualquer vetor y que resolva a equação $Ay=0$. A solução trivial $y=0$ é excluida. O espaço nulo é importante para encontrar os auto vetores, entre outras aplicações.
- Rank é um inteiro não negativo associado a uma matriz. O rank reflete a maior quantidade de colunas (ou linhas) que formam um conjunto linearmente independente. Matrizes com um rank menor que o maximo possivel são chamadas de rank reduzidas ou singulares.
- Deslocar uma matriz quadrada adicionando uma constante a diagonal assegura o rank completo
- Uma das muitas aplicações do rank é determinar se um vetor esta no espaço coluna da matriz.
- A determinante é um numero associado a matriz quadrada. Ela é 0 para todas as matrizes de rank reduzido e diferente de zero para as de rank completo
- A caracteristica polinomial transforma uma matriz quadrada, deslocada por λ, em uma equação igual a determinante