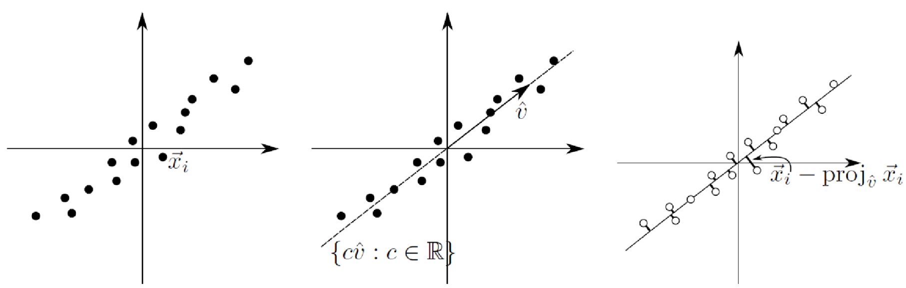
Metodi Numerici per l'Informatica

Anthony

4 apr 2023

1 Asse principale

Consideriamo dei dati in $\mathbb{R}^2$. Supponiamo di voler trovare un vettore $\mathbf{v}$ che minimizza la distanza tra i punti.



(a) Input data (b) Principal axis (c) Projection error

Stiamo cercando una retta, in modo analogo a abbiamo visto con la regressione lineare, ma il criterio di minimizzazione è diverso: vogliamo minimizzare la somma di tutte le norme dei complementi ortogonali.

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \text{proj}_{\mathbf{v}}\mathbf{x}_i\|_2^2 \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Il vincolo sulla norma non è necessario, ma semplifica come risolvere il problema di ottimizzazione. Il vettore $\mathbf{v}$ che minimizza l'errore si chiama *asse principale*. Osserviamo che possiamo traslare il problema in notazione matriciale:

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \text{proj}_{\mathbf{v}}\mathbf{x}_i\|_2^2 \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

$$= \min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \frac{\mathbf{x}_i^T\mathbf{v}}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\|_2^2 \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Dato che la norma è uguale a 1:

$$= \min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{v})\mathbf{v}\|_2^2 \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Espandiamo la norma al quadrato:

$$= \min_{\mathbf{v}} \sum_i (\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{v})\mathbf{v})^T(\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{v})\mathbf{v}) \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Effettuiamo i prodotti e semplici passaggi algebrici:

$$= \min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - 2(\mathbf{x}_i^T\mathbf{v})(\mathbf{x}_i^T\mathbf{v}) + (\mathbf{x}_i^T\mathbf{v})^2\|\mathbf{v}\|_2^2) \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

$$= \min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - 2(\mathbf{x}_i^T\mathbf{v})^2 + (\mathbf{x}_i^T\mathbf{v})^2\|\mathbf{v}\|_2^2) \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

$$= \min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - (\mathbf{x}_i^T\mathbf{v})^2) \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Semplifichiamo eliminando il valore che non dipende da $\mathbf{v}$:

$$= \min_{\mathbf{v}} - \sum_i (\mathbf{x}_i^T\mathbf{v})^2 \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Ora sia $\mathbf{X}$ la matrice che contiene i vettori $\mathbf{x}_i$ come colonne:

$$= \min_{\mathbf{v}} -\|\mathbf{X}^T\mathbf{v}\|_2^2 \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

$$= \max_{\mathbf{v}} \|\mathbf{X}^T\mathbf{v}\|_2^2 \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Che può anche essere riscritto come:

$$\max_{\mathbf{v}} \mathbf{v}^T\mathbf{X}\mathbf{X}^T\mathbf{v} \quad \text{t.c. } \|\mathbf{v}\|_2 = 1$$

Il massimizzatore che risolve il problema $\mathbf{v}^*$ è il *componente principale* dei dati contenuti nella matrice $\mathbf{X}$.

2 Autovettori e autovalori

Un autovettore $\mathbf{x}$ di una matrice quadrata $\mathbf{A}$ è un qualsiasi vettore che soddisfa la seguente espressione:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \text{per qualche valore } \lambda \text{ anche complesso}$$

Il valore λ è l'*autovalore* dell'autovettore $\mathbf{x}$. Intuitivamente, gli autovettori di una mappa lineare $\mathbf{A}$ sono quei vettori che, quando applichiamo $\mathbf{A}$, vengono solo scalati del valore λ, non vengono ruotati.

La scala di un autovettore non è importante In particolare:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda c\mathbf{x}$$

Per qualunque c, otteniamo sempre lo stesso λ. Per questa ragione ci limiteremo quindi a cercare gli autovettori di norma 1.

Similarità di trasformazioni lineari Consideriamo la seguente trasformazione lineare invertibile:

$$\mathbf{ATx} = \lambda\mathbf{Tx}$$

Stiamo dicendo che $\mathbf{Tx}$ è un autovettore di $\mathbf{A}$ con autovalore λ. Ora proviamo ad applicare la trasformazione lineare $\mathbf{T}^{-1}$ a sinistra:

$$\mathbf{T}^{-1}\mathbf{ATx} = \lambda\mathbf{x}$$

Possiamo osservare che $\mathbf{x}$ è un autovettore di $\underbrace{\mathbf{T}^{-1}\mathbf{AT}}_{\mathbf{B}}$ con autovalore λ. Poiché le trasformazioni lineari $\mathbf{A}$ e $\mathbf{B}$ hanno gli stessi autovalori, diciamo che esse sono *simili* o *co-spettrali* o *iso-spettrali*.

Autovalori di matrici ortogonali Osserviamo ora una proprietà importante delle matrici ortogonali, esse hanno tutti autovalori ± 1:

$$\begin{gathered}
\mathbf{Qx} = \lambda\mathbf{x} \\
= \|\mathbf{Qx}\|_2^2 = \|\lambda\mathbf{x}\|_2^2 \\
= (\mathbf{Qx})^T\mathbf{Qx} = |\lambda|^2\|\mathbf{x}\|_2^2 \\
= \mathbf{x}^T\mathbf{Q}^T\mathbf{Qx} = |\lambda|^2\|\mathbf{x}\|_2^2 \\
= \mathbf{x}^T\mathbf{x} = |\lambda|^2\|\mathbf{x}\|_2^2 \\
= \|\mathbf{x}\|_2^2 = |\lambda|^2\|\mathbf{x}\|_2^2 \\
= 1 = |\lambda|^2 \\
\lambda = \pm 1
\end{gathered}$$

Autovalori di matrici diagonali o triangolari superiori Gli autovalori di tali matrici sono i valori sulla diagonale principale.

Autovalori di matrici simmetriche Ricordiamo che una matrice $\mathbf{A}$ è simmetrica se $\mathbf{A} = \mathbf{A}^T$.
Possiamo affermare che tutti gli autovalori di una matrice simmetrica sono reali. Consideriamo la coppia autovalore-autovettore $(\lambda, \mathbf{x})$ tale che $\|\mathbf{x}\|_2^2 = 1$:

$$\begin{aligned}
\lambda\mathbf{x}^T\mathbf{x} &= (\lambda\mathbf{x})^T\mathbf{x} \\
&= (\mathbf{Ax})^T\mathbf{x} \\
&= \mathbf{x}^T\mathbf{A}^T\mathbf{x} \\
&= \mathbf{x}^T(\mathbf{Ax}) \\
&= \overline{(\mathbf{Ax})^T\mathbf{x}} \quad \text{in cui } \overline{z} \text{ è il coniugato di } z \\
&= \overline{(\lambda\mathbf{x})^T\mathbf{x}} \\
&= \overline{\lambda}\mathbf{x}^T\mathbf{x}
\end{aligned}$$

Consideriamo ora due distinti autovettori $\mathbf{x}_i, \mathbf{x}_j$ con $\lambda_i \neq \lambda_j$:

$$(\mathbf{A}\mathbf{x}_i)^T\mathbf{x}_j = \mathbf{x}_i^T\mathbf{A}\mathbf{x}_j$$

$$(\lambda_i\mathbf{x_i})^T\mathbf{x}_j = \mathbf{x}_i^T\lambda_j\mathbf{x}_j$$

Questo significa che $\mathbf{x}_i^T\mathbf{x}_j = 0$, ovvero $\mathbf{x}_i$ e $\mathbf{x}_j$ sono ortogonali.

Autovettori di matrici che commutano Consideriamo due matrici $\mathbf{A}$ e $\mathbf{B}$. Possiamo dire che:

$$\mathbf{AB} = \mathbf{BA} \Leftrightarrow \mathbf{A} \text{ e } \mathbf{B} \text{ hanno gli stessi autovettori}$$

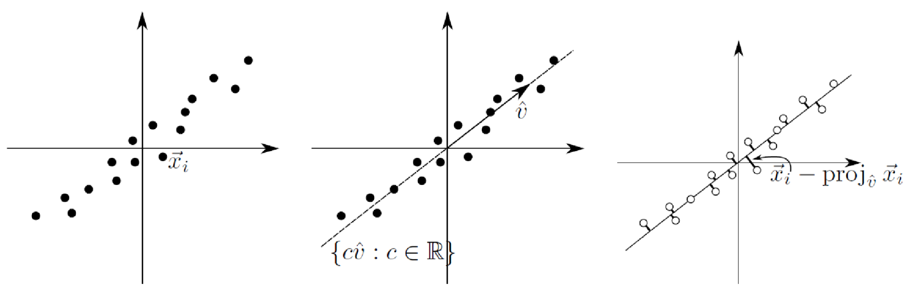
Possiamo anche osservare che non è detto che $\mathbf{A}$ e $\mathbf{B}$ abbiano gli stessi autovalori:

$$\mathbf{v}^T\mathbf{A}\mathbf{v} = \lambda \neq \mu = \mathbf{v}^T\mathbf{B}\mathbf{v}$$

Teorema di Courat L'autovettore associato all'autovalore massimo si chiama *autovettore principale*.

3 Asse principale (continua)

Torniamo al nostro problema: dobbiamo fittare una retta tra dei punti minimizzando l'errore quadratico medio perpendicolare alla retta:



(a) Input data (b) Principal axis (c) Projection error

Per risolvere il problema dobbiamo trovare il vettore $\mathbf{v}$.

$$\max_{\mathbf{v}} \mathbf{v}^T \underbrace{\mathbf{X}\mathbf{X}^T}_{\text{simmetrica}} \mathbf{v}$$

3.1 Teorema min-max

Se $\mathbf{A}$ è una matrice simmetrica, allora il suo autovalore massimo è dato dalla seguente equazione:

$$\max_{\mathbf{v}} \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2}$$

con $\mathbf{v}$ il corrispettivo autovettore. Più in generale:

$$\lambda_{\min} \leq \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_2^2} \leq \lambda_{\max}$$

In cui il quoziente $\frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_2^2}$ è chiamato *quoziente di Rayleigh*.

4 Teorema spettrale

L'insieme di autovalori $\{\lambda_i\}$ di una matrice $\mathbf{A}$ è chiamata *spettro*. Possiamo scrivere l'equazione degli autovalori come:

$$\mathbf{A}\mathbf{X} = \mathbf{X}\boldsymbol{\Lambda}$$

Se $\mathbf{A}$ è simmetrica, allora $\mathbf{X}$ è una matrice ortogonale di autovettori e $\boldsymbol{\Lambda}$ è la matrice diagonale con autovalori reali. L'equazione appena mostrata possiamo scriverla anche nel seguente modo:

$$\mathbf{X}^T \mathbf{A} \mathbf{X} = \boldsymbol{\Lambda}$$

E tale equazione è chiamata la *decomposizione spettrale* di $\mathbf{A}$. Osserviamo che essa è molto simile al problema dell'asse principale:

$$\max_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} \quad \text{t.c. } \|\mathbf{x}\|_2 = 1 \quad \text{asse principale}$$

Modifichiamo il problema per risolvere per tutti gli autovettori e i rispettivi autovalori:

$$\min_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X}) \quad \text{t.c. } \mathbf{X}^T \mathbf{X} = \mathbf{I} \quad \text{decomposizione spettrale}$$

In questo problema stiamo minimizzando con il vincolo che $\mathbf{X}$ sia ortogonale. Tale quesito non è convesso, ovvero non è banale trovare una soluzione globale al problema ma, per questo in problema in particolare, esistono algoritmi che permettono di risolverlo globalmente.

5 Trovare gli autovalori

Autovettore massimo (power iteration) Un algoritmo molto semplice per trovare l'autovalore più grande è quello della *normalized-iteration*, mostrato in seguito:

Algorithm 1 Normalized-Iteration

```
function NORMALIZED-ITERATION(A)
    v⃗ ← Arbitrary(n)
    for k ← 1, 2, 3, ... do
        w⃗ ← Av⃗
        v⃗ ← w⃗/‖w⃗‖          ▷ la normalizzazione è atta a ridurre errori numerici
    end for
    return v⃗
end function
```

Autovalore minimo (inverse iteration) Per trovare l'autovettore minimo osserviamo che:

$$\mathbf{Ax} = \lambda\mathbf{x} \Rightarrow \mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Da ciò ne concludiamo che possiamo semplicemente applicare il power method a $\mathbf{A}^{-1}$. Tuttavia, esistono tecniche più efficienti che non prevedono l'inversa di $\mathbf{A}$, come la decomposizione LU.

5.1 Iterazione QR

Abbiamo già visto nella sezione 2. che se $\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}$, allora $\mathbf{A}$ e $\mathbf{B}$ hanno lo stesso spettro, ovvero sono matrici simili. Per cui potremmo prima ottenere una decomposizione $\mathbf{QR}$ di $\mathbf{A}$:

$$\mathbf{A} = \mathbf{QR}$$

E poi sovrascrivere $\mathbf{A}$, mantenendo ancora gli stessi autovalori:

$$\mathbf{A} \leftarrow \mathbf{Q}^T\mathbf{AQ}$$

Ora osserviamo:

$$\mathbf{Q}^T\mathbf{AQ} = \mathbf{Q}^T(\mathbf{QR})\mathbf{Q} = (\mathbf{Q}^T\mathbf{Q})\mathbf{RQ} = \mathbf{RQ}$$

In altre parole, questa operazione preserva gli autovalori:

$$\mathbf{A} \leftarrow \mathbf{RQ}$$

Se continuiamo ad effettuare questa operazione in modo iterativo, otteniamo l'algoritmo *QR-Iteration*:

Algorithm 2 QR-Iteration

```
function QR-ITERATION(A ∈ ℝ^{n×n})
    for k ← 1, 2, 3, ... do
        Q, R ← Fattorizzazione-QR(A)
        A ← RQ
    end for
    return diag(R)
end function
```

Facciamo alcune osservazioni:

1. Tutte le matrici intermedie generate dall'algoritmo $\mathbf{A}_1, \mathbf{A}_2, \ldots$ hanno gli stessi autovalori
2. Le iterazioni convergono, ovvero a un certo punto $\mathbf{A}_k \approx \mathbf{A}_{k-1}$
3. A convergenza, otteniamo $\mathbf{Q}_\infty^T \mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{A}_\infty$, il che implica $\mathbf{Q}_\infty \mathbf{R}_\infty = \mathbf{R}_\infty \mathbf{Q}_\infty$ e $\mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{Q}_\infty \mathbf{A}_\infty$
4. Se $\mathbf{A}_\infty \mathbf{x} = \lambda \mathbf{x}$ allora possiamo scrivere:

$$\lambda \mathbf{x} = \mathbf{A}_\infty \mathbf{x} = \mathbf{Q}_\infty \mathbf{R}_\infty \mathbf{x} = \mathbf{R}_\infty \mathbf{Q}_\infty \mathbf{x} = \pm \mathbf{R}_\infty \mathbf{x}$$

   L'ultima parte dell'uguaglianza deriva dal fatto che gli autovalori di $\mathbf{Q}$ sono ± 1 poiché essa è ortogonale. In altre parole, gli autovalori di $\mathbf{A}_\infty$, e quindi anche quelli di $\mathbf{A}$, sono uguali agli elementi sulla diagonale di $\mathbf{R}_\infty$ a parte il segno.

6 Shifting

Data una matrice $\mathbf{A}$ con autovalori $\{\lambda_i\}$ e autovettori $\{\mathbf{x}_i\}$:

$$(\mathbf{A} - \sigma \mathbf{I}) \mathbf{x}_i = \mathbf{A} \mathbf{x}_i - \sigma \mathbf{x}_i = \lambda_{\mathbf{x}_i} - \sigma \mathbf{x}_i = (\lambda_i - \sigma) \mathbf{x}_i$$

Che significa che gli autovalori di $\mathbf{A} - \sigma \mathbf{I}$ sono $\lambda_i - \sigma$, ovvero effettuando uno shift sulla matrice otteniamo lo stesso shift anche sui suoi autovalori.
Possiamo osservare che se σ è vicino all'autovalore di $\mathbf{A}$, allora $\mathbf{A} - \sigma \mathbf{I}$ ha un autovalore vicino allo 0. Possiamo usare questo fatto per stimare porzioni dello spettro di $\mathbf{A}$:

1. Effettuiamo un ipotesi su un autovalore
2. Computiamo la matrice shiftata $\mathbf{B} = \mathbf{A} - \sigma \mathbf{I}$
3. Applichiamo l'inverse iteration su $\mathbf{B}$ per trovare l'autovalore minimo.