

Probabilità e Statistica in Alta Dimensione: Esercitazione 1 (Johnson-Lindenstrauss)

18 Novembre 2024

Ogni studente generi un dataset sintetico costituito da N vettori in \mathbb{R}^d le cui componenti sono distribuite normalmente, $u^1, u^2, \dots, u^N \in \mathbb{R}^d$, con $N = 10$ e $d = 300$.

Si definiscano le soglie $\epsilon = 2/5$ e $\delta = 1/2$.

Per $m < d$ sia A una matrice $d \times m$, con entrate i.i.d. gaussiane standardizzate. Si definisca la funzione $F : \mathbb{R}^d \mapsto \mathbb{R}^m$,

$$F := \frac{A}{\sqrt{m}} \quad (1)$$

1. Task teorico 1: applicazione di Johnson Lindenstrauss embedding:

Applicando il lemma di Johnson Lindenstrauss si stimi la dimensione m più piccola affinché $F(u^1), \dots, F(u^N)$ soddisfi la relazione

$$1 - \delta \leq \frac{\|F(u^i) - F(u^j)\|^2}{\|u^i - u^j\|^2} \leq 1 + \delta \quad (2)$$

per ogni $i \neq j$ con probabilità almeno $1 - \epsilon$. Si denoti d_{min} tale scelta ottimale della dimensione m .

2. Task numerico 1:

Assumendo che $m = d_{min}$, si determini numericamente un insieme di vettori $v^1, v^2, \dots, v^N \in \mathbb{R}^m$ tali che per ogni i, j con $i \neq j$ si abbia

$$1 - \delta \leq \frac{\|v^i - v^j\|^2}{\|u^i - u^j\|^2} \leq 1 + \delta. \quad (3)$$

Si riporti sul foglio di consegna una tabella con i vettori v^1, v^2, \dots, v^N

3. Task numerico 2:

Si rappresenti un grafo i cui nodi corrispondono ai vettori u^1, u^2, \dots, u^N tale che c'è un arco indiretto *colorato in blu* tra u^i e u^j se e solo se

$$\|u^i - u^j\| \leq \sqrt{2d},$$

e un arco indiretto *colorato in rosso* tra u^i e u^j se e solo se

$$\|v^i - v^j\| \leq \sqrt{2d}.$$

La struttura del grafo con archi rossi dovrebbe essere simile alla struttura del grafo con archi blu.

La consegna è individuale. Il foglio deve essere consegnato in formato PDF entro una settimana dall'esercitazione ed inviato tramite email ai docenti. Allegare lo script in Matlab per completezza.