Johnson-Lindenstrauss lema
Kratek opis

Klara Doler in Eva Winkler

17. 12. 2021

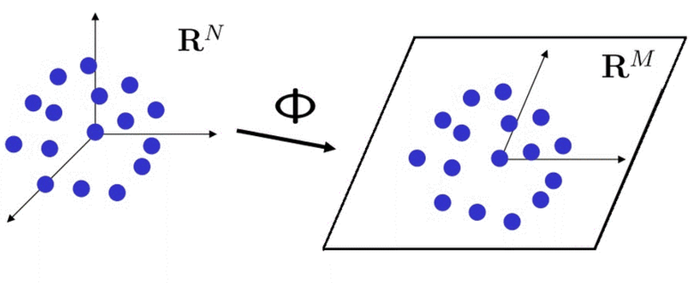
1 Definicije in leme

Johnson-Lindenstraussova lema
Naj bo $0 < \varepsilon < 1$ in $n \in \mathbb{N}$. Predspostavimo, da je $k > \frac{8\log(n)}{\varepsilon^2}$. Potem za vsak nabor m točk iz množice $X \subset \mathbb{R}^d$ obstaja taka preslikava $f : \mathbb{R}^d \to \mathbb{R}^k$, da za vse $u, v \in X$ velja:

$$(1-\varepsilon)\|u-v\|^2 \leqslant \|f(u)-f(v)\|^2 \leqslant (1+\varepsilon)\|u-v\|^2$$

Zgornje lahko zapišemo tudi kot:

$$(1+\varepsilon)^{-1}\|f(u)-f(v)\|^2 \leqslant \|u-v\|^2 \leqslant (1-\varepsilon)^{-1}\|f(u)-f(v)\|^2$$



Slika 1: Slika prikazuje preslikavo točk s $f : \mathbb{R}^N \to \mathbb{R}^M$, za $N > M$

Johnson-Lindenstraussova distribucija
Sorodna lema je distribucijska Johnson-Lindenstraussova lema. Ta lema pravi, da za katerakoli $0 < \varepsilon$, $\delta < \frac{1}{2}$ in pozitivno celo število d, obstaja porazdelitev na $\mathbb{R}^{k\times d}$, iz katere je sestavljena matrika A, tako da za $k = O(\varepsilon - 2\log(\frac{1}{\delta}))$ in za kateri koli enotski vektor x $\in \mathbb{R}^d$ velja spodnja trditev:

$$P(|\|Ax\|_2^2 - 1| > \varepsilon) < \delta$$

Vidimo lahko, da lahko lemo JL pridobimo iz distribucijske različice tako, da nastavimo $x = \frac{u-v}{\|u-v\|^2}$ in $\delta < \frac{1}{n^2}$ za nek par $u, v \in X$.

Definicija 1.1 Končno razsežen vektorski prostor nad $\mathbb{R}$ opremljen s skalarnim produktom imenujemo evklidski prostor.

Definicija 1.2 V n-razsežnem evklidskem prostoru $\mathbb{R}^n$ je dolžina vektorja $\vec{x} = [x_1, x_2, \cdots x_n]$ določena z

$\vec{x} = \sqrt{(x_1)^2 + (x_2)^2 + (x_3)^2 + \cdots + (x_n)^2}$.

Imenujemo jo evklidska norma.

2 Opis problema in navodilo

V projektni nalogi se bova ukvarjali z Johnson-Lindenstraussovo lemo, poimenovano po Williamu B. Johnsonu in Joramu Lindenstraussu. Lema pravi, da je mogoče množico točk iz visokodimenzionalnega prostora projecirati v podprostor dimenzije $O(\frac{log(n)}{\varepsilon^2})$ tako, da se razdalja med poljubnima točkama z veliko verjetnostjo spremeni za največ ε. V najini nalogi bova preverili veljavnost leme.

3 Opis dela

Za izvedbo najinega problema bova uporabili programski jezik *Python*. Programirali bova v $Cocalc - u$, ki ima vgrajene funkcije za delo z matrikami. Izvedli bova poskuse, s katerimi bova preverili, ali distribucijska JL lema drži:

- izbrali bova vzorec točk v prostoru določene dimenzije, ki je odvisen od ε,
- naredili bova projekcijo izbranih točk na podprostor,
- primerjale bova razdaljo med točkami v podprostoru z originalno razdaljo,
- za konec bova še pogledali, kakšna je verjetnost, da je razdalja med točkami spremenjena za največ ε (ter npr. $\frac{\varepsilon}{2}$).