

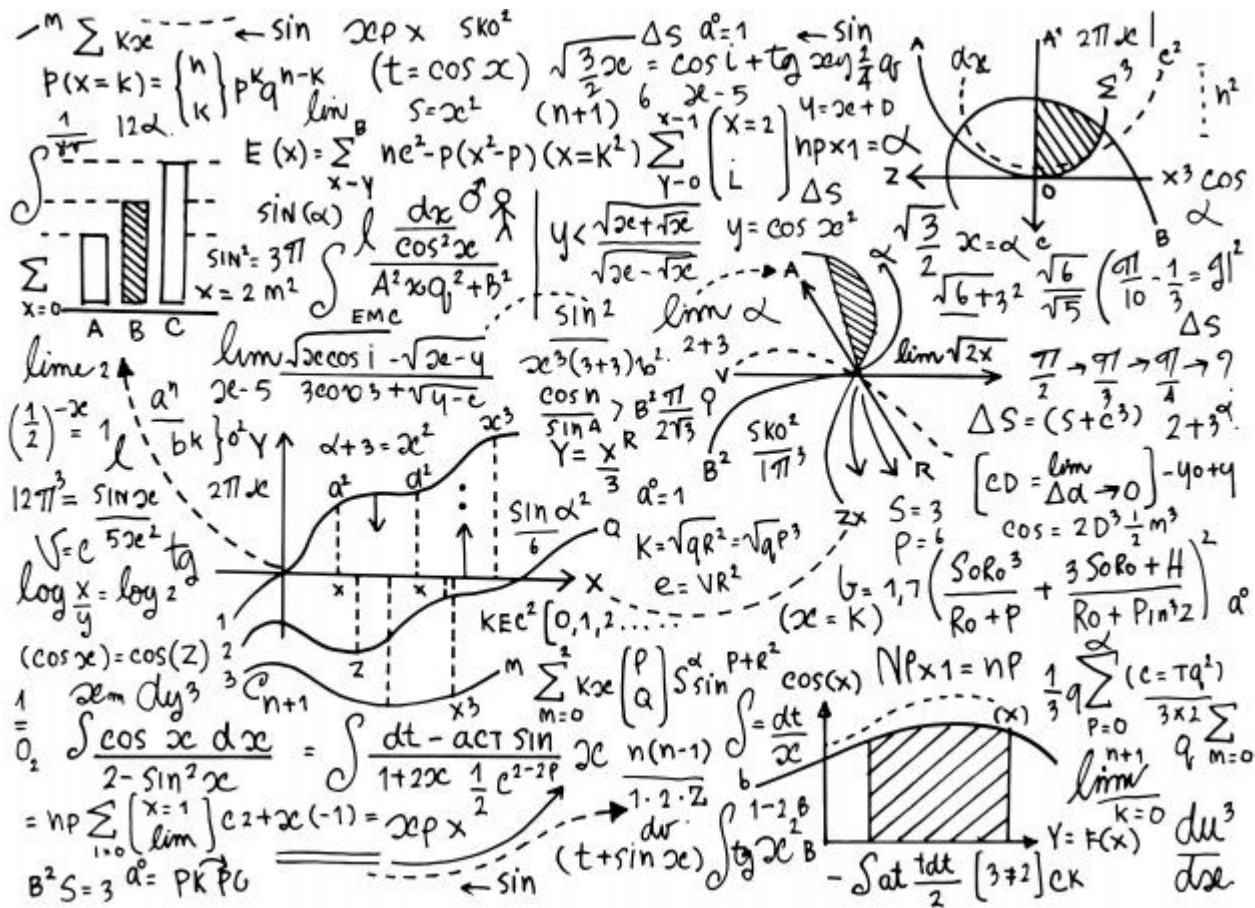
Problemes S4

Outline

Sessió 2: Classificació

Sessió 3: Backprop

Sessió 4: Memorització



Què es feia abans?

Outline

Sessió 2: Intro + Classificació

Sessió 3: Nets + Backprop

Sessió 4: KNN + Memorització



Què voldriem fer ara?

Entregues

Sessió 2: Intro + Classificació

Regresor Logistic + SVM

Sessió 3: Nets + Backprop

Feedforward + CNN

Sessió 4: KNN + Memorització
NN search (raw data + features)

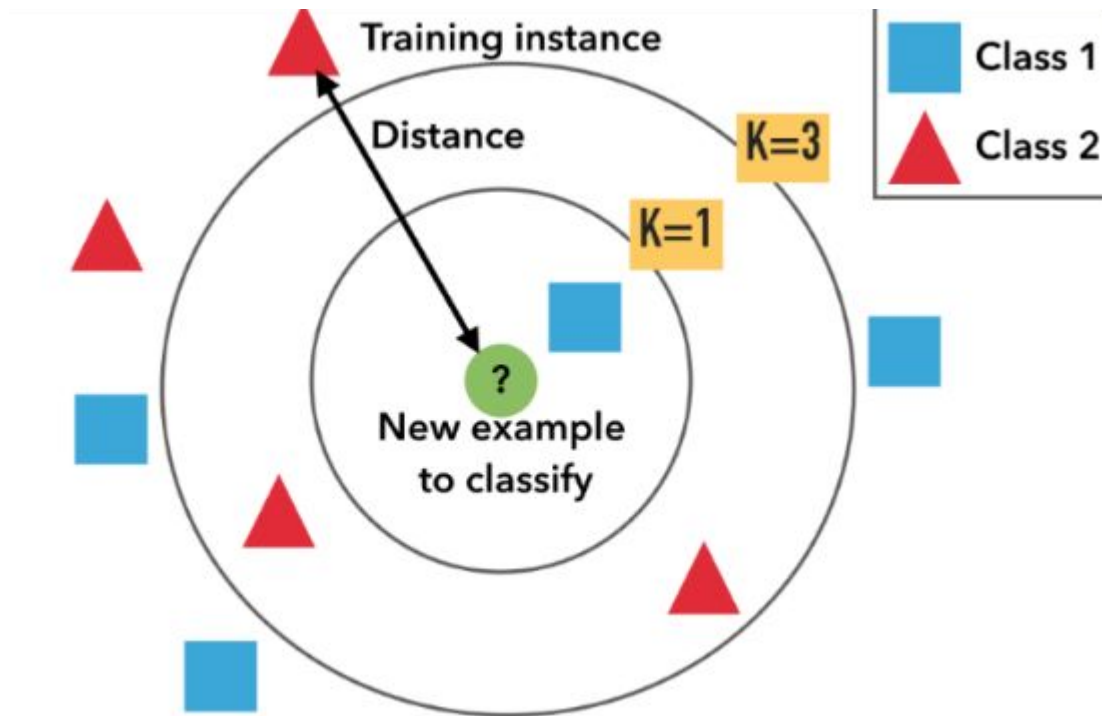


Haureu d'entregar un informe sobre Jupyter Notebook amb el codi explicant el que heu fet

Sessió 4: KNN, memorització

Búsqueda dels K-veïns més propers

KNN



Nearest Neighbour

- Brute-force
- Approximate
 - sklearn: KD-Tree
 - sklearn: Ball-Tree
 - annoy (Approximate Nearest Neighbors Oh Yeah): used in spotify

D'altres implementacions més eficients (però més difícils de instal·lar..)

- NMSLIB(Non-Metric Space Library) used in Amazon Elasticsearch & Yandex
- NGT(Neighborhood Graph and Tree for Indexing High-dimensional Data): developed at yahoo
- FAISS (Fair AI Similarity Search): developed at facebook
- SCANN (Scalable Nearest Neighbors): developed at google

Brute-force

for i in query item:

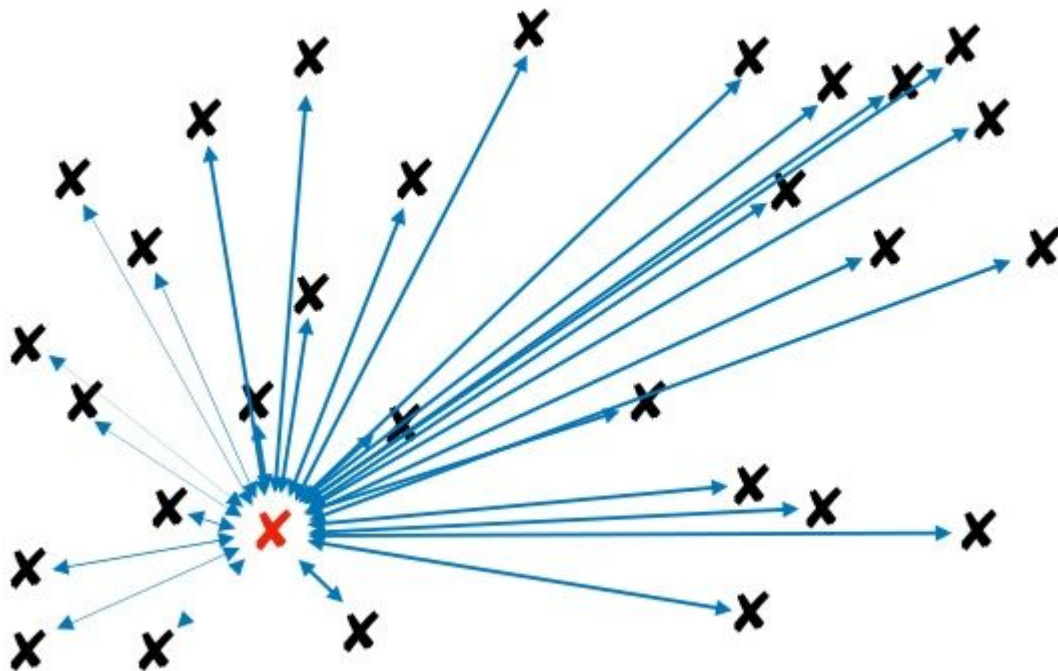
for j in train item:

compute distance(i,j)

time complexity:

$$O(d * n^2)$$

fast for small n (<100)



Sklearn: KDtree

Distribute data in a tree

Split the data,

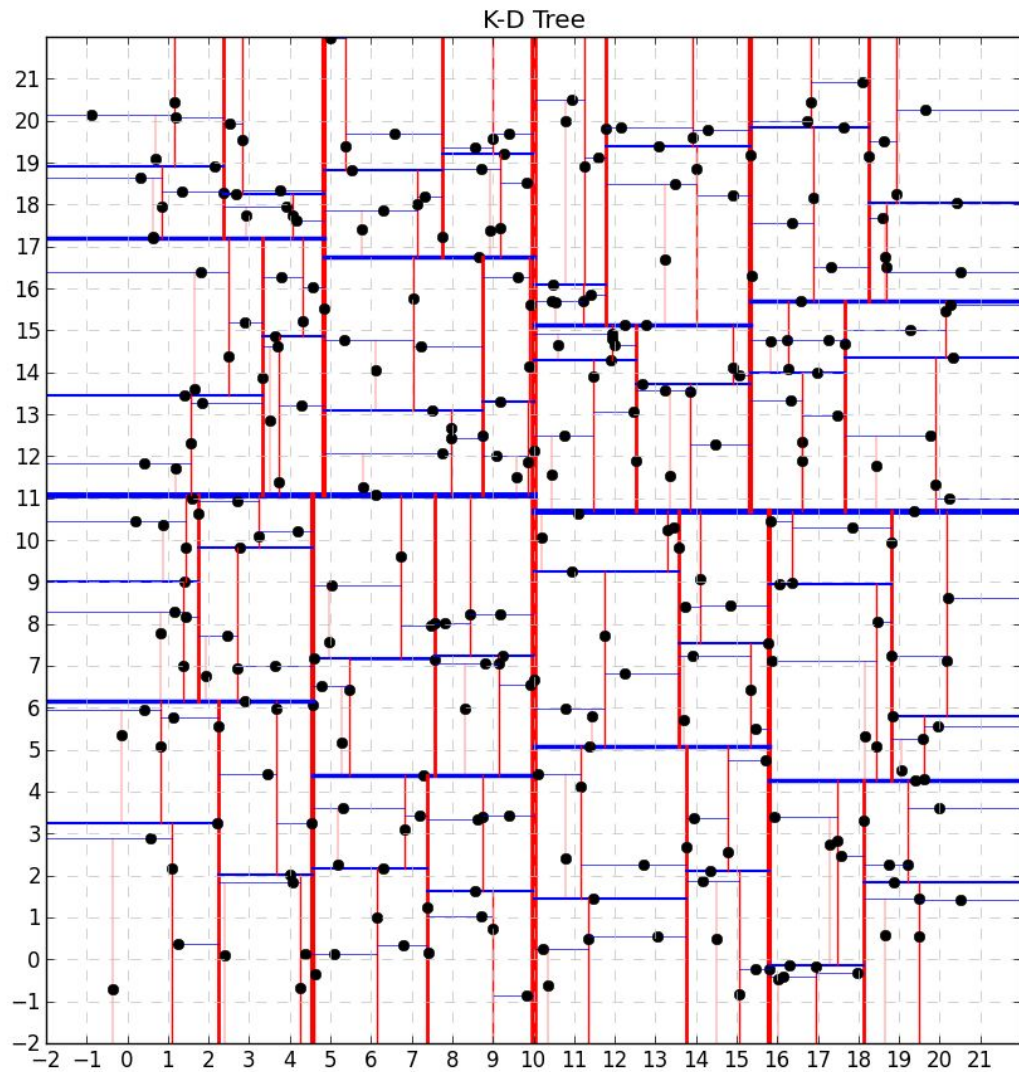
one dimension each time

time complexity:

$$O(d * n * \log(n))$$

faster with n

costly for large d



Sklearn: BallTree

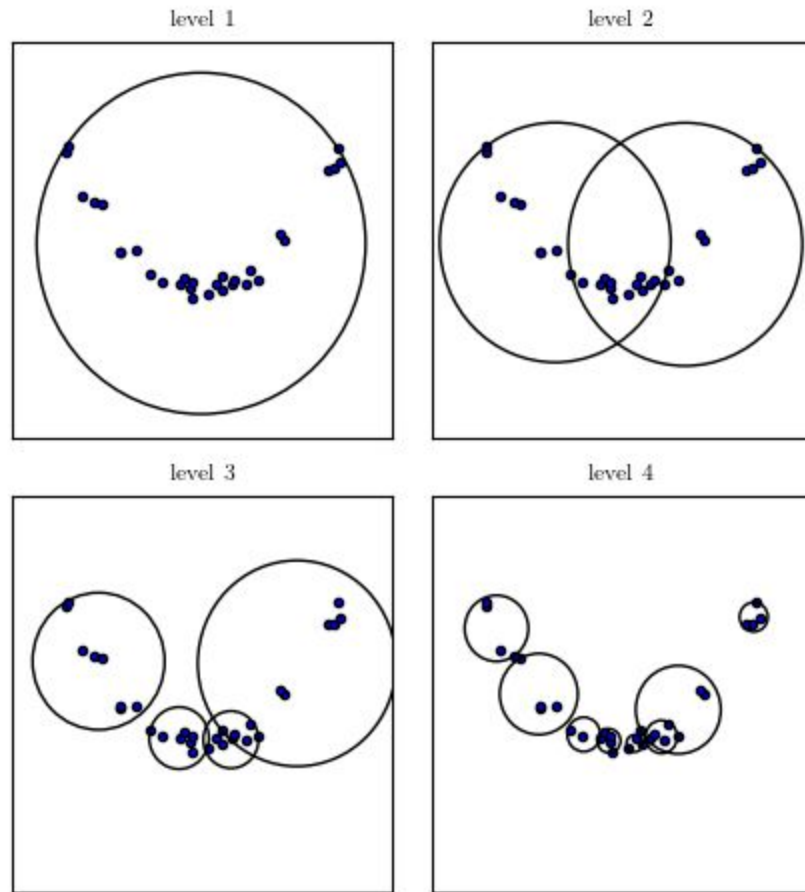
ball trees partition data
of nesting hyper-spheres

time complexity:

$$O(d * n * \log(n))$$

faster with n
efficiency related to data

Ball-tree Example



Annoy

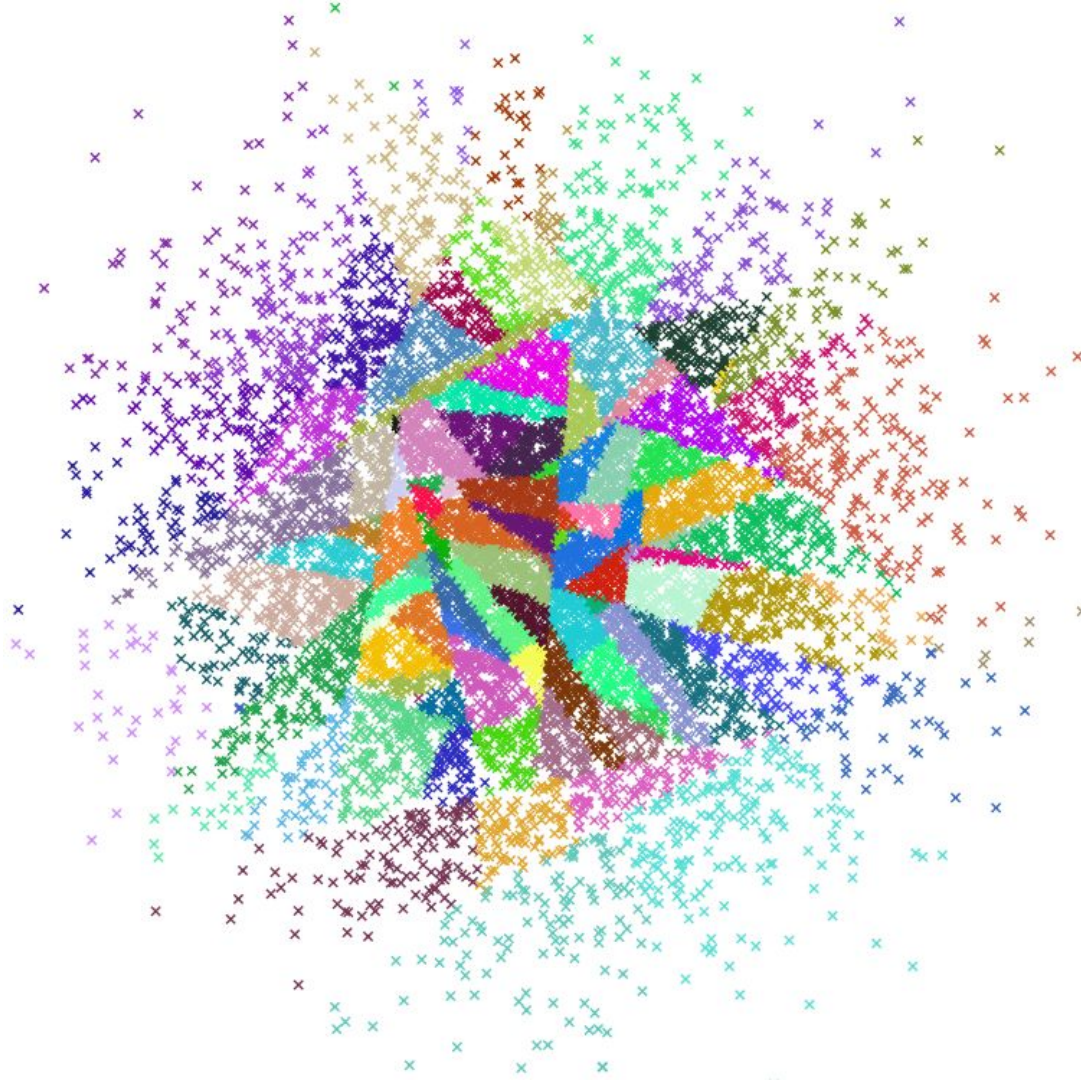
multiple trees

with random hyper-planes

time complexity:

$$O(d * n * \log(n))$$

the evil is in the details..



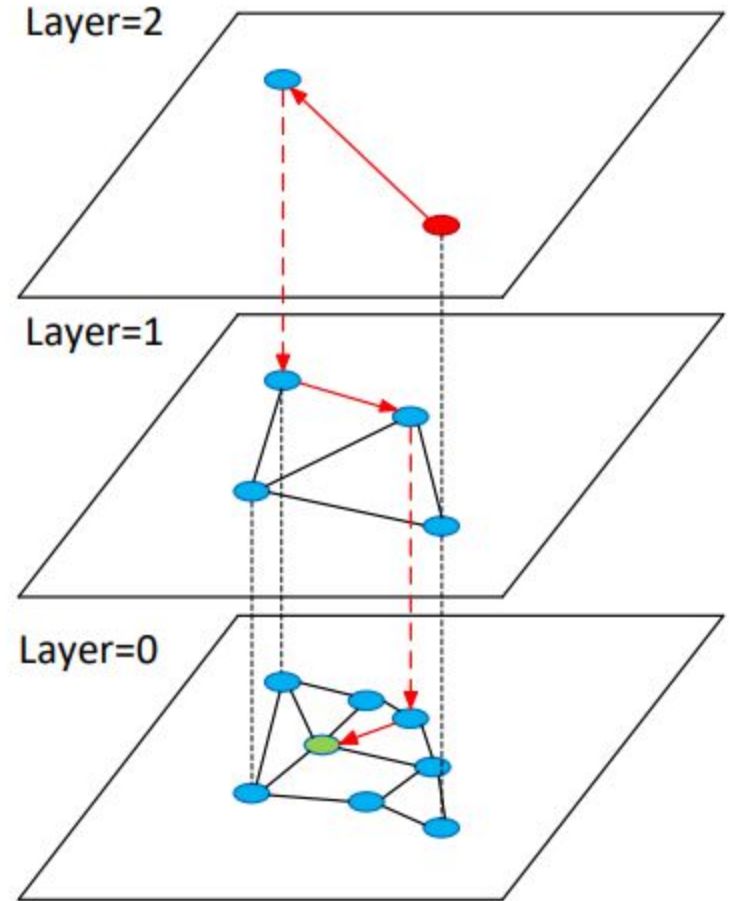
NMSLib

Hierarchical Navigable Small World graph

time complexity:

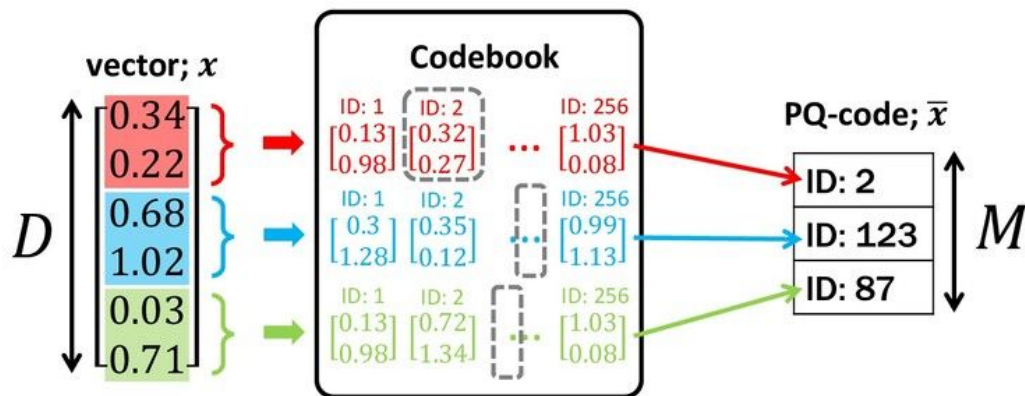
$$O(d * n * \log(n))$$

but, not only time is important..
space complexity is also relevant

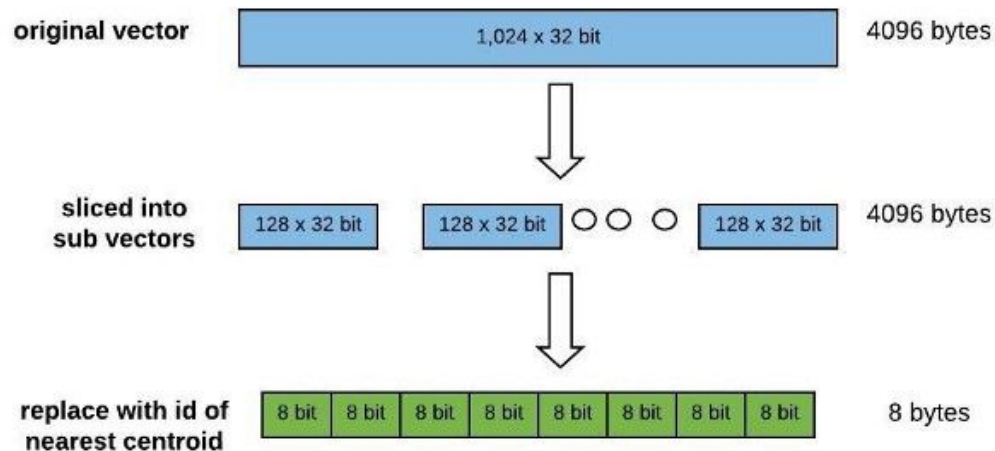


FAISS

Product Quantization: **Memory efficient**



FAISS



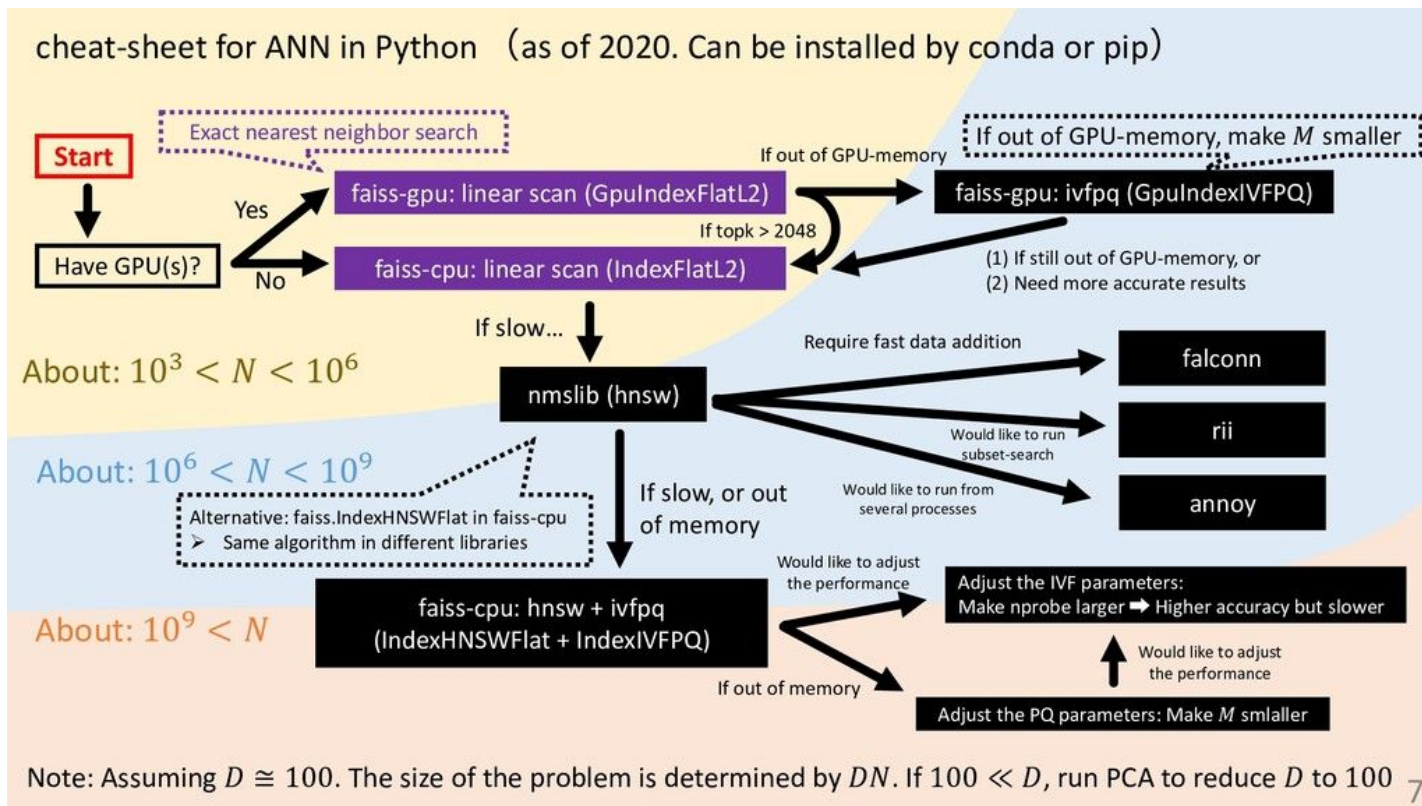
Summary

	search time	1-R@1	index size	index build time
Flat-CPU	9.100 s	1.0000	512 MB	0 s
nmslib (hnsw)	0.081 s	0.8195	512 + 796 MB	173 s
IVF16384,Flat	0.538 s	0.8980	512 + 8 MB	240 s
IVF16384,Flat (Titan X)	0.059 s	0.8145	512 + 8 MB	5 s
Flat-GPU (Titan X)	0.753 s	0.9935	512 MB	0 s

The database consists of 1,000,000 SIFT descriptors (128D) extracted from image patches.

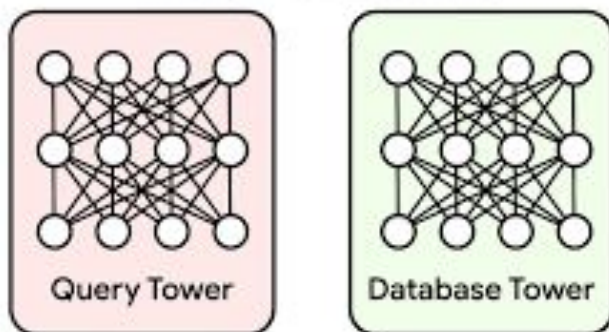
Guia interessant de llegir: <https://towardsdatascience.com/comprehensive-guide-to-approximate-nearest-neighbors-algorithms-8b94f057d6b6>

Summary



ScaNN

Model Architecture



Què farem?

Data

Fashion-MNIST

60.000 training samples

10.000 testing samples

each sample is 28x28 pixels

784 dimensions

10 categories

T-Shirt/Top

Trouser

Pullover

Dress

Coat

Sandals

Shirt

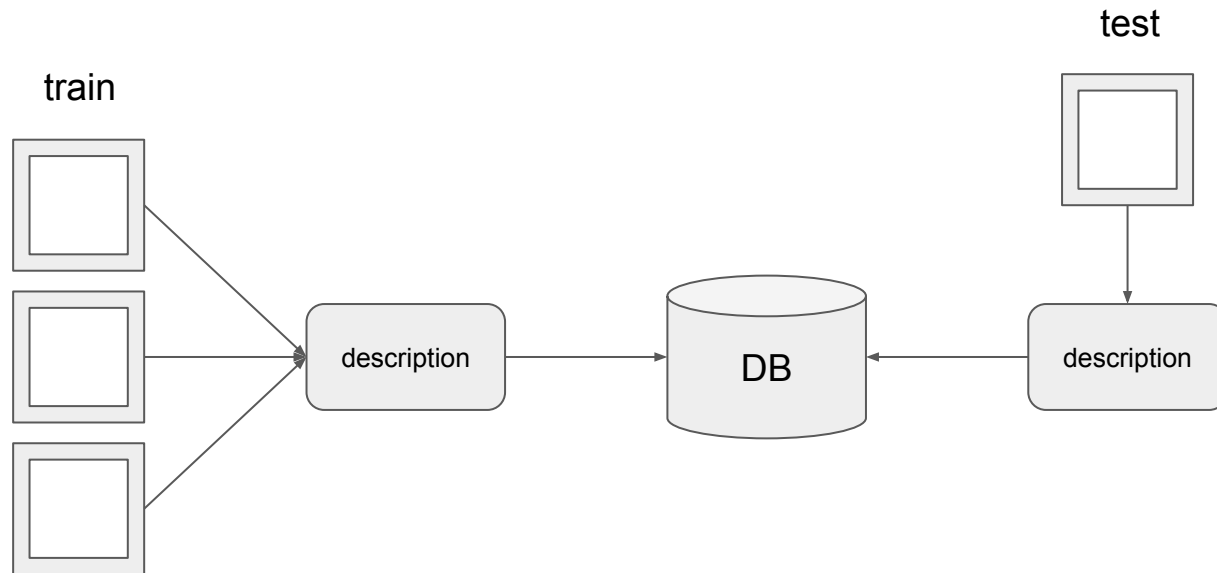
Sneaker

Bag

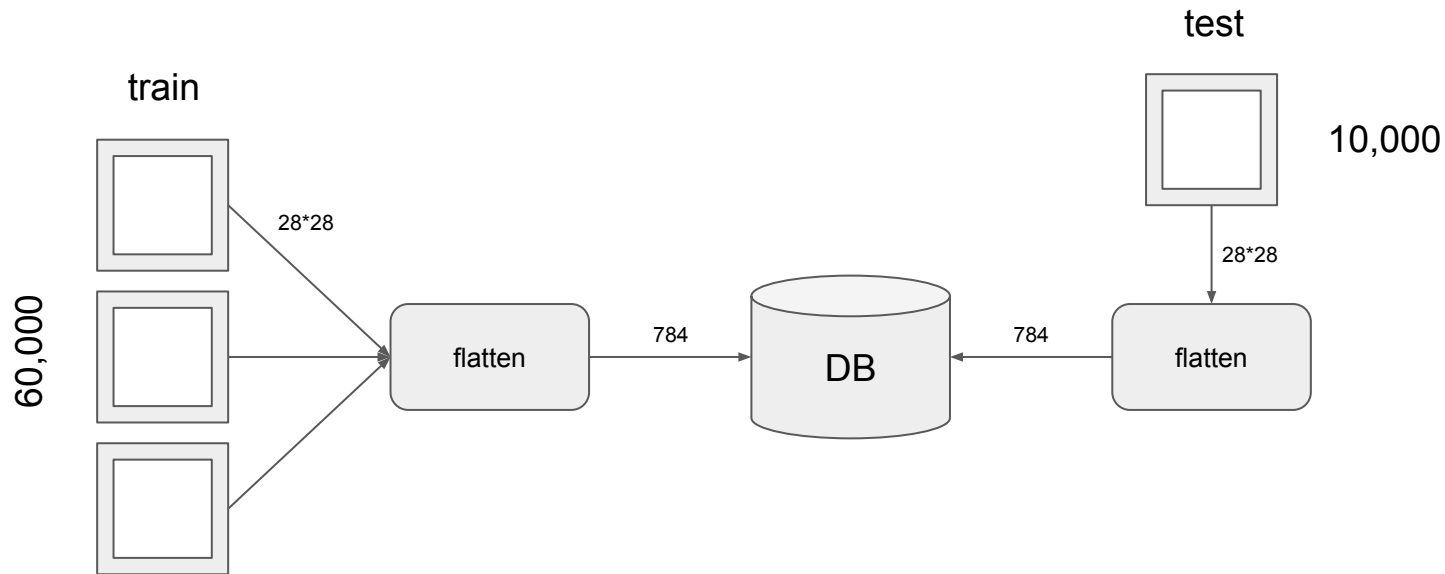
Ankle boots



Què farem?

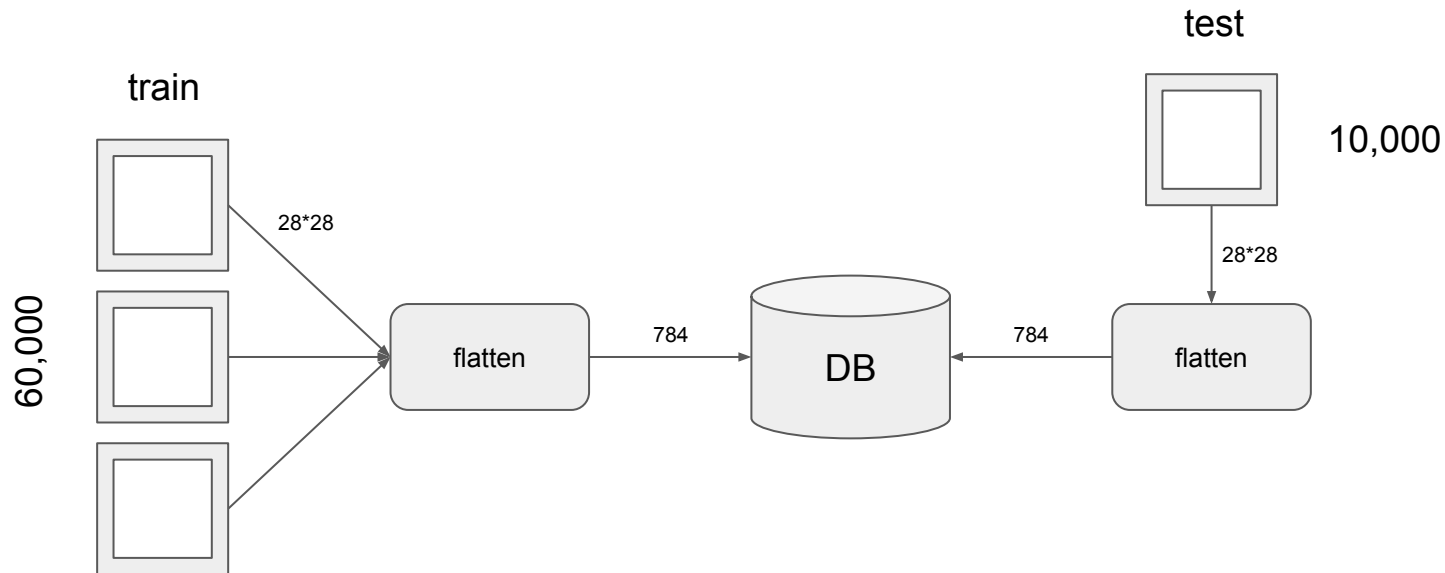


Què farem?



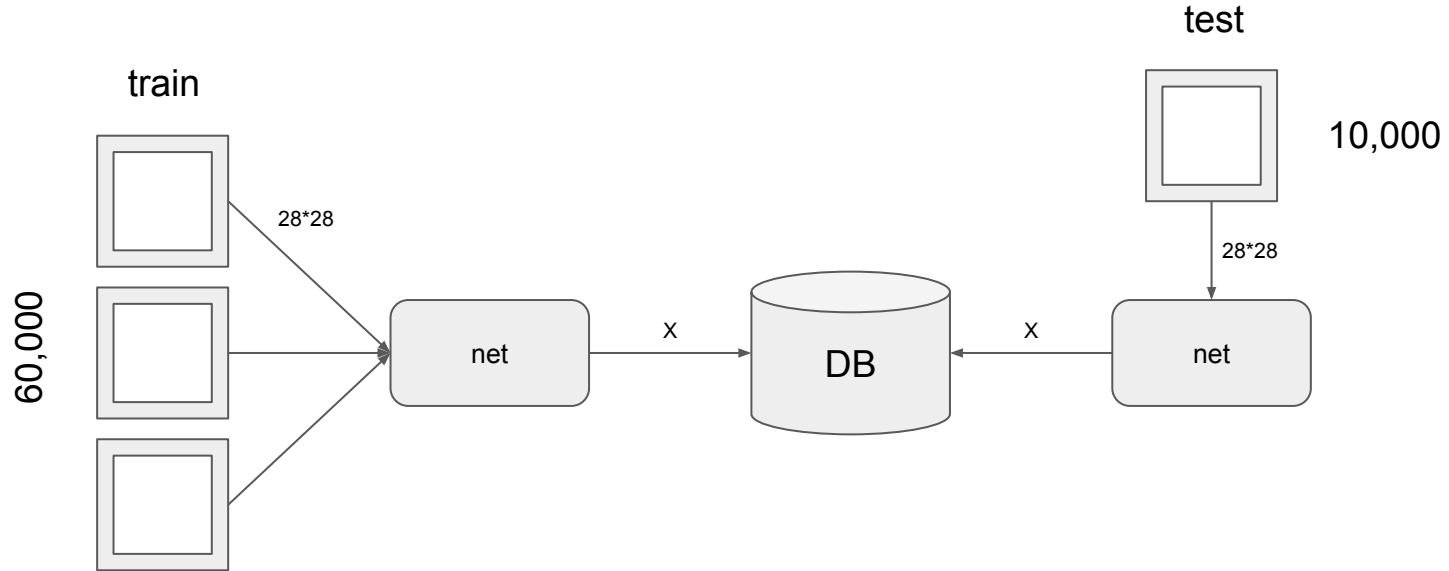
Apartat A. Búsqueda Brute Force

Què farem?



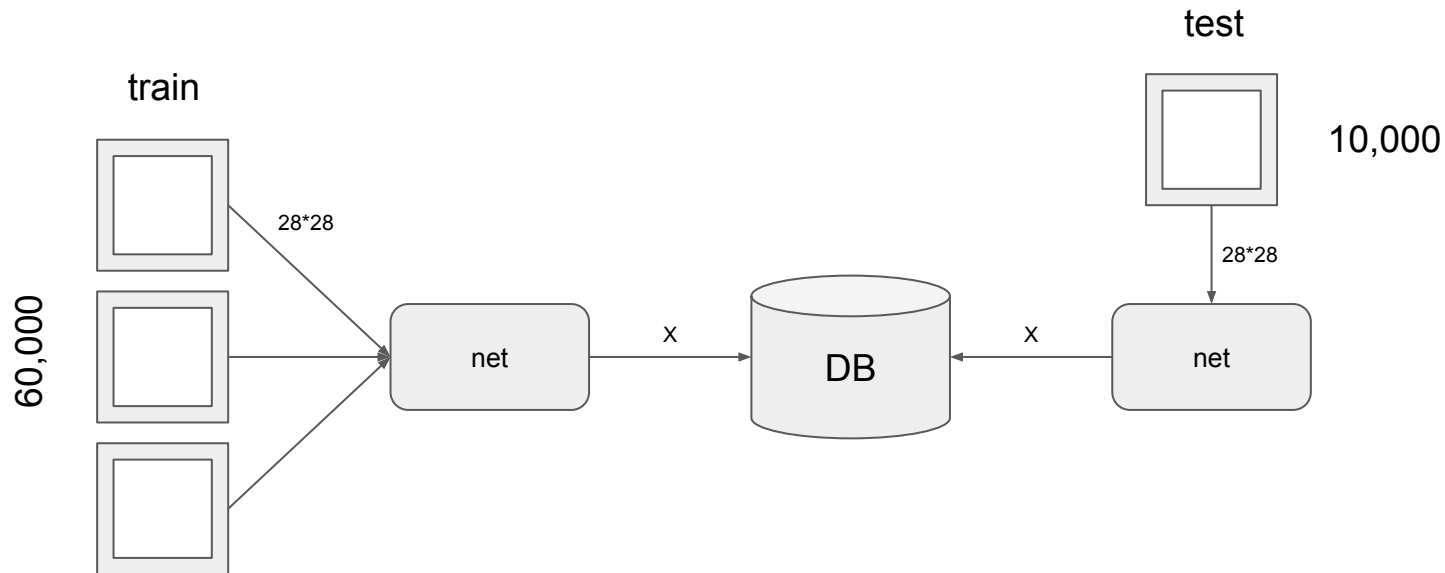
Apartat B. Búsqueda Aproximada

Què farem?



Apartat C. Net Encoding. Búsqueda Brute Force

Què farem?



Apartat D. Net Encoding. Búsqueda Aproximada

Apartat A. Búsqueda Brute Force (2pts)

PREGUNTA: Implementa el mètode `query` de la classe `BruteForce`.

PREGUNTA: Executa la búsqueda amb diferents `k`.

Apartat B. Búsqueda Aproximada (4pts)

PREGUNTA: Implementa els mètodes `build` i `query` de les 3 classes anteriors.

PREGUNTA: Compara els resultats del bruteforce i els 3 ANN.

PREGUNTA: Executa cerques amb varis paràmetres dels ANN.

Apartat C. Net Encoding (2pts)

PREGUNTA: Aprèn un model i extreu les característiques de la penúltima capa.

PREGUNTA: Modifica la dimensionalitat del model de la capa `fc2` i reaprèn el model.

Apartat D. Net Encoding. Búsqueda BruteForce i Aproximada (3pts)

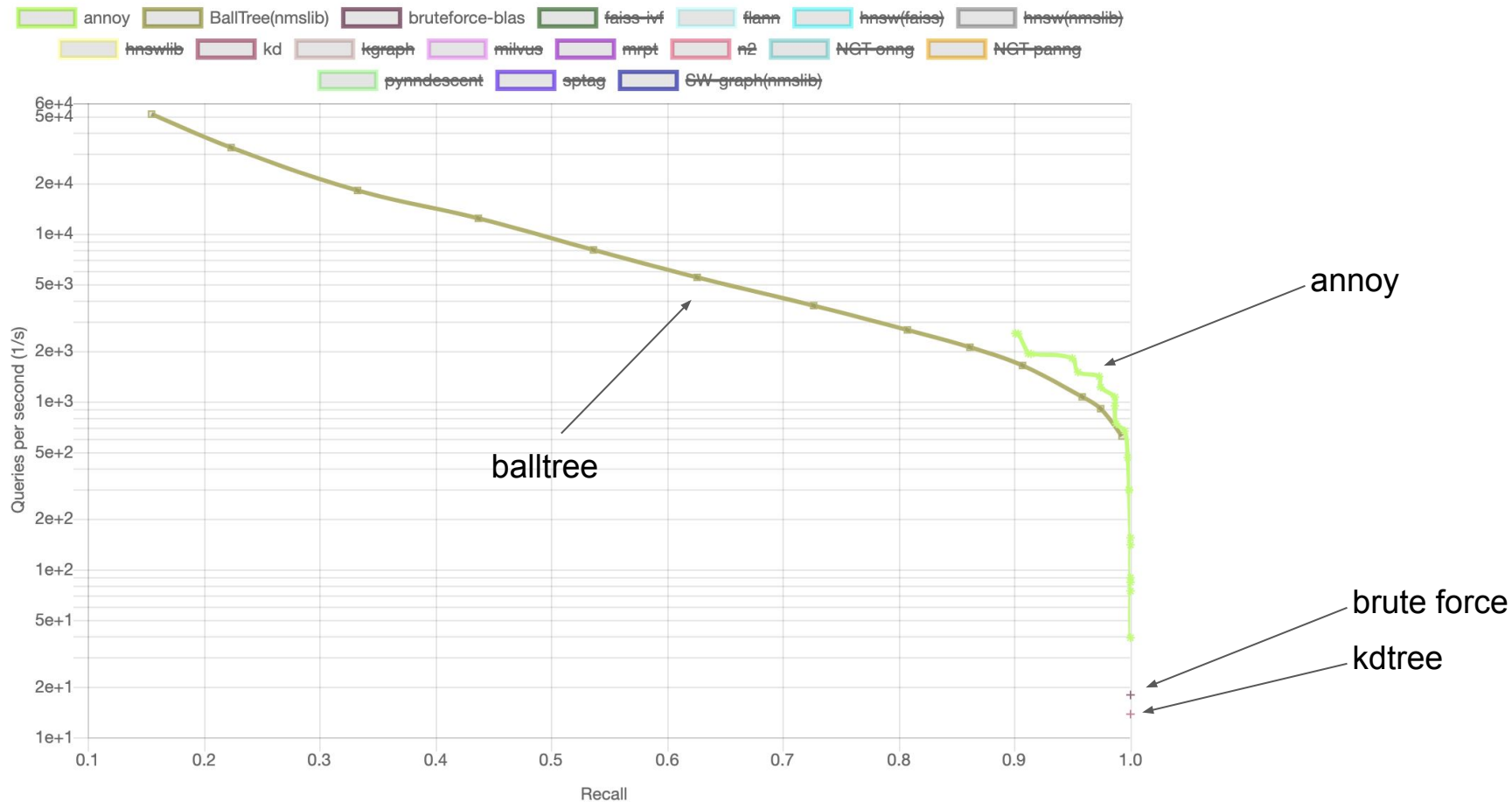
PREGUNTA: Compara els resultats del bruteforce i els 3 ANN utilitzant les característiques de la xarxa

PREGUNTA: Explica les diferències respecte utilitzant les dades originals.

PREGUNTA: Executa cerques amb varis paràmetres dels ANN.

PREGUNTA: Mostra una gràfica mostrant les queries/s respecte el recall que aconseguixen.

Recall-Queries per second (1/s) tradeoff - up and to the right is better



Recall-Queries per second (1/s) tradeoff - up and to the right is better

