classmate
Date
Page

Revisting KNN:

use images of size 8×8

does not calcale exact calculation & grad descent
~~appo~~ approximation of nearest neighbours

* Decision Boundary of the classifier

→ The decision boundary is the line that seperates the classes in the feature space.
→ How well the model was trained.
(took into acc non-linear complexities).
Helps to see the complexity of the learned model.
→ Helps to visualize how examples will be classified for the entire feature space.
→ The more examples that are stored, the more complex the decision boundaries can become.

k-NN → smoother more continuous decision boundaries
1 NN → picking up noise ⇒ learning noise also
⇒ overfitting

How much should be value of k?
small k ⇒ small boundaries / non-smooth decision boundaries
⇒ overfit. (may lead to non-smooth decision boundaries)
⇒ Creates many small regions for each class.

large k ⇒ creates fewer regions
⇒ usually leads to smoother decision boundaries
(NOTE: too smooth ⇒ underfit)

Go back to prev works of people & then decide value of k.

choosing k ⇒ data dependent (3-11) & heuristic based
⇒ use cross-validation.
⇒ NOTE: k too small or too ~~bad~~ big is bad

How to understand the distances

$$(p_1-q_1)^2 + (p_2-q_2)^2 = d^2$$

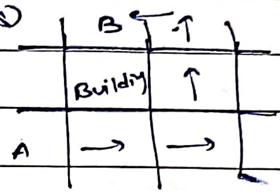
$$q_1^2 + q_2^2 = r^2$$

2) Manhattan Distance [L1] (very high dim, more features)

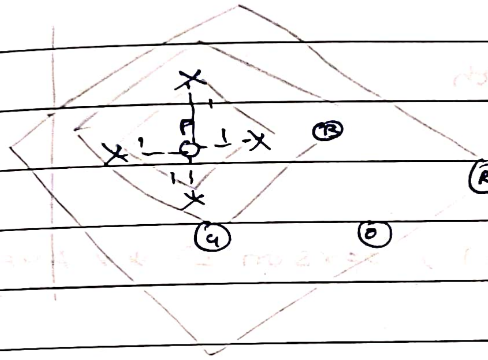
(travelling in a grid) Point A to Point B

(numerical values)

$$d(P, Q) = \sum_{i=1}^{d} |(p_i - q_i)|$$



Iso-surface



0 1 0 0 1

$$(p_1-q_1) + (p_2-q_2) = d$$

(distance b/w points should be equal)

Equal distances

3) Minkowski Distance

$$d(P, Q)^r = \sum_{i=1}^{d} |p_i - q_i|^r$$

when $r = 1 \rightarrow$ Manhattan

$r = 2 \rightarrow$ euclidean

4) Mahalanobis Distance

Takes into acc correlation of points

(mainly used for outlier detection, whether it is close to a cluster).

Spread & correlation of cluster.

$$D^2 = (x-\mu)^T C^{-1} (x-\mu)$$

Book : Hastie & Tibshrani, (CH-13)

(Each point is its own neighbour.

it is its own neighbour.

=> training error = 0 (∵ label is provided)

As # of neighbours ↑

test data error is decreasing slowly & increases a little

K=1 overfitting K=6/7 => BEST

constant

generalizing as K ↑

Bayes Error : Violet line -> Best that classifier can do (max it could achieve)

KNN -> smart to chose best value of k

Trial & error => similar to result obtained by classifier.

cross-validation.

* Computing the distances

-> The K-NN algo requires computing distances of the text exampl and each of the training examples.

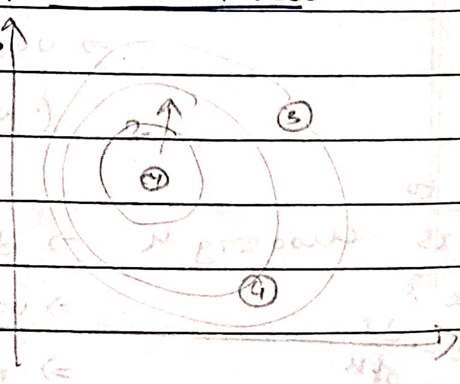
-> The choice depends on the type of the features in the data.

1) Euclidean distance $[l_2]$ (more computation) ISO-surfaces

continuous data

$$d(P, Q)^2 = (P-Q)^T (P-Q)$$

$$d(P, Q)^2 = \sum_{i=1}^{d=10} (p_i - q_i)^2$$



How to understand the distances

$(p_1-q_1)^2 + (p_2-q_2)^2 = d^2$

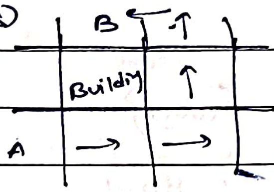
$q_1^2 + q_2^2 = r^2$

2) Manhattan Distance [L_1] (very high dim, more features)

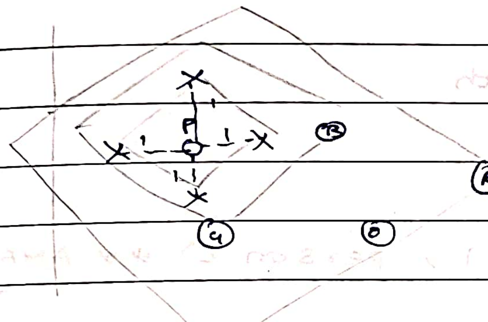
(travelling in a grid) Point A to Point B

(numerical values)

$$d(P, Q) = \sum_{i=1}^{d} |(p_i - q_i)|$$



Iso-surface



$(p_1-q_1) + (p_2-q_2) = d$

(distance b/w points should be equal)

Equal distances

3) Minkowski Distance

$$d(P, Q)^r = \sum_{i=1}^{d} |p_i - q_i|^r$$

when $r = 1 \rightarrow$ Manhattan

$r = 2 \rightarrow$ euclidean

4) Mahalanobis Distance

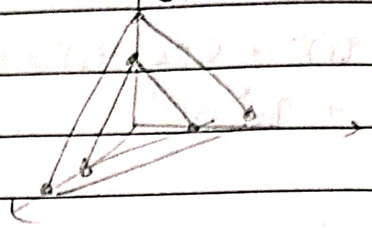
Takes into acc correlation of points

(mainly used for outlier detection, whether it is close to a cluster).

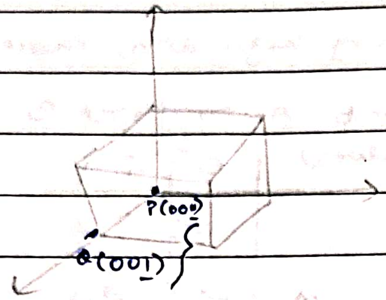
Spread & correlation of cluster.

$$D^2 = (x-\mu)^T C^{-1} (x-\mu)$$

5) Hamming distance (categorical / binary data).



$$d(P,Q) = \sum_{i=1}^{d} I(p_i \neq q_i)$$



How many positions are they diff.

looks at binary data

at how many positions are they not equal.

∴ HD=1

Practical Ex: Two words of same length

TEXT DATA
VECTORS

Monday
Sunday ⇒ HD=2

DNA sequencing (of person 1, person 2) ** AMAZING USAGE!

6) Cosine distance

(Textual data)

$$S(P,Q) = \cos\theta = \frac{P \cdot Q}{\|P\|\|Q\|}$$

amount of similarity between two datapoints.

Real-life Application: Document 1, Document 2

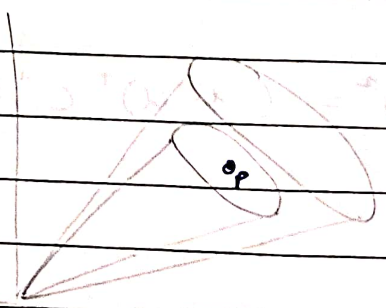
TEXT DOCUMENTS.

how similar are they to each other.

similarity ↑ ⇒ distance is less (close to each other).

$$S(P,Q) = \cos\theta = \frac{P \cdot Q}{\|P\|\|Q\|}$$

$$d(P,Q) = 1 - S(P,q)$$



Sckit KNN → Distance metric