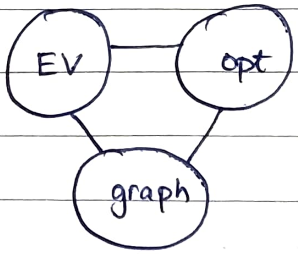
Date 3.3.2020
Page 1

OPTIMIZATION METHODS FOR ML-1



max — min

opt. $\omega^T A \omega$ — λ.

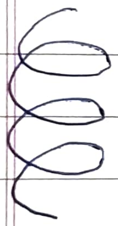
EV.

$\|\omega\| = 1$

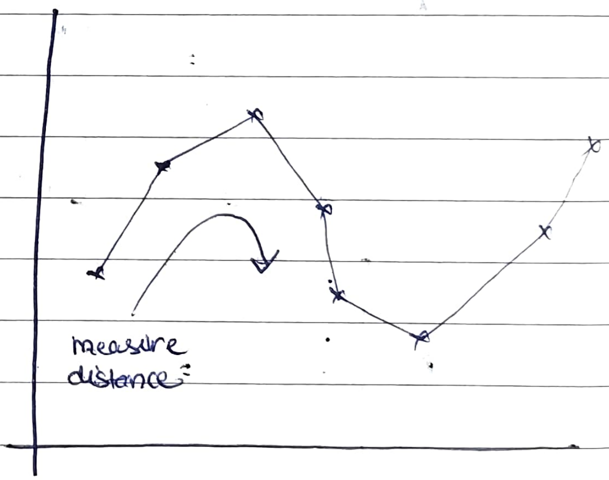
ω is the EV of A.

⑤ max. $\frac{\omega^T A \omega}{\omega^T \omega}$ set $\omega^T\omega = 1$

"on the manifold" $d(x_1, x_2)$ = shortest path b/w x_1, x_2.



follow the spring



k-nearest points

vs

within a distance

- circle

find $y_1 \dots y_n$ in 2D st. $d(y_i + y_j) \approx d(x_i, x_j)$

2D — 3D/4D

on wrap paper — popular in data visualization.

min $[d(y_i, y_j) - d(x_i, x_j)]^2$

see higher dimension

need a transformation from x_i to y_i such that it has similar local geometry.

fn. – ML.

$c = \{ x_i \mid \omega_i > 0 \}$

hyper parameter

EV OPTIMIZATION : MORE EXAMPLES

1. max. $\omega^T A \omega$ where, $A \in \mathbb{R}^{d \times d}$
st. $\omega^T \omega = 1$ $\omega \in \mathbb{R}^d$

2. max. $Tr(W^T A W)$ where, $A \in \mathbb{R}^{d \times d}$
st. $W^T W = 1$ $W \in \mathbb{R}^{d \times d}$

3. min. $\| X - \omega \omega^T X \|_F^2$ where $X \in \mathbb{R}^{d \times n}$
st. $\omega^T \omega = 1$ $\omega \in \mathbb{R}^d$
$\| \cdot \|_F \Rightarrow$ frobenius norm.

4. min. $\| X - W W^T X \|_F^2$ where $X \in \mathbb{R}^{d \times n}$,
st. $W^T W = 1$ $\omega \in \mathbb{R}^d$
$\| \cdot \|_F \Rightarrow$ frobenius norm.

5. max. $\dfrac{\omega^T A \omega}{\omega^T \omega}$

GRAPHS INDUCED ON DATA

- appreciate and represent the data as a graph.
  - data points as vertices
  - Some relationship (say distance or similarity) as the edge.

- nearest neighbour graph
  - K-nearest neighbour graphs.
  - epsilon neighbourhoods.

- data on a low-dimensional manifold
  - consider low-dimensional nonlinear manifolds.

- simple trick : Similar to ISOMAP
  - construct a neighbourhood graph.
  - $d(i,j)$ is the shortest path or the distance on the manifold.

RATIO CUTS AND CLUSTERING

$$\text{cut}(A, \bar{A}) = \sum_{i \in A} \sum_{j \in \bar{A}} w_{ij}$$

can lead to unwanted solution (such as one tiny cluster corresponding to an outlier).

ratio cut:

objective

$$\frac{cut(A,\bar{A})}{|A|} + \frac{cut(A,\bar{A})}{|\bar{A}|} \Leftrightarrow \sum_{ij}(f_i - f_j)^2$$

$$\Leftrightarrow f^T L f$$

wher, f_i is $\frac{\sqrt{|\bar{A}|}}{\sqrt{|A|}}$ if $i \in A$

$-\frac{\sqrt{|A|}}{\sqrt{|\bar{A}|}}$ if $i \in \bar{A}$

this is equivalent to optimization of:

$$f^T L f$$

st. $f^T f = 1$

note: the constraint on f is a "relaxation" of what we really want to optimize.

CLUSTERING / CUT OBJECTIVE

ratio cut

$$\frac{cut(A,B)}{|A|} + \frac{cut(B,A)}{|B|}$$

normalized cut

$$\frac{cut(A,B)}{cut(A,A) + cut(A,B)} + \frac{cut(B,A)}{cut(B,A) + cut(B,B)}$$

min-max cut

$$\frac{cut(A,B)}{cut(A,A)} + \frac{cut(B,A)}{cut(B,B)}$$

comments

- if clusters are

- well separated, all three give very similar and accurate results.

- marginally separated, NormCut and MinMax Cut give better results.

- overlapping significantly, MinMaxCut tend to give more compact and balanced clusters.

extensions to multi-way cuts (beyond this course).

K-MEANS FOR CLUSTERING

$$\min. \; J = \sum_{j=1}^{K} \sum_{i=1}^{N_j} \| x_j - \frac{1}{N_j} \sum_{p=1}^{N_j} x_p \|$$

NP hard to minimize in general.

- choose k centers at random
- assign points to closest center
- compute new centers are cluster centroids.

k-means++ is lloyd's with smarter initialization.