OM-S20-14: Optimization for ML - I

C. V. Jawahar

IIIT Hyderabad

http://preon.iiit.ac.in/om_quiz

03 Mar 2020

EV Optimization: More Examples

$$(1) \quad \text{Maximize} \quad w^T A w \quad \text{Subject to: } w^T w = 1$$

where $A \in R^{d \times d}$ and $w \in R^d$

$$(2) \quad \text{Maximize} \quad Tr(W^T A W) \quad \text{Subject to} \quad W^T W = I$$

where $A \in R^{d \times d}$ and $W \in R^{d \times d}$

$$(3) \quad \text{Minimize} \quad ||X - ww^T X||_F^2 \quad \text{Subject to: } w^T w = 1$$

where $X \in R^{d \times n}$, $w \in R^d$ and $|| \cdot ||_F$ denotes Frobenius norm.

$$(4) \quad \text{Minimize} \quad ||X - WW^T X||_F^2 \quad \text{Subject to:} \quad W^T W = I$$

where $X \in R^{d \times n}$, $w \in R^{d \times d}$ and $|| \cdot ||_F$ denotes Frobenius norm.

$$(5) \quad \text{Maximize} \quad \frac{w^T A w}{w^T w}$$

Graphs Induced on Data

- Appreciuate 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

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

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

$$\frac{Cut(A, \bar{A})}{|A|} + \frac{Cut(A, \bar{A})}{|\bar{A}|} \Leftrightarrow \sum_{ij} w_{ij}(f_i - f_j)^2 \Leftrightarrow f^T L f$$

where f_i is $\frac{\sqrt{|\bar{A}|}}{\sqrt{|A|}}$ if $i \in A$ and $-\frac{\sqrt{|A|}}{\sqrt{|\bar{A}|}}$ if $i \in \bar{A}$
This is equivalent to optimization of :

$$f^T L f \text{ subject to: } f^T f = 1$$

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 MinMaxCut 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

Minimize:

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

NP-hard to minimize in general.

- Choose k centers at random.
- Assign points to closest center.
- Compute new centers are clusters centroids.

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