SMAI-M20-L10: LSI etc.

C. V. Jawahar

IIIT Hyderabad

August 31, 2020

Class Review Questions

1. If $A = UDV^T$, then $A^T A$ is
   - (i) always a rank deficient matrix (ii) VD^2V^T (iii) UD^2U^T (iv) A square matrix (v) is always full rank (vi) can some times be full rank (vi) none of the above etc.
2. Consider a set of general vectors $\mathbf{a}_i \in R^d$. (assume all elements are some random numbers in the range of $[0, 1]$) $\mathbf{b}$ is another such vector. Consider the matrix:

$$\mathbf{A} = \sum_{i=1}^{k} \alpha_i \mathbf{a}_i \mathbf{a}_i^T + \sum_{i=k+1}^{d} \beta_i \mathbf{b}\mathbf{b}^T$$

Comment on the effective rank of $\mathbf{A}$ for various values of α_i and β_i

Announcements:

Post Your queries about:

- Chapters 2, 3 5 in the book.
- Class Reviews in L01-L08
- Micro-Lecture Videos L01-L08

Blank

Recap:

- Problem Space:
  - Learn a function $y = f(\mathbf{W}, \mathbf{x})$ from the data.
    - for classification
    - for regression
  - Learn useful features
- Supervised Learning:
  - Notion of Training and Testing
  - Notion of Loss Function and Optimization
  - Need of Generalization and Worry of Overfitting
  - Occam's razor and role of model complexity
  - Estimating error on validation set.
- Classification Algorithms:
  - Nearest Neighbour Algorithm
  - Linear Classification; Linear Regression
  - Decide as ω_1 if $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$ else ω_2
  - Performance Metrics
- Mathematical Foundations: Linear Algebra, Probability, Optimization
  - SVD, Eigen Decomposition
  - MLE

This Lecture:

man

$[\,0 \mid 0 \mid 0 \mid 0 \mid 1 \mid \dots \mid 0 \mid 0 \mid 0\,]$

Micro-Lecture Videos

1. Rank of Term-Document Matrix and LSI
   - Q: How do we compare documents? words?
   - Appreciate "one-hot" representation of words.
2. Bias and Variance
   - Yet another balancing act to do while designing ML solution.
3. Loss Function
   - Not all intuitive loss functions are in use due to practical difficulties.

w d |V|

Doc

N

$\frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i, y_i))$

w

Blank

Questions? Comments?

Blank

Matrix completion: We know completion of rank 1 matrix as:

Given A, rank = 1

Find B

$$A = \begin{bmatrix} 7 & ? & ? \\ ? & 8 & ? \\ ? & 12 & 6 \\ ? & ? & 2 \\ 21 & 6 & ? \end{bmatrix} \Rightarrow \begin{bmatrix} 7 & 2 & 1 \\ 28 & 8 & 4 \\ 42 & 12 & 6 \\ 14 & 4 & 2 \\ 21 & 6 & 3 \end{bmatrix}$$

(0) B

i, j available

Example Problem Formulation:

$$\min_{B} \sum_i \sum_j (A_{ij} - B_{ij})^2 \quad s.t \quad rank(B) = 1$$

B

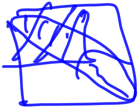
Q: What is this summation over? What is the min over?

The rank constraint is not easy to enforce. See alternate formulations based on Nuclear norm[1]. i.e., L1 norm over Diag(D) instead of L0 norm.

[1] Nuclear norm of a matrix is just the sum of the singular values of the matrix

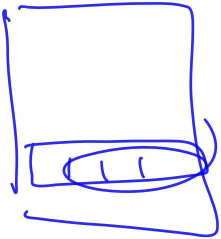
Blank

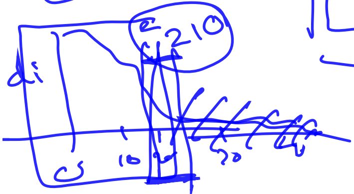
$$\sum_{i,j,\text{visible}} (A_{ij} - B_{ij})^2 \quad \text{s.t.} \quad \text{rank}(B) = 1$$



Ed ~ 300

350





Given A with missing elements and rank const.

Find B

Discussion Point - II

LSI has many applications [2]

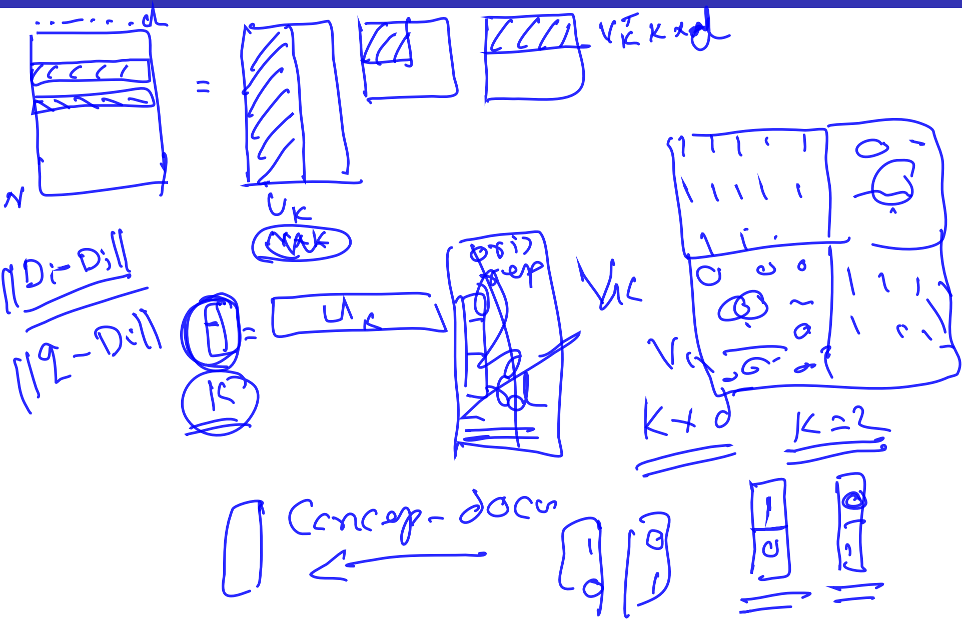
- Compare the documents in the low-dimensional space (data clustering, document classification).
- Find similar documents across languages, after analyzing a base set of translated documents (cross-language information retrieval).
- Find relations between terms (synonymy and polysemy).
- Given a query of terms, translate it into the low-dimensional space, and find matching documents (information retrieval).
- Find the best similarity between small groups of terms, in a semantic way.
- Expand the feature space of machine learning / text mining systems
- Analyze word association in text corpus

Q: **How does it help in Synonymy** ?

Q: **How does it help in Clustering documents** ?

[2] Read: https://en.wikipedia.org/wiki/Latent_semantic_analysis

Blank

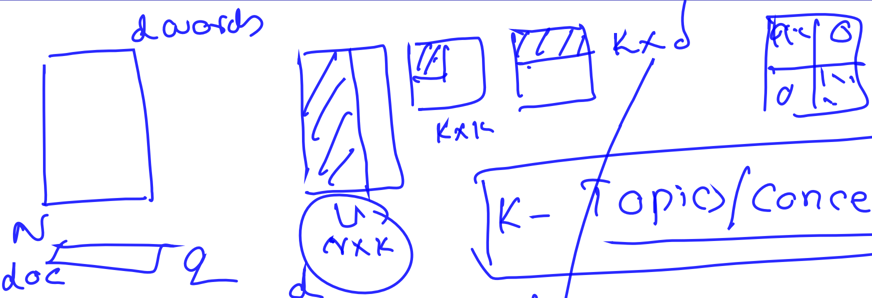


Discussion Point - III

Appreciate the fact that there is an important balancing act to be done to get good solutions. Why?

1. overfitting and underfitting
2. bias and variance
3. empirical error and model complexity

Blank



$D_1 \dots D_N \in \mathbb{R}^d$

K - Topics/Concepts

$q \in \mathbb{R}^d$

$q' \in \mathbb{R}^k$ = K V_k

Term span

$q \in \mathbb{R}^d$

$\min_i \| D_i - q \|$

What Next:? (next three)

- Applications and insights into of SVD and Eigen Decomposition
- Supervised Learning: Regularization and Loss Functions
- Bayesian View and Optimal Classification
- PCA and Dimensionality Reduction

$x' = U^T x$