SMAI-M20-07:Examples of Optimization Formulations

C. V. Jawahar

IIIT Hyderabad

August 24, 2020

Recap

- Two problems of interest:
  - Learn a function $y = f(\mathbf{W}, \mathbf{x})$ from the data.
  - Learn Feature Transformation as a step to find useful representations.
- Algorithms:
  - Nearest Neighbour Algorithm
  - Linear Classification
  - Decide as ω_1 if $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$ else ω_2
  - Performance Metrics
- Vectors, Matrices and Data Matrix
  - Comparison of Vectors.
  - Matrix vector multiplication. Differentiation.
  - Properties, Rank, Eigen Values and Eigen Vectors
- Supervised Learning:
  - Notion of Training and Testing
  - Notion of Loss Function
  - Role of Optimization
  - Need of generalization
  - Worry of Overfitting

This Lecture:

- This Lecture:
  - MLE
  - Linear Regression
  - Eigen Vector based optimization
- Problems:
  - How optimization arise?
  - Objective/Loss and Problem Statement
- Solutions:
  - Linear Regression: Differentiation of quadratic function. Unconstrained.
  - Constrained Optimization Problems that lead to solution as eigen vectors
  - Probabilistic formulations, intuitive solutions, interpretations.
  - Closed form solutions (today); iterative in future.

Blank

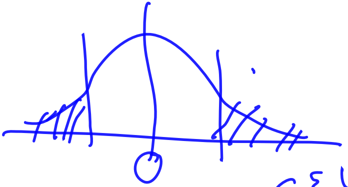
Questions? Comments?

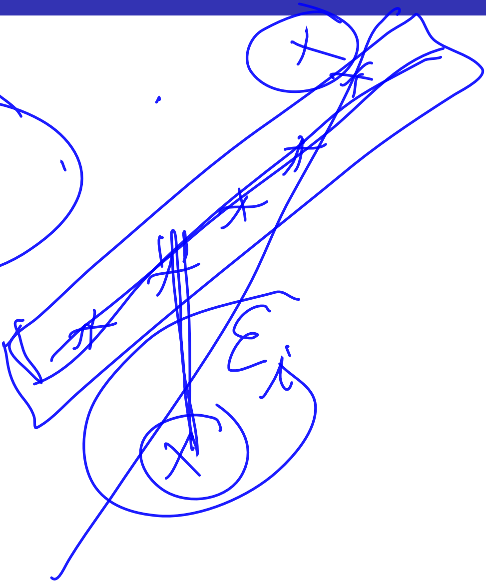
Blank

$$\sum_i (y_i - w^T x_i)$$

N λ

MSE





$$\varepsilon_i = y_i - w^T x_i$$

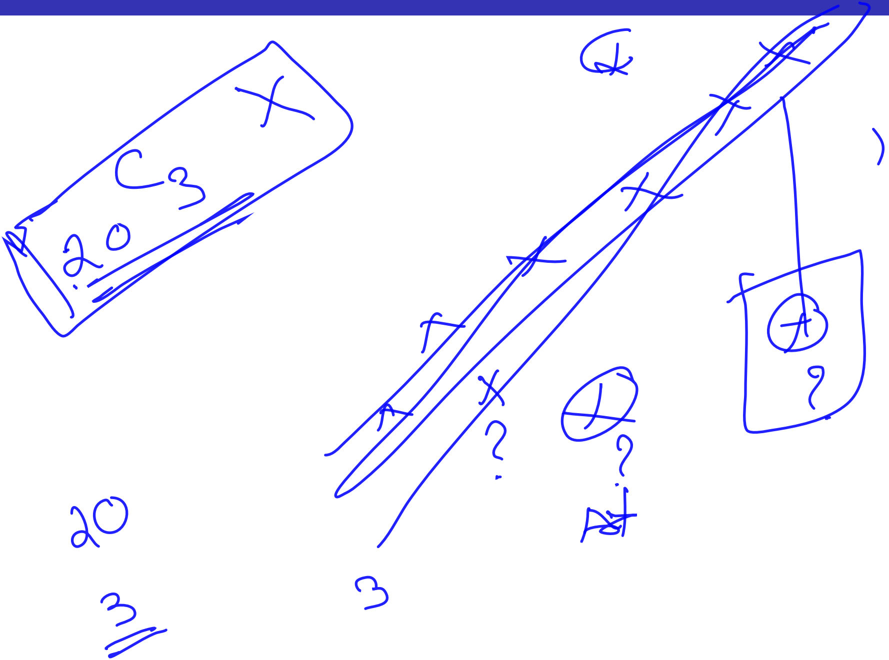
Discussions Point -I

We are interested in fitting a solution in presence of outliers. A student came up with an original, yet simple idea. (i) Detect outliers from the data set $\mathcal{D}$ and remove them (ii) Build models on what is left out $\mathcal{D}'$. Here are the formal steps:

1. Find $\mathbf{w}$, the solution for linear regression problem with data set as $\mathcal{D}$.
2. Remove all samples where $|y_i - \mathbf{w}^T\mathbf{x}_i| > \theta$ from $\mathcal{D}$ to create an inlier set $\mathcal{D}'$.
3. Now find a new $\mathbf{w}$ with $\mathcal{D}'$.

- Student tested it with θ as very small (say zero) and very large (say infinity). In both cases, the algorithm came out to be ineffective. What could be the reason?
- Another student helped to solve the problem by coming with a nice intuitive way of finding θ. It did reasonably well. Suggest a simple and effective way to address outliers (or finding θ).

Blank



Discussion Point - II

We are intersted in solving an overdetermined system of homogenous equations

$$\mathbf{Ax} = \mathbf{0}$$

where $\mathbf{A}$ is $m \times n$. $m > n$

Problem is formulated as: $n \times n$

$(Ax)^T(Ax)$ $\quad x^T x = 1$

$$\arg\min_{\mathbf{x}} ||\mathbf{Ax}|| \quad \text{Subject to: } ||\mathbf{x}|| = 1$$

$x^T A^T A x$

Solution to this problem is:

- (a) Eigen Vector corresponding to the largerst eigen value of $\mathbf{AA}^T$.
- (b) Eigen Vector corresponding to the largerst eigen value of $\mathbf{A}^T\mathbf{A}$.
- (c) Eigen Vector corresponding to the smallest eigen value of $\mathbf{AA}^T$.
- (d) Eigen Vector corresponding to the smallest eigen value of $\mathbf{A}^T\mathbf{A}$.

Read later[1]

[1] https://cw.fel.cvut.cz/old/_media/courses/a3m33iro/overdetermined_homogenous_linear_equations.pdf

Blank

$x^T A^T A x - \lambda (x^T x - 1)$

$= 0$

$A^T A x - \lambda x$

λx

x is e.vec of $A^T A$

$\lambda x^T x$

smaller

Discussion Point - III

- We know the $d \times N$ data matrix $\mathbf{X}$ with every column as data elements. Show that $\mathbf{A} = \mathbf{X}\mathbf{X}^T$ is Symmetric. Show that $\mathbf{A}$ is PSD. [2]
- If $\mathbf{A}$ is real symmetric PSD matrix, we know $\mathbf{A} = \mathbf{V\Lambda V^T}$, then what is $\mathbf{V^TAV}$? What is this process called?

$$V^T A V = V^T V \, \Lambda \, V^T V$$

$$\boxed{V^T A V = \Lambda}$$

[2] A matrix A is PSD if $z^T A z \geq 0 \ \forall z \in R^d$.

Blank

Review Question - I (one, none or more correct)

Consider a square matrix $\mathbf{A}$ with eigen values λ_i and eigen vectors $\mathbf{v}_i$. Then for $\mathbf{A}^T$,

- (a) Eigen values and eigen vectors are the same as that of $\mathbf{A}$.
- (b) Eigen values are the same. Eigen vectors are $\mathbf{v}^T$.
- (c) Eigen values are $\frac{1}{\lambda_i}$
- (d) We can comment for symmetric matrix $\mathbf{A}$. But not for other square matrices.
- (e) None of the above.

Blank

Review Question - II (one, none or more correct)

Consider the problem of finding the minima of $f(x) = x^2 - 2x + 4$

$$\min_{x} x^2 - 2x + 4$$

The optimum value of the function f^* is:

- (a) 1
- (b) 2
- (c) 3
- (d) 4
- (e) -2
- (f) none of the above

Blank

Review Question - III (one, none or more correct)

Consider a data matrix $\mathbf{X}$ where every column is mean subtracted samples. We compute $\mathbf{A} = \mathbf{X}\mathbf{X}^{\mathbf{T}}$. Then $\mathbf{A}$

- is a $d \times d$ matrix
- is a $N \times N$ matrix
- is a symmetric matrix
- is a scaled version of the covariance matrix
- none of the above

Blank

Review Question - IV (one, none or more correct)

The inverse of an upper triangular matrix is:

- (a) Upper triangular
- (b) Lower triangular
- (c) Need not be triangular.
- (c) Will not exist
- (d) None of the above

Blank

Review Question - V (one, none or more correct)

The the eigen values of $\mathbf{A}$ are λ_i, what are the eigen values of $\alpha\mathbf{A}$, where α is a scalet.

- (a) λ_i itself.
- (b) $\frac{\alpha}{\lambda_i}$
- (c) $\alpha\lambda_i$
- (d) $\frac{1}{\alpha}\lambda_i$
- (e) None of the above.

Blank

What Next: (two sessions?)

- SVD, Rank and Data Matrix
- More into Supervised Learning and the associated issues
- Bayesian Optimal Classification