Lecture 2 5232 Part 3

Application of Convex Prog.

(powerful)

the Application of Weak Duality $\Rightarrow$ set the lower Bound of PRIMAL Problem

$\downarrow$ use this to check whether we achieve optimal

$\downarrow$ convince someone our current solution is relatively good

---

① Conic Program

$$\boxed{\begin{array}{ll} \min\limits_{x} & c^T x \\ \text{s.t} & Ax = b \\ & x \in K \end{array}}$$

formulation

$K \rightarrow \begin{cases} \text{non-negative tilt} \\ \text{SOCP} \\ \text{PSD Matrix} \end{cases}$

(a) Diet Problem $\rightarrow$ LP(special Conic Prog)

(not Std form) Model $\rightarrow$

$$\begin{cases} \min & \sum c_j x_j \\ \text{s.t} & \sum\limits_j a_{ij} x_j \geq b_i \;\; \} m \uparrow \\ & x_j \geq 0 \end{cases}$$

a_{ij} : type j provide a_{ij} nutrient i

b_i : at least b_i nutrient i

C_j : cost of type j

$\downarrow$ Std form of LP

$\Rightarrow$ $\boxed{\begin{array}{ll} \min & C^T x \\ \text{s.t} & Ax = b \\ & x \geq 0 \end{array}}$

$\Downarrow$ $\boxed{\text{Add slackness}}$

Std form

$$\begin{cases} \min & \tilde{C}^T \tilde{x} \\ \text{s.t} & \tilde{A} \tilde{x} = \tilde{b} \\ & \tilde{x} \geq 0 \end{cases}$$

$$\begin{cases} \tilde{x} = \begin{pmatrix} x \\ z \end{pmatrix} & z \rightarrow \text{slackness} \\ \tilde{A} = (A \mid I_m) \\ \tilde{b} = b \end{cases}$$

(b) (SOCP) $\boxed{\text{Portfolio Selection}}$

$x_i \to$ how much hold stock i

$\Rightarrow$ overall return $= p^T x = \sum p_i x_i$

Question :
$$\begin{cases} \min\limits_{x} & \text{Risk} \\ \text{s.t} & \text{return} \geq \text{threshold} \\ & \sum x_i \leq \text{Budget} \quad (\text{Total Money}) \\ & \text{no short} \Leftrightarrow x \geq 0 \end{cases}$$

covariance matrix of stock price

formulation $\longrightarrow$

$$\begin{aligned} \min_{x} \quad & x^T \Sigma x \qquad (\Sigma \succeq 0) \\ \text{s.t} \quad & p^T x \geq r_{min} \\ & 1^T x \leq B \\ & x \geq 0 \end{aligned}$$

SOCP

$$\begin{aligned} \min_{(x,t)} \quad & \tilde{t} \\ \text{s.t} \quad & x^T \Sigma x \leq \tilde{t}^2 \\ & \hat{x} = \Sigma^{\frac{1}{2}} x \end{aligned} \Bigg] \approx \underbrace{\{(\tilde{x}, \tilde{t}) : \tilde{x}^T \tilde{x} \leq \tilde{t}^2\}}_{\text{S-O Cone}}$$

$$\begin{aligned} & p^T x \leq r_{min} \\ & \mathbb{1}^T x \leq B \\ & x \geq 0 \end{aligned}$$

(c) Minima volume covering ellipsoid

ellipsoid $= \{x : (x-c)^T Q (x-c) \leq 1\} \subseteq \mathbb{R}^n$

$Q \in PSD$

$$\text{Volume} = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \frac{1}{\sqrt{\det(Q)}} \propto \frac{1}{\sqrt{\det(Q)}}$$

const

Question: min Volume

s.t ellipsoid contains $\{a_i\}$

Application:

search for Outlier in Dataset

$$\Leftrightarrow \left[\begin{array}{ll} \min\limits_{(c,Q)} & (\det Q)^{-\frac{1}{2}} \\ \text{s.t} & (a_i-c)^T Q (a_i-c) \leq 1 \\ & Q \succeq 0 \end{array}\right.$$

not so convex (c, Q)

transform $M = Q^{\frac{1}{2}}$ & $z = Q^{\frac{1}{2}} c$

$-\log(\cdot) \Rightarrow$ convex

$$\Leftrightarrow \begin{array}{ll} \min & -\log \det M \\ \text{s.t} & \boxed{(Ma_i - z)^T (Ma_i - z) \leq 1} \\ & M \succeq 0 \end{array}$$

$\boxed{v^T v \leq 1}$

$v = Ma_i - z$

$(a_i-c)^T Q (a_i-c)$

$= (a_i-c)^T M^T M (a_i-c)$

$= (Ma_i - z)^T (Ma_i - z)$

$$\Leftrightarrow \begin{array}{ll} \min & -\log \det(M) \\ \text{s.t} & v^T v \leq 1 \\ & v = Ma_i - z \\ & M \succeq 0 \end{array}$$

Section 2 Regularizd Least Square (R-LSQ)

Method 1: OLS (Ordinary LS) (Believe in Linear Gaussian Mode)

→ Formulation

Notation

$X_i = \begin{pmatrix} 1 \\ * \\ \vdots \\ * \end{pmatrix}$ $1 \to$ offset

$$\min_\beta \sum_i \| y_i - \beta^T x_i \|_2^2$$

$\Downarrow$

to achieve $\hat\beta$

$\Rightarrow$ predictor $y = \hat\beta^T x$

→ Property :
① No Bias
② High Variance ⇒ $\boxed{\text{not robust}}$ → for change of (x_i), $\hat\beta$ may change a lot!
③ ☆ Difficult to interpret → $\boxed{\text{few confidence}}$

Another drawback: How to deal with those small weights?

observation : put some small value to 0

$\Downarrow$

improve performance!

Method 2 : Subset Selection (Modification of OLS)

↳ $\boxed{\text{unstable}}$

Method 3 : Ridge Regression → $\boxed{\text{NO SHRINK TO 0}}$

$$\min_\beta \sum_i \| y_i - \beta^T x_i \|_2^2 + \lambda \|\beta\|_2^2$$

$\downarrow$

$$\hat\beta = (X^T X + \lambda I)^{-1} X^T y$$

Method 4: LASSO → $\boxed{\text{SOCP}}$ → $\boxed{\text{convex program}}$

$$\min_\beta \sum_i \| y_i - \beta^T x_i \|_2^2 + \lambda \|\beta\|_1$$

① not always

⇒ Highlight: In most cases, β has a tendency to have 0 enty!

⇙ interpretability

⇓ $\boxed{\ell_1\text{-norm}}$

↳ Perfect Property of Interpretation → SPARSE

↳ compare with Ridge Reg.

$\boxed{\text{Great Property}}$ of Convex Program

② the solution will change slightly if we change $(x_i)_{i=1}^{N}$ slightly!

⇙ stability ⇑ $\boxed{\text{convex}}$

↳ if we set threshold, the change will be greatly! (Like Subset Selection)

Sec 3 Compressed Sensing

Q1: $y = Ax$ $\dim(y) \ll \dim(x)$

⇒ if we have y, then we have infinite x!

Q2: if we know x is sparse? → Difficulty → we do not know where the sparsity is.

↓

transform to Optimization Problem! (Convex)

↓

$\min \|x\|^0$

$\boxed{\text{(not norm)}}$

$\|\cdot\|^0$ → count # of non-zero

s.t $Ax=y$ $\quad$ entries $\quad$ Non-convex

$\rightarrow$ NP HARD

Intuition: $\quad$ Image $\xrightarrow{\text{sparse}}$ Sparsed Image

$\rightarrow$ intuition $\rightarrow$ (Convex) Best Relaxation of $\|\cdot\|^0$

Modification: $\begin{cases} \min & \|x\|_1 \\ \text{s.t} & Ax=y \end{cases}$ $\Rightarrow$ tend to have 0 elements $\quad \Rightarrow$ LP !

$\hat{x}$

Fact: in many cases, [Modification] will give the same as [Primal]

x^*

Thm: $\quad y^* = Ax^*$ $\quad$ ① $x^* \in \mathbb{R}^p$ $\quad$ ② x^* has s non-zero entries

Random Gaussian Measurement Matrix

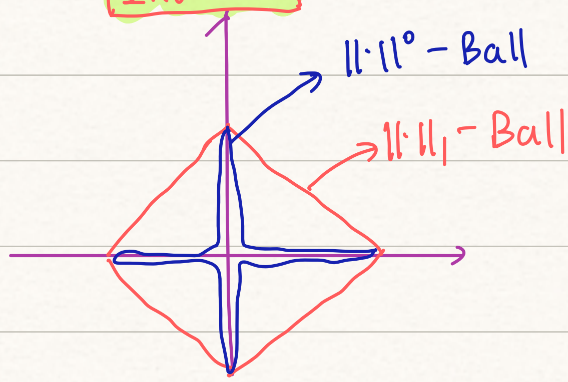
$y = Ax^*$ $\quad a_{ij} \in N(0,1)$

then # of measurements exceed $O(s\log(p/s))$

$\hat{y}$ attains by [Modification] has high probability

to achieve $\hat{x} = x^*$

Intuition



Application: $\begin{cases} \text{our interest : } x^* + \boxed{\text{sparse prior}} \\ \text{observation: } y^* = A x^* \end{cases}$

$\Rightarrow$ Target: try to recover $\hat{x} \approx x^*$

high-quality re-construction

Noise Cases:

observation

noise

Model: $y = A x^* + z$

$\downarrow$ formulate

$$\min_x \|x\|_1$$
$$\text{s.t } \|y - Ax\|_2 \le \varepsilon$$

size of noise z

$\downarrow$ Simplification

$$\hat{x} = \arg\min_x \frac{1}{2}\|y - Ax\|_2^2 + \lambda \|x\|_1$$

Question

$\begin{cases} \min \|x\|_0 \\ \text{s.t } Ax = y \end{cases}$

① Sparsity of signal reasonable?

② Gaussian Measurement Matrix appropriate? (thm problem)

③ Pratical Methods work?

Recall: $\begin{cases} \min_x \|x\|_1 \\ \text{s.t } Ax = y \end{cases} \longrightarrow$ tractability & optimality

Q: $\begin{cases} \min\limits_{x} \ \|x\|_1 \\ \text{s.t} \ \ Ax=y \end{cases} \longleftrightarrow \begin{cases} \min\limits_{x} \ \|x\|_0 \\ \text{s.t} \ \ Ax=y \end{cases}$

$\downarrow \hat{x}_1 \qquad\qquad \downarrow \hat{x}_2$

How to know $\hat{x}_1 = \hat{x}_2$?

A: Generally Difficult!

(under certain condition on A & x^*, there are theorems to prove this)

$\boxed{\text{Matrix Completion}}$ → rating movies (watched)

⇓

infer ratings on the movies that they have not watched

$\boxed{\text{Maths Abstraction:}}$

Ω → observed date

⇓

M_{ij} , $(i,j) \in \Omega$

⇓

infer M_{ij} $(i,j) \notin \Omega$

Intuition: low rank property

⇓

similar people have similar taste

⇓

formulation : $\begin{cases} \min\limits_{X} & \text{rank}(X) \\ \text{s.t} & X_{ij} = M_{ij} \quad \text{for all } (i,j) \in \Omega \end{cases}$

$\Downarrow$

$\boxed{\text{NP - Hard}}$

→ A good surrogate : (of rank(X))

↓ (nuclear norm for matrix)

$\|X\|_* = \sum \sigma_i(X)$ (convex function)

$\Downarrow$

$\boxed{\text{Surrogation}}$

Semidefinite Program $\begin{cases} \min\limits_{X} & \|X\|_* \\ \text{s.t} & X_{ij} = M_{ij} \quad \text{for all } (i,j) \in \Omega \end{cases}$

↓

since $\|X\|_* = \min\limits_{W_1, W_2} \frac{1}{2}\left(tr(W_1) + tr(W_2)\right)$

s.t $\begin{pmatrix} W_1 & X \\ X^T & W_2 \end{pmatrix} \succcurlyeq 0$

where W_1 & W_2 are opt. variables

[Thm]: X^* (true matrix) → rank r

location of observed entries → Ω

suppose observe $\gtrsim m \cdot n \cdot r$ entries

→ Then $\|\cdot\|_*$ recovers the true matrix with high probability

can all info be captured by low rank approx. ?

$\boxed{\min \quad \text{rank}(X)}$ → $\boxed{\text{assumption}}$

$\Downarrow$

$\boxed{\text{maybe not}}$

$\min_{X}$

s.t. $X_{ij} = M_{ij}$ for all $(i,j) \in \Omega$

Question:

① Is low-rank approximation suitable ?

② Is nuclear-norm a good surrogation ?

③ Good Numerical Algorithms ?

maybe not since the condition of theorem may not hold !

$\Rightarrow$ Trade-off between Capture & Tractability

$\Downarrow$

$\begin{cases} \text{low-rank assumption} \\ \|\cdot\|_* \text{ approximation} \end{cases}$

$\Rightarrow$ Bear In Mind:

When assumption, how much we will lose ?