$y = X\beta + \varepsilon$ (1), $\varepsilon \sim N(0, \sigma^2 I_n)$, $X \in \mathbb{R}^{n \times (m+1)}$, $\beta \in \mathbb{R}^{m+1}$ (sparse), $y \in \mathbb{R}^n$

Εύρεση ενεργών μεταβλητών:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{\rho} \|X\beta - y\|_\rho^\rho + \lambda \frac{1}{q} \|\beta\|_q^q \right\}, \quad 0 < \rho, q < 2 \qquad (2)$$

$\lambda > 0$ $\quad \|X\beta - y\|_\rho \rightarrow$ error term

$\|\beta\|_q \rightarrow$ regularization term

Αφού β sparse θέλω q: να ενισχύει το sparsity.

ρ: $\|X\beta - y\|_\rho \rightsquigarrow$ measures the residual

▷ Επιλογή ρ: ανάλογα με τον τύπο του θορύβου

- Λευκός Γκαουσιανός $\rightarrow \rho = 2$
- salt and pepper $\rightarrow 0 < \rho < 1$ πιο robust σε outliers

▷ Επιλογή q: επιδέχεται ώστε να ενισχυθεί το sparsity

- $q = 1$ προσέγγιση της lo νόρμας απ' την l_1 (η lo νόρμα μετράει το πλήθος των μη μηδενικών εισόδων αλλά: minimization της lo νόρμας δεν ανήκει στα → δύσκολο ⇒ προσέγγιση της lo-νόρμας απ' την l_1).
- $0 < q < 1$ προσεγγίζουν την lo-νόρμα καλύτερα: The smaller q the better approximation of lo-norm by l_q-norm. Όμως $0 < q < 1 \rightarrow$ non convex

• Το λ: ελέγχει το trade-off data fitting and sparsity

▷ Επιλογή λ: • μικρό $\lambda \rightsquigarrow$ οδηγεί σε dense solution αλλά ακριβή ανάκτηση των σεσ. ενών y.

• μεγάλο $\lambda \rightsquigarrow$ sparse solution αλλά μπορεί να μην έχω ακριβή ανάκτηση των σεσ. ενών y.

Χρησιμοποιούμε το μοντέλο παλινδρόμησης (2) για να λύσουμε το πρόβλημα (1).

Σκοπός: εύρεση ενεργών μεταβλητών μέσω των μη μηδενικών εκτιμήσεων του β που παίρνουμε από τη λύση του $\arg\min_{\beta} \left\{ \frac{1}{\rho} \|X\beta - y\|_\rho^\rho + \lambda \frac{1}{q} \|\beta\|_q^q \right\}$

▷ Στο Real Data model (it has been studied in almost every paper that deals with analysis of SSDs):

- Επειδή ο θόρυβος δεν είναι Gaussian → το $\rho = 2$ όχι καλή επιλογή
- Με βάση ευρετικές μεθόδους: $\rho < 1$ και πιο συγκεκριμένα: $\rho = 0{,}8$
- Για ενίσχυση sparsity: $q = 0{,}1$

Το ρ → data fidelity

Το q → sparsity

Το λ → regularization weight → ισορροπεί (i) το error $X\beta - y$
(ii) το sparsity $\|\beta\|_q$

$y = X\beta + \varepsilon$: Το β_j δείχνει: (i) πόσο συμβάλλει η στήλη X_j στο y
(ii) αν συμβάλλει θετικά ή αρνητικά
(iii) την ένταση με την οποία συμβάλλει.

$\beta_j \simeq 0 \Rightarrow$ η X_j δεν επηρεάζει το $y \Rightarrow$ μη ενεργός παράγοντας

$\beta_j \neq 0 \Rightarrow$ η X_j είναι σημαντική $\Rightarrow$ ενεργός παράγοντας

X → στήλες >> γραμμές → το β είναι η πιο αραιή λύση που εξηγεί σωστά τα δεδομένα

Συνήθη ρ, q για πειράματα: $\rho = [0.8, 1, 1.5, 2]$

$q = [0.1, 0.3, 0.5, 0.8, 1]$

Παράδειγμα

Έστω $X = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}, y = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$

και τρέχω τη μέθοδο lp-lq για $\lambda = [0.01, 0.1, 1, 10]$ και έστω ότι πήρα τα παρακάτω αποτελέσματα (υποθετικά δεν το έψαξα)

λ	b_1	b_2	b_3
0.01	1.95	0.98	0.05
0.1	1.90	0.95	0.01
1	1.70	0.85	0.00
10	1.20	0.60	0.00

Παρατηρήσεις: $b_1, b_2 \rightarrow$ μεγάλα για $\lambda = 0.01, 0.1, 1$

$b_3 \rightarrow$ μικρό καθώς $\lambda \uparrow$

Θέλω: sparse $b \Rightarrow$ οι μη σημαντικές συνιστώσες (b_3) να $\rightarrow 0$ και οι σημαντικές (b_1, b_2) να παραμένουν αρκετά μεγάλες

Άρα $\lambda = 0.1$ φαίνεται καλή επιλογή

($\lambda = 0.01 \rightarrow$ όλα τα $b_j \neq 0 \rightarrow$ παραμένει θόρυβος

$\lambda = 1 \rightarrow$ οι b_1, b_2 αρχίζουν να $\downarrow$

$\lambda = 10 \rightarrow$ οι b_1, b_2 (-επιώνονται πολύ $\Rightarrow$ χάνεται πληροφορία).

- Μικρό λ καλό για ακριβές fitting αλλά μπορεί να μη δώσει sparse b
- Μεγάλο λ καλό για επιλογή μεταβλητών αλλά μπορεί να χάνω πληροφορία