Μείωση διάστασης στην παλινδρόμηση

Δημήτρης Κουγιουμτζής

17 Δεκεμβρίου 2019

Εισαγωγή

Για τη μείωση μεταβλητών εισόδου σε μοντέλα παλινδρόμησης:

Εισαγωγή

Για τη μείωση μεταβλητών εισόδου σε μοντέλα παλινδρόμησης:

1. Γραμμικοί μετασχηματισμοί:
   - παλινδρόμηση κύριων συνιστωσών (principal component regression),
   - παλινδρόμηση με μερικά ελάχιστα τετράγωνα (partial least squares regression).

Εισαγωγή

Για τη μείωση μεταβλητών εισόδου σε μοντέλα παλινδρόμησης:

1. Γραμμικοί μετασχηματισμοί:
   - παλινδρόμηση κύριων συνιστωσών (principal component regression),
   - παλινδρόμηση με μερικά ελάχιστα τετράγωνα (partial least squares regression).

2. Τεχνικές στάθμισης μεταβλητών:
   - παλινδρόμηση ridge

Εισαγωγή

Για τη μείωση μεταβλητών εισόδου σε μοντέλα παλινδρόμησης:

1. Γραμμικοί μετασχηματισμοί:
   - παλινδρόμηση κύριων συνιστωσών (principal component regression),
   - παλινδρόμηση με μερικά ελάχιστα τετράγωνα (partial least squares regression).

2. Τεχνικές στάθμισης μεταβλητών:
   - παλινδρόμηση ridge

3. Τεχνικές επιλογής μεταβλητών:
   - βηματική παλινδρόμηση,
   - LASSO.

Πολλαπλή γραμμική παλινδρόμηση

Y: εξαρτημένη μεταβλητή (μεταβλητή απόκρισης)
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες (επεξηγηματικές) μεταβλητές

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon,$$
$$\mathrm{E}[\epsilon] = 0, \quad \sigma_e^2 = \mathrm{Var}[\epsilon]$$

Πολλαπλή γραμμική παλινδρόμηση

Y: εξαρτημένη μεταβλητή (μεταβλητή απόκρισης)
$X_1, X_2, \dots, X_K$: K ανεξάρτητες (επεξηγηματικές) μεταβλητές

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon,$$

$$\mathrm{E}[\epsilon] = 0, \quad \sigma_e^2 = \mathrm{Var}[\epsilon]$$

Προσοχή στην ύπαρξη πολυ-συγραμμικότητητας multiple collinearity (multicollinearity):
κάποιες από τις $X_1, X_2, \dots, X_K$ είναι ισχυρά συσχετισμένες.

Πολλαπλή γραμμική παλινδρόμηση

Y: εξαρτημένη μεταβλητή (μεταβλητή απόκρισης)
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες (επεξηγηματικές) μεταβλητές

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon,$$
$$\mathrm{E}[\epsilon] = 0, \quad \sigma_e^2 = \mathrm{Var}[\epsilon]$$

Προσοχή στην ύπαρξη πολυ-συγραμμικότητας multiple collinearity (multicollinearity):
κάποιες από τις $X_1, X_2, \ldots, X_K$ είναι ισχυρά συσχετισμένες.

Πρώτη προσέγγιση: βηματική παλινδρόμηση (stepwise regression) αλλά…

Πολλαπλή γραμμική παλινδρόμηση

Y: εξαρτημένη μεταβλητή (μεταβλητή απόκρισης)
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες (επεξηγηματικές) μεταβλητές

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon,$$
$$\mathrm{E}[\epsilon] = 0, \quad \sigma_e^2 = \mathrm{Var}[\epsilon]$$

Προσοχή στην ύπαρξη πολυ-συγραμμικότητας multiple collinearity (multicollinearity):
κάποιες από τις $X_1, X_2, \ldots, X_K$ είναι ισχυρά συσχετισμένες.

Πρώτη προσέγγιση: βηματική παλινδρόμηση (stepwise regression) αλλά...
τείνει να υπερ-προσαρμόζεται στα δεδομένα (καλύτερη προσαρμογή στο δείγμα εκμάθησης αλλά χειρότερη πρόβλεψη σε νέο σύνολο δεδομένων).

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$

$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon$, $\quad \epsilon$ τυχαίο σφάλμα

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^{\mathsf{T}}$

$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους: $n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon, \quad \epsilon$ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^{\mathsf{T}}$,
$U^{\mathsf{T}}U = I_n$, $V^{\mathsf{T}}V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon, \quad \epsilon$ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^\mathsf{T}$,
$U^\mathsf{T} U = I_n$, $V^\mathsf{T} V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Λύση ελαχίστων τετραγώνων (ordinary least squares (OLS)):

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon, \quad \epsilon$ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^\mathsf{T}$,
$U^\mathsf{T} U = I_n$, $V^\mathsf{T} V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Λύση ελαχίστων τετραγώνων (ordinary least squares (OLS)):

$$\boxed{\mathbf{b}_{\mathrm{OLS}} = V\Sigma^{-1}U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \dots, y_n]^\mathsf{T}$
$X_1, X_2, \dots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \dots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon, \quad \epsilon$ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^\mathsf{T}$,
$U^\mathsf{T} U = I_n$, $V^\mathsf{T} V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Λύση ελαχίστων τετραγώνων (ordinary least squares (OLS)):

$$\boxed{\mathbf{b}_{\mathrm{OLS}} = V\Sigma^{-1}U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

Κανονικοποίηση (regularization) της λύσης OLS για $\mathbf{b}$

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon$, $\quad \epsilon$ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^\mathsf{T}$,
$U^\mathsf{T} U = I_n$, $V^\mathsf{T} V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Λύση ελαχίστων τετραγώνων (ordinary least squares (OLS)):

$$\boxed{\mathbf{b}_{\mathrm{OLS}} = V\Sigma^{-1}U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

Κανονικοποίηση (regularization) της λύσης OLS για $\mathbf{b}$

$$\boxed{\mathbf{b} = V\Sigma^{-1}\Lambda U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon$, $\quad \epsilon$ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^\mathsf{T}$,
$U^\mathsf{T} U = I_n$, $V^\mathsf{T} V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Λύση ελαχίστων τετραγώνων (ordinary least squares (OLS)):

$$\boxed{\mathbf{b}_{\mathrm{OLS}} = V\Sigma^{-1}U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

Κανονικοποίηση (regularization) της λύσης OLS για $\mathbf{b}$

$$\boxed{\mathbf{b} = V\Sigma^{-1}\Lambda U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

όπου $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ λέγονται παράγοντες φίλτρου (filter factors)

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon$, ϵ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^\mathsf{T}$,
$U^\mathsf{T} U = I_n$, $V^\mathsf{T} V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Λύση ελαχίστων τετραγώνων (ordinary least squares (OLS)):

$$\boxed{\mathbf{b}_{\mathrm{OLS}} = V\Sigma^{-1}U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

Κανονικοποίηση (regularization) της λύσης OLS για $\mathbf{b}$

$$\boxed{\mathbf{b} = V\Sigma^{-1}\Lambda U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

όπου $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ λέγονται παράγοντες φίλτρου (filter factors)

Μείωση διάστασης $\equiv$ κανονικοποίηση της λύσης του $\mathbf{b}$

Μετασχηματισμός και προβολή

Y: εξαρτημένη μεταβλητή και παρατηρήσεις της: $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$
$X_1, X_2, \ldots, X_K$: K ανεξάρτητες μεταβλητές και παρατηρήσεις τους:
$n \times K$ πίνακας δεδομένων $X = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$

Μοντέλο γραμμικής παλινδρόμησης: $\mathbf{y} = X\mathbf{b} + \epsilon$, $\quad \epsilon$ τυχαίο σφάλμα

Ανάλυση ιδιάζουσων τιμών (SVD) του X: $X = U\Sigma V^\mathsf{T}$,
$U^\mathsf{T} U = I_n$, $V^\mathsf{T} V = I_r$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(n, K)$ τάξη του X

Λύση ελαχίστων τετραγώνων (ordinary least squares (OLS)):

$$\boxed{\mathbf{b}_{\mathrm{OLS}} = V\Sigma^{-1}U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

Κανονικοποίηση (regularization) της λύσης OLS για **b**

$$\boxed{\mathbf{b} = V\Sigma^{-1}\Lambda U^\mathsf{T}\mathbf{y} = \sum_{i=1}^{r} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i}$$

όπου $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ λέγονται παράγοντες φίλτρου (filter factors)

Μείωση διάστασης $\equiv$ κανονικοποίηση της λύσης του **b**

Οι μέθοδοι κανονικοποίησης εισάγουν μεροληψία στην λύσης OLS του **b** αλλά μειώνει τη διασπορά.

Παλινδρόμηση κύριων συνιστωσών

Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

Παλινδρόμηση κύριων συνιστωσών

Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

1. Περιστροφή της φυσικής βάσης του R^K: με PCA (ή SVD) υπολογίζονται οι K κύριες συνιστώσες (principal components, PCs), $X = U \Sigma V^{\mathsf{T}}$ (όπως για OLS).

Παλινδρόμηση κύριων συνιστωσών

Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

1. Περιστροφή της φυσικής βάσης του R^K: με PCA (ή SVD) υπολογίζονται οι K κύριες συνιστώσες (principal components, PCs), $X = U\Sigma V^\mathsf{T}$ (όπως για OLS).
2. Προβολή στον υποχώρο R^k, $k \leq r$: Επιλέγονται οι k πρώτες PCs από κάποιο κριτήριο, π.χ. ποσοστό διασποράς που εξηγείται.

Παλινδρόμηση κύριων συνιστωσών

Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

1. Περιστροφή της φυσικής βάσης του $\mathbf{R}^K$: με PCA (ή SVD) υπολογίζονται οι K κύριες συνιστώσες (principal components, PCs), $X = U\Sigma V^\top$ (όπως για OLS).
2. Προβολή στον υποχώρο $\mathbf{R}^k$, $k \leq r$: Επιλέγονται οι k πρώτες PCs από κάποιο κριτήριο, π.χ. ποσοστό διασποράς που εξηγείται.
3. Εκτίμηση του μοντέλου παλινδρόμησης στις επιλεγμένες k PCs και εύρεση του διανύσματος συντελεστών.

Παλινδρόμηση κύριων συνιστωσών

Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

1. Περιστροφή της φυσικής βάσης του $\mathbf{R}^K$: με PCA (ή SVD) υπολογίζονται οι K κύριες συνιστώσες (principal components, PCs), $X = U\Sigma V^\intercal$ (όπως για OLS).
2. Προβολή στον υποχώρο $\mathbf{R}^k$, $k \leq r$: Επιλέγονται οι k πρώτες PCs από κάποιο κριτήριο, π.χ. ποσοστό διασποράς που εξηγείται.
3. Εκτίμηση του μοντέλου παλινδρόμησης στις επιλεγμένες k PCs και εύρεση του διανύσματος συντελεστών.
4. Μετασχηματισμός του διανύσματος συντελεστών πίσω στον αρχικό χώρο των K ανεξάρτητων μεταβλητών.

Παλινδρόμηση κύριων συνιστωσών

Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

1. Περιστροφή της φυσικής βάσης του $\mathbf{R}^K$: με PCA (ή SVD) υπολογίζονται οι K κύριες συνιστώσες (principal components, PCs), $X = U\Sigma V^\mathsf{T}$ (όπως για OLS).
2. Προβολή στον υποχώρο $\mathbf{R}^k$, $k \leq r$: Επιλέγονται οι k πρώτες PCs από κάποιο κριτήριο, π.χ. ποσοστό διασποράς που εξηγείται.
3. Εκτίμηση του μοντέλου παλινδρόμησης στις επιλεγμένες k PCs και εύρεση του διανύσματος συντελεστών.
4. Μετασχηματισμός του διανύσματος συντελεστών πίσω στον αρχικό χώρο των K ανεξάρτητων μεταβλητών.

Η εκτίμηση PCR του $\mathbf{b}$ ως προς του παράγοντες φίλτρου:

Παλινδρόμηση κύριων συνιστωσών

Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

1. Περιστροφή της φυσικής βάσης του $\mathbf{R}^K$: με PCA (ή SVD) υπολογίζονται οι K κύριες συνιστώσες (principal components, PCs), $X = U\Sigma V^\mathsf{T}$ (όπως για OLS).
2. Προβολή στον υποχώρο $\mathbf{R}^k$, $k \leq r$: Επιλέγονται οι k πρώτες PCs από κάποιο κριτήριο, π.χ. ποσοστό διασποράς που εξηγείται.
3. Εκτίμηση του μοντέλου παλινδρόμησης στις επιλεγμένες k PCs και εύρεση του διανύσματος συντελεστών.
4. Μετασχηματισμός του διανύσματος συντελεστών πίσω στον αρχικό χώρο των K ανεξάρτητων μεταβλητών.

Η εκτίμηση PCR του $\mathbf{b}$ ως προς του παράγοντες φίλτρου:

$$\boxed{\mathbf{b}_{\mathrm{PCR}} = \sum_{i=1}^{k} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i} \qquad \lambda_{k+1} = \ldots = \lambda_r = 0$$

Παλινδρόμηση κύριων συνιστωσών

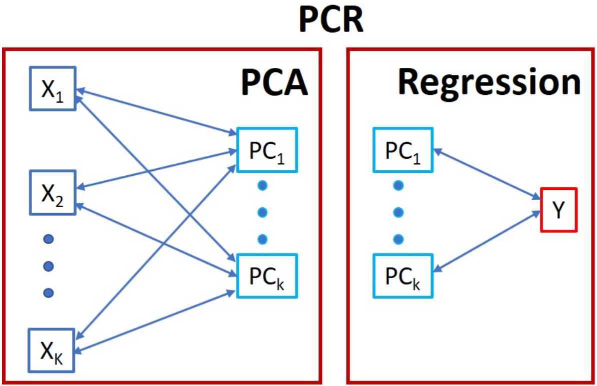
Η παλινδρόμηση κύριων συνιστωσών (principal component regression, PCR) βασίζεται στην ανάλυση κύριων συνιστωσών (PCA):

1. Περιστροφή της φυσικής βάσης του $\mathbb{R}^K$: με PCA (ή SVD) υπολογίζονται οι K κύριες συνιστώσες (principal components, PCs), $X = U\Sigma V^\mathsf{T}$ (όπως για OLS).
2. Προβολή στον υποχώρο $\mathbb{R}^k$, $k \leq r$: Επιλέγονται οι k πρώτες PCs από κάποιο κριτήριο, π.χ. ποσοστό διασποράς που εξηγείται.
3. Εκτίμηση του μοντέλου παλινδρόμησης στις επιλεγμένες k PCs και εύρεση του διανύσματος συντελεστών.
4. Μετασχηματισμός του διανύσματος συντελεστών πίσω στον αρχικό χώρο των K ανεξάρτητων μεταβλητών.

Η εκτίμηση PCR του $\mathbf{b}$ ως προς του παράγοντες φίλτρου:

$$\boxed{\mathbf{b}_{\mathrm{PCR}} = \sum_{i=1}^{k} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^\mathsf{T}\mathbf{y})\mathbf{v}_i} \qquad \lambda_{k+1} = \ldots = \lambda_r = 0$$

<u>Υπόθεση για την επιλογή k</u>: Η προβολή του $\mathbf{y}$ στις $r-k$ τελευταίες στήλες του U είναι κάτω από το επίπεδο του θορύβου.


PCR
PCA
Regression
X1
X2
XK
PC1
PCk
PC1
PCk
Y

Πλεονεκτήματα του PCR:

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.
- Αντιμετωπίζει την πολυ-συγγραμμικότητα. Οι ισχυρά συσχετισμένες μεταβλητές θα αντικατασταθούν με μια PC.

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.
- Αντιμετωπίζει την πολυ-συγγραμμικότητα. Οι ισχυρά συσχετισμένες μεταβλητές θα αντικατασταθούν με μια PC.
- Έχει μια μόνο ελεύθερη παράμετρο, k.

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.
- Αντιμετωπίζει την πολυ-συγραμμικότητα. Οι ισχυρά συσχετισμένες μεταβλητές θα αντικατασταθούν με μια PC.
- Έχει μια μόνο ελεύθερη παράμετρο, k.
- Μπορεί να εφαρμοσθεί και για $K > n$.

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.
- Αντιμετωπίζει την πολυ-συγγραμμικότητα. Οι ισχυρά συσχετισμένες μεταβλητές θα αντικατασταθούν με μια PC.
- Έχει μια μόνο ελεύθερη παράμετρο, k.
- Μπορεί να εφαρμοσθεί και για $K > n$.

Μειονεκτήματα του PCR:

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.
- Αντιμετωπίζει την πολυ-συγραμμικότητα. Οι ισχυρά συσχετισμένες μεταβλητές θα αντικατασταθούν με μια PC.
- Έχει μια μόνο ελεύθερη παράμετρο, k.
- Μπορεί να εφαρμοσθεί και για $K > n$.

Μειονεκτήματα του PCR:

- Στη μείωση διάστασης δεν συμμετέχει η εξαρτημένη μεταβλητή. Οι PCs του PCA μπορεί να μην έχουν προβλεπτική ισχύ αντίστοιχη με αυτήν του συνόλου των K ανεξαρτήτων μεταβλητών. Το $\mathbf{y}$ μπορεί ακόμα και να είναι κάθετο στον υποχώρο που ορίζεται από τις k πρώτες στήλες του U.

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.
- Αντιμετωπίζει την πολυ-συγραμμικότητα. Οι ισχυρά συσχετισμένες μεταβλητές θα αντικατασταθούν με μια PC.
- Έχει μια μόνο ελεύθερη παράμετρο, k.
- Μπορεί να εφαρμοσθεί και για $K > n$.

Μειονεκτήματα του PCR:

- Στη μείωση διάστασης δεν συμμετέχει η εξαρτημένη μεταβλητή. Οι PCs του PCA μπορεί να μην έχουν προβλεπτική ισχύ αντίστοιχη με αυτήν του συνόλου των K ανεξαρτήτων μεταβλητών. Το $\mathbf{y}$ μπορεί ακόμα και να είναι κάθετο στον υποχώρο που ορίζεται από τις k πρώτες στήλες του U.
- Η υπόθεση πως συνιστώσες μικρής διασποράς έχουν και μικρή προβλεπτική ισχύ σε σχέση με συνιστώσες μεγάλης διασποράς δεν είναι πάντα ορθή.

Πλεονεκτήματα του PCR:

- Μπορεί να μειώσει αποτελεσματικά τη διάσταση, $k \ll r$. Με αυτόν τον τρόπο το PCR μειώνει την υπερ-προσαρμογή που οφείλεται στην ύπαρξη πολλών μεταβλητών στο μοντέλο.
- Αντιμετωπίζει την πολυ-συγραμμικότητα. Οι ισχυρά συσχετισμένες μεταβλητές θα αντικατασταθούν με μια PC.
- Έχει μια μόνο ελεύθερη παράμετρο, k.
- Μπορεί να εφαρμοσθεί και για $K > n$.

Μειονεκτήματα του PCR:

- Στη μείωση διάστασης δεν συμμετέχει η εξαρτημένη μεταβλητή. Οι PCs του PCA μπορεί να μην έχουν προβλεπτική ισχύ αντίστοιχη με αυτήν του συνόλου των K ανεξαρτήτων μεταβλητών. Το $\mathbf{y}$ μπορεί ακόμα και να είναι κάθετο στον υποχώρο που ορίζεται από τις k πρώτες στήλες του U.
- Η υπόθεση πως συνιστώσες μικρής διασποράς έχουν και μικρή προβλεπτική ισχύ σε σχέση με συνιστώσες μεγάλης διασποράς δεν είναι πάντα ορθή.
- Όταν οι ανεξάρτητες μεταβλητές είναι σε διαφορετικές κλίμακες τιμών θα πρέπει να τυποποιηθούν πριν την εφαρμογή του PCA.

Παλινδρόμηση με μερικά ελάχιστα τετράγωνα

Η παλινδρόμηση με μερικά ελάχιστα τετράγωνα (principal least square, PLS) προσπαθεί να μετασχηματίσει το X ώστε να μεγιστοποιηθεί η συνδιασπορά του με το Y.

Παλινδρόμηση με μερικά ελάχιστα τετράγωνα

Η παλινδρόμηση με μερικά ελάχιστα τετράγωνα (principal least square, PLS) προσπαθεί να μετασχηματίσει το X ώστε να μεγιστοποιηθεί η συνδιασπορά του με το Y.

$X = TP^{\mathsf{T}} + E$,
T: $n \times k$ πίνακας των k κρυφών διανυσμάτων (scores) του X.
P: $K \times k$ πίνακας των k μετασχηματισμένων κρυφών μεταβλητών (loadings). E: $n \times K$ πίνακας υπολοίπων

Παλινδρόμηση με μερικά ελάχιστα τετράγωνα

Η παλινδρόμηση με μερικά ελάχιστα τετράγωνα (principal least square, PLS) προσπαθεί να μετασχηματίσει το X ώστε να μεγιστοποιηθεί η συνδιασπορά του με το Y.

$X = TP^{\mathsf{T}} + E$,
T: $n \times k$ πίνακας των k κρυφών διανυσμάτων (scores) του X.
P: $K \times k$ πίνακας των k μετασχηματισμένων κρυφών μεταβλητών (loadings). E: $n \times K$ πίνακας υπολοίπων

Έχουν αναπτυχθεί διάφοροι αλγόριθμοι για την υλοποίηση του PLS.

Παλινδρόμηση με μερικά ελάχιστα τετράγωνα

Η παλινδρόμηση με μερικά ελάχιστα τετράγωνα (principal least square, PLS) προσπαθεί να μετασχηματίσει το X ώστε να μεγιστοποιηθεί η συνδιασπορά του με το Y.

$X = TP^{\mathsf{T}} + E$,
T: $n \times k$ πίνακας των k κρυφών διανυσμάτων (scores) του X.
P: $K \times k$ πίνακας των k μετασχηματισμένων κρυφών μεταβλητών (loadings). E: $n \times K$ πίνακας υπολοίπων

Έχουν αναπτυχθεί διάφοροι αλγόριθμοι για την υλοποίηση του PLS.

Η εκτίμηση PLS του $\mathbf{b}$ ως προς τους παράγοντες φίλτρου (μετασχηματίζοντας τον υποχώρο R^k στον χώρο R^r που ορίζεται από τη βάση των ιδιάζοντων διανυσμάτων):

Παλινδρόμηση με μερικά ελάχιστα τετράγωνα

Η παλινδρόμηση με μερικά ελάχιστα τετράγωνα (principal least square, PLS) προσπαθεί να μετασχηματίσει το X ώστε να μεγιστοποιηθεί η συνδιασπορά του με το Y.

$X = TP^{\mathsf{T}} + E$,
T: $n \times k$ πίνακας των k κρυφών διανυσμάτων (scores) του X.
P: $K \times k$ πίνακας των k μετασχηματισμένων κρυφών μεταβλητών (loadings). E: $n \times K$ πίνακας υπολοίπων

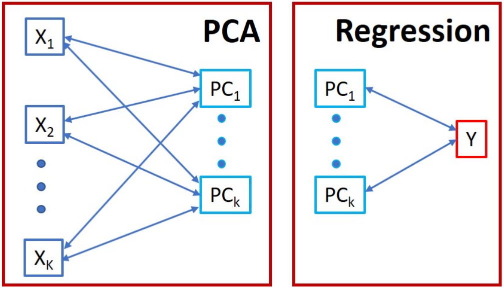
Έχουν αναπτυχθεί διάφοροι αλγόριθμοι για την υλοποίηση του PLS.

Η εκτίμηση PLS του $\mathbf{b}$ ως προς τους παράγοντες φίλτρου (μετασχηματίζοντας τον υποχώρο R^k στον χώρο R^r που ορίζεται από τη βάση των ιδιάζοντων διανυσμάτων):

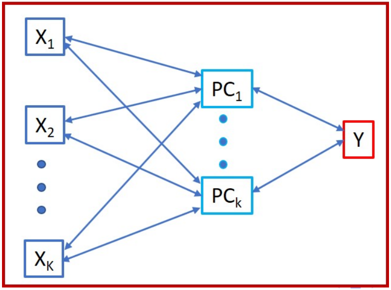
$$\boxed{\mathbf{b}_{\mathrm{PLS}} = \sum_{i=1}^{r} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^{\mathsf{T}}\mathbf{y})\mathbf{v}_i} \qquad \lambda_i = 1 - \prod_{j=1}^{q}\left(1 - \frac{\sigma_i^2}{\theta_j}\right),$$

όπου $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_r$ είναι οι Ritz τιμές (πολύπλοκες εκφράσεις)
Ενώ τα λ_i για $i = k+1, \ldots, r$ δεν είναι 0, έχουμε $\|\mathbf{b}_{\mathrm{PLS}}\|_2 \leq \|\mathbf{b}_{\mathrm{PCR}}\|_2$

PCR



PLS



Πλεονεκτήματα της PLS:

Πλεονεκτήματα της PLS:

- Μπορεί να εφαρμοσθεί και για πολλαπλές εξαρτημένες μεταβλητές.

Πλεονεκτήματα της PLS:

- Μπορεί να εφαρμοσθεί και για πολλαπλές εξαρτημένες μεταβλητές.
- Αντιμετωπίζει αποτελεσματικά τη συγραμμικότητα.

Πλεονεκτήματα της PLS:

- Μπορεί να εφαρμοσθεί και για πολλαπλές εξαρτημένες μεταβλητές.
- Αντιμετωπίζει αποτελεσματικά τη συγραμμικότητα.
- Στη μείωση διάστασης (κανονικοποίηση) συμμετέχει η εξαρτημένη μεταβλητή και για αυτό το μοντέλο έχει μεγάλη προβλεπτική ισχύ.

Πλεονεκτήματα της PLS:

- Μπορεί να εφαρμοσθεί και για πολλαπλές εξαρτημένες μεταβλητές.
- Αντιμετωπίζει αποτελεσματικά τη συγραμμικότητα.
- Στη μείωση διάστασης (κανονικοποίηση) συμμετέχει η εξαρτημένη μεταβλητή και για αυτό το μοντέλο έχει μεγάλη προβλεπτική ισχύ.

Μειονεκτήματα PLS:

Πλεονεκτήματα της PLS:

- Μπορεί να εφαρμοσθεί και για πολλαπλές εξαρτημένες μεταβλητές.
- Αντιμετωπίζει αποτελεσματικά τη συγραμμικότητα.
- Στη μείωση διάστασης (κανονικοποίηση) συμμετέχει η εξαρτημένη μεταβλητή και για αυτό το μοντέλο έχει μεγάλη προβλεπτική ισχύ.

Μειονεκτήματα PLS:

- Είναι δύσκολο να ερμηνεύσουμε του συντελεστές των k μετασχηματισμένων μεταβλητών στον πίνακα P (loadings).

Πλεονεκτήματα της PLS:

- Μπορεί να εφαρμοσθεί και για πολλαπλές εξαρτημένες μεταβλητές.
- Αντιμετωπίζει αποτελεσματικά τη συγραμμικότητα.
- Στη μείωση διάστασης (κανονικοποίηση) συμμετέχει η εξαρτημένη μεταβλητή και για αυτό το μοντέλο έχει μεγάλη προβλεπτική ισχύ.

Μειονεκτήματα PLS:

- Είναι δύσκολο να ερμηνεύσουμε του συντελεστές των k μετασχηματισμένων μεταβλητών στον πίνακα P (loadings).
- Η κατανομή των εκτιμητών $\mathbf{b}$ και $\hat{y}$ δεν είναι γνωστές και δε μπορεί να γίνει παραμετρικά η συμπερασματολογία (διαστήματα εμπιστοσύνης και έλεγχοι υπόθεσης).

Παλινδρόμηση ridge

Η παλινδρόμηση ridge (RR) συρρικνώνει του συντελεστές παλινδρόμησης (τα στοιχεία του $\mathbf{b}$) προς το 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_2$.

Παλινδρόμηση ridge

Η παλινδρόμηση ridge (RR) συρρικνώνει του συντελεστές παλινδρόμησης (τα στοιχεία του **b**) προς το 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_2$.

RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\mathrm{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

όπου $\mu > 0$ είναι η παράμετρος κανονικοποίησης που ελέγχει την επίδραση του όρου ποινής.

Παλινδρόμηση ridge

Η παλινδρόμηση ridge (RR) συρρικνώνει του συντελεστές παλινδρόμησης (τα στοιχεία του $\mathbf{b}$) προς το 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_2$.

RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\text{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

όπου $\mu > 0$ είναι η παράμετρος κανονικοποίησης που ελέγχει την επίδραση του όρου ποινής.

Η εκτίμηση RR του $\mathbf{b}$ ως προς τους παράγοντες φίλτρου:

Παλινδρόμηση ridge

Η παλινδρόμηση ridge (RR) συρρικνώνει του συντελεστές παλινδρόμησης (τα στοιχεία του $\mathbf{b}$) προς το 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_2$.

RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\text{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

όπου $\mu > 0$ είναι η παράμετρος κανονικοποίησης που ελέγχει την επίδραση του όρου ποινής.

Η εκτίμηση RR του $\mathbf{b}$ ως προς τους παράγοντες φίλτρου:

$$\boxed{\mathbf{b}_{\text{RR}} = \sum_{i=1}^{r} \frac{\lambda_i}{\sigma_i}(\mathbf{u}_i^T\mathbf{y})\mathbf{v}_i} \qquad \lambda_i = \frac{\sigma_i^2}{\sigma_i^2+\mu}$$

+ Το RR αποδίδει καλά ως προς την ακρίβεια πρόβλεψης.

+ Το RR αποδίδει καλά ως προς την ακρίβεια πρόβλεψης.
− Η εφαρμογή του RR βασίζεται κύρια στην επιλογή του μ.

+ Το RR αποδίδει καλά ως προς την ακρίβεια πρόβλεψης.

− Η εφαρμογή του RR βασίζεται κύρια στην επιλογή του μ.

Κατάλληλη επιλογή για μ είναι η εκτίμηση της διασποράς των υπολοίπων από την εκτίμηση της παλινδρόμησης με OLS, που ισοδύναμα δίνεται ως

$$\mu = s_e^2 = \frac{1}{n-r} \sum_{i=r+1}^{n} (\mathbf{u}_i^\mathsf{T} \mathbf{y})^2$$

+ Το RR αποδίδει καλά ως προς την ακρίβεια πρόβλεψης.

— Η εφαρμογή του RR βασίζεται κύρια στην επιλογή του μ.

Κατάλληλη επιλογή για μ είναι η εκτίμηση της διασποράς των υπολοίπων από την εκτίμηση της παλινδρόμησης με OLS, που ισοδύναμα δίνεται ως

$$\mu = s_e^2 = \frac{1}{n-r} \sum_{i=r+1}^{n} (\mathbf{u}_i^T \mathbf{y})^2$$

— Το RR δεν επιτελεί επιλογή μεταβλητών (μείωση διάστασης).

+ Το RR αποδίδει καλά ως προς την ακρίβεια πρόβλεψης.

− Η εφαρμογή του RR βασίζεται κύρια στην επιλογή του μ.

Κατάλληλη επιλογή για μ είναι η εκτίμηση της διασποράς των υπολοίπων από την εκτίμηση της παλινδρόμησης με OLS, που ισοδύναμα δίνεται ως

$$\mu = s_e^2 = \frac{1}{n-r} \sum_{i=r+1}^{n} (\mathbf{u}_i^T \mathbf{y})^2$$

− Το RR δεν επιτελεί επιλογή μεταβλητών (μείωση διάστασης).

− Είναι δύσκολο να ερμηνευθεί το RR.

LASSO

LASSO

Το RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\text{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

Οι συντελεστές στο διάνυσμα $\mathbf{b}_{\text{RR}}$ συρρικνώνονται αλλά δε γίνονται 0.

LASSO

Το RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\mathrm{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

Οι συντελεστές στο διάνυσμα $\mathbf{b}_{\mathrm{RR}}$ συρρικνώνονται αλλά δε γίνονται 0.

Η μέθοδος LASSO (Least Absolute Selection and Shrinkage Operator) θέτει τους συντελεστές του $\mathbf{b}$ σε 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_1$.

LASSO

Το RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\mathrm{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

Οι συντελεστές στο διάνυσμα $\mathbf{b}_{\mathrm{RR}}$ συρρικνώνονται αλλά δε γίνονται 0.

Η μέθοδος LASSO (Least Absolute Selection and Shrinkage Operator) θέτει τους συντελεστές του $\mathbf{b}$ σε 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_1$.

Η LASSO ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\mathrm{LASSO}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_1, \quad \|\beta\|_1 = \sum_{i}^{r} |\beta_i|$$

$\mu = 0 \Rightarrow$ ΟΛΣ $\quad \mu = \infty \Rightarrow \mathbf{b}_{\mathrm{LASSO}} = \mathbf{0}$

LASSO

Το RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\text{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

Οι συντελεστές στο διάνυσμα $\mathbf{b}_{\text{RR}}$ συρρικνώνονται αλλά δε γίνονται 0.

Η μέθοδος LASSO (Least Absolute Selection and Shrinkage Operator) θέτει τους συντελεστές του $\mathbf{b}$ σε 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_1$.

Η LASSO ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\text{LASSO}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_1, \quad \|\beta\|_1 = \sum_{i}^{r} |\beta_i|$$

$\mu = 0 \Rightarrow$ ΟΛΣ $\quad \mu = \infty \Rightarrow \mathbf{b}_{\text{LASSO}} = \mathbf{0}$

\+ Η LASSO εφαρμόζει επιλογή μεταβλητών (μείωση διάστασης) σύμφωνα με την τιμή του $\mu > 0$ και είναι εύκολο να ερμηνευθεί.

LASSO

Το RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\text{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

Οι συντελεστές στο διάνυσμα $\mathbf{b}_{\text{RR}}$ συρρικνώνονται αλλά δε γίνονται 0.

Η μέθοδος LASSO (Least Absolute Selection and Shrinkage Operator) θέτει τους συντελεστές του $\mathbf{b}$ σε 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_1$.

Η LASSO ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\text{LASSO}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_1, \quad \|\beta\|_1 = \sum_{i}^{r} |\beta_i|$$

$\mu = 0 \Rightarrow$ ΟΛΣ $\quad \mu = \infty \Rightarrow \mathbf{b}_{\text{LASSO}} = \mathbf{0}$

+ Η LASSO εφαρμόζει επιλογή μεταβλητών (μείωση διάστασης) σύμφωνα με την τιμή του $\mu > 0$ και είναι εύκολο να ερμηνευθεί.

+ Η LASSO αποδίδει καλά ως προς την ακρίβεια πρόβλεψης.

LASSO

Το RR ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\mathrm{RR}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_2^2$$

Οι συντελεστές στο διάνυσμα $\mathbf{b}_{\mathrm{RR}}$ συρρικνώνονται αλλά δε γίνονται 0.

Η μέθοδος LASSO (Least Absolute Selection and Shrinkage Operator) θέτει τους συντελεστές του $\mathbf{b}$ σε 0 επιβάλλοντας ποινή στο $\|\mathbf{b}\|_1$.

Η LASSO ελαχιστοποιεί τη συνάρτηση κόστους

$$\mathbf{b}_{\mathrm{LASSO}} = \arg\max_{\beta} \|\mathbf{y} - X\beta\|_2^2 + \mu\|\beta\|_1, \quad \|\beta\|_1 = \sum_{i}^{r} |\beta_i|$$

$\mu = 0 \Rightarrow$ ΟΛΣ $\quad \mu = \infty \Rightarrow \mathbf{b}_{\mathrm{LASSO}} = \mathbf{0}$

+ Η LASSO εφαρμόζει επιλογή μεταβλητών (μείωση διάστασης) σύμφωνα με την τιμή του $\mu > 0$ και είναι εύκολο να ερμηνευθεί.

+ Η LASSO αποδίδει καλά ως προς την ακρίβεια πρόβλεψης.

− Η επίδοση της LASSO βασίζεται στην επιλογή της παραμέτρου μ.

5. Δημιουργήστε τον πίνακα δεδομένων X μεγέθους $n \times p$ από $p = 5$ μεταβλητές που ακολουθούν εκθετική κατανομή με διαφορετικές μέσες τιμές η κάθε μια (ελεύθερη επιλογή). Δημιουργήστε το διάνυσμα απόκρισης $\mathbf{y}$ από τη σχέση $\mathbf{y} = X\beta + \epsilon$, όπου το διάνυσμα συντελεστών β έχει μόνο δύο μη-μηδενικά στοιχεία, π.χ. $\beta = [0, 2, 0, -3, 0]^{\mathrm{T}}$, και η τυχαία μεταβλητή θορύβου, που δίνει τις τιμές στο ϵ, ακολουθεί κανονική κατανομή με δεδομένη τυπική απόκλιση, π.χ. $\sigma_\epsilon = 5$.

   (α') Εκτιμήστε το μοντέλο πολλαπλής γραμμικής παλινδρόμησης με κάθε μια από τις μεθόδους OLS, PCR, PLS, RR και LASSO.

   (β') Για κάθε μέθοδο σχηματίστε το διάγραμμα διασποράς των παρατηρούμενων και εκτιμώμενων τιμών της εξαρτημένης μεταβλητής καθώς και το διάγραμμα των τυποποιημένων σφαλμάτων παλινδρόμησης.

   (γ') Συγκρίνετε τις εκτιμήσεις του β με κάθε μια μέθοδο.

6. Στην Άσκηση 5.8 εκτιμήσαμε το μοντέλο γραμμικής παλινδρόμησης με την OLS και τη βηματική παλινδρόμηση, όπου χρησιμοποιήθηκαν τα δεδομένα στο αρχείο `physical.txt`. Στα ίδια δεδομένα εκτιμήστε το μοντέλο γραμμικής παλινδρόμησης με τις μεθόδους PCR, PLS, RR και LASSO. Συγκρίνετε τα αποτελέσματα με αυτά των OLS και βηματικής παλινδρόμησης.