

Αριθμητικές Μεθοδοι στην Στατιστική

ΕΝΟΤΗΤΑ 1: Αριθμητική Γραμμική Άλγεβρα - Εφαρμογές

Α. Ν. Γιαννακόπουλος

Ο.Π.Α

Εαρινό Εξάμηνο 2026

- 1 Προβλήματα ελαχιστων τετραγώνων
- 2 Singular value decomposition
- 3 Ομαλοποίηση *Tikhonov*
- 4 Προσομοίωση κανονικών κατανομων

Προβλήματα ελαχίστων τετραγώνων

Μια σημαντική εφαρμογή της αριθμητικής γραμμικής άλγεβρας είναι στην επίλυση προβλημάτων ελαχίστων τετραγώνων.

Εφαρμογή 1: Γραμμική παλινδρόμηση

Ας υποθέσουμε το μοντέλο

$$y_i = x_{i1}b_1 + \cdots + x_{ip}b_p + \epsilon_i = x_i b + \epsilon_i,$$

όπου y_i είναι παρατηρήσεις της εξαρτώμενης μεταβλητής και $(x_{i1}, \dots, x_{ip})$ οι παρατηρήσεις της ανεξάρτητης μεταβλητής.

Το συνολικό πλήθος των παρατηρήσεων είναι $i = 1, \dots, n$.

$b_1, \dots, b_p$ είναι οι παράμετροι του γραμμικού μοντέλου που θέλουμε να υπολογίσουμε.

Οι παρατηρήσεις μας μπορεί να γραφούν σε συμπαγή μορφή πινάκων ως

$$y = Xb + e, \quad \text{cov}(e) = \sigma^2 I_n$$

Ο πίνακας X ονομάζεται πίνακας σχεδιασμού (design matrix)

Η ελαχιστοποίηση του μεσου τετραγωνικού σφάλματος του υποδείγματος είναι ισοδυναμη με την ελαχιστοποίηση της συνάρτησης

$$S(b) = (y - Xb)^T (y - Xb).$$

Η κανονική εξίσωση είναι η

$$X^T X b = X^T y$$

Πολυωνυμική παλινδρομηση

Ας υποθέσουμε ότι έχουμε δεδομένα (x_i, y_i) , $i = 1, \dots, n$ τα οποία θεωρούμε οτι προέρχονται απο το πολυωνυμικό υπόδειγμα

$$y_i = p(x_i) = c_0 + c_1 x_i + \dots + c_{m-1} x_i^{m-1},$$

για κάποιο $m < n$.

Η επιλογή ενός τέτοιου πολυωνύμου γίνεται επιλύοντας το πρόβλημα βελτιστοποίησης

$$\min_{c_0, c_1, \dots, c_{m-1}} \sum_{i=1}^n |p(x_i) - y_i|^2$$

Αν το υποδειγμα ήταν ακριβές θα μπορούσε να γραφεί στην μορφή $r = 0$ όπου

$$r = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{m-1} \\ 1 & x_2 & \cdots & x_2^{m-1} \\ \cdots & \cdots & \ddots & \cdots \\ 1 & x_n & \cdots & x_n^{m-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \cdots \\ c_{m-1} \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

Η καλύτερη επιλογή του μοντέλου είναι ελαχιστοποιώντας την νόρμα του υπολοίπου r , δηλαδή επιλύοντας ένα πρόβλημα ελαχιστων τετραγώνων.

Ο πίνακας που παράγεται απο τα x_i ονομάζεται πίνακας Vandermonde

Προβλήματα ελαχίστων τετραγώνων

Τα προβλήματα ελαχίστων τετραγώνων εμφανίζονται πολύ συχνά την στατιστική, την μηχανική μάθηση και τα εφαρμοσμένα μαθηματικά.

Η γενική μορφή ενός τέτοιου προβλήματος είναι δεδομένου ενός πίνακα $A \in \mathbb{R}^{n \times m}$ και ενός διανύσματος $b \in \mathbb{R}^n$ να βρούμε ένα διάνυσμα $x \in \mathbb{R}^m$, τέτοιο ώστε

$$\|Ax - b\| = \min_{z \in \mathbb{R}^m} \|Az - b\|.$$

Η λύση του προβλήματος δίνεται απο το συστημα γραμμικών εξισώσεων

$$A^T Ax = A^T b$$

Η λύση του παραπάνω συστήματος ονομάζεται λύση ελαχίστων τετραγώνων του συστήματος $Ax = b$ και είναι μοναδική αν ο πίνακας $A^T A$ είναι μη ιδιόμορφος (ή ισοδύναμα ο πίνακας A έχει τάξη $\text{rank}(A) = \min(m, n)$).

$$A \in \mathbb{R}^{n \times m}, n \geq m, \text{rank}(A) = m$$

Στην περίπτωση αυτή υπάρχει ο $(A^T A)^{-1}$ και

$$x = (A^T A)^{-1} A^T b.$$

Ο πίνακας $A^\dagger = (A^T A)^{-1} A^T$ ονομάζεται ο Moore - Penrose ψευδοαντίστροφος του A .

Η λύση στο υπερκαθορισμένο πρόβλημα ελαχίστων τετραγώνων $Ax = b$ είναι η $x = A^\dagger b$.

Πως αποδεικνύεται η κανονική εξίσωση;

Ας ορίσουμε την συνάρτηση $F : \mathbb{R}^m \rightarrow \mathbb{R}$,

$$z \mapsto F(z) = \|Az - b\|^2 = \langle Az - b, Az - b \rangle$$

και ας υποθέσουμε ότι το ελαχιστό της επιτυγχάνεται στο $z = x$.

Ας μετακινηθούμε σε ένα γειτονικό σημείο $x + \epsilon z$ και ας υπολογίσουμε

$$\begin{aligned} \frac{1}{\epsilon}(F(x + \epsilon z) - F(x)) &= 2\langle Ax - b, Ax \rangle + O(\epsilon) \\ &= 2\langle A^T Ax - A^T b, z \rangle + O(\epsilon), \quad \forall z \in \mathbb{R}^m, \epsilon > 0. \end{aligned}$$

Εφ'οσον $F(x) = \min_{z \in \mathbb{R}^m} F(z)$, θα έχουμε

$$0 \leq \frac{1}{\epsilon}(F(x + \epsilon z) - F(x)) = 2\langle A^T Ax - A^T b, z \rangle + O(\epsilon), \quad \forall z \in \mathbb{R}^m, \epsilon > 0,$$

και περνώντας στο όριο καθώς $\epsilon \rightarrow 0^+$

$$0 \leq 2\langle A^T Ax - A^T b, z \rangle, \quad \forall z \in \mathbb{R}^m.$$

Για οποιοδήποτε $z \in \mathbb{R}^m$ παίρνουμε το παραπάνω και για $-z \in \mathbb{R}^m$ συνεπώς

$$\langle A^T A x - A^T b, z \rangle = 0, \quad \forall z \in \mathbb{R}^m,$$

άρα

$$A^T A x - A^T b = 0$$

δηλ. η κανονική εξίσωση.

$$A \in \mathbb{R}^{n \times m}, n < m, \text{rank}(A) = n$$

Στην περίπτωση αυτή ο πίνακας AA^T είναι μη ιδιόμορφος και μπορούμε να ορίσουμε τον ψευδοαντίστροφο

$$A^\dagger = A^T(AA^T)^{-1},$$

ο οποίος είναι και ο ψευδοαντίστροφος για τον A^T .

Επίλυση προβλημάτων ελαχίστων τετραγώνων

- Η επίλυση προβλημάτων ελαχίστων τετραγώνων μπορεί να γίνει μέσω
- 1 της εξίσωσης $A^T A x = A^T b$ με την χρήση της διάσπασης Cholesky .
 - 2 της διάσπασης QR
 - 3 της διάσπασης σε ιδιάζουσες τιμές SVD .
- Οι μέθοδοι 2 και 3 είναι πολλές φορές προτιμότερες.

Σύντομη υπενθύμιση βασικών εννοιών απο την γραμμική άλγεβρα

Πριν προχωρήσουμε θα υπενθυμίσουμε ορισμένες βασικές έννοιες απο την γραμμική άλγεβρα που είναι πολύ χρησιμες στην επίλυση προβλημάτων ελαχίστων τετραγώνων.

1. Διάσπαση Cholesky

- Κάθε πίνακας A **συμμετρικός και θετικά ορισμένος** μπορεί να διασπαστεί ως

$$A = LL^T$$

2. Διάσπαση QR .

- Κάθε πίνακας $A \in \mathbb{R}^{n \times m}$, $n \geq m$ μπορεί να παραγοντοποιηθεί ως

$$A = Q \begin{pmatrix} R \\ 0_{n-m,m} \end{pmatrix} = Q \begin{pmatrix} U \\ 0_{n-m,m} \end{pmatrix} = [\hat{Q} | \bar{Q}] \begin{pmatrix} U \\ 0_{n-m,m} \end{pmatrix},$$

ή ισοδύναμα $A = \hat{Q}\hat{R}$, με Q ορθογώνιοι πίνακες και R άνω τριγωνικοί.

Επίλυση με την διάσπαση QR

Ας θυμηθούμε την διάσπαση QR ενός πίνακα $A \in \mathbb{R}^{n \times m}$, $n \geq m$,

$$A = Q \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix} = [\hat{Q} \mid \bar{Q}] \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix}$$

με διαστάσεις

$$A^{(n \times m)} = Q^{(n \times n)} \begin{bmatrix} \hat{R}^{(m \times m)} \\ 0^{(n-m \times m)} \end{bmatrix} = [\hat{Q}^{(n \times m)} \mid \bar{Q}^{(n \times n-m)}] \begin{bmatrix} \hat{R}^{(m \times m)} \\ 0^{(n-m \times m)} \end{bmatrix}$$

ή η ανηγμένη μορφή

$$A = \hat{Q} \hat{R}.$$

Τότε

$$A^T A = \hat{R}^T \hat{Q}^T \hat{Q} R = R^T R$$

και η κανονική εξίσωση γίνεται

$$R^T R x = R^T \hat{Q}^T b$$

Αν ο πίνακας A είναι full rank τότε $r_{ii} > 0$ οπότε το πρόβλημα γράφεται

$$R x = \hat{Q}^T b.$$

Το πρόβλημα αυτό είναι ένα άνω τριγωνικό σύστημα το οποίο μπορεί να λυθεί με backsubstitution

Η εξίσωση $R x = \hat{Q}^T b$ μπορεί να εξαχθεί χρησιμοποιώντας την έννοια της προβολής παρακάμπτοντας την κανονική εξίσωση

Ελάχιστα τετράγωνα με την χρήση της διάσπασης QR

- 1 Υπολογίζουμε την ανηγμένη διάσπαση QR του A

$$A = \hat{Q}\hat{R}$$

- 2 Υπολογίζουμε το διάνυσμα

$$\hat{Q}^T b$$

- 3 Λύνουμε το άνω τριγωνικό σύστημα

$$\hat{R}x = \hat{Q}^T b$$

χρησιμοποιώντας backsubstitution

$$\begin{aligned}
A &= QR, \\
A &\in \mathbb{R}^{n \times m}, \quad Q \in \mathbb{R}^{n \times n}, \quad R \in \mathbb{R}^{n \times m}, \\
Q &\text{ ορθογώνιος, } R \text{ άνω τριγωνικός}
\end{aligned}$$

$$\begin{aligned}
\|Ax - b\|^2 &= \|QRx - b\|^2 = \langle QRx - b, QRx - b \rangle = \langle QRx - b, QRx - \underbrace{QQ^T}_{I} b \rangle \\
&= \langle Q^T(QRx - b), Rx - Q^T b \rangle = \langle Rx - \underbrace{Q^T b}_c, Rx - \underbrace{Q^T b}_c \rangle = \|Rx - c\|^2 \\
&= \left\| \begin{pmatrix} \hat{R} \\ 0 \end{pmatrix} x - \begin{pmatrix} \hat{c} \\ \bar{c} \end{pmatrix} \right\|^2 = \|\hat{R}x - \hat{c}\|^2 + \|\bar{c}\|^2
\end{aligned}$$

Η λύση στο πρόβλημα των ελαχίστων τετραγώνων είναι η

$$\hat{R}x = \hat{c}$$

η οποία αντιστοιχεί σε ένα άνω τριγωνικό σύστημα το οποίο μπορεί να επιλυθεί χρησιμοποιώντας back substitution

$$A \in \mathbb{R}^{n \times m}, n \geq m, \text{rank}(A) = m$$

- $A = QR$
- $\hat{R} \in \mathbb{R}^{m \times m}$ το άνω τριγωνικό (μη μηδενικό) κομμάτι του R .
- $c = Q^T b \in \mathbb{R}^n$ και $\hat{c} \in \mathbb{R}^m$ το άνω μέρος του c .
- Επιλύουμε το $\hat{R}x = \hat{c}$ μέσω back substitution

3. Singular value decomposition (SVD)

Ας πάρουμε ένα πίνακα $A \in \mathbb{R}^{n \times m}$ τον οποίο και μπορούμε να κατανοήσουμε ως ένα γραμμικό μετασχηματισμό από τον $\mathbb{R}^m$ στον $\mathbb{R}^n$.

Ας θυμηθούμε τις έννοιες του πυρήνα (kernel) και του range του A ,

$$\text{Ker}(A) = \{z \in \mathbb{R}^m \mid Az = 0\},$$

$$\text{Ran}(A) = \{y \in \mathbb{R}^n \mid \exists z \in \mathbb{R}^m \text{ με } Az = y\},$$

$\text{rank}(A)$ = ο αριθμός των γραμμικά ανεξάρτητων στηλών του A .

Ισχύουν τα ακόλουθα

$$\text{rank}(A) = \text{rank}(A^T),$$

$$m = \dim(\text{Ker}(A)) + \text{rank}(A).$$

Επειδή

$$\begin{aligned} \text{Ker}(A) &= \text{Ker}(A^T A), \\ \text{Ker}(A^T) &= \text{Ker}(A A^T), \end{aligned}$$

έχουμε ότι

$$\text{rank}(A) = \text{rank}(A^T) = \text{rank}(\underbrace{A^T A}_{m \times m}) = \text{rank}(\underbrace{A A^T}_{n \times n}).$$

Οι πίνακες $A^T A$ και $A A^T$ είναι τετραγωνικοί, συμμετρικοί και θετικά ορισμένοι.

- Το πρόβλημα ιδιοτιμών

$$A^T A v = \lambda v,$$

έχει m ιδιοτιμες

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0,$$

με αντίστοιχα ιδιοδιανύσματα $v_i \in \mathbb{R}^m$, που ικανοποιούν $v_j^T v_i = \delta_{ij}$.

Απο τις ιδιοτιμες αυτές οι πρώτες $r = \text{rank}(A) = \text{rank}(A^T A)$ είναι θετικές και οι υπόλοιπες μηδενικές.

- Το πρόβλημα ιδιοτιμών

$$A A^T v = \lambda v,$$

έχει n ιδιοτιμες

$$\lambda'_1 \geq \lambda'_2 \geq \dots \geq \lambda'_n \geq 0,$$

με αντίστοιχα ιδιοδιανύσματα $v'_i \in \mathbb{R}^n$, που ικανοποιούν $v'_j{}^T v'_i = \delta_{ij}$.

Απο τις ιδιοτιμες αυτές οι πρώτες $r = \text{rank}(A) = \text{rank}(A A^T)$ είναι θετικές και οι υπόλοιπες μηδενικές.

Έχει ενδιαφέρον να δούμε πως συνδέονται μεταξύ τους τα (λ_i, v_i) με τα (λ'_j, v'_j) .

Επίσης θα δούμε πως μπορούμε να χρησιμοποιήσουμε τα (λ_i, v_i) για να παράγουμε μια διάσπαση του A .

Ορίζουμε τα

$$\sigma_j = \sqrt{\lambda_j}, \quad 1 \leq j \leq m,$$

$$u_j = \frac{1}{\sigma_j} A v_j, \quad 1 \leq j \leq r = \text{rank}(A) = \text{rank}(A^T A)$$

Μπορούμε να δούμε ότι

$$A A^T u_j = \frac{1}{\sigma_j} A \underbrace{A^T A v_j}_{=\lambda_j v_j} = \frac{\lambda_j}{\sigma_j} A v_j = \lambda_j u_j, \quad 1 \leq j \leq r,$$

και (με παρομοιο τρόπο)

$$u_i^T u_j = \delta_{ij}, \quad 1 \leq i, j \leq r.$$

Συνοψίζουμε,

Έστω $A \in \mathbb{R}^{n \times m}$ και έστω $r = \text{rank}(A) = \text{rank}(A^T A)$.

Το πρόβλημα ιδιοτιμών

$$A^T A v = \lambda v$$

έχει ιδιοτιμες $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_m = 0$ με αντιστοιχα ορθοκανονικά ιδιοδιανύσματα $v_j^T v_i = \delta_{ij}$ ενώ τα

$$\sigma_j = \sqrt{\lambda_j}, \quad 1 \leq j \leq m,$$

$$u_j = \frac{1}{\sigma_j} A v_j, \quad 1 \leq j \leq r = \text{rank}(A) = \text{rank}(A^T A),$$

είναι λύσεις του προβλήματος ιδιοτιμών

$$A A^T u = \sigma^2 u$$

με τις ίδιες ιδιότητες.

Συμπληρώνουμε το $[u_1, \dots, u_r]$ σε μία ορθοκανονική βάση του $\mathbb{R}^n$, $[u_1, \dots, u_r, u_{r+1}, \dots, u_n]$ που αποτελείται απο ιδιοδιανύσματα του AA^T έτσι ώστε

$$\begin{aligned} Av_j &= \sigma_j u_j, \quad 1 \leq j \leq r, \\ Av_j &= 0, \quad r+1 \leq j \leq m. \end{aligned}$$

Χρησιμοποιώντας τους πίνακες

$$\begin{aligned} V &= [v_1, \dots, v_m] \in \mathbb{R}^{n \times n}, \\ U &= [u_1, \dots, u_n] \in \mathbb{R}^{m \times m}, \\ S &= \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) = (\sigma_j \delta_{ij}) \in \mathbb{R}^{n \times m}, \end{aligned}$$

η παραπάνω σχέση γράφεται

$$AV = US$$

ή χρησιμοποιώντας την ορθογωνιότητα

$$A = USV^T$$

Καταλήξαμε λοιπόν στο ακόλουθο θεώρημα

Θεώρημα (SVD)

Για κάθε $A \in \mathbb{R}^{n \times m}$ υπάρχει η διάσπαση

$$A = USV^T$$

όπου $U \in \mathbb{R}^{n \times n}$, $V \in \mathbb{R}^{m \times m}$ ορθογώνιοι πίνακες, ενώ ο πίνακας $S = (\sigma_j \delta_{ij}) \in \mathbb{R}^{n \times m}$ είναι διαγώνιος.

Οι πίνακες U , V δεν είναι απαραίτητα μοναδικοί, ενώ τα σ_j ονομάζονται **ιδιάζουσες τιμές (singular values)** του A (αυτές είναι!).

Ισοδύναμα μπορούμε να γράψουμε

$$A = \sum_{i=1}^m \sigma_i u_i v_i^T = \sum_{i=1}^{r=\text{rank}(A)} \sigma_i u_i v_i^T,$$

οπότε ορίζοντας

$$\begin{aligned}\hat{U} &= [u_1, \dots, u_r] \in \mathbb{R}^{n \times r}, \\ \hat{V} &= [v_1, \dots, v_r] \in \mathbb{R}^{m \times r}, \\ \hat{S} &= \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r},\end{aligned}$$

παιρνουμε την ανηγμένη (reduced) SVD

$$A = \hat{U} \hat{S} \hat{V}^T.$$

Ο αριθμητικός υπολογισμός της SVD προϋποθέτει την δυνατότητα επίλυσης προβλημάτων ιδιοτιμών (λίγο υπομονή!).

4. Ο ψευδοαντίστροφος (pseudo-inverse)

Ο ψευδοαντίστροφος είναι μια γενίκευση της έννοιας του αντιστρόφου για πίνακες που δεν είναι τετραγωνικοί, $A \in \mathbb{R}^{n \times m}$.

Μπορεί να οριστεί μέσω της διάσπασης SVD του πίνακα $A = USV^T$.

Ορισμός (Ψευδοαντίστροφος Moore- Penrose)

Αν $A = USV^T$ ο ψευδοαντίστροφος A^+ ορίζεται ως

$$A^+ = VS^+U^T$$

όπου

$$S^+ = \begin{bmatrix} \frac{1}{\sigma_1} & & & 0 & \dots & 0 \\ & \ddots & & & \vdots & \dots \\ & & \frac{1}{\sigma_r} & 0 & & 0 \\ \vdots & & \vdots & & \vdots & \vdots \\ 0 & \ddots & & & \vdots & \vdots \end{bmatrix} \in \mathbb{R}^{m \times n}$$



H. E. Moore



Σχήμα: H. E. Moore and R. Penrose

Θεώρημα (Ιδιότητες του ψευδοαντίστροφου)

Ισχύουν τα ακόλουθα:

1

$$AA^+ = (AA^+)^T, \quad A^+A = (A^+A)^T, \quad AA^+A = A, \quad A^+AA^+ = A^+$$

και απο την παραπάνω συνθήκη προκύπτει η μοναδικότητα του A^+ .

2 Αν $A \in \mathbb{R}^{n \times n}$ αντιστρέψιμος τότε $A^+ = A^{-1}$.

3 Αν $A \in \mathbb{R}^{n \times m}$ και $\text{rank}(A) = m$ τότε

$$A^+ = (A^T A)^{-1} A^T.$$

4

$P_1 := AA^+ : \mathbb{R}^n \rightarrow \mathbb{R}^n$, είναι η ορθογώνια προβολή στο $\text{Ran}(A)$,
 $P_2 := A^+A : \mathbb{R}^m \rightarrow \mathbb{R}^m$, είναι η ορθογώνια προβολή στο $\text{Ran}(A^T)$,

Η τελευταία ιδιότητα μας επιτρέπει την χρήση του ψευδοαντίστροφου στην επίλυση γραμμικών εξισώσεων όταν αυτές δεν έχουν μοναδική λύση.

Ο ψευδοαντίστροφος αποτελεί ένα κριτήριο επιλογής μιας λύσης απο την απειρία των πιθανών λύσεων, χρησιμοποιώντας ως κριτήριο επιλογής την ελαχιστοποίηση της Ευκλείδειας νορμας της λύσης.

Θεώρημα (Ψευδοαντίστροφος και λύση γραμμικών εξισώσεων)

Ας θεωρήσουμε το σύστημα $Ax = b$.

- 1 Το σύστημα έχει λύση αν και μονο αν $AA^+b = b$.
- 2 Αν υπάρχουν λύσεις, η γενική λύση είναι της μορφής

$$x = A^+b + (I - A^+A)w$$

όπου w είναι ένας αυθαίρετος πίνακας παραμέτρων, και η επιλογή $w = 0$ αντιστοιχεί πάντοτε στην ειδική λύση $x = A^+b$.

- 3 Η λύση $x^* = A^+b$ είναι η λύση ελάχιστης νόρμας δηλ. η λύση του προβλήματος ελαχιστοποίησης

$$\|x^*\| = \min_{x \in \mathbb{R}^m : Ax=b} \|x\|$$

Ας θεωρήσουμε το σύστημα $Ax = b$ στην περίπτωση που δεν υπάρχουν λύσεις.

Μπορούμε να ζητήσουμε προεγγιστικές λύσεις ψάχνοντας το $\hat{x} \in \mathbb{R}^m$ για το οποίο

$$\|A\hat{x} - b\| = \min_{x \in \mathbb{R}^m} \|Ax - b\|.$$

Η λύση αυτή ονομάζεται **λύση ελαχίστων τετραγώνων** του συστήματος και δεν είναι απαραίτητο να είναι μοναδική (εκτός και αν $\text{rank}(A) = m$).

Θεώρημα (Ψευδοαντίστροφος και λύση ελαχίστων τετραγώνων)

- Η λύση ελαχίστων τετραγώνων η οποία ελαχιστοποιεί και την νόρμα είναι η $\hat{x}^* = A^+ b$.

Με άλλα λόγια το A^+b είναι η λύση ελαχίστου νόρμας του $Ax = b$ είτε το σύστημα έχει λύση είτε όχι!

- Αν έχει λύση τότε $A(A^+b) = b$ και είναι η λύση ελάχιστης νόρμας.
- Αν δεν έχει λύση τότε A^+b είναι η λύση ελαχίστων τετραγώνων, ελάχιστης νόρμας.

Επίλυση με την χρήση *SVD*

Μία εναλλακτική μέθοδος επίλυσης είναι με την χρήση της μεθόδου *SVD* δηλ. με την διάσπαση

$$A = \hat{U} \hat{S} \hat{V}^T$$

όπου $\hat{S}$ $m \times m$ διαγώνιος, $\hat{U}$, $n \times m$, $\hat{V}^T$ $m \times m$ unitary.

Θα δούμε συντομα πως μπορεί να γίνει αυτή η διάσπαση.

Στην περίπτωση αυτή

$$A^T A = \hat{V} \hat{S} \hat{U}^T \hat{U} \hat{S} \hat{V}^T = \hat{V} \hat{S}^2 \hat{V}^T$$

η κανονική εξίσωση γίνεται

$$\hat{V} \hat{S}^2 \hat{V}^T x = \hat{V} \hat{S} \hat{U}^T b,$$

και πολλαπλασιάζοντας με τον $\hat{V}^T$,

$$\hat{S} \hat{V}^T x = \hat{U}^T b$$

Η τελευταία εξίσωση είναι ένα πολύ ευκολο σύστημα εφόσον ο $\hat{\Sigma}$ είναι διαγώνιος και ο V^T unitary.

Η εξίσωση αυτή μπορεί να παραχθεί και παρακάμπτοντας την κανονική εξίσωση χρησιμοποιώντας την έννοια της προβολής.

1 Υπολογίζουμε την ανηγμένη *SVD*

$$A = \hat{U} \hat{\Sigma} \hat{V}^T$$

2 Υπολογίζουμε το διάνυσμα $\hat{U}^T b$

3 Λύνουμε το διαγώνιο σύστημα $\hat{S} y = \hat{U}^T b$

4 $x = \hat{V} y$

$$A = USV^T,$$

$$U \in \mathbb{R}^{n \times n}, V \in \mathbb{R}^{m \times m} \text{ ορθογώνιοι}$$

$$S \in \mathbb{R}^{n \times m} \text{ διαγώνιος που περιέχει τις ιδιάζουσες τιμές του } A.$$

Ο A δεν είναι απαραίτητα πλήρους τάξης (full rank) $r = \text{rank}(A)$.

$$\begin{aligned} \|b - Ax\|^2 &= \langle b - Ax, b - Ax \rangle = \langle b - USV^T, b - USV^T \rangle = \\ &= \langle \underbrace{UU^T}_I b - USV^T x, \underbrace{UU^T}_I b - USV^T x \rangle = \langle U^T UU^T b - U^T USV^T x, U^T b - SV^T x \rangle \\ &= \langle \underbrace{U^T b}_c - S \underbrace{V^T x}_y, \underbrace{U^T b}_c - S \underbrace{V^T x}_y \rangle = \langle c - Sy, c - Sy \rangle \end{aligned}$$

Ο πίνακας S περιέχει και μηδενικά στοιχεία, $\sigma_{r+1} = \dots = \sigma_n = 0$.

$$\begin{aligned}\|b - Ax\|^2 &= \|c - Sy\|^2 = \sum_{j=1}^r (c_j - \sigma_j y_j)^2 + \sum_{j=r+1}^n c_j^2 \\ &= \sum_{j=1}^r |\sigma_j v_j^T x - u_j^T b|^2 + \sum_{j=r+1}^n |u_j^T b|^2\end{aligned}$$

Επιλέγουμε το x έτσι ώστε

$$y_j = v_j^T x = \frac{1}{\sigma_j} u_j^T b, \quad 1 \leq j \leq r,$$

ενώ όλες οι υπόλοιπες μεταβλητές y_j , $1 \leq j \leq m$ είναι ελεύθερες.

Κατά συνέπεια η γενική λύση του προβλήματος είναι η

$$x = \sum_{j=1}^r \frac{u_j^T b}{\sigma_j} v_j + \sum_{j=r+1}^m a_j v_j = x^+ + \sum_{j=r+1}^m a_j v_j$$

με

$$x^+ := \sum_{j=1}^r \frac{u_j^T b}{\sigma_j} v_j,$$

την λύση ελαχίστου νόρμας εφόσον $\|x\|^2 = \|x^+\|^2 + \sum_{j=r+1}^m a_j^2 \geq \|x^+\|^2$.

Η λύση αυτή συνδέεται με τον ψευδοαντίστροφο του $A = USV^T$, τον $A^+ = VS^+U^T$.

Συνεπώς

$$A^+b = VS^+U^Tb = VS^+c = x^+.$$

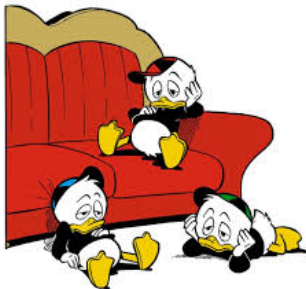
Το πρόβλημα είναι ότι δεν ξέρουμε πως να λύσουμε το βήμα 1



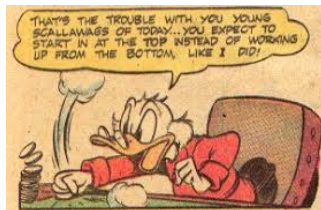
Το πρόβλημα είναι ότι δεν ξέρουμε πως να λύσουμε το βήμα 1



Ας καταφύγουμε προσωρινά στις έτοιμες συναρτήσεις του octave



Αυτό όμως θα είναι μια προσωρινή λύση γιατί η φιλοσοφία μας είναι διαφορετική!



Όμως πολύ σύντομα θα δούμε τον τρόπο!



Ομαλοποίηση Tikhonov

Πολλές φορές για λόγους αριθμητικής ακρίβειας (θα γίνουμε σαφέστεροι πολύ σύντομα!) αντί για το πρόβλημα ελαχίστων τετραγώνων

$$\min \|Ax - b\|$$

προτιμούμε να λύσουμε το **κανονικοποιημένο** πρόβλημα

$$\min_{x \in \mathbb{R}^m} \|Ax - b\|^2 + \gamma \|x\|^2, \quad \gamma > 0,$$

ή ακόμα και το γενικευμένο πρόβλημα

$$\min_{x \in \mathbb{R}^m} \|Ax - b\|^2 + \gamma \|Lx\|^2, \quad \gamma > 0,$$

με L δεδομένο πίνακα.

Ο όρος $\gamma \|x\|^2$ (ή $\gamma \|Lx\|^2$ ονομάζεται **όρος ποινικοποίησης** και η παραπάνω διαδικασία ομαλοποίηση Tikhonov (Tikhonov regularization)

Το ομαλοποιημένο πρόβλημα

$$\min_{x \in \mathbb{R}^m} \|Ax - b\|^2 + \gamma \|x\|^2, \quad \gamma > 0,$$

μπορεί να γραφεί στην μορφή ενός τυπικού προβλήματος ελαχίστων τετραγώνων παρατηρώντας ότι

$$\|Ax - b\|^2 + \gamma \|x\|^2 = \left\| \begin{pmatrix} A \\ \sqrt{\gamma} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|^2$$

οπότε ορίζοντας τα

$$A_\gamma := \begin{pmatrix} A \\ \sqrt{\gamma} I \end{pmatrix} \in \mathbb{R}^{(n+m) \times m}, \quad \bar{b} := \begin{pmatrix} b \\ 0 \end{pmatrix} \in \mathbb{R}^{n+m},$$

μπορούμε να γράψουμε το ομαλοποιημένο πρόβλημα Tikhonov σαν το τυπικό πρόβλημα ελαχίστων τετραγώνων

$$\min_{x \in \mathbb{R}^m} \|A_\gamma x - \bar{b}\|^2 = \min_{x \in \mathbb{R}^m} \|A_\gamma x - \bar{b}\|^2, \quad \gamma > 0,$$

Η διαφορά του ομαλοποιημένου προβλήματος με το αρχικό πρόβλημα είναι ότι ο πίνακας A_γ τώρα έχει $\text{rank}(A_\gamma) = m$ ακόμα και αν $\text{rank}(A) = r < m$!

Η λύση του ομαλοποιημένου προβλήματος μπορεί να γίνει επίσης με την χρήση της SVD και μπορούμε να δείξουμε ότι αν $\text{rank}(A) = r$ και $A = \hat{U}\hat{S}\hat{V}^T$ είναι η ανηγμένη SVD του πίνακα A τότε το ομαλοποιημένο κατά Tikhonov πρόβλημα έχει λύση

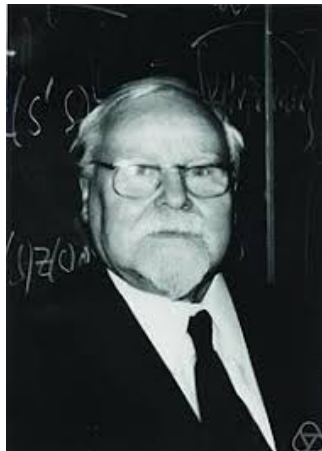
$$x_\gamma^+ = \sum_{j=1}^r \frac{\sigma_j u_j^T b}{\sigma_j^2 + \gamma} v_j,$$

το οποίο στο όριο $\gamma \rightarrow 0$ τείνει στο $x^+ = A^+ b$.

Το σφάλμα των 2 λύσεων είναι

$$\|x_\gamma^+ - x^+\| \leq \frac{\gamma}{\sigma_r(\sigma_r^2 + \gamma)} \|b\|$$

Το ομαλοποιημένο πρόβλημα είναι πολύ πιο ευстаθές από το αρχικό (θα επανέλθουμε σε αυτό συντομα)



Η ομαλοποίηση Tychonov χρησιμοποιείται ευρύτατα σε εφαρμογές όπως η λύση αντιστρόφων προβλημάτων, στην ιατρική απεικόνιση, την ανάλυση εικόνας κλπ.

Προσομοίωση κανονικών κατανομών

Η πολυμεταβλητή κανονική κατανομή καθορίζεται πλήρως από το διάνυσμα των μέσων τιμών και τον πίνακα συνδιακύμανσης.

Θα λέμε ότι η $X = (X_1, \dots, X_d)^T$ ακολουθεί την πολυμεταβλητή κανονική κατανομή με μέσο το διάνυσμα μ και πίνακα συνδιακύμανσης S , αν ακολουθεί την από κοινού πυκνότητα πιθανότητας

$$f(x) = \frac{1}{(2\pi)^{d/2} |\det(S)|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu)^T S^{-1} (x - \mu) \right), \quad x \in \mathbb{R}^d.$$

Τα σημεία του $\mathbb{R}^d$ τα οποία είναι ισοπίθανα είναι αυτά που ικανοποιούν την εξίσωση $\frac{1}{2} (x - \mu)^T S^{-1} (x - \mu) = C$ δηλαδή βρίσκονται επάνω σε ελλειψοειδή.

Οι συντελεστές του πίνακα $S = (\sigma_{ij})$ αντιστοιχούν στις συνδιακυμάνσεις

$$\sigma_{ij} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])].$$

Είναι επίσης χρήσιμο να δώσουμε την μορφή της χαρακτηριστικής συνάρτησης για την X , η οποία είναι

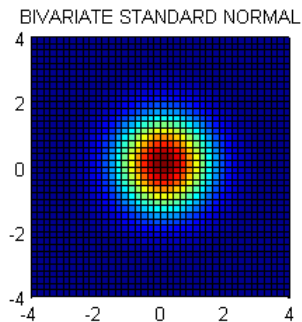
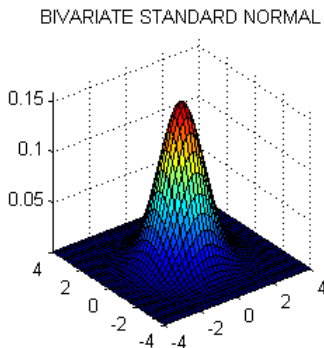
$$\phi_X(u) = \mathbb{E}[e^{iu \cdot X}] = \exp(i\mu^T u - \frac{1}{2}u^T S u), \quad u \in \mathbb{R}^d.$$

Στην ειδική περίπτωση όπου $\mu = 0$ και $S = I$ λέμε ότι η X ακολουθεί την τυπική κανονική κατανομή, και στην περίπτωση αυτή οι διαφορετικές συνιστώσες του διανύσματος X_i και X_j είναι ανεξαρτητες τυχαίες μεταβλητές.

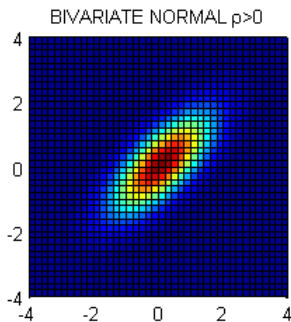
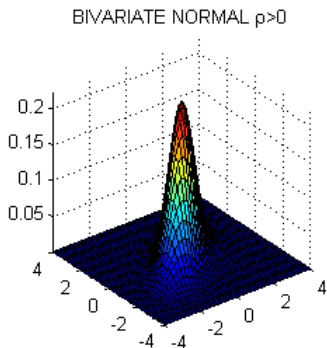
Προσοχή, το ότι οι X_i και X_j είναι ασυσχέτιστες είναι ισοδύναμο με την ανεξάρτησή τους είναι ένα φαινόμενο το οποίο ισχύει ειδικά για την κανονική κατανομή!

Αν υποθέσουμε ότι η $Z = (Z_1, \dots, Z_d)$ ακολουθεί την τυπική κανονική κατανομή, τότε η $X = \mu + AZ$ όπου $\mu \in \mathbb{R}^{d \times 1}$, $A \in \mathbb{R}^{d \times d}$ ακολουθεί την κανονική κατανομή με μέσο μ και πίνακα συνδιακύμανσης $S = A^T A$.

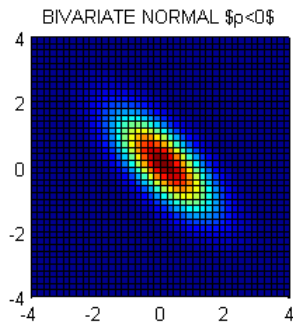
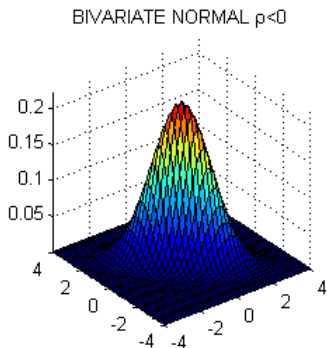
Αυτό μπορεί να αποδειχθεί πολύ εύκολα κάνοντας χρήση της χαρακτηριστικής συνάρτησης.



Σχήμα: Κανονική κατανομή $\rho = 0$



Σχήμα: Κανονική κατανομή $\rho > 0$



Σχήμα: Κανονική κατανομή $\rho < 0$

Προσομοίωση της μονοδιάστατης κανονικής κατανομής

Αλγόριθμος (Box-Muller)

- Παράγουμε $U_1, U_2 \sim \mathcal{U}(0, 1)$, *i.i.d.*
- Υπολογίζουμε τα

$$\begin{aligned}X_1 &= \sqrt{-2 \ln(U_1)} \cos(2\pi U_2), \\X_2 &= \sqrt{-2 \ln(U_1)} \sin(2\pi U_2).\end{aligned}$$

- $X_1, X_2 \sim N(0, 1)$, *i.i.d*

Αλγόριθμος (Polar Marsaglia)

- Παράγουμε $U_1, U_2 \sim \mathcal{U}(0, 1)$, *i.i.d.*
- Θέτουμε $V_i = 2U_i - 1$, $i = 1, 2$ και $S = V_1^2 + V_2^2$.
- Αν $S \leq 1$, παράγουμε $U \sim \mathcal{U}(0, 1)$ και υπολογίζουμε τα

$$X_1 = V_1 \sqrt{-2 \ln(U/S)},$$

$$X_2 = V_2 \sqrt{-2 \ln(U/S)}.$$

αλλιώς ξαναγυρνάμε στο βήμα 1.

- $X_1, X_2 \sim N(0, 1)$, *i.i.d*

Προσομοίωση της πολυδιάστατης κανονικής

Ας υποθέσουμε ότι θέλουμε να προσομοιώσουμε σενάρια απο την πολυδιάστατη κανονική κατανομή $N_d(\mu, S)$.

Αλγόριθμος

- 1 Υπολογίζουμε την διάσπαση Cholesky του πίνακα συνδιακύμανσης

$$S = B B^T.$$

- 2 Παράγουμε την $Z_1, \dots, Z_d \sim N(0, 1)$, *i.i.d.* και θέτουμε $Z = (Z_1, \dots, Z_d)^T$.
- 3 Παράγουμε το $X = \mu + BZ$.

Χρησιμοποιώντας την χαρακτηριστική συνάρτηση μπορούμε να μελετήσουμε τους γραμμικούς μετασχηματισμούς τυχαίων μεταβλητών που ακολουθούν την πολυμεταβλητή κανονική κατανομή.

Αν $X \sim N_d(\mu, S)$ με $X \in \mathbb{R}^{d \times 1}$ και $A \in \mathbb{R}^{k \times d}$, $a \in \mathbb{R}^{k \times 1}$ έχουμε ότι $AX + a \sim N_k(A\mu + a, ASA^T)$.

Σαν ειδική μορφή αυτής της σχέσης στην περίπτωση όπου $A \in \mathbb{R}^{d \times 1}$ παίρνουμε την κατανομή των γραμμικών συνδυασμών των συνιστωσών μιας πολυδιάστατης κανονικής κατανομής, και συγκεκριμένα αν $A = \lambda = (\lambda_1, \dots, \lambda_d)$, τότε

$$\sum_{i=1}^d \lambda_i X_i \sim N\left(\sum_{i=1}^d \lambda_i \mu_i, \sum_{i=1}^d \sum_{j=1}^d \lambda_i \lambda_j \sigma_{ij}\right).$$

Μάλιστα η X ακολουθεί την πολυμεταβλητή κανονική κατανομή αν και μόνον αν για κάθε διάνυσμα $\lambda \in \mathbb{R}^{d \times 1}$ ισχύει ότι η τυχαία μεταβλητή $\lambda^T X$ ακολουθεί την μονοδιάστατη κανονική κατανομή.

Η ιδιότητα όλοι οι γραμμικοί συνδυασμοί μιας πολυμεταβλητής κανονικής κατανομής να συνεχιζουν να ακολουθούν την κανονική κατανομή μεταφέρεται και στις περιθώριες και τις δεσμευμένες κατανομες της.

Συγκεκριμένα, αν $X \sim N_d(\mu, S)$ και σπάσουμε την X ως $X = (X_1^T, X_2^T)^T$, με $X_1 \in \mathbb{R}^{k \times 1}$ και $X_2 \in \mathbb{R}^{(d-k) \times 1}$, σπάζοντας επίσης και τα $\mu = (\mu_1^T, \mu_2^T)^T$ και

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

τότε οι περιθώριες κατανομές των X_1 και X_2 είναι επίσης κανονικές κατανομές με $X_1 \sim N_k(\mu_1, S_{11})$ και $X_2 \sim N_{d-k}(\mu_2, S_{22})$.

Ομοίως και οι δεσμευμένες κατανομές $P(X_2 \mid X_1 = x_1)$ ακολουθούν την κανονική και συγκεκριμένα

$$X_2 \mid x_1 \sim N_{d-k}(\mu_2 + S_{21}S_{11}^{-1}(x_1 - \mu_1), S_{22} - S_{21}S_{11}^{-1}S_{12}).$$

Παράδειγμα

Έχουμε 3 τίτλους που οι αποδόσεις τους $X = (X_1, X_2, X_3)$ ακολουθούν την κανονική κατανομή με

$$\mu = (0.3, 0.2, 0.1)^T$$

και πίνακα διασποράς

$$S = \begin{pmatrix} 4 & 0.1 & -0.2 \\ 0.1 & 2 & -0.1 \\ -0.2 & -0.1 & 1 \end{pmatrix}$$

Ποια είναι η περιθώρια κατανομή της (X_1, X_2) και της X_3 ;

Χωρίζουμε τον μέσο και την διασπορά ως

$$\mu = ((0.3, 0.2)^T, 0.1)^T,$$

και

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix}$$

με

$$S_{11} = \begin{pmatrix} 4 & 0.1 \\ 0.1 & 2 \end{pmatrix}, \quad S_{12} = \begin{pmatrix} -0.2 \\ -0.1 \end{pmatrix}, \quad S_{22} = (1).$$

Τότε οι περιθώριες δίνονται απο

$$(X_1, X_2)^T \sim N_2 \left(\begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix}, \begin{pmatrix} 4 & 0.1 \\ 0.1 & 2 \end{pmatrix} \right), \quad X_3 \sim N_1(0.1, 1).$$

Παράδειγμα

Ο Δρ. Χ, διαχειριστής κινδύνου της εταιρείας Α, έχει στην διάθεση του τους 3 τίτλους του προηγούμενου παραδείγματος και σκέπτεται να τοποθετηθεί στον τίτλο 3.

Απο παρατηρήσεις της αγοράς γνωρίζει ότι οι αποδόσεις των τίτλων 1 και 2 ήταν 0.4 και 0.8 αντίστοιχα.

Ποια θα είναι η καλύτερη πρόβλεψη του βάσει της πληροφορίας αυτής για την απόδοση του τίτλου 3;

Θα πρέπει να χρησιμοποιήσει την δεσμευμένη κατανομή $R_3 \mid R_1 = 0.4, R_2 = 0.8$.

Θέτοντας $X_1 = (R_1, R_2)^T$ και $X_2 = R_3$, και επειδή η $X = (R_1, R_2, R_3) = (X_1, X_2)$ ακολουθεί την κανονική κατανομή, ομοίως και οι δεσμευμένες κατανομές $P(X_2 \mid X_1 = x_1)$ ακολουθούν την κανονική και συγκεκριμένα

$$X_2 \mid x_1 \sim N_{d-k}(\mu_2 + S_{21}S_{11}^{-1}(x_1 - \mu_1), S_{22} - S_{21}S_{11}^{-1}S_{12}).$$

με $\mu_1 = (0.3, 0.2)^T$, $\mu_2 = 0.1$, $q_1 = (0.4, 0.8)^T$, και

$$S_{11} = \begin{pmatrix} 4 & 0.1 \\ 0.1 & 2 \end{pmatrix}, \quad S_{12} = \begin{pmatrix} -0.2 \\ -0.1 \end{pmatrix}, \quad S_{22} = (1).$$

Μπορούμε να υπολογίσουμε ότι

$$\begin{aligned}\mu_2 + S_{21}S_{11}^{-1}(x_1 - \mu_1) &= 0.066, \\ S_{22} - S_{21}S_{11}^{-1}S_{12} &= 0.9855,\end{aligned}$$

συνεπώς η καλύτερη πρόβλεψη για την απόδοση θα είναι 0.066 και η αξιοπιστία της πρόβλεψης αυτής θα καθορίζεται από την διασπορά 0.9855.

Αναφορές

R. Piziak, R. L. Odell, Matrix theory from generalized inverses to Jordan form, Chapman and Hall, 2007

L. N. Trefethen, and D. Bau III, Numerical Linear Algebra, SIAM 1997

W. Ford, Numerical Linear Algebra with Applications, using Matlab, Academic Press, 2015

H. Wendland, Numerical Linear Algebra, An introduction, CUP, 2017