Λογισμος II, $\mathbb{R}^n$

Α. Ν. Γιαννακόπουλος
Τμήμα Στατιστικής

Ο.Π.Α

Εαρινό Εξάμηνο 2018

Το $\mathbb{R}^n$.

$$\mathbb{R}^n = \{x = (x_1, \cdots, x_n) \ : \ x_i \in \mathbb{R}, \ i = 1, \cdots, n\} = \mathbb{R} \times \cdots \times \mathbb{R}$$

Το σύνολο των αντικειμένων που χρειάζονται n πραγματικούς αριθμούς για να περιγραφούν πλήρως.

Παράδειγμα

Η αναλυτική βαθμολογία ενός αποφοιτου του τμήματος στατιστικής του ΟΠΑ ο οποίος για να πάρει πτυχίο πρέπει να περάσει 30 μαθήματα είναι ένα στοιχείο του $\mathbb{R}^{30}$.

Παράδειγμα

Ένα σημείο της αίθουσας που για να περιγραφεί χρειάζεται μήκος, πλάτος και ύψος είναι ένα στοιχείο του $\mathbb{R}^3$.

Απόσταση δύο στοιχείων του $\mathbb{R}^n$

Έστω $x = (x_1, \cdots, x_n)$ και $y = (y_1, \cdots, y_n)$ δυο στοιχεια του $\mathbb{R}^n$.

Για να ταυτίζονται αυτά θα πρέπει

$$x_i = y_i, \;\; \forall \;\; i = 1, \cdots, n.$$

Στην περιπτωση αυτή γράφουμε

$$x = y \;\; \text{στον} \;\; \mathbb{R}^n$$

Αν υπάρχει τουλάχιστον ένα $i = 1, \cdots, n$ τέτοιο ώστε $x_i \neq y_i$ τότε

$$x \neq y \quad \text{στον} \quad \mathbb{R}^n.$$

Για δύο οποιαδηποτε $x, y \in \mathbb{R}^n$ ένας τροπος (όχι ο μοναδικός) να ορίσουμε την απόσταση τους ειναι ως

$$d(x,y) = \|x-y\| = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

Αν $n=1$ τότε $d(x,y) = |x-y|$.

Η απόσταση αυτή ονομάζεται ευκλείδια απόσταση στον $\mathbb{R}^n$.

Η ποσότητα

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2},$$

ονομάζεται μήκος του διανύσματος x.

Ιδιότητες της $d(\cdot, \cdot)$

1. $d(x, y) \geq 0, \ \forall \, x, y \in \mathbb{R}^n$ και $d(x, y) = 0$ αν και μόνο αν $x = y$.
2. $d(x, y) = d(y, x), \ \forall \, x, y \in \mathbb{R}^n$.
3. $d(x, y) \leq d(x, z) + d(z, y), \ \forall \, x, y, z \in \mathbb{R}^n$. (τριγωνική ανισότητα)

Απόσταση και εσωτερικό γινόμενο

Για οποιαδήποτε δύο διανύσματα $x = (x_1, \cdots, x_n), y = (y_1, \cdots, y_n)$ στον $\mathbb{R}^n$ μπορούμε να ορίσουμε το εσωτερικό γινόμενο

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i .$$

Μπορούμε ευκολα να δούμε οτι

$$d(x, y) = \sqrt{\langle x - y, x - y \rangle}.$$

Ιδιότητες του εσωτερικού γινομένου

1. $\langle x, y \rangle = \langle y, x \rangle$, $\forall\ x, y \in \mathbb{R}^n$.
2. $\langle x, x \rangle \geq 0$, $\forall\ x \in \mathbb{R}^n$ και $\langle x, x \rangle = 0$ αν και μόνο αν $x = 0$.
3. Για οποιαδηποτε $x, y, z \in \mathbb{R}^n$ και $\lambda_1, \lambda_2 \in \mathbb{R}$ ισχύει ότι

$$\langle \lambda_1 x \lambda_2 y, z \rangle = \lambda_1 \langle x, z \rangle + \lambda_2 \langle y, z \rangle,$$
$$\langle x, \lambda_1 y + \lambda_2 z \rangle = \lambda_1 \langle x, y \rangle + \lambda_2 \langle x, z \rangle$$

Η ανισότητα Cauchy-Schwarz

Θεώρημα

Για κάθε $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$ *στο* $\mathbb{R}^n$ *ισχύει οτι*

$$|\langle x, y\rangle| \leq \|x\| \, \|y\|,$$

ή ισοδύναμα

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leq \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \left(\sum_{i=1}^{n} y_i^2\right)^{1/2}.$$

Η ισότητα ισχύει αν και μόνο αν υπάρχει $\lambda \in \mathbb{R}$ *τέτοιο ώστε* $y = \lambda x$ *δηλαδή,*

$$y_i = \lambda x_i, \quad i = 1, \cdots, n.$$

Η ανισότητα Cauchy-Schwarz έχει πολλές και χρησιμες εφαρμογές

Παράδειγμα (Απόδειξη της τριγωνικής ιδιότητας)

Αρκει να δειξουμε ότι $\|x+y\| \leq \|x\| + \|y\|$ *για κάθε* $x, y \in \mathbb{R}^n$.

$$\|x+y\| \leq \|x\| + \|y\| \iff \|x+y\|^2 \leq (\|x\| + \|y\|)^2$$

Το αριστερό μέλος δίνει

$$\|x+y\|^2 = \langle(x+y, x+y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2.$$

Το δεξιό μελος δίνει

$$(\|x\| + \|y\|)^2 = \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2.$$

Συγκρίνοντας και λαμβάνοντας υπόψιν την ανισότητα Cauchy-Schwarz καταλήγουμε στο επιθυμητό αποτέλεσμα.

Ο νόμος του παραλληλογράμμου

Για κάθε $x, y \in \mathbb{R}^n$ ισχυει

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Ορθογώνια διάσπαση

Έστω $x, y \in \mathbb{R}^n$ με $x \neq 0$.
Ορίζουμε

$$c = \frac{\langle x, y\rangle}{\|y\|^2},$$

$$z = x - \frac{\langle x, y\rangle}{\|y\|^2} y = x - c\,y.$$

Ισχύει ότι

$$\langle z, y\rangle = 0,$$

$$x = c\,y + z.$$

Εφαρμογές της ανισοτητας Cauchy-Schwarz στην βελτιστοποίηση και την στατιστικη

Παράδειγμα

Θεωρείστε τους αριθμούς a_i, $i = 1, \cdots, n$ *γνωστούς.*
Ποια ειναι η ελάχιστη τιμή της παράστασης $I = \sum_{i=1}^{n} a_i x_i$ *για όλες τις πιθανες επιλογές των αριθμών* x_i, $i = 1, \cdots, n$ *αν θέλουμε να ικανοποιείται ο περιορισμός* $\sum_{i=1}^{n} x_i^2 = 1$ *και για ποιά* x_i *επιτυγχάνεται;*

Η ανισότητα Cauchy -Schwarz δινει

$$|\sum_{i=1}^{n} a_i x_i| \leq (\sum_{i=1}^{n} a_i^2)^{1/2} (\sum_{i=1}^{n} x_i^2)^{1/2} = (\sum_{i=1}^{n} a_i^2)^{1/2}$$

δηλαδη,

$$-(\sum_{i=1}^{n} a_i^2)^{1/2} \leq \sum_{i=1}^{n} a_i x_i \leq (\sum_{i=1}^{n} a_i^2)^{1/2}.$$

Η μικροτερη τιμή που μπορει να πάρει η παράσταση I είναι ίση με $-(\sum_{i=1}^{n} a_i^2)^{1/2}$.

Επίσης απο την ανισότητα Cauchy-Schwarz έχουμε ότι για να επιτευχθεί η ισότητα θα πρέπει να υπάρχει $\lambda \in \mathbb{R}$ τέτοιο ωστε

$$x_i = \lambda\, a_i, \quad i = 1, \cdots, n.$$

Αυτό μας δίνει ότι

$$\sum_{i=1}^{n} x_i^2 = \lambda^2 \sum_{i=1}^{n} a_i^2,$$

συνεπώς λόγω του περιορισμού $\sum_{i=1}^{n} x_i^2 = 1$ καταλήγουμε ότι

$$\lambda = \pm \frac{1}{\sum_{i=1}^{n} a_i^2}.$$

Άρα, η ελάχιστη τιμή της παράστασης I επιτυγχάνεται στο

$$x_i = \frac{a_i}{\sum_{i=1}^{n} a_i^2}, \quad i = 1, \cdots, n.$$

Παράδειγμα

Έστω $\Omega = \{\omega_1, \cdots, \omega_n\}$ *ένας διακριτός δειγματικός χώρος και* $p = (p_1, \cdots, p_n)$ *μια διακριτή κατανομή στον* Ω.
Έστω $X, Y : \Omega \to \mathbb{R}$ *δύο διακριτές τυχαιες μεταβλητές τις οποίες μπορούμε να κατανοήσουμε σαν διανύσματα*

$$X = (x_1, \cdots, x_n),$$
$$Y = (y_1, \cdots, y_n).$$

Δειξτε ότι η ανισότητα Cauchy-Schwarz σημαινει ότι

$$|\mathbb{E}[(X - \mathbb{E}[X])\,[(Y - \mathbb{E}[Y])\,]| \leq (\mathbb{E}[(X - \mathbb{E}[X])^2])^{1/2}(\mathbb{E}[(Y - \mathbb{E}[Y])^2])^{1/2}$$

ή ισοδύναμα

$$|Cov(X, Y)| \leq V\sqrt{Var(X)}\,\sqrt{Var(Y)}.$$

Βέλτιστοι γραμμικοί εκτιμητές (BLUE)

Σύμφωνα με την θεωρία μας, η ποσότητα Z θα πρέπει να είναι ίση προς μια σταθερά c που θέλουμε να εκτιμήσουμε.
Δυστυχώς στην διάθεση μας έχουμε μόνο J παρατηρήσεις της Z, Z_i οι οποίες όμως υπόκεινται σε τυχαία σφάλματα ϵ_i τα οποία θεωρούμε ότι ικανοποιούν τις ιδιότητες

$$\mathbb{E}[\epsilon_i] = 0, \quad i = 1, \cdots, J,$$
$$\mathbb{E}[\epsilon_i \epsilon_j] = \sigma_i^2 \, \delta_{i,j}, \quad \sigma_i^2 > 0, \quad i, j = 1, \cdots, J,$$

Πως θα μπορούσατε να χρησιμοποιήσετε τις μετρήσεις Z_i, $i = 1, \cdots, J$ για να εκτιμήσετε την σταθερά c;

Οι μετρήσεις μας θα ερχονται απο ένα υπόδειγμα της μορφής

$$Z_i = c + \epsilon_i, \quad i = 1, \cdots, J,$$

όπου το c είναι άγνωστο.
Θα εκτιμήσουμε το c απο μια εκτιμήτρια της μορφής

$$\hat{c} = \sum_{i=1}^{N} b_i Z_i,$$

για κατάλληλη επιλογή των σταθερών b_i.
Η εκτιμήτρια αυτή είναι τυχαια γιατί εξαρτάται απο τις παρατηρήσεις.
Όπως και αν έχουν τύχει τα σφάλματα θα πρέπει να ικανοποιεί

$$\mathbb{E}[\hat{c}] = c \quad \text{αμεροληψία}$$

καθώς και να ελαχιστοποίεί την ποσότητα

$$\mathbb{E}[(c - \hat{c})^2]$$

Η συνθήκη της αμεροληψίας δίνει

$$\sum_{i=1}^{J} b_i = 1.$$

Μπορούμε επίσης να δειξουμε ότι

$$\mathbb{E}[(c-\hat{c})^2] = \mathbb{E}[\sum_{i=1}^{J}\sum_{j=1}^{J} b_i b_j \epsilon_i \epsilon_j] = \sum_{i=1}^{J} b_i^2 \sigma_i^2.$$

Το πρόβλημα της εύρεσης της BLUE ανάγεται στο πρόβλημα βελτιστοποίησης

$$\min_{b_1,\cdots,b_J} \sum_{i=1}^{J} \sigma_i^2 \, b_i^2,$$

υπο τον περιορισμό

$$\sum_{i=1}^{J} b_i = 1.$$

Το πρόβλημα αυτό μπορεί να λυθεί με την χρηση της ανισότητας Cauchy-Schwarz και να δείξουμε ότι

$$\hat{c} = \lambda \sum_{i=1}^{J} \frac{Z_i}{\sigma_i^2},$$

$$\lambda = \frac{1}{\sum_{i=1}^{J} \frac{1}{\sigma_i^2}}.$$

Μπορούμε να γράψουμε

$$1 = \sum_{i=1}^{J} b_i = \sum_{i=1}^{J} (b_i \sigma_i) \frac{1}{\sigma_i},$$

και εφαρμόζοντας την ανισότητα Cauchy-Schwarz ,

$$1 = \sum_{i=1}^{I} (b_i \sigma_i) \frac{1}{\sigma_i} \le \left(\sum_{i=1}^{J} b_i^2 \sigma_i^2 \right)^{1/2} \left(\sum_{i=1}^{J} \frac{1}{\sigma_i^2} \right)^{1/2},$$

Άρα η μικρότερη τιμή που μπορεί να πάρει η ποσότητα $I = \sum_{i=1}^{J} b_i^2 \sigma_i^2$ ειναι ίση προς $\frac{1}{\sum_{i=1}^{J} \frac{1}{\sigma_i^2}}$ και η ισότητα επιτυγχάνεται αν υπάρχει $\lambda \in \mathbb{R}$ τέτοιο ώστε

$$b_i \sigma_i = \lambda \frac{1}{\sigma_i}, \quad i = 1, \cdots, J.$$

Καθετοτητα διανυσματων στον $\mathbb{R}^n$

Ορισμός

Δύο διανύσματα $x, y \in \mathbb{R}^n$ *θα λέγονται κάθετα* $(x \perp y)$ *αν*

$$\langle x, y \rangle = 0.$$

Θεώρημα

Αν $x \perp y$ *τότε*

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Καλύτερη προσέγγιση και καθετότητα

Θεώρημα

Ας υποθέσουμε ότι εχουμε m γραμμικά ανεξαρτητα διανύσματα $z_1, \cdots, z_m \in \mathbb{R}^n$, $m \leq n$ και ένα διάνυσμα $x \in \mathbb{R}^n$ και ζητούμε να βρούμε την καλύτερη προσέγγιση για το x της μορφής

$$x^* = \sum_{i=1}^{m} b_i z_i, \quad b_i \in \mathbb{R}.$$

Το στοχείο x^* καθορίζεται απο τις σχέσεις ορθογωνιότητας

$$\langle x - x^*, z_i \rangle = 0, \quad i = 1, \cdots, m.$$

Ψάχνουμε $x^* = \sum_{i=1}^m b_i z_i$ τέτοιο ώστε

$$\|x - x^*\| \leq \|x - y\|, \ \ \forall\, y \in V := span(z_1, \cdots, z_m).$$

Έχουμε ότι

$$\begin{aligned} \|x-y\|^2 &= \|x - x^* + x^* - y\|^2 \\ &= \|x - x^*\|^2 + 2\langle x - x^*, x^* - y\rangle + \|x^* - y\|^2. \end{aligned}$$

Εφοσον, $x^*, y \in V$ ισχύει και $x^* - y \in V$ αρα $x^* - y = \sum_{i=1}^m c_i z_i$ για κάποια $c_i \in \mathbb{R}$ συνεπώς,

$$\begin{aligned} \langle x - x^*, x^* - y\rangle &= \langle x - x^*, \sum_{i=1}^m c_i z_i\rangle \\ &= \sum_{i=1}^m c_i \langle x - x^*, z_i\rangle = 0, \end{aligned}$$

Άρα,

$$\|x - y\|^2 = \|x - x^*\|^2 + \|x^* - y\|^2 \geq \|x - x^*\|^2.$$

Η ανισότητα Hölder

Θεώρημα

Αν ορίσουμε την ποσοτητα $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, *για* $p \geq 1$ *τότε*

$$|\sum_{i=1}^n x_i y_i| \leq \|x\|_p \, \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1,$$

με την ισότητα να ισχύει αν και μόνο αν

$$\left(\frac{x_i}{\|x\|_p}\right)^p = \left(\frac{y_i}{\|y\|_q}\right)^q, \quad i = 1, \cdots, n.$$

Η ανισότητα Minkowski

Η ανισότητα του Hölder μας επιτρέπει μια γενίκευση της τριγωνικής ανισότητας.

Θεώρημα

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p, \ \ \forall\, x, y \in \mathbb{R}^n, \ \ p \geq 1.$$

Απόδειξη.

Για κάθε $i = 1, \cdots, n$,

$$|x_i + y_i|^p \leq |x_i|\,|x_i + y_i|^{p-1} + |y_i|\,|x_i + y_i|^{p-1}.$$

Αθροίζουμε σε ολα τα i,

$$\sum_{i=1}^{n} |x_i + y_i|^p \leq \sum_{i=1}^{n} |x_i|\,|x_i + y_i|^{p-1} + \sum_{i=1}^{n} |y_i|\,|x_i + y_i|^{p-1},$$

και εφαρμόζουμε 2 φορές την Hölder. □

Γενικεύσεις της εννοιας της απόστασης στον $\mathbb{R}^n$.

Στο $\mathbb{R}^n$ μπορούμε να γενικεύσουμε την έννοια της απόστασης

$$d(x,y) = \left(\sum_{i=1}^{n} w_i |x_i - y_i|^p \right)^{1/p},$$

για κάποια επιλογή βαρών $w_1, \cdots, w_n$.

Η επιλογή $p = 2$ είναι πολύ σημαντική γιατι τότε η απόσταση σχετίζεται με ενα εσωτερικό γινόμενο και ικανοποιεί ιδιότητες οπως π.χ. ο νόμος του παραλληλογράμμου και το Πυθαγόρειο θεώρημα.

Η περίπτωση του $\mathbb{R}^3$

Στην περίπτωση αυτή μποοούμε να γράψουμε

$$\langle x, y \rangle = \|x\| \, \|y\| \cos(\theta),$$

όπου θ η γωνία μεταξύ των διανυσμάτων x και y.

Το εξωτερικό γινόμενο στο $\mathbb{R}^3$

Αν έχουμε δύο διανύσματα $x = x_1 e_1 + x_2 e_2 + x_3 e_3$ και $y = y_1 e_1 + y_2 e_2 + y_3 e_3$ στον $\mathbb{R}^3$, μπορούμε να ορίσουμε ένα τριτο διάνυσμα ως

$$\begin{aligned} x \times y &= \begin{vmatrix} e_1 & e_2 & e_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \\ &= (x_2 y_3 - x_3 y_2) e_1 + (x_3 y_1 - x_1 y_3) e_2 + (x_1 y_2 - x_2 y_3) e_3. \end{aligned}$$

Ισχυει ότι

$$x \times y = -y \times x,$$
$$x \times x = 0$$