ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΕΚΠΑ

Ανάλυση 2: Εργασία, Νοέμβρης 2023

Να απαντήσετε στα παρακάτω ερωτήματα (τα ερωτήματα A6 και A7 είναι προαιρετικά)

Μέθοδος Ελαχίστων Τετραγώνων: *Έστω σύνολο πειραματικών δεδομένων* (x_i, y_i), $i = 1, 2, \ldots, n$, $x_i \neq x_j$ *αν* $i \neq j$. *Θέλουμε να βρούμε τις τιμές των παραμέτρων* m *και* b, *έτσι ώστε η ευθεία* $y = mx + b$ *να ελαχιστοποιεί το άθροισμα τετραγώνων των αποκλίσεων* e_i:

$$f(m, b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2 := \sum_{i=1}^{n} e_i^2$$

A1: Δείξτε ότι αν $n = 2$, δηλαδή έχουμε μόνο ένα ζεύγος σημείων από πειραματικά δεδομένα (x_1, y_1) και (x_2, y_2), τότε η συγκεκριμένη μέθοδος παράγει την ευθεία που διέρχεται από τα (x_1, y_1) και (x_2, y_2).

A2: Δείξτε ότι οι εξισώσεις $\partial f/\partial b = 0$ και $\partial f/\partial m = 0$ που δίνουν το (μοναδικό) κρίσιμο σημείο είναι ισοδύναμες με τις:

$$m\left(\sum x_i\right) + nb = \left(\sum y_i\right)$$

και

$$m\left(\sum x_i^2\right) + b\left(\sum x_i\right) = \left(\sum x_i y_i\right)$$

όπου όλα τα αθροίσματα είναι από $i = 1$ έως $i = n$.

A3: Αν $y = \hat{m}x + \hat{b}$ είναι η ευθεία με την καλύτερη προσαρμογή στα πειραματικά δεδομένα σύμφωνα με την μέθοδο ελαχίστων τετραγώνων, δείξτε ότι

$$\sum_{i=1}^{n} (y_i - \hat{m}x_i - \hat{b}) = 0$$

δηλαδή οι θετικές και αρνητικές αποκλίσεις αλληλοαναιρούνται.

A4: Χρησιμοποιώντας το κριτήριο δεύτερης παραγώγου δείξτε ότι το κρίσιμο σημείο της f είναι σημείο ελαχίστου. *Υπόδειξη:* Μπορείτε να χρησιμοποιήσετε μαθηματική επαγωγή η την ανισότητα Cauchy-Schwartz.

Γενικεύουμε την μέθοδο ελαχίστων τετραγώνων για το γραμμικό μοντέλο $y = X\theta$, *όπου* $y \in \mathbb{R}^n$ *και* $X \in \mathbb{R}^{n \times q}$. *Η μέθοδος παρέχει εκτίμηση του διανύσματος παραμέτρων* $\theta \in \mathbb{R}^q$ *ελαχιστοποιώντας την συνάρτηση:*

$$f(\theta) = \|y - X\theta\|^2 = (y - X\theta)^T (y - X\theta)$$

A5: Δείξτε ότι αν οι στήλες του πίνακα X είναι γραμμικά ανεξάρτητες, τότε $\det(X^T X) \neq 0$ και $\hat{\theta} = (X^T X)^{-1} X^T y$ είναι η (μοναδική) βέλτιστη λύση του προβλήματος ελαχιστοποίησης. Δείξτε επίσης ότι $X\hat{\theta} = \Pi y$, όπου $\Pi = X(X^T X)^{-1} X^T$ είναι ο τελεστής ορθογώνιας προβολής του διανύσματος y στον υπόχωρο $\mathcal{R}(X) := \{X\theta : \theta \in \mathbb{R}^q\}$ του $\mathbb{R}^n$.

A6* (Αναδρομικός αλγόριθμος ελαχίστων τετραγώνων): Εξετάζουμε μία παραλλαγή της μεθόδου για το γραμμικό μοντέλο $y = X\theta$. Η μέθοδος τροποποιεί την τρέχουσα εκτίμηση του διανύσματος θ, $\hat{\theta}_{n-1}$

(διαθέσιμη σε χρόνο $t = n - 1$) καθώς η πληροφορία του μοντέλου ανανεώνεται (σε χρόνο $t = n$). Γράφουμε τις εξισώσεις του μοντέλου (σε χρόνο $t = n$) ως:

$$Y_n = X_n \theta + E_n \ \text{ όπου } \ Y_n = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \ X_n = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}, \ E_n = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Η εκτίμηση ελαχίστων τετραγώνων την χρονική στιγμή $t = n$ είναι: $\hat{\theta}_n = (X_n^T X_n)^{-1} X_n^T Y_n$. Ορίζουμε: $P_n := (X_n^T X_n)^{-1}$. Δείξτε ότι

$$\hat{\theta}_n = \hat{\theta}_{n-1} + P_n x_n (y_n - x_n^T \hat{\theta}_{n-1})$$

όπου

$$P_n = P_{n-1} - \frac{P_{n-1} x_n x_n^T P_{n-1}}{1 + x_n^T P_{n-1} x_n}$$

Υπόδειξη: Δείξτε ότι γιά πίνακες A, B και C με συμβατές διαστάσεις, ισχύει ότι:

$$(A + BC)^{-1} = A^{-1} - A^{-1} B (I + C A^{-1} B)^{-1} C A^{-1}$$

υποθέτοντας ότι οι πίνακες A και $A + BC$ είναι μη ιδιάζοντες.

Στατιστικές ιδιότητες εκτίμησης με την μέθοδο ελαχίστων τετραγώνων: Εδώ υποθέτουμε ότι το γραμμικό μοντέλο είναι πάλι της μορφής $y = X\theta + e$, αλλά ότι τα στοιχεία του διανύσματος e είναι τυχαίες μεταβλητές που ικανοποιούν: (i) $\mathbb{E}(e) = 0$ και (ii) $\mathbb{E}(ee^T) = \sigma^2 I_n$, όπου $\mathbb{E}(\cdot)$ είναι ο τελεστής μέσης τιμής (μαθηματικής προσδοκίας), δηλαδή ότι οι τυχαίες μεταβλητές e_i έχουν μηδενική μέση τιμή, είναι στατιστικά ασυσχέτιστες και έχουν κοινή διασπορά σ^2. Η εκίμηση ελαχίστων τετραγώνων $\hat{\theta} = (X^T X)^{-1} X^T y$ είναι επομένως διάνυσμα τυχαίων μεταβλητών.

A7* Δείξτε ότι: (i) $\hat{\theta}$ είναι αμερόληπτη εκτιμήτρια του θ, δηλαδή ότι $\mathbb{E}(\hat{\theta}) = \theta$. (ii) $\mathrm{Cov}(\hat{\theta}) = \sigma^2 (X^T X)^{-1}$, όπου $\mathrm{Cov}(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)(\hat{\theta} - \theta)^T)$ είναι ο πίνακας συνδιακύμανσης της εκτίμησης θ. (iii) Αν β είναι αυθαίρετη γραμμική αμερόληπτη εκτιμήτρια του θ της μορφής $\beta = By$, τότε $\mathrm{Cov}(\hat{\theta}) \preceq \mathrm{Cov}(\beta)$ (δηλ. $\mathrm{Cov}(\beta) - \mathrm{Cov}(\hat{\theta}) \succeq 0$).