

ΠΕΡΙΛΗΨΗ

Ευαγγελία Μητσάκη

Ανάλυση συσχετισμένων δεδομένων με ridge παλινδρόμηση

Η πολυσυγγραμμικότητα είναι σύνηθες φαινόμενο στην ανάλυση παλινδρόμησης κατά το οποίο δύο ή περισσότερες ερμηνευτικές μεταβλητές παρουσιάζουν υψηλή συσχέτιση (αλλά όχι πλήρη) μεταξύ τους, κάνοντας έτσι δύσκολη την ερμηνεία της κάθε μεταβλητής. Σε αυτή την εργασία ασχολούμαστε με τη θεωρία της πολυσυγγραμμικότητας καθώς και τους τρόπους που έχουν προταθεί για να αναγνωρίζουμε την παρουσία της. Προκειμένου να χειριστούμε τα πολυσυγγραμμικά δεδομένα παρουσιάζουμε διάφορες μεθόδους όπως τη μέθοδο των κυρίων συνιστωσών (principal components), μεθόδους επιλογής μεταβλητών (variable selection) και τη μέθοδο της μεροληπτικής εκτίμησης (biased estimation).

Ωστόσο αυτή η εργασία επικεντρώνεται στη ridge παλινδρόμηση. Προτεινόμενη αρχικά από τους Hoerl and Kennard, η ridge παλινδρόμηση έχει αποδειχθεί χρήσιμη μέθοδος αντιμετώπισης του προβλήματος της πολυσυγγραμμικότητας στα πλαίσια του γραμμικού μοντέλου παλινδρόμησης. Παρουσιάζουμε τους ridge εκτιμητές και τις ιδιότητές τους καθώς και τρόπους επιλογής της ridge σταθεράς. Θα αναφερθούμε επίσης σε διαφορετικές προσεγγίσεις της ridge παλινδρόμησης καθώς και εφαρμογές της σε περιπτώσεις άλλες εκτός της γραμμικής παλινδρόμησης. Ακόμα, παρουσιάζουμε τη χρήση της influence analysis σε συσχετισμένες μεταβλητές. Δίνουμε παραδείγματα με πραγματικά δεδομένα όπου κρίνεται αναγκαίο. Τέλος, θέλοντας να δείξουμε την εφαρμογή της ridge παλινδρόμησης σε δεδομένα και να συγκρίνουμε τα χαρακτηριστικά των ridge εκτιμητών χρησιμοποιούμε προσομοιώσεις (Monte Carlo simulations).