## Μη παραμετρική παλινδρόμηση ΙΙ: splines και penalized παλινδρόμηση.

(Από Wassermann ΚΑΙ σημειώσεις του P. Breheny, University of Kentucky

Βασική ιδέα: Για  **εκτίμησε την  ως εκείνη τη συνάρτηση που ελαχιστοποιεί την**



Το** θα παίξει εδώ το ρόλο της παραμέτρου εξομάλυνσης**:

* : με νοιάζει μόνο το fit, και άρα  θα είναι μια συνάρτηση που προσαρμόζει τέλεια στα δεδομένα 
*  ενδιάμεσο: προτιμώ  με «μικρή» διακυμαντικότητα και ας προσαρμόζει κάπως «χειρότερα» στα δεδομένα
* : προτιμώ  χωρίς καμία καμπυλότητα ευθεία γραμμή και άρα κάνω γραμμική παλινδρόμηση.

Λύση τα splines. Διακρίνονται σε δύο είδη

* Regression splines: πιο κοντά σε *παραμετρική* παλινδρόμηση
* Smoothing splines: «*πραγματική*» *μη παραμετρική* παλινδρόμηση

### Regression splines

Ιδέα: γράψε την  σα γραμμικό συνδυασμό μιας κατάλληλης «βάσης» συναρτήσεων 

Π.χ.

*  τα **πολυώνυμα** οστού βαθμού.
* Ή  ***σταθερές* συναρτήσεις σε *προεπιλεγμένα διαστήματα*** , δηλαδή  για . Άρα 
* Ή  ***γραμμικές* συναρτήσεις σε *προεπιλεγμένα διαστήματα***  , δηλαδή  και  για . Άρα 
* Ή  ***γραμμικές* συναρτήσεις σε *προεπιλεγμένα διαστήματα*** , αλλά **με περιορισμούς που να εξασφαλίζουν συνέχεια**:

Πάρε π.χ.  ,  ,  

Πόσες συναρτήσεις χρειάστηκα (βαθμοί ελευθερίας)

* ( διαστήματα επί 2 παράμετροι αν διάστημα - περιορισμοί 
* Ή ισοδύναμα 







* ***Spline***: ***πολυώνυμα  βαθμού το καθένα*** σε  διαστήματα, που έχουν τις ***πρώτες  παραγώγους συνεχείς στα όρια των διαστημάτων*** (κόμβοι).
* Π.χ. για *γραμμικά* splines (προηγούμενο Bullet) πάρε .
* Για *κυβικά* splines (*cubic splines , το στάνταρ*…) πάρε : Πολυώνυμα 3 βαθμού σε κάθε διάστημα , με συνεχείς τις δύο πρώτες παραγώγους (κλίση καμπυλότητα) , στους κόμβους.

Γενικό : βάση: , και τέλος  , δηλαδή  συναρτήσεις

* Λίγα διαστήματα  ή και χαμηλός βαθμός **υψηλότερο Bias** εκτιμήτριας
* Πολλά διαστήματα / υψηλός βαθμός **υψηλή διακύμανση** εκτιμήτριας

Slides 15&16 Breheny

Quadratic splines



cubic splines

Συνήθως παίρνουμε : ***κυβικά (cubic) splines*** 

**Πρόβλημα**: υψηλή αστάθεια στα άκρα των  (boundaries)

**Βελτίωση**: αν απαιτήσουμε πρόσθετα ότι εκτός boundaries  η συνάρτηση θα συμπεριφέρεται γραμμικά: ***natural cubic spline*** (4 ακόμα περιορισμοί )

Slides 16,21,22 Breheny

Natural cubic splines



Natural cubic splines, 6df



Slides 23 ,24 Breheny





Πιο βολική βάση για τα natural cubic splines τα ***B-splines*** (βλ. Wassermann)

### Εκτίμηση στα regression splines

Θέτω  ο πίνακας σχεδιασμού και

  ο πίνακας προβολής .

Άρα οι εκτιμήτριες ελαχίστων τετραγώνων δίνονται από:

: 

Παίρνουμε:

*  ,όπου 
*  , δηλαδή  με 
* 

Slide 27 Βreheny



**Πρόβλημα στα regression splines**:

* Εξαρτώνται τα πάντα από την (αυθαίρετη ) ***επιλογή των κόμβων***
* ***Όχι πραγματικά μη –παραμετρικό***: έχουμε περιορισμένο αριθμό παραμέτρων

### Smoothing-splines

Επανερχόμαστε στην αρχική διατύπωση: βρες ** που ελαχιστοποιεί την**

  (\*)

To  είναι «τιμωρία» (penalty) για ανώμαλη (μη ομαλή) συμπεριφορά (roughness penalty) κανονικοποίηση (regularization)

**Θεώρημα:** Ανάμεσα τις συναρτήσεις που είναι δύο φορές παραγωγίσιμες, η \* ελαχιστοποιείται από ένα ***natural cubic spline*** με κόμβους στα 

Έστω  μία βάση για τα natural cubic splines και άρα με 

 ο  πίνακας .

Τότε

και  με 

Και εκτιμάμε τα  **ελαχιστοποιώντας την**

 , με ,

καθώς

*  και
* 

**Θεώρημα.** Η λύση στο παραπάνω πρόβλημα δίνεται από



**Απόδειξη**.



=

=

Παραγωγίζοντας ως προς  και θέτοντας ίσο με 0 παίρνουμε:





Δηλαδή  .

Από το παραπάνω παίρνουμε:



με . Βλέπουμε λοιπόν ότι η εκτίμηση με splines είναι ένας γραμμικός εξομαλυντής (linear smoother) με πίνακα εξομάλυνσης smoothing matrix τον .

Ισχύουν λοιπόν τα συνήθη, δηλαδή με 

* , δηλαδή 
* , δηλαδή  , όπου  η ι-οστή γραμμή του 
* 
* , με  και 

Slides 38, 39, 40 Breheny

****

****



**Σημείωση.** Ο Silvermann έδειξε ότι τα splines προσεγγίζονται από NW με

 με ,

 την πυκνότητα των  και .