# ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΜΑΘΗΣΗΣ

Ακαδημαϊκό Έτος 2020-2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

***3ο Σετ Ασκήσεων***



| | | |
|---|---|---|
| Φώτιος-Παναγιώτης | Μπασαμάκης | 1046975 |

## ΠΡΟΒΛΗΜΑ 1:

Έχω μια μεταβλητή ομοιόμορφα κατανεμημένη στο διάστημα [0 , 1] καθώς και 1000 υλοποιήσεις της συγκεκριμένης μεταβλητής. Για αυτό το λόγο γνωρίζω θεωρητικά ότι η συνάρτηση πυκνότητας πιθανότητας της θα είναι η εξής:

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & 0 < x \text{ ή } x > 1 \end{cases}$$

Σκοπός μου είναι να προσεγγίσω αυτή τη συνάρτηση πυκνότητας πιθανότητας με πυρήνες kernel.

Αρχικά γνωρίζουμε ότι η συνάρτηση dirac δ(χ) έχει τις παρακάτω ιδιότητες:

- $\int \Phi(x)\delta(x)dx = \Phi(0)$
- $\int \delta(x)dx$=1
- $f(x) = \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$

Αναζητώ μια συνάρτηση η οποία προσεγγίζει την δ(χ) και έτσι ορίζω τον πυρήνα Kernel

$K(x,h) = g\left(\frac{x}{h}\right)\left(\frac{1}{h}\right)$ όπου προσεγγίζει την δ(χ) όταν το h τείνει στο 0

Λαμβάνοντας υπόψη τις παρακάτω ιδιότητες:

$$\int \Phi(x)\delta(x - x0)dx = \Phi(x0)$$

$$\frac{G(x1,y) + G(x2,y) + \cdots + G(x_N,y)}{N} = E_X[G(X,y)]$$

(Όπου $E_X[G(X,y)]$ ο μέσος όρος που προσεγγίζεται από την παραπάνω σχέση λόγω νόμου των μεγάλων αριθμών )

Η συνάρτηση πυκνότητας πιθανότητας προσεγγίζεται μέσω των πυρήνων ως εξής:

$$f'(y) = \int f'(x)\delta(x - y)dx = \int f'(x)K(x - y, h)dx = E_X[K(x - y, h)]$$

Όπου τελικά

$$f'(x) = \frac{K(x - x1, h) + K(x - x2, h) + \cdots + G(x - x_N, h)}{N}$$

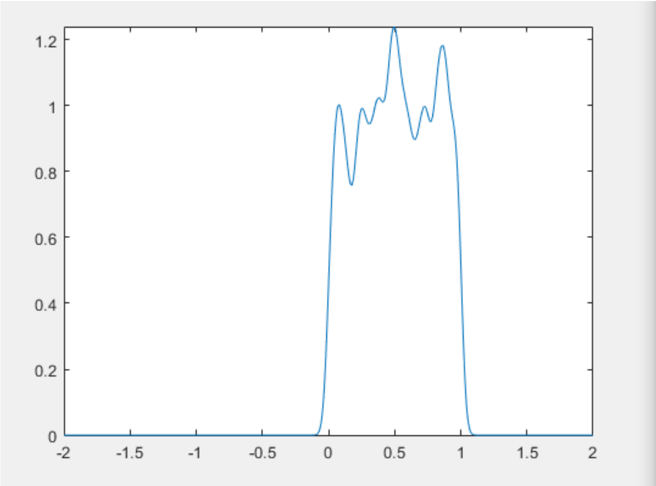
Βάση λοιπόν της παραπάνω λογικής αναπτύχθηκε κώδικας στο matlab για προσέγγιση της συνάρτησης πυκνότητας πιθανότητας χρησιμοποιώντας τον παρακάτω πυρήνα kernel

$$\mathsf{K}(x,h) = \frac{1}{\sqrt{2\pi h}} e^{-\frac{1}{2h}x^2}.$$

Τα αποτελέσματα για διάφορες τιμές του h είναι τα εξής:

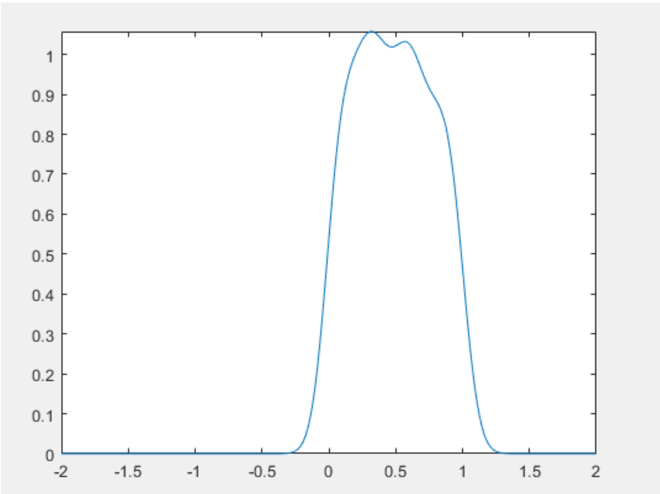
h = 0.001

gaussian



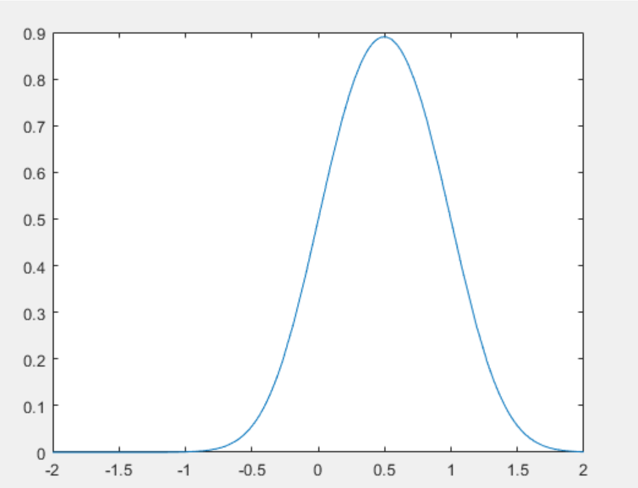

h = 0.01

gaussian



h = 0.1

gaussian



Παρατηρώ ότι όσο μικραίνει το h τόσο πιο γρήγορα ανεβαίνει η συνάρτηση στο 0 και πέφτει στο 1 αλλά ταλαντώνεται στην μέγιστη τιμή της ενώ για μεγαλύτερα h η μέγιστη τιμή είναι περισσότερο σταθερή.

**<u>ΠΡΟΒΛΗΜΑ 2:</u>**

α) Για να χρησιμοποιήσω το Representer Theorem ορίζω την γραμμική θήκη Ω (span{z1,z2,..,zn}) η οποία είναι γραμμικός υπόχωρος του V και αποτελείται από στοιχεία που είναι γραμμικοί συνδυασμοί των z1,z2,..,zn.

Έστω επίσης ότι υπάρχει x το οποίο δεν ανήκει στον Ω αλλά ανήκει στο V. Θέλω να βρω έναν αντιπρόσωπο του $x$ τον $\acute{x}$ στο Ω.

Προφανώς $\acute{x} = a_1 z_1 + \cdots + a_n z_n$ και $(x - \acute{x})$ Πρέπει να είναι κάθετο στο Ω (ώστε το $\acute{x}$ να είναι το διάνυσμα με την μικρότερη απόσταση από το x από όλα τα διανύσματα του Ω).Αυτό σημαίνει ότι το εσωτερικό γινόμενο $< x - \acute{x}, y > = 0$ Με y να είναι όλα τα διανύσματα που ανήκουν στον Ω.

Αν το παραπάνω εσωτερικό γινόμενο είναι 0 για $y = (z_1 \ldots z_n)$ τότε μπορώ να θεωρήσω το $\acute{x}$ σαν αντιπρόσωπο του $x$ στο Ω.

Αντίστοιχα με τις συναρτήσεις $\Phi(x)$ και $\Phi'(x)$ θα έχω:

$\Phi(x) = a_1 K(x, z_1) + \cdots + a_n K(x, z_n)$

$\Phi'(x) = b_1 K(x, x_1) + \cdots + b_n K(x, x_n)$ όπου χ1,…,χn οι υλοποιήσεις που έχω.

Το K(x,y) ονομάζεται πυρήνας mercer και είναι μια θετικά ορισμένη συνάρτηση.

Το εσωτερικό γινόμενο των παραπάνω 2 συναρτήσεων είναι:

$< \Phi(x) - \Phi'(x), K(x, x_\iota) > = 0$

Επίσης ισχύει ότι:

$\Phi(x) = \sum_{i=1}^{n} a_i K(x, z_i)$

$$< \Phi(x), K(x, x_0) > = \sum_{i=1}^{n} a_i K(x, x_i) K(x, x_0) = \sum_{i=1}^{n} a_i K(x_i, x_0) = \Phi(x_0)$$

Και με χρήση αυτών η σχέση του εσωτερικού γινόμενου των συναρτήσεων γίνεται

$< \Phi(x), K(x, x_\iota) > = < \Phi'(x), K(x, x_\iota) > \delta\eta\lambda.\ \Phi(x_\iota) = \Phi'(x_\iota)$

Άρα μπορώ αντί για το $\Phi(x_\iota)$ να αντικαθιστώ με τον αντιπρόσωπο του στο υπόχωρο Ω του V που περιέχει μόνο γραμμικούς συνδυασμούς.

Στην συνάρτηση $\Phi'(x)$ έχω άγνωστους μόνο τους συντελεστές $a_i$ οπότε αν παραγωγίσω ως προς αυτούς μπορώ να λύσω γραμμικό σύστημα για να τους υπολογίσω.

β) Για τον όρο $\|\Phi(x)\|^2$ ισχύει ότι:

$\|\Phi(x)\|^2 = \|\Phi(x) + \Phi'(x) - \Phi'(x)\|^2 = \|\Phi(x)\|^2 + \|\Phi(x) - \Phi'(x)\|^2$

$+2 < \Phi(x) - \Phi'(x), \Phi'(x) >$

Το παραπάνω εσωτερικό γινόμενο όμως είναι 0 (αρχή ορθογωνιότητας) διότι τα διανύσματα είναι κάθετα μεταξύ τους (αφού το $\Phi'(x)$ ανήκει στον Ω και $\Phi(x) - \Phi'(x)$ είναι κάθετο στον Ω )

Αρά προκύπτει ότι :

$\|\Phi(x)\|^2 \geq \|\Phi'(x)\|^2$ που σημαίνει ότι μπορώ να αντικαταστήσω το $\Phi(x)$ με $\Phi'(x)$ και στον τελευταίο όρο του προβλήματος βελτιστοποίησης και πλέον έχω μόνο να βρω τους συντελεστές .