Ονοματεπώνυμο: Σκαρπέτης Ιωάννης

Τμήμα: Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

Τομέας: Υπολογιστών

Έτος Σπουδών: 4ο

**Εργασία Μηχανικής Μάθησης 3**

**Πρόβλημα 1**

Καλούμαστε να αξιολογήσουμε τις δυνατότητες της πρώτης μεθόδου kernel στο να προσεγγίζει πυκνότητες πιθανότητας. Για τον σκοπό αυτό δημιουργούμε 1000 υλοποιήσεις μιας τυχαίας μεταβλητής ομοιόμορφα κατανεμημένης στο διάστημα [0, 1]. Θα προσεγγίσουμε την πυκνότητα πιθανότητας χρησιμοποιώντας το Gaussian kernel:

**A picture containing text, watch, clock, gauge

Description automatically generated**

Θα αναπαραστήσουμε την προσέγγιση αυτή για διαφορετικές τιμές του h.

**Επίλυση**

Ξεκινάμε με τη δημιουργεία τυχαίας μεταβλητής Χ = [ χ1, χ2, χ3, ... , χ1000 ], και γνωρίζουμε πως:

(1)

Εάν στη θέση του Χ βάλουμε μια ντετερμινιστική συνάρτηση G(x), παίρνουμε.

(2)

Εάν τώρα δημιουργήσουμε μια ντετερμινιστική συνάρτηση G(x,y), δηλαδή μία που να εξαρτάται και από το y παίρνουμε τα παρακάτω.

(3)

Η παραπάνω εξίσωση μας δίνει τον στοχαστικό μέσο όρο ως προς το Χ.

Γνωρίζουμε πως: (4)

Επειδή όμως έχουμε την K(x-y, h) η οποία προσεγγίζει το δ γράφεται ως εξής:

(5)

Στο παρόν πρόβλημα η τυχαία μεταβλητή Χ έχει πυκνότητα πιθανότητας f(x). Οπότε αντικαθιστούμε στην φ(χ) με την f(x).

Άρα καταλήγουμε:

(6)

Αναπτύσσοντας την σχέση (3):

(7)

Συνδυάζοντας τις (6),(7) παίρνουμε:

Λόγω ύπαρξης συμμετρίας παίρνουμε:

(8)

Στο πρόβλημα μας βασικές προυποθέσεις αποτελούν Κ(x,h) ≥ 0 , .

Εάν οι παραπάνω ισχύουν η που αντιστοιχεί στην ισότητα της εξίσωσης (8) μπορεί να λειτουργήσει ώς π.π της Τ.Μ.

Ο κώδικας του προβλήματος 1 δίνεται στο τέλος της αναφοράς (**Κώδικας 1**).

Τρέχοντας τον **κώδικα 1** παίρνουμε τα εξής αποτελέσματα.

Chart, line chart

Description automatically generated

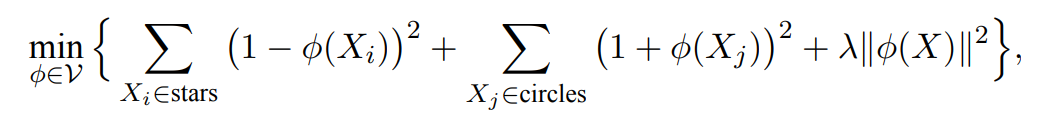
Παρατηρούμε πως για πολύ μικρό h το γράφιμα μας στο τμήμα ενδιαφέροντος [0,1] παρουσιάζει πολλές κυματώσεις. Όσο αυξάνουμε το h παρατηρούμε πως οι κυματώσεις ελαττώνονται. Ιδιαίτερη προσοχή να δώσουμε στα γραφίματα που αντιστοιχούν στο h = 0.005 και h = 0.001 τα οποία μας παρουσιάζουν την κοντινότερη συμπεριφορά στην συνάρτηση πυκνότητας πιθανότητας μίας ομοιόμορφης κατανομής. Συμπερασματικά, με χρήση αυτής της μεθόδου πυρηνών μπορούμε ικανοποιητικά να προσεγγίσουμε την συνάρτηση πυκνότητας πιθανότητας.

**Πρόβλημα 2**

Μας δίνεται το Matlab αρχείο δεδομένων data32.mat το οποίο περιέχει δύο μήτρες: stars και circles. Κάθε μια είναι μια λίστα από δι-διάστατα (2-D) διανύσματα. Κάθε 2-D διάνυσμα αντιστοιχεί σε ένα σημείο στον 2-D επίπεδο και έχει σαν ετικέτα (label) “star” ή “circle”. Μας ζητείται να δημιουργήσουμε έναν classifier ο οποίος να διακρίνει μεταξύ των 2 συνόλων, και να αναπαραστήσουμε το διαχωριστικό σύνορο.

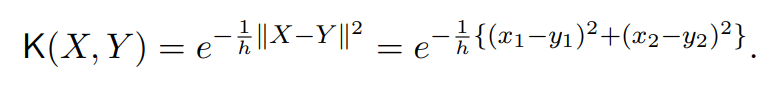
Καλούμε ϕ(X), X = [x1, x2] ⊺ το μετασχηματισμό που επιθυμούμε να εφαρμόσουμε στα δεδομένα ο οποίος θα υλοποιεί την ακόλουθει αντιστοίχηση: αντιστοιχίζουμε την αριθμητική ετικέτα “1” στο stars και την “−1” στο circles.

Σκόπος του προβλήματος είναι να προσδιορίσουμε την συνάρτηση ϕ(X), ώστε να καταφέρουμε να επιλύσουμε το εξής πρόβλημα βελτιστοποίησης.

 (1)

Όπου V είναι ο διανυσματικός χώρος που ορίζεται με τη βοήθεια του γκαουσιανού πυρήνου.

Στο παρόν πρόβλημα χρησιμοποιούμε ως gaussian kernel το εξής:



**Επίλυση**

α) Για την επίλυση του παραπάνω προβλήματος βελτιστοποίησης θα κάνουμε χρήση του representer theorem. Σύμφωνα με το τελευταίο, μας δίνεται η δυνατότητα στα πρώτα δύο αθροίσματα της (1) να αντικαταστήσουμε την φ(Χ) με:

A picture containing text, clock, watch

Description automatically generated

Όπου είναι η ορθογώνια προβολή της φ(Χ) στον γραμμικό υποχώρο που δημιουργείται από τους πυρήνες. Καλούμαστε να αποδείξουμε ότι ισχύει αυτή η δυνατότητα αντικατάστασης.

Απόδειξη

* Ξεκινάμε με Χ = χ1,χ2,χ3, ... , χΝ .
* Άρα έχουμε K(X , ), K(X , ), … , K(X , ) όπου το εκάστοτε Κ(Χ, Χi) αντιστοιχεί σε ένα Zi. (Ζ1, Ζ2, ... , ΖΝ στοιχεία ενός χώρου V)
* Ορίζουμε γραμμική θήκη Ω = -> γραμμικός υποχώρος του V.
* Αφού η συνάρτηση φ(Χ) ανοίκει στο V πρέπει να βρούμε μία συνάρτηση η οποία θα είναι ο αντιπρόσωπος της φ(Χ) στο Ω.
* Καθώς το Ω αποτελεί γραμμική θήκη το είναι υποχρεωτικά της μορφής:

= K(X , ) + K(X,)+ … + K(X , ) (2)

Σημείωση:

Το παραπάνω αποτελεί μορφή. Συνεπώς στο πρόβλημα μας, εάν οι συντελεστές αι – αN είναι και οι βι – βΝ στην δεύτερη περίπτωση.

* Τώρα από την αρχή ορθογωνιότητας παίρνουμε

1. < φ(X) - , > = 0 (3)
2. < φ(X), > = < , > (4)

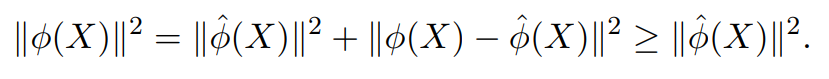
* Σύμφωνα με την ιδιότητα ότι < φ(X), > = φ() από την εξίσωση (4) παίρνουμε το εξής:

* Άρα μπορούμε να αντικαταστήσουμε την της εξίσωσης (2) στην εξίσωση (1), κάνοντας μία μικρή τροποποίηση για το πρόβλημα μας.

A picture containing text, clock, watch

Description automatically generated

β) Θα χρειαστούμε επίσης να αναλύσουμε τον όρο χρησιμοποιόντας την Αρχή της Ορθογωνιότητας. Πιο συγκεκριμένα μας ζητείται να δείξουμε το παρακάτω.



Αναλύοντας το παίρνουμε:

=  0

=

Το εσωτερικό μας γινόμενο είναι μηδέν λόγω Ορθογωνιότητας.

Άρα έχουμε το εξής:

= (5)

Είναι λοιπόν προφανές πως ισχύει η ανισότητα:

(6)

Αφού το παρόν πρόβλημα είναι ένα πρόβλημα βελτιστοποίησης και σκοπός μας είναι να ελαχιστοποιήσουμε την εξίσωση (1) μπορούμε εφόσον έχουμε αποδείξει την (6) να αντικαταστήσουμε τον όρο στην εξίσωση (1) με τον όρο , καθώς στη χειρώτερη περίπτωση αυτοί οι δύο θα είναι ίσοι και σε κάθε άλλη περίπτωση ο όρος είναι καλύτερος για μας, καθώς είναι μικρότερος.

γ) Καλούμαστε να υπολογίσουμε τους συντελεστές αι, βj καθώς αυτοί είναι πλεόν οι άγνωστοι τους οποίους πρέπει να βελτιστοποιήσουμε. Αρχικά, αρχικοποιούμε τους πίνακες με τυχαίες τιμές στο διάστημα [-1,1]. Για κάθε ζεύγος τιμών που έχουμε, το περνάμε μέσα από την και ανάλογα με το αποτέλεσμα που παίρνουμε αν είναι κοντά στο  θα προσθέτουμε στο αντίστοιχο a[star\_pair\_index] τον αριθμό 1 διαφορετικά άμα είμαστε στο τότε αφαιρούμε το 1 απο το β[circle\_pair\_index]. Δηλαδή ανάλογα το output που παίρνουμε αφαιρούμε το output που θα θέλαμε να δεχτούμε ανάλογα την προσέγγιση των υπολογισμών μας. Καταλήγουμε στους συντελεστές:

ai = [74.44467041357453, 274.0682713782122, 323.96762650133155, 2.4942558594509414, 15.964330642383604, 12.79425810375826, 1.3373818103500277, 0.20577428375132434, 3.2970513005223294, 3.096091038940009, 6.895848710080397, 1.5590512479210228, 3.13501520966261, 3.6456691825230507, -0.018307549507942333, 23.688518455099135, 0.7442513085152382, 21.108700461164666, 113.31340311202894, 22.17145215800749, 21.04747891698562]

bj = [-158.56307490589438, -450.88648535223865, -51.82632278717688, 24.546519410895534, -0.22267223877141618, -16.76227737952788, -0.5233780664767917, -1.3997563275181473, -0.9049850359874865, -2.738012148191091, -0.7230228076933511, -5.930103482090549, -3.754582015432547, -4.659611013989951, -2.4863788657589785, -11.425000739933875, -0.02181362750317084, -33.03113121516923, -22.517615633174337, -107.20238249021737, -27.590078255688837]

Σημείωση:

Οι παραπάνω συντελεστές είναι αποτέλεσμα, του αλγορύθμου για 50 εποχές εκπαίδευσης l = 0.1 και h = 0.1.

Πλέον έχουμε υπολογίσει τη βέλτιστη και τώρα καλούμαστε να κατηγοιοποιήσουμε ένα νέο δεδομένο ανάλογα με το αποτέλεσμα της όταν της το τροφοδοτούμε. Περιμένουμε το αποτέλεσμα να μην είναι ακριβώς 1 ή -1.

Ο κώδικας του προβλήματος δίνεται στο τέλος της αναφοράς (Κώδικας 2).

Δίνοντας τα δεδομένα των stars και circles να μας τα κάνει predict, παίρνουμε:

Για l = 0.1 και h = 0.1

A picture containing text, keyboard, electronics

Description automatically generatedA picture containing calendar

Description automatically generated

A picture containing text, keyboard

Description automatically generatedΓια l = 0.5 και h = 0.5

A picture containing text, keyboard

Description automatically generated

Στα αριστερά των αποτελεσμάτων είναι για κάθε ζεύγος η τιμή που δίνεται από την

μετά από 50 εποχές εκπαίδευσης, ενώ δεξία αναγράφεται για κάθε σημείο (χ,y) ποιό θα ήταν το ιδανικό του αποτέλεσμα. Παρατηρούμε πως έχουμε γενικά καλύτερα αποτελέσματα για μικρότερα l και h.

ε) Καλούμαστε τώρα να βρούμε το διαχωριστικό σύνορο που χωρίζει τα δεδομένα μας. Για τον σκοπό αυτό θα χρησιμοποιήσουμε το average των predicted αποτελεσμάτων και των εκάστοτε stars, circles. Το σύνορο που υπολογίζουμε φαίνεται παρακάτω, τόσο σε 2-D όσο και σε 3-D.

Σημείωση:

Chart, line chart

Description automatically generated Τα παρακάτω είναι για l = 0.1 και h = 0.1

Chart, line chart

Description automatically generated

**Πρόβλημα 3**