MAI4CAREU
Master programmes in Artificial Intelligence 4 Careers in Europe

Πανεπιστήμιο Κύπρου - Τεχνητή Νοημοσύνη

MAI612 - ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διάλεξη 8: Μέθοδοι με βάση τον πυρήνα 2

Βασίλης Βασιλειάδης, PhD
Χειμερινό Εξάμηνο 2022/23

CYENS
CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάλεξη 8: Μέθοδοι με βάση τον πυρήνα 2

Μαθησιακά αποτελέσματα

Θα καταλάβετε:

1. Σχετικά με την ακριβή παρεμβολή έναντι της προσέγγισης
2. Τι είναι οι Radial Basis Functions (RBF) και πώς μπορούν να χρησιμοποιηθούν σε δίκτυα RBF
3. Πώς να εκπαιδεύσετε ένα δίκτυο RBF για ακριβή παρεμβολή
4. Πώς να εκπαιδεύσετε ένα δίκτυο RBF για την προσέγγιση
5. Πώς να βελτιστοποιήσετε αυτόματα τις παραμέτρους του RBF χρησιμοποιώντας gradient descent
6. Κανονικοποιημένα δίκτυα RBF
7. Δίκτυα RBF για ταξινόμηση
8. Το πρόβλημα του XOR
9. Τα βασικά στοιχεία για το μοντέλο Gaussian Process

Radial Basis Function Networks

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μέθοδοι πυρήνων για παλινδρόμηση

- Η τοποθέτηση ενός πυρήνα σε κάθε σημείο εκπαίδευσης επιτρέπει την **ακριβή παρεμβολή**
  - το μοντέλο δίνει **ακριβώς** τα σωστά αποτελέσματα για όλα τα δεδομένα κατάρτισης

- Τι θα γίνει αν έχουμε ένα **τεράστιο** σετ εκπαίδευσης;
- Τι γίνεται αν τα σημεία δεδομένων είναι **θορυβώδη**;

- **Κατά προσέγγιση**:
  - έχουμε λιγότερους πυρήνες από τα σημεία δεδομένων εκπαίδευσης
  - το μοντέλο συσχετίζει καλά τα διανύσματα κατάρτισης με τα επιθυμητά αποτελέσματα, αλλά δεν παράγει τις ακριβώς σωστές εξόδους

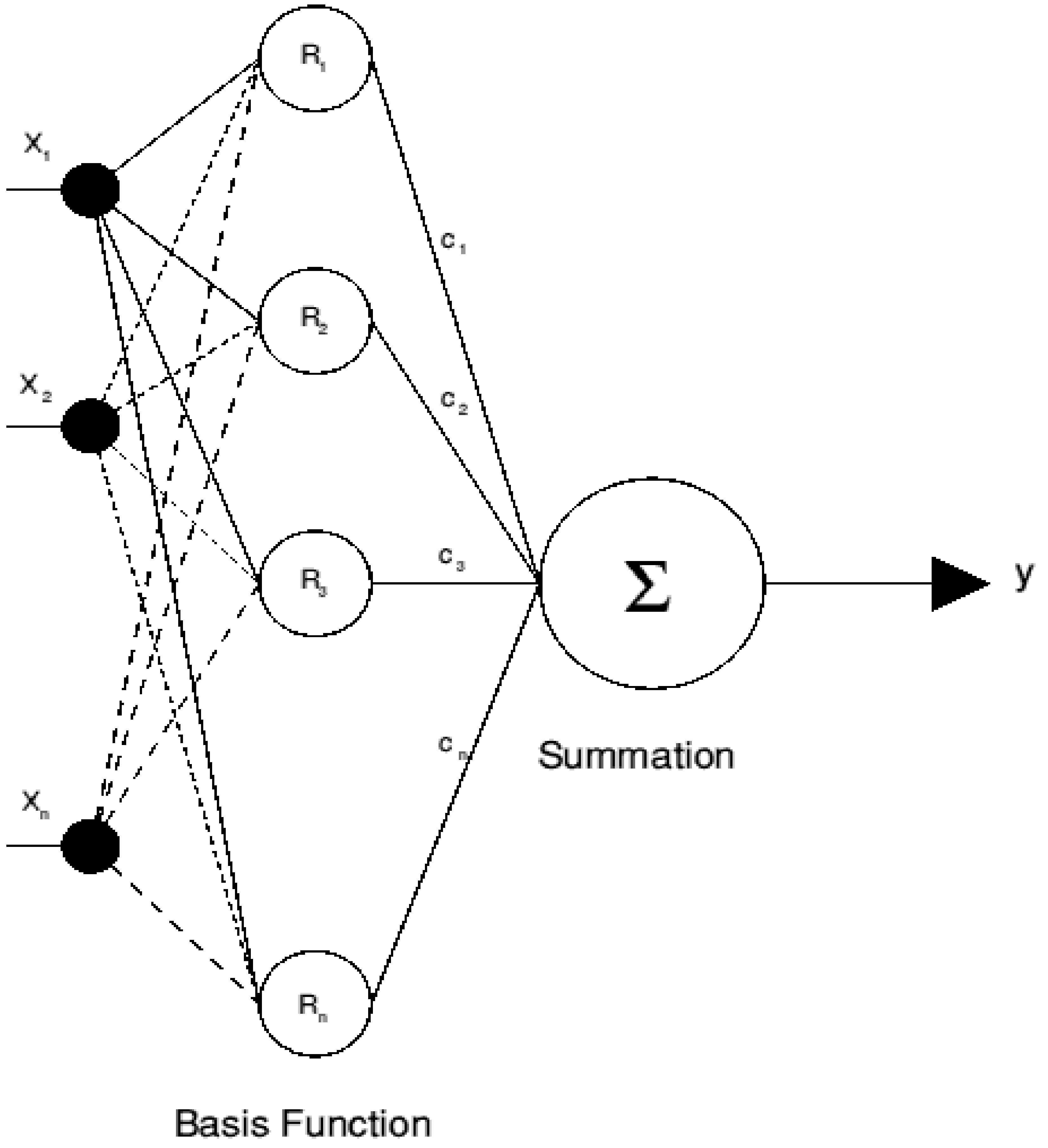
- **Radial Basis Function (RBF) Networks**: μοντέλα παρεμβολής και προσέγγισης

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Radial Basis Function Networks

- Πολυεπίπεδη αρχιτεκτονική δικτύου
- Στρώμα εισαγωγής:
  - διανέμει το διάνυσμα εισόδου σε όλους τους κρυμμένους κόμβους
- Κρυφό στρώμα:
  - κάθε κόμβος είναι ένας πυρήνας που περιέχει ένα **κέντρο** και μια **συνάρτηση βάσης**
  - εκτελεί μια μη γραμμική αντιστοίχιση από το χώρο εισόδου (n) τυπικά σε ένα υψηλότερο διαστατικό χώρο (k>n)
- Στρώμα παραγωγής:
  - πολλαπλασιάζει την έξοδο κάθε κρυφού στρώματος με έναν συντελεστή και αθροίζει τις προκύπτουσες τιμές



Co-financed by the European Union
Connecting Europe Facility

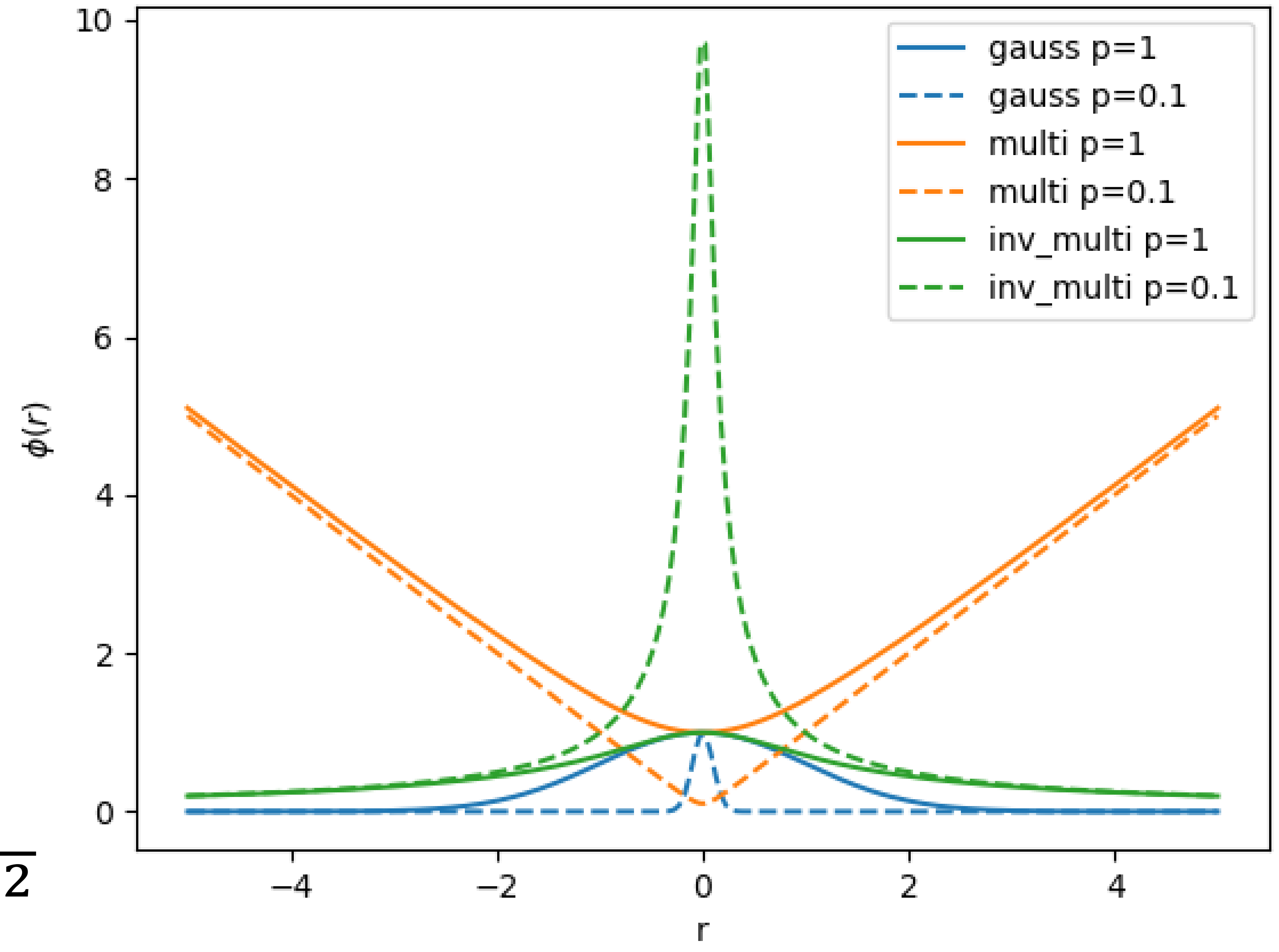
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Radial Basis Function (RBF)

Συνάρτηση φ πραγματικής αξίας της οποίας η τιμή εξαρτάται μόνο από την **απόσταση** μεταξύ της εισόδου $\boldsymbol{x}$ και κάποιου σταθερού σημείου $\boldsymbol{\mu}$, έτσι ώστε $\varphi(\boldsymbol{x}) := \varphi(r)$, $r = \|\boldsymbol{x} - \boldsymbol{\mu}\|$όπου, και η απόσταση είναι συνήθως Ευκλείδεια απόσταση.

Παραδείγματα

- Gaussian: $\varphi(r) = \exp\left(-{r^2}/{p^2}\right)$
- Πολυτετραγωνικό: $\varphi(r) = \sqrt{r^2 + p^2}$
- Αντίστροφο πολυτετραγωνικό: $\varphi(r) = {1}/{\sqrt{r^2+p^2}}$



ΠΗΓΗ

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κύρια ιδέα του δικτύου RBF

$$\varphi_1(\boldsymbol{x}) = k\left(\boldsymbol{x}, \boldsymbol{R}^{(1)}\right)$$
$$\varphi_2(\boldsymbol{x}) = k\left(\boldsymbol{x}, \boldsymbol{R}^{(2)}\right)$$
$$\varphi_3(\boldsymbol{x}) = k\left(\boldsymbol{x}, \boldsymbol{R}^{(3)}\right)$$

Δεδομένου ότι $\boldsymbol{x}$ η εκτιμώμενη έξοδος $\hat{y}$ είναι ο γραμμικός συνδυασμός συντελεστών και χαρακτηριστικών.

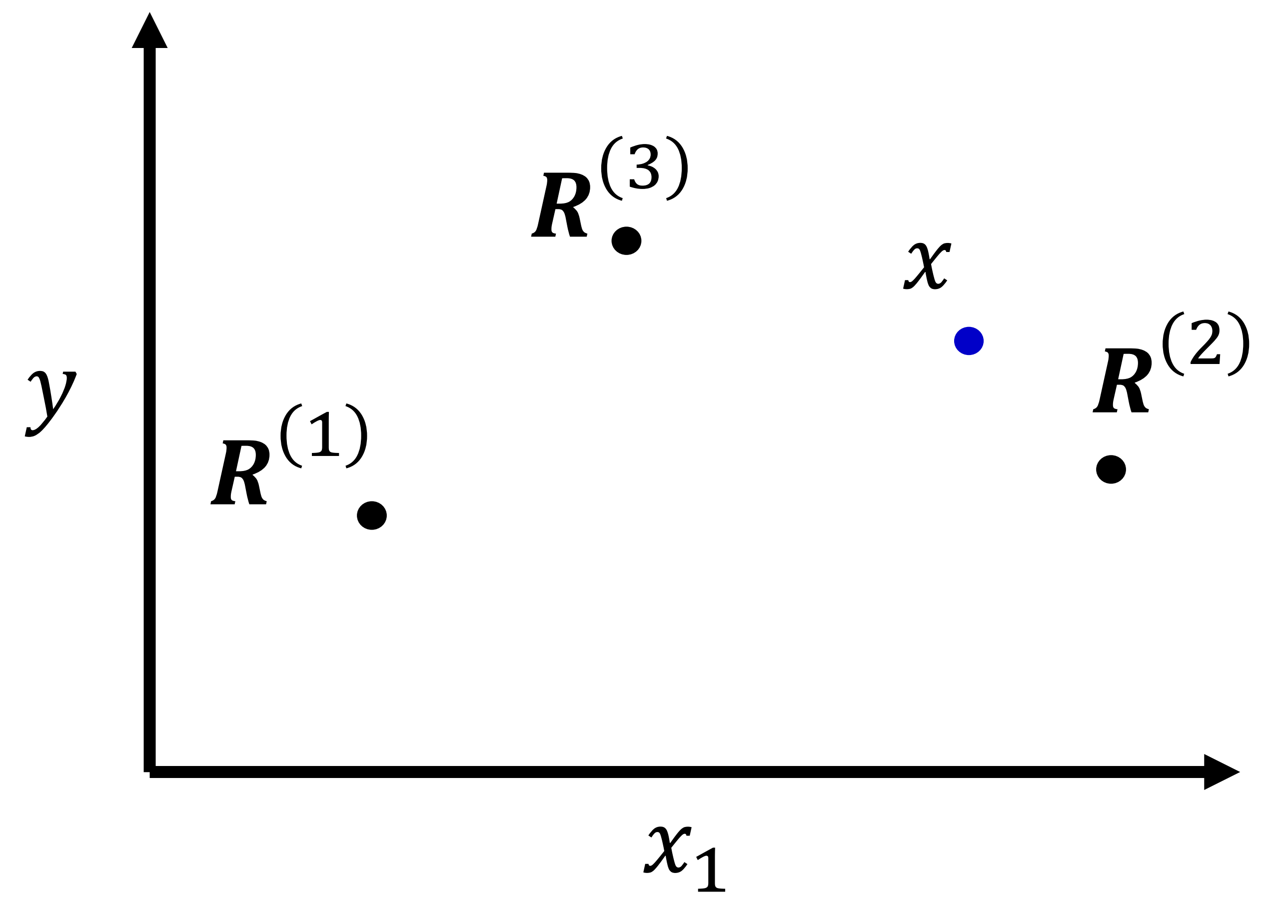
$$\hat{y} = \sum_{i=1}^{k} c_i\, \varphi_i(\boldsymbol{x})$$

Τα χαρακτηριστικά βασίζονται στην **εγγύτητα** στο κέντρο $\boldsymbol{R}^{(i)}$:

$$\varphi_i(\boldsymbol{x}) = k\left(\boldsymbol{x}, \boldsymbol{R}^{(i)}\right)$$

Συγκρίνετε με απλή γραμμική παλινδρόμηση:

$$\hat{y} = \sum_{i=1}^{k} c_i x_i \quad where\ k = n$$



Υποθέτοντας ένα Gaussian πυρήνα, το οποίο έχει το μεγαλύτερο και το οποίο έχει τη μικρότερη αξία;

Gaussian RBF πυρήνας και εγγύτητα

$$\varphi_1(\boldsymbol{x}) = k\left(\boldsymbol{x}, \boldsymbol{R}^{(1)}\right) = \exp\left(-\frac{\left\|\boldsymbol{x} - \boldsymbol{R}^{(1)}\right\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{j=1}^{n}\left(x_j - R_j^{(1)}\right)^2}{2\sigma^2}\right)$$

$\boldsymbol{R}^{(1)}$ είναι το n-διαστατο κέντρο

Εάν $\boldsymbol{x} \approx \boldsymbol{R}^{(1)}$:

$$\varphi_1(\boldsymbol{x}) \approx \exp\left(-\frac{0^2}{2\sigma^2}\right) \approx 1$$

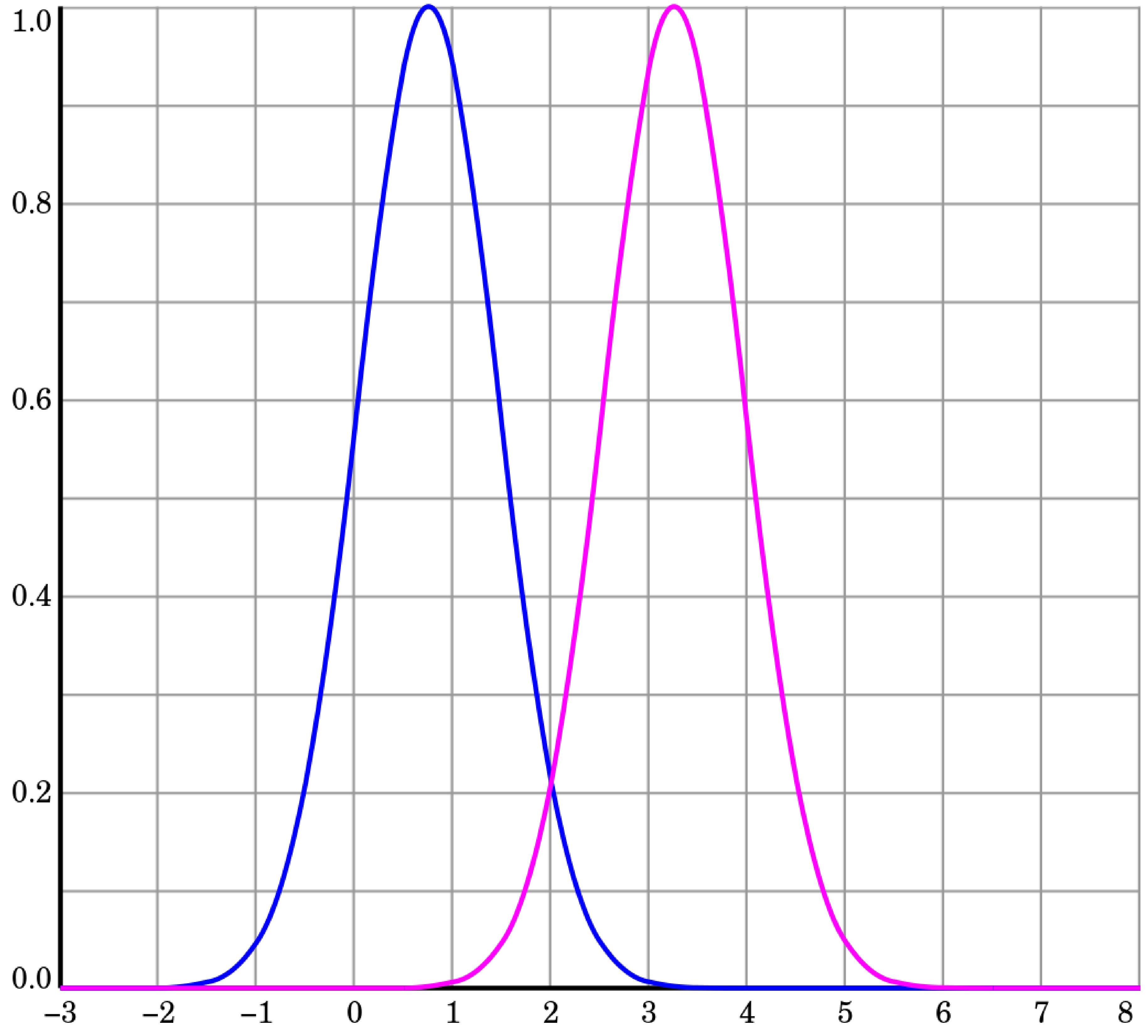
Εάν $\boldsymbol{x}$ απέχει πολύ από $\boldsymbol{R}^{(1)}$:

$$\varphi_1(\boldsymbol{x}) \approx \exp\left(-\frac{(large\ number)^2}{2\sigma^2}\right) \approx 0$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Πυρήνας RBF: κεντρικές συντεταγμένες



ΠΗΓΗ

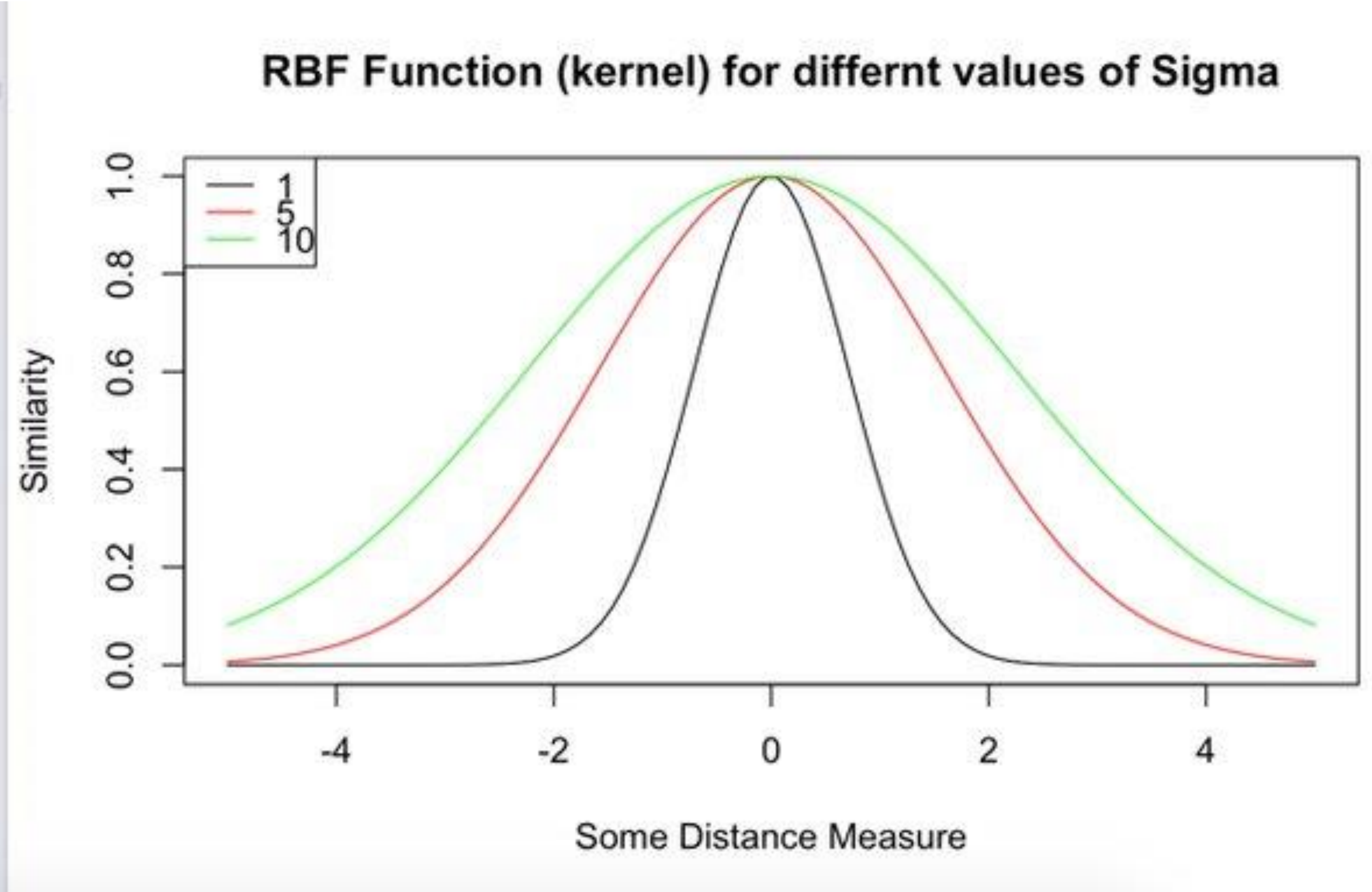
Κρατώντας το σ σταθερό

$$R^{(1)} = 0.75$$
$$R^{(2)} = 3.25$$

Κατά το σχεδιασμό ενός δικτύου RBF είναι ζωτικής σημασίας να καλύπτεται ολόκληρος ο χώρος δεδομένων με πυρήνες.

Η επίδραση του κάθε πυρήνα μειώνεται με την απόσταση.

Πυρήνας RBF: επίδραση του σ



[ΠΗΓΗ]

σ = πλάτος του Gaussian

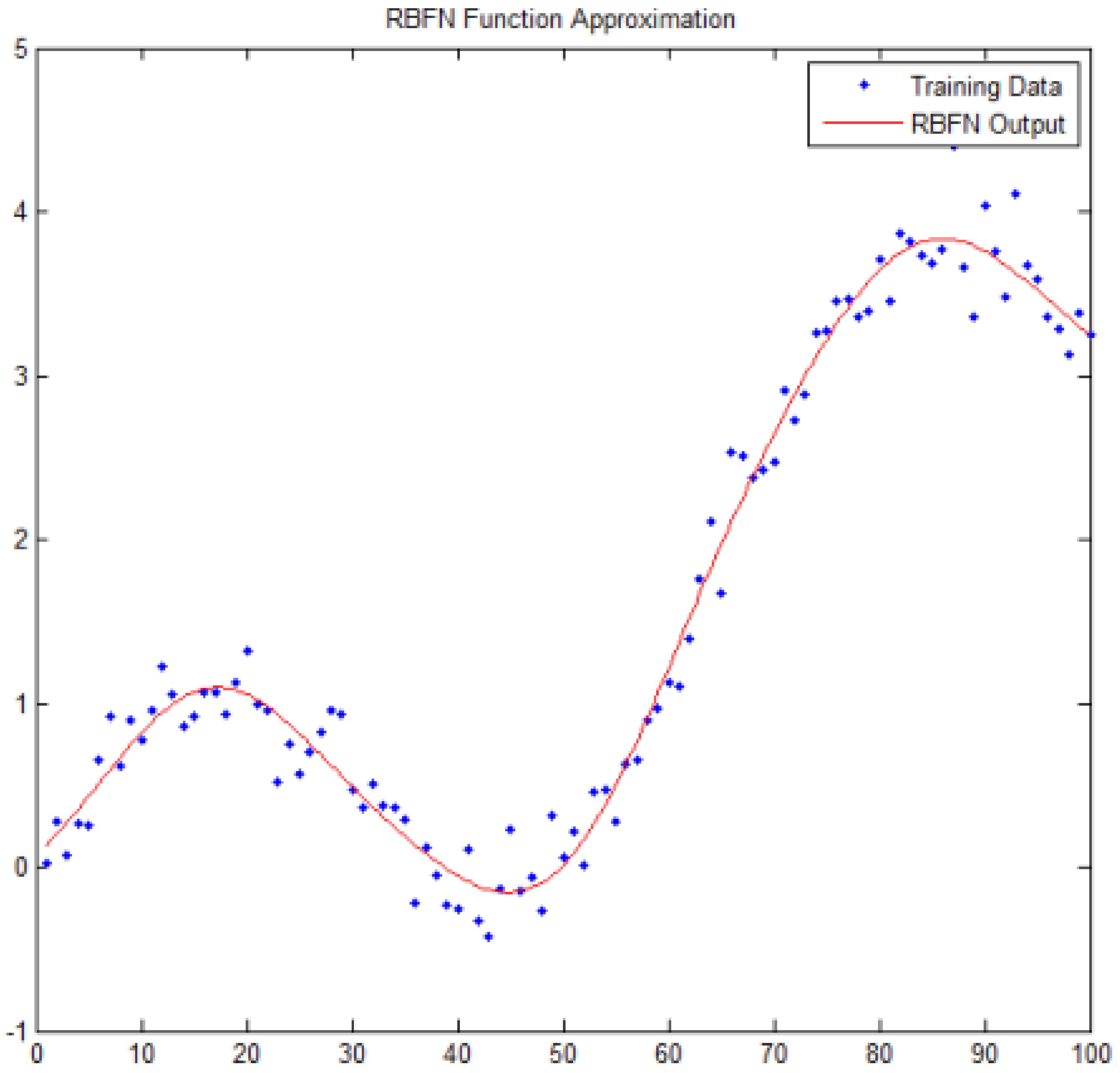
Μεγάλες σ: Τα χαρακτηριστικά ποικίλουν πιο ομαλά

Υψηλότερη προκατάληψη, χαμηλότερη διακύμανση

Μικροί σ: Τα χαρακτηριστικά ποικίλλουν λιγότερο ομαλά

Χαμηλότερη προκατάληψη, υψηλότερη διακύμανση

Τι μαθαίνει ένα εκπαιδευμένο δίκτυο RBF



ΠΗΓΗ

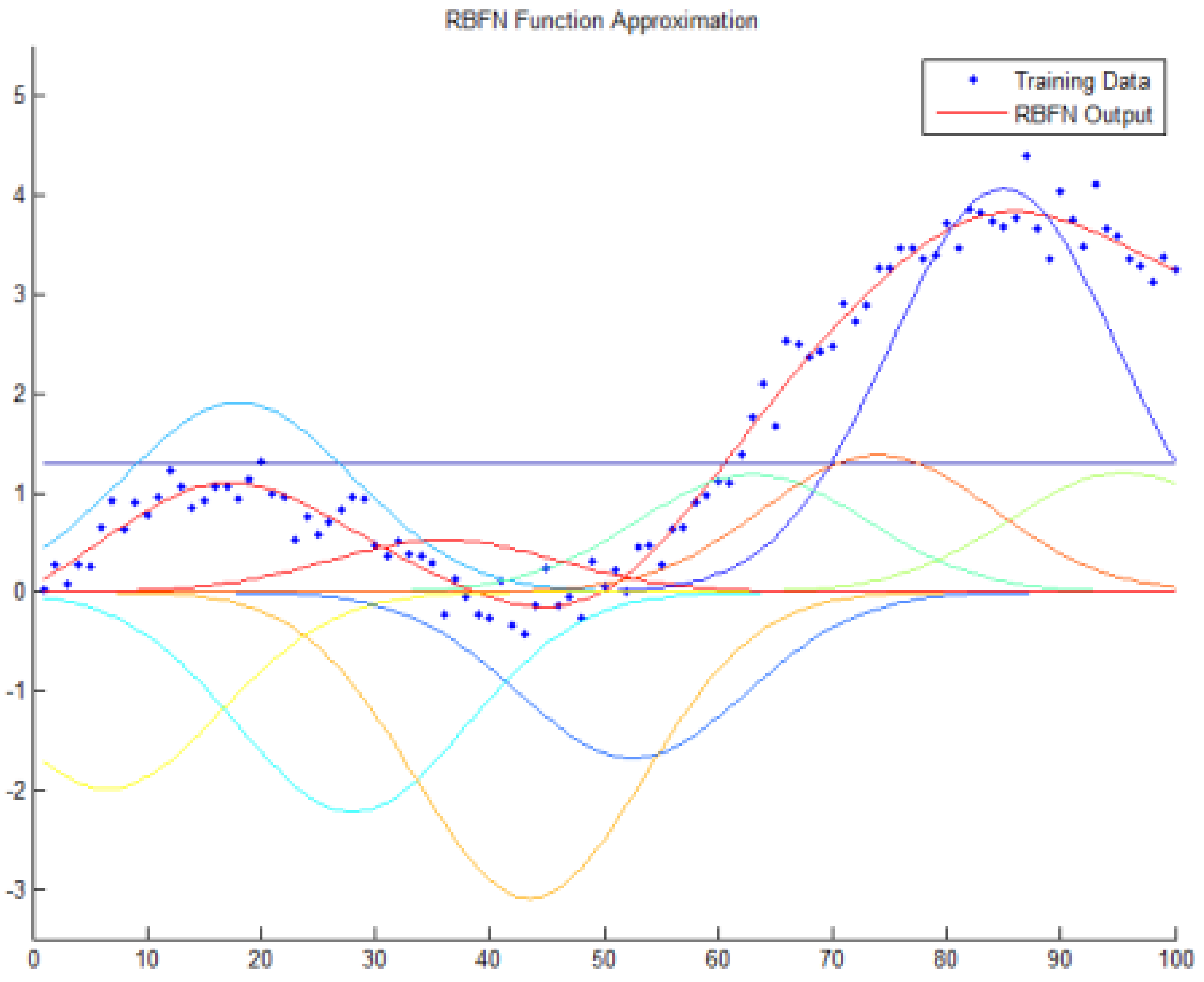
Παράδειγμα

1D εισαγωγή

Σύνολο δεδομένων 100 σημείων

RBF με 10 βασικές λειτουργίες

Πώς μαθαίνει ένα εκπαιδευμένο δίκτυο RBF



ΠΗΓΗ

Η παραγωγή του δικτύου RBF είναι το σταθμισμένο άθροισμα 10 Gaussians

Κάθε Gaussian έχει ένα διαφορετικό κέντρο

Το ύψος καθορίζεται από τον μορφωμένο συντελεστή (βάρος) ο οποίος μπορεί να είναι αρνητικός

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Εκπαίδευση δικτύου RBF για παρεμβολή

1. Δημιουργία κέντρων m, όπου κάθε κέντρο τοποθετείται σε ένα σημείο εκπαίδευσης: $\boldsymbol{R}^{(i)} = \boldsymbol{x}^{(i)}$
2. Ορισμός σ σε κάποια κατάλληλη τιμή
3. Παράμετροι υπολογισμού **γ**

Για κάθε εκπαιδευτικό ζεύγος $(\boldsymbol{x}^{(i)}, y^{(i)})$ έχουμε μία εξίσωση:

$$y^{(1)} = c_1 k\left(\boldsymbol{x}^{(1)}, \boldsymbol{R}^{(1)}\right) + c_2 k\left(\boldsymbol{x}^{(1)}, \boldsymbol{R}^{(2)}\right) + \cdots + c_m k\left(\boldsymbol{x}^{(1)}, \boldsymbol{R}^{(m)}\right)$$
$$y^{(2)} = c_1 k\left(\boldsymbol{x}^{(2)}, \boldsymbol{R}^{(1)}\right) + c_2 k\left(\boldsymbol{x}^{(2)}, \boldsymbol{R}^{(2)}\right) + \cdots + c_m k\left(\boldsymbol{x}^{(2)}, \boldsymbol{R}^{(m)}\right)$$

...

$$y^{(m)} = c_1 k\left(\boldsymbol{x}^{(m)}, \boldsymbol{R}^{(1)}\right) + c_2 k\left(\boldsymbol{x}^{(m)}, \boldsymbol{R}^{(2)}\right) + \cdots + c_m k\left(\boldsymbol{x}^{(m)}, \boldsymbol{R}^{(m)}\right)$$

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Εκπαίδευση δικτύου RBF για παρεμβολή

Αναπαριστούν εξισώσεις σε μορφή πίνακα:

$$\underbrace{\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} k(\boldsymbol{x}^{(1)}, \boldsymbol{R}^{(1)}) & \dots & k(\boldsymbol{x}^{(1)}, \boldsymbol{R}^{(m)}) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^{(m)}, \boldsymbol{R}^{(1)}) & \dots & k(\boldsymbol{x}^{(m)}, \boldsymbol{R}^{(m)}) \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}}_{\boldsymbol{c}}$$

Λύση για **γ**:

$$\boldsymbol{c} = (A^T A)^{-1} A^T \boldsymbol{y}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Εκπαίδευση δικτύου RBF για προσέγγιση με τη χρήση σταθερών κέντρων

1. Επιλέξτε k<m κατάλληλα κέντρα:
- ισότιμα απομακρυσμένα
- τυχαία δειγματοληψία (π.χ. ομοιόμορφα στο διάστημα ή από σύνολο παραδειγμάτων)
- χρήση αλγορίθμων ομαδοποίησης (π.χ. k-means clustering)

2. Ορισμός σ σε κάποια κατάλληλη τιμή
- χειροκίνητα
- προσδιορίζεται από την ακτίνα της συστάδα (θα μπορούσε να έχει διαφορετικές σ για κάθε κέντρο)

3. Παράμετροι υπολογισμού **γ**

Εκπαίδευση δικτύου RBF για προσέγγιση με τη χρήση σταθερών κέντρων

Σε μορφή πίνακα:

$$\underbrace{\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} k(\boldsymbol{x}^{(1)}, \boldsymbol{R}^{(1)}) & \dots & k(\boldsymbol{x}^{(1)}, \boldsymbol{R}^{(k)}) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^{(m)}, \boldsymbol{R}^{(1)}) & \dots & k(\boldsymbol{x}^{(m)}, \boldsymbol{R}^{(k)}) \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}}_{\boldsymbol{c}}$$

Λύση για **γ**:

$$\boldsymbol{c} = (A^T A)^{-1} A^T \boldsymbol{y}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Πλάτος του Gaussian (σ)

Εργασία παρεμβολής:

Μικρό σ: ανώμαλες παρεμβολές

Μεγαλύτερο σ: επέκταση της επιρροής κάθε δείγματος

Πολύ μεγάλο σ: υπερβολική παγκόσμια επιρροή κάθε δείγματος

Καθήκον προσέγγισης:

Για να επιτευχθεί μια καλή γενίκευση, το πλάτος θα πρέπει να είναι μεγαλύτερο από τη μέση απόσταση μεταξύ των παρακείμενων διανυσμάτων, αλλά μικρότερο από την απόσταση σε όλο το χώρο εισόδου.

Εκπαίδευση δικτύου RBF για προσέγγιση με τη χρήση ρυθμιζόμενων κέντρων

- Μπορούμε να εκπαιδεύσουμε τους συντελεστές c, τα κέντρα R, και τις διακυμάνσεις σ
  - Αυτό σημαίνει ότι οι κεντρικές συντεταγμένες και οι ακτίνες των βασικών λειτουργιών θα μπορούσαν να προσαρμοστούν καλύτερα στα δεδομένα.

- Μπορούμε να χρησιμοποιήσουμε (Στοχαστική) βαθμιαία κάθοδο για να επιτευχθεί αυτό
  - Ανάγκη: συνάρτηση απώλειας, παράγωγα της ζημίας w.r.t. c, R και σ

Εκπαίδευση δικτύου RBF για προσέγγιση με τη χρήση ρυθμιζόμενων κέντρων

1. Επιλέξτε τον αριθμό (k) και τις αρχικές συντεταγμένες των κέντρων (R).
2. Επιλέξτε την αρχική τιμή της παραμέτρου εξάπλωσης (σ) για κάθε κέντρο (R).
3. Αρχικοποιήστε τα βάρη/συντελεστές (γ) σε μικρές τυχαίες τιμές [-1,1].
4. Επαναλάβετε μέχρι να ικανοποιηθούν τα κριτήρια διακοπής (π.χ., μέγιστος αριθμός εποχών)
    1. Δώστε μια είσοδο $(\boldsymbol{x}^{(p)}, y^{(p)})$ στο δίκτυο και υπολογίστε την έξοδο του δικτύου

$$\widehat{y^{(p)}} = c_0 + \sum_{h=1}^{k} c_h \varphi(\boldsymbol{x}^{(p)}, \boldsymbol{R}^{(h)}, \sigma_h) \text{ που } \varphi\left(\boldsymbol{x}^{(p)}, \boldsymbol{R}^{(h)}, \sigma_h\right) = \exp\left(-\frac{\left\|\boldsymbol{x}^{(p)} - \boldsymbol{R}^{(h)}\right\|^2}{2\sigma_h^2}\right)$$

    2. Ενημέρωση των παραμέτρων του δικτύου (γ, **R**, σ) (επόμενη διαφάνεια)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Εκπαίδευση δικτύου RBF για προσέγγιση με τη χρήση ρυθμιζόμενων κέντρων

Ενημέρωση γ:

$$c_0 := c_0 - \eta_c \boxed{\varepsilon^{(p)}}$$

$$c_h := c_h - \eta_c \boxed{\varepsilon^{(p)} \varphi(\boldsymbol{x}^{(p)}, \boldsymbol{R}^{(h)}, \sigma_h)} \rightarrow \frac{\partial L}{\partial c}$$

$$\varepsilon^{(p)} = \widehat{y^{(p)}} - y^{(p)}$$

$$\mathrm{L}(\boldsymbol{c}, \boldsymbol{R}, \boldsymbol{\sigma}) = \frac{1}{2}[\varepsilon^{(p)}]^2$$

Ενημέρωση $\boldsymbol{R}$:

$$\boldsymbol{R}^{(h)} := \boldsymbol{R}^{(h)} - \eta_R \boxed{\varepsilon^{(p)} c_h\, \varphi\left(\boldsymbol{x}^{(p)}, \boldsymbol{R}^{(h)}, \sigma_h\right) (\boldsymbol{x}^{(p)} - \boldsymbol{R}^{(h)})/\sigma_h^2} \rightarrow \frac{\partial L}{\partial R}$$

Ενημέρωση σ:

$$\sigma_h := \sigma_h - \eta_\sigma \boxed{\varepsilon^{(p)} c_h\, \varphi\left(\boldsymbol{x}^{(p)}, \boldsymbol{R}^{(h)}, \sigma_h\right) \left\|\boldsymbol{x}^{(p)} - \boldsymbol{R}^{(h)}\right\|^2 / \sigma_h^3} \rightarrow \frac{\partial L}{\partial \sigma}$$

Ποσοστά εκμάθησης

$0 < \eta_c \leq 1$

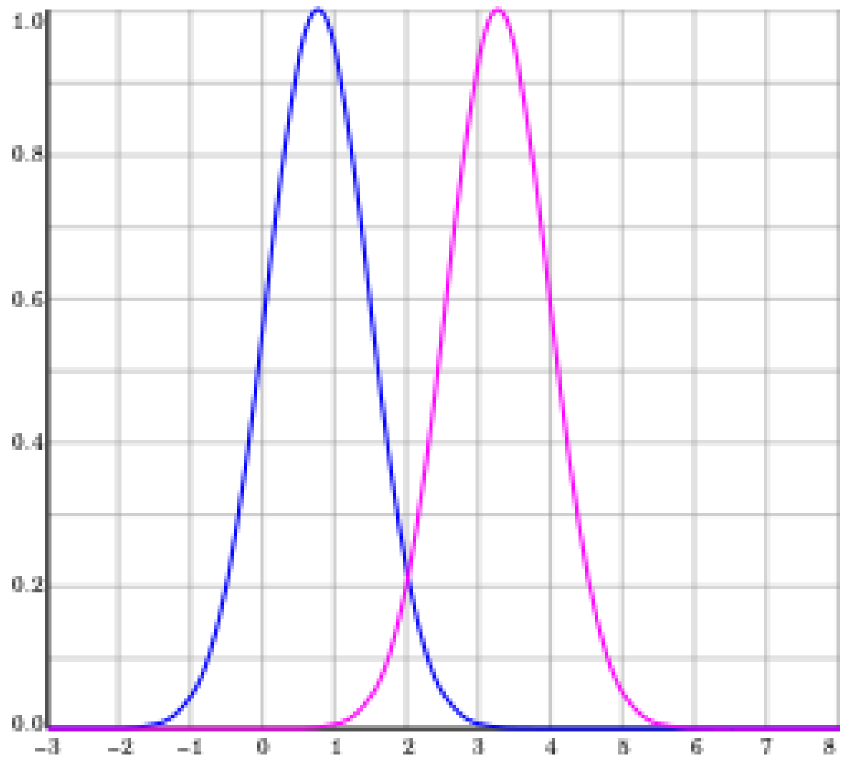
$0 < \eta_{\mathrm{R}} \leq 1$

$0 < \eta_\sigma \leq 1$

Κανονικοποιημένο δίκτυο RBF

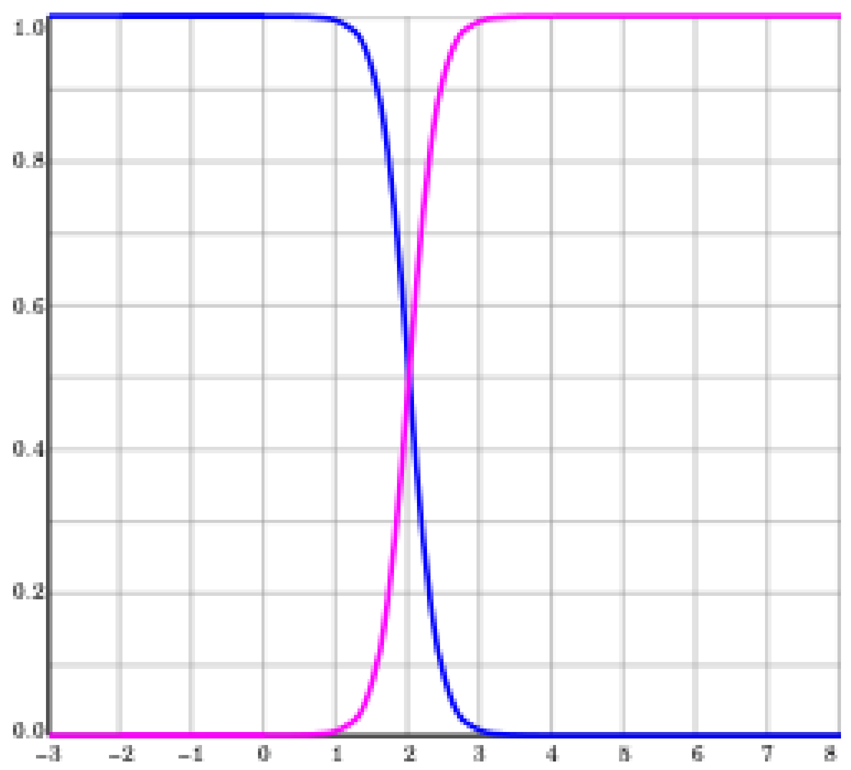
Μη κανονικοποιημένοι:

$$\hat{y} = \sum_{i=1}^{k} c_i \, \varphi_i(\boldsymbol{x})$$



Κανονικοποιημένο:

$$\hat{y} = \frac{\sum_{i=1}^{k} c_i \, \varphi_i(\boldsymbol{x})}{\sum_{i=1}^{k} \varphi_i(\boldsymbol{x})}$$



Μη κανονικοποιημένοι:

- Πιο τοπικά

Ομαλοποιημένος:

- μπορεί να έχουν καλύτερη γενίκευση

Και οι δύο είναι καθολικοί προσεγγιστές λειτουργίας: δεδομένων αρκετών κρυφών μονάδων, μπορούν να προσεγγίσουν οποιαδήποτε συνεχή λειτουργία.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Δίκτυα RBF για ταξινόμηση

Μπορούμε να χρησιμοποιήσουμε τα δίκτυα RBF για την ταξινόμηση τροφοδοτώντας την παραγωγή στη logistic/sigmoid συνάρτηση:

$\hat{y} = \sigma\left(\sum_{i=1}^{k} c_i\, \varphi_i(\boldsymbol{x})\right)$ που $\sigma(z) = \frac{1}{(1+e^{-z})}$

Αυτό είναι το ίδιο με το logistic regression, όπου τα χαρακτηριστικά υπολογίζονται χρησιμοποιώντας τον πυρήνα Gaussian RBF.

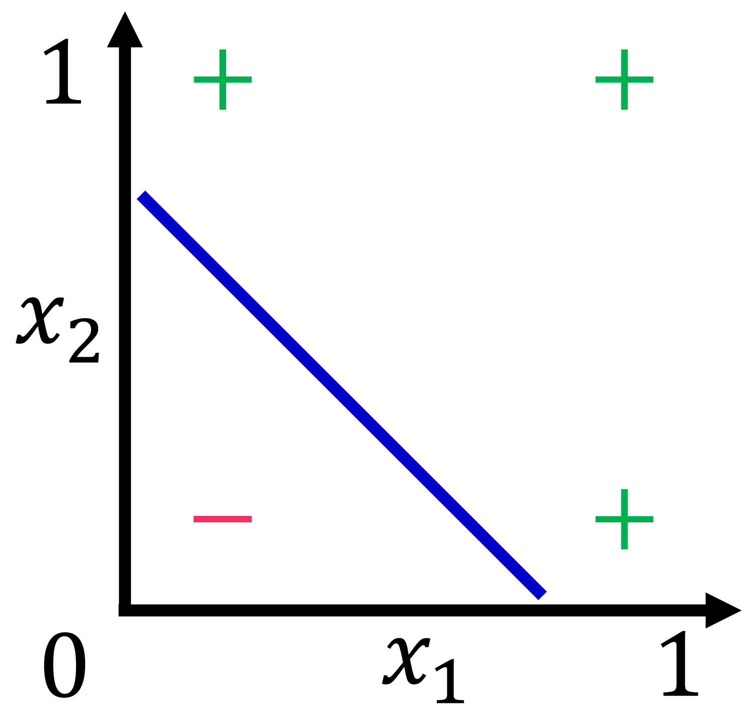
- Δημιουργεί μη γραμμικά όρια αποφάσεων

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

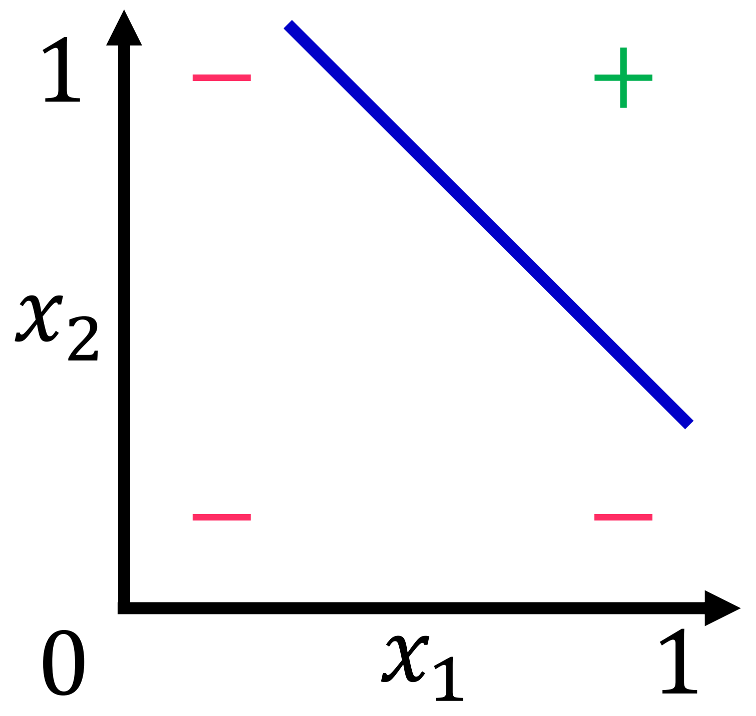
Πρόβλημα XOR

OR



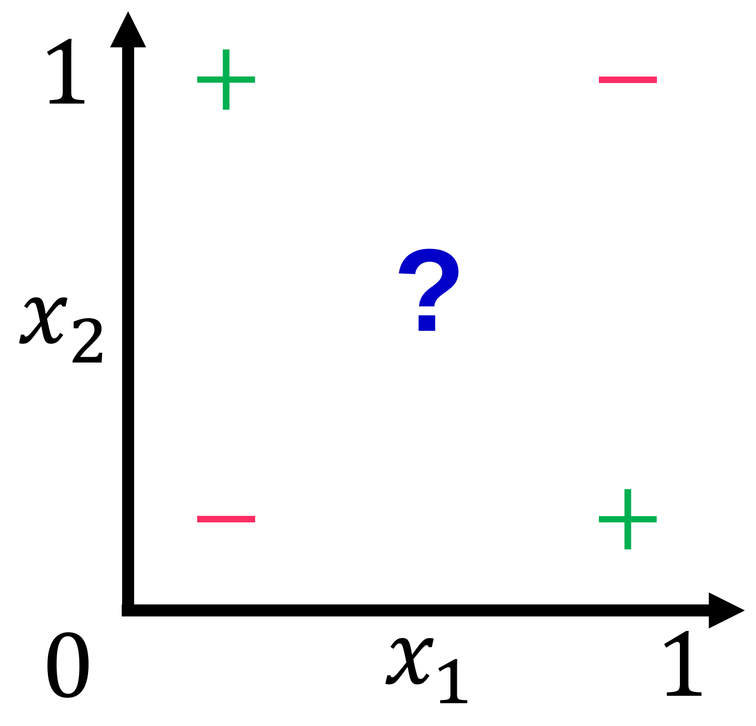
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

AND



x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

XOR



x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

OR, AND: γραμμικά διαχωρίσιμα

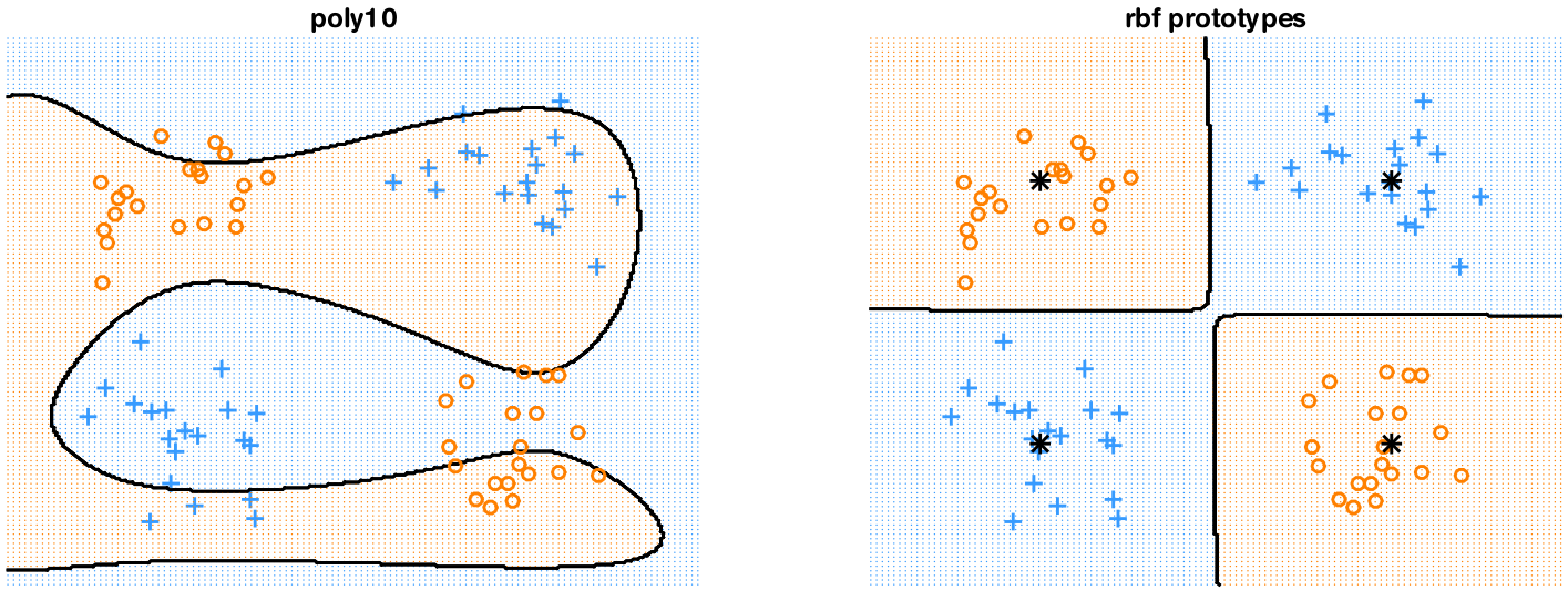
TO XOR: **μη** γραμμικά διαχωρίσιμο

Ενθαρρύνει την ανάγκη για μη γραμμικό μετασχηματισμό των εισόδων

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Θορυβώδες πρόβλημα XOR



(α) (β)

Πηγή: Murphy, K. (2022). Probabilistic Machine Learning - An Introduction. MIT Press

Τοποθέτηση γραμμικού ταξινομητή λογιστικής παλινδρόμησης με τη χρήση:

a) βαθμός 10 πολυωνυμική επέκταση
   - ➢ δεν μπορούν να διαχωρίσουν τις τάξεις
b) Πυρήνας RBF με κέντρα που προσδιορίζονται από τα 4 μαύρα σημεία
   - ➢ λύνει εύκολα το πρόβλημα

Gaussian ανελίξεις

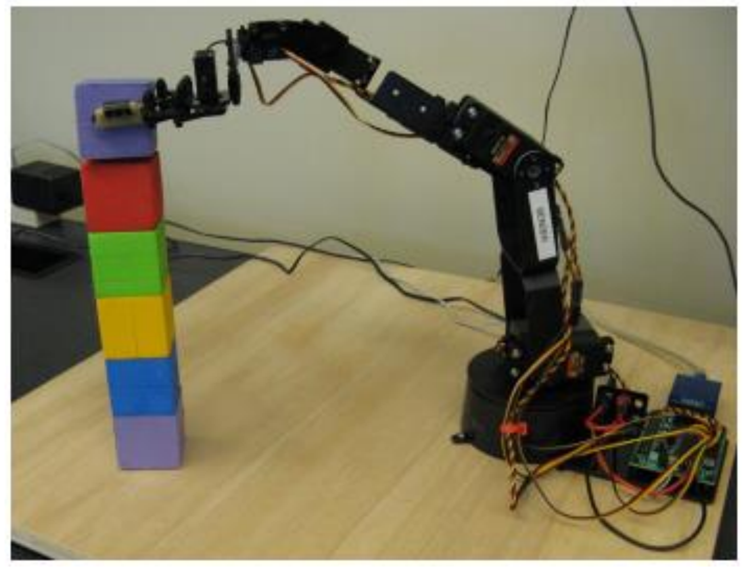
Co-financed by the European Union
Connecting Europe Facility

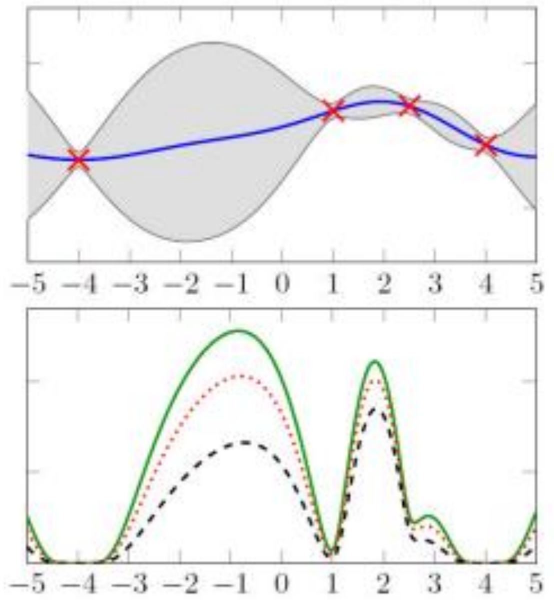
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

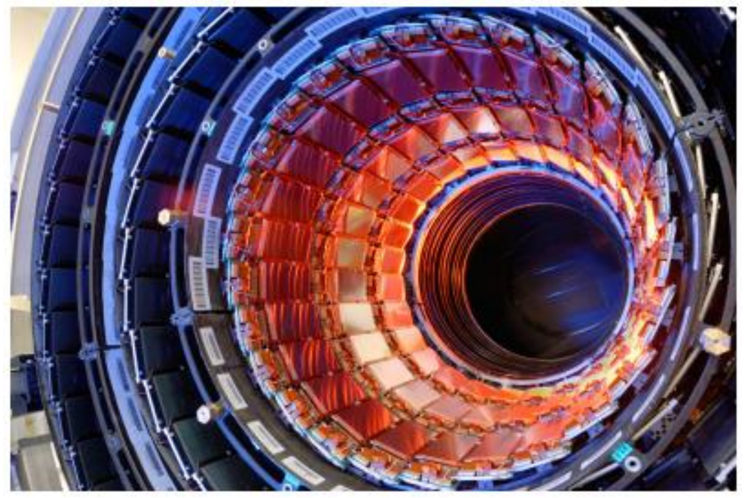
Gaussian ανελίξεις

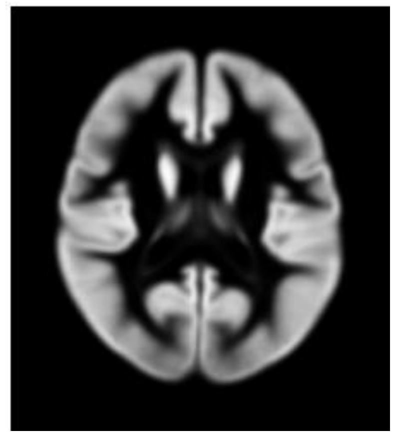
- **Gaussian ανελίξεις**: πιθανολογική μέθοδος που δίνει εμπιστοσύνη στην προβλεπόμενη λειτουργία
- Χρησιμοποιείται κυρίως σε εργασίες παλινδρόμησης (ωστόσο, έχει χρησιμοποιηθεί επίσης σε εργασίες ταξινόμησης και ομαδοποίησης)

- **Μερικές περιοχές εφαρμογής:**
  - Ρομποτική
  - Πειραματικός σχεδιασμός
  - Γεωστατιστική
  - Μοντελοποίηση και πρόβλεψη χρονοσειρών
  - Φυσική υψηλής ενέργειας
  - Ιατρικές εφαρμογές







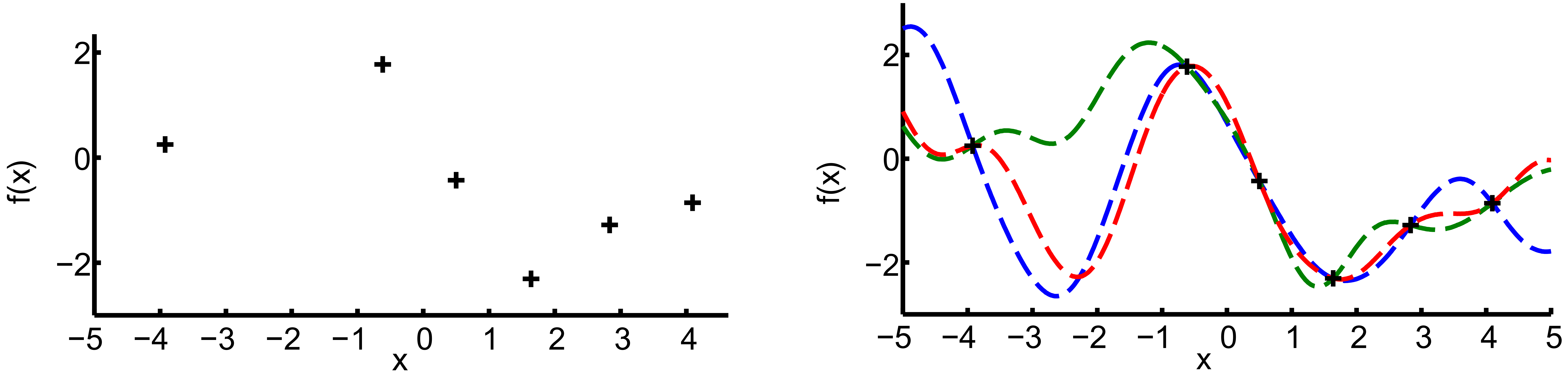


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Gaussian ανελίξεις κύρια ιδέα

- Για ένα δεδομένο σύνολο σημείων εκπαίδευσης, υπάρχουν δυνητικά απείρως πολλές λειτουργίες που ταιριάζουν στα δεδομένα.

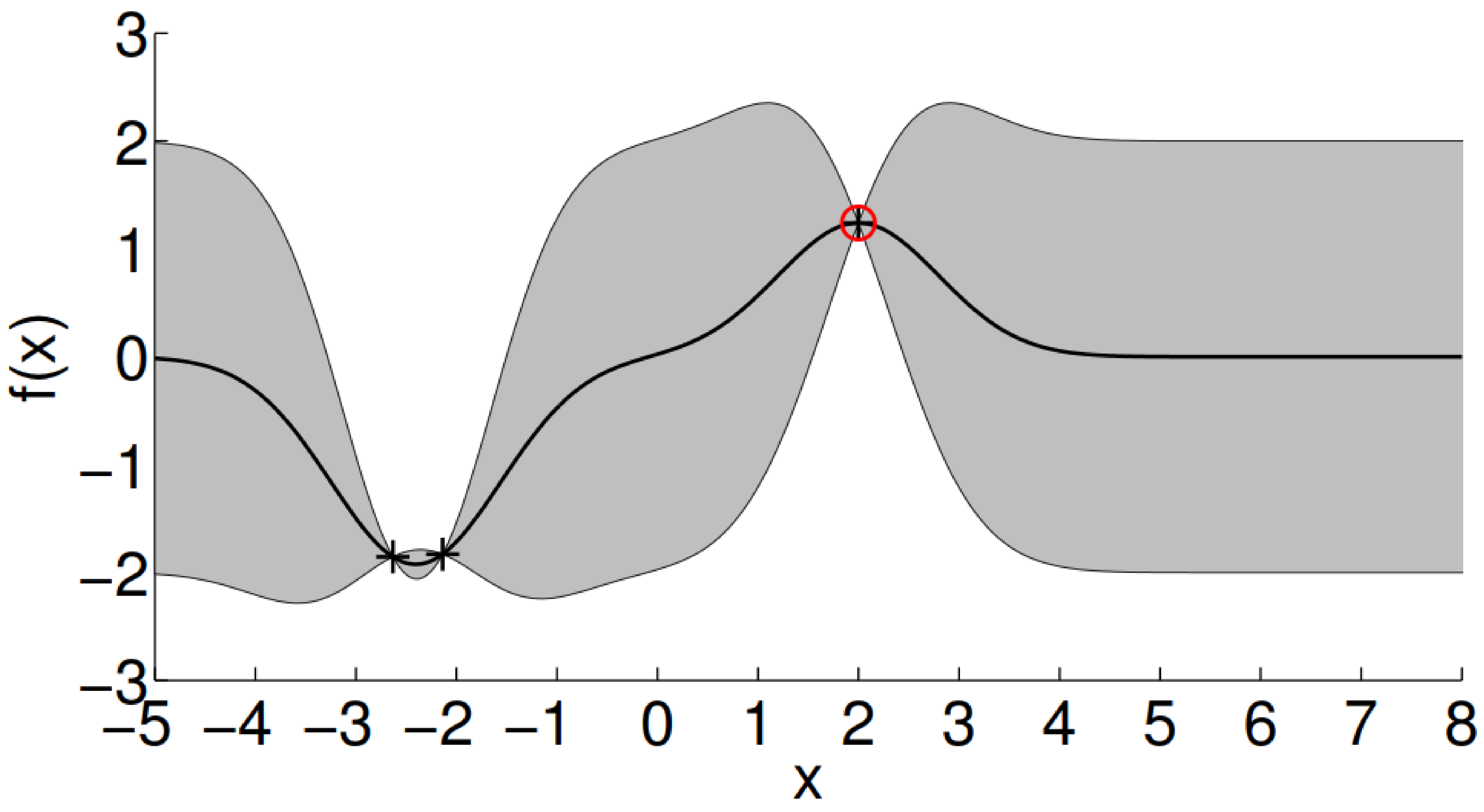


- Πώς μπορούμε να το αντιμετωπίσουμε αυτό;

Τα αριθμητικά στοιχεία προσαρμόστηκαν από: Deisenroth, M. P. & Rasmussen, C. E. (2011) PILCO: Μια προσέγγιση με βάση το μοντέλο και την αποδοτική χρήση των δεδομένων στην αναζήτηση πολιτικής

Gaussian ανελίξεις κύρια ιδέα

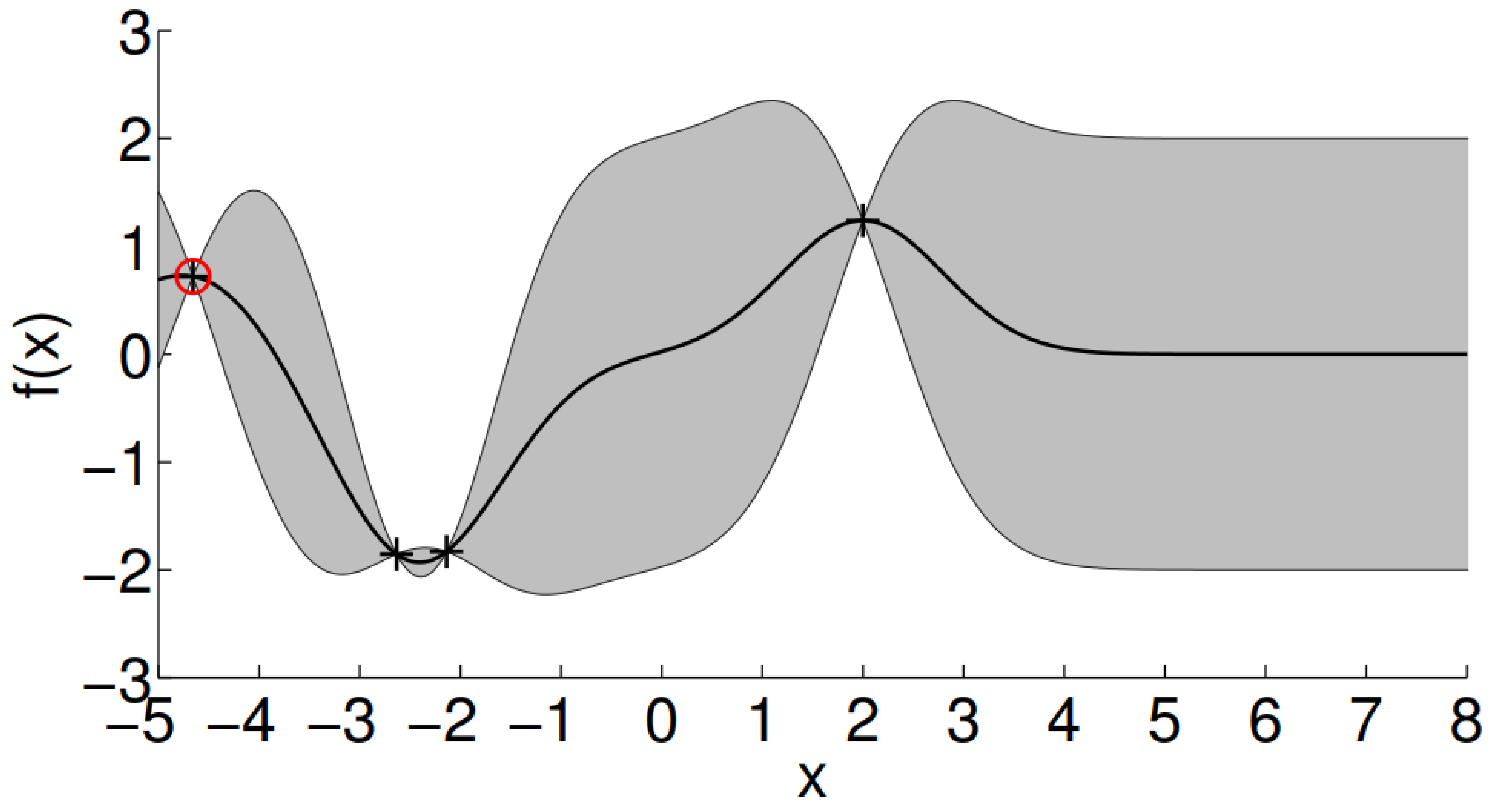
- Οι GPs εκχωρουν μια πιθανότητα σε κάθε μία από αυτές τις λειτουργίες
- Ο μέσος όρος αυτής της κατανομής πιθανοτήτων (μαύρη γραμμή) αντιπροσωπεύει τον πιο πιθανό χαρακτηρισμό των δεδομένων
- Πιθανολογική προσέγγιση μας επιτρέπει να ενσωματώσουμε την εμπιστοσύνη της πρόβλεψης στο αποτέλεσμα παλινδρόμησης (εμφανίζεται ως σκιασμένη περιοχή)



ΠΗΓΗ

Gaussian ανελίξεις κύρια ιδέα

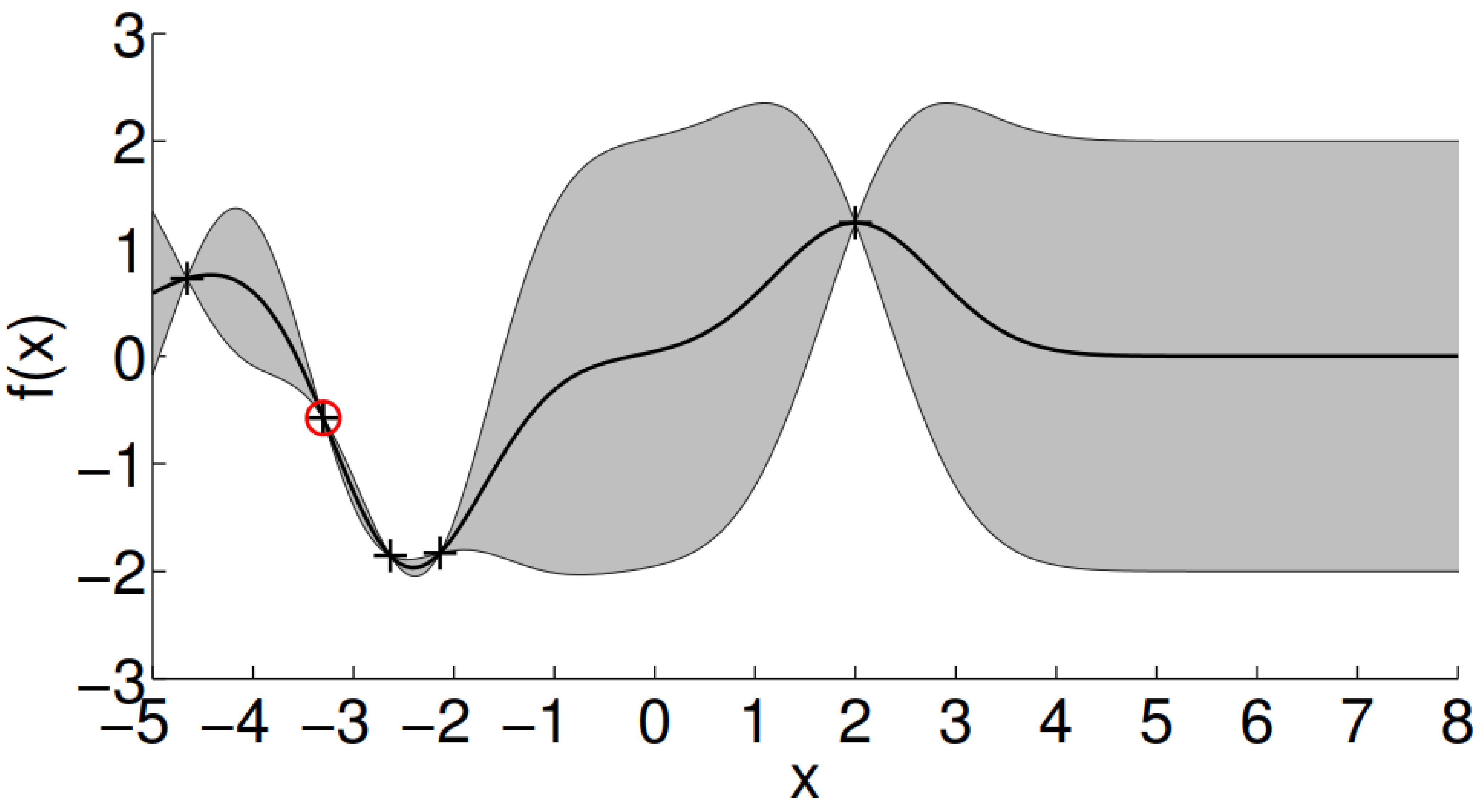
- Καθώς προσθέτουμε περισσότερα σημεία δεδομένων, προσαρμόζουμε τη διανομή



ΠΗΓΗ

Gaussian ανελίξεις κύρια ιδέα

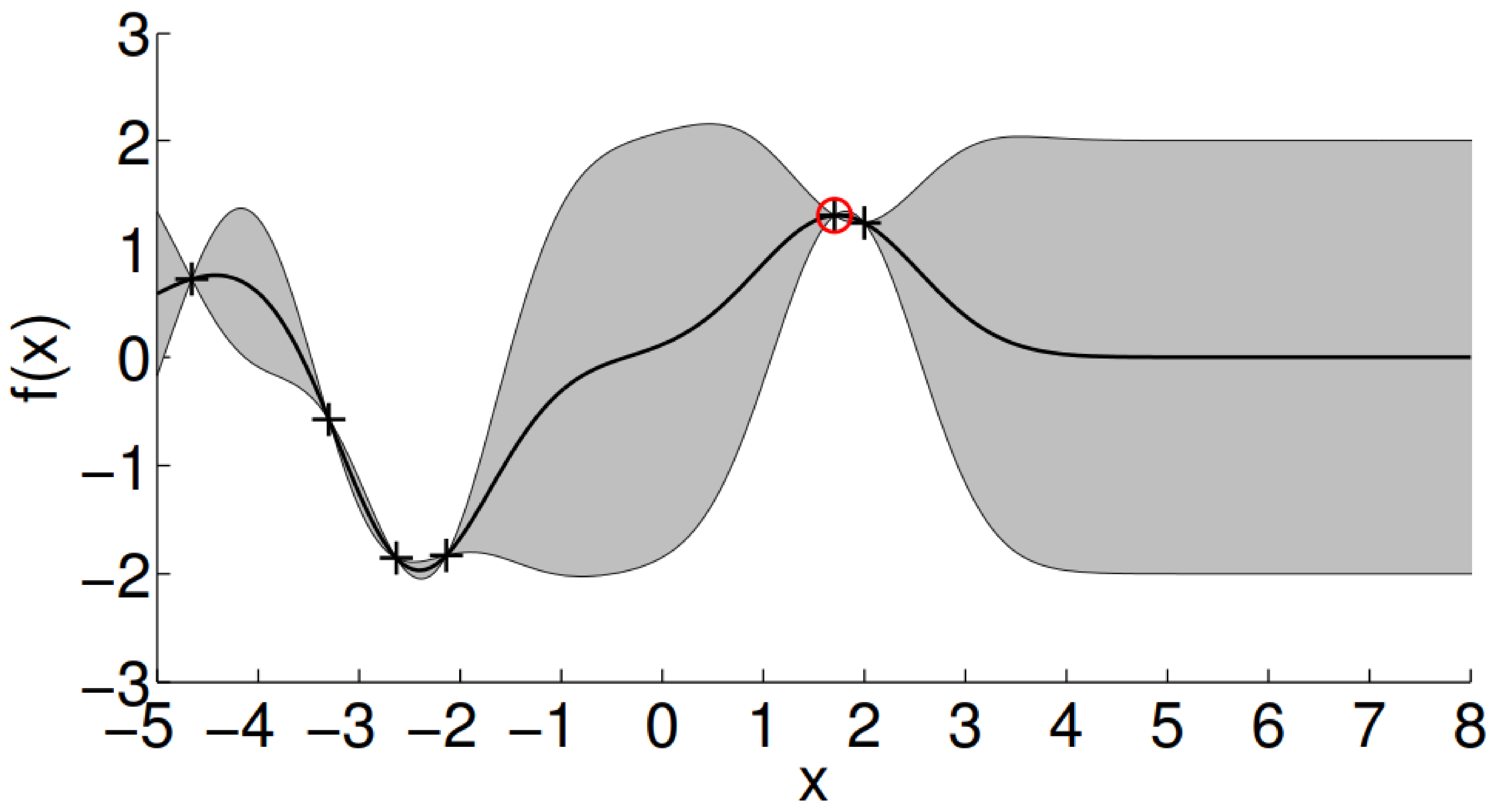
- Καθώς προσθέτουμε περισσότερα σημεία δεδομένων, προσαρμόζουμε τη διανομή



ΠΗΓΗ

Gaussian ανελίξεις κύρια ιδέα

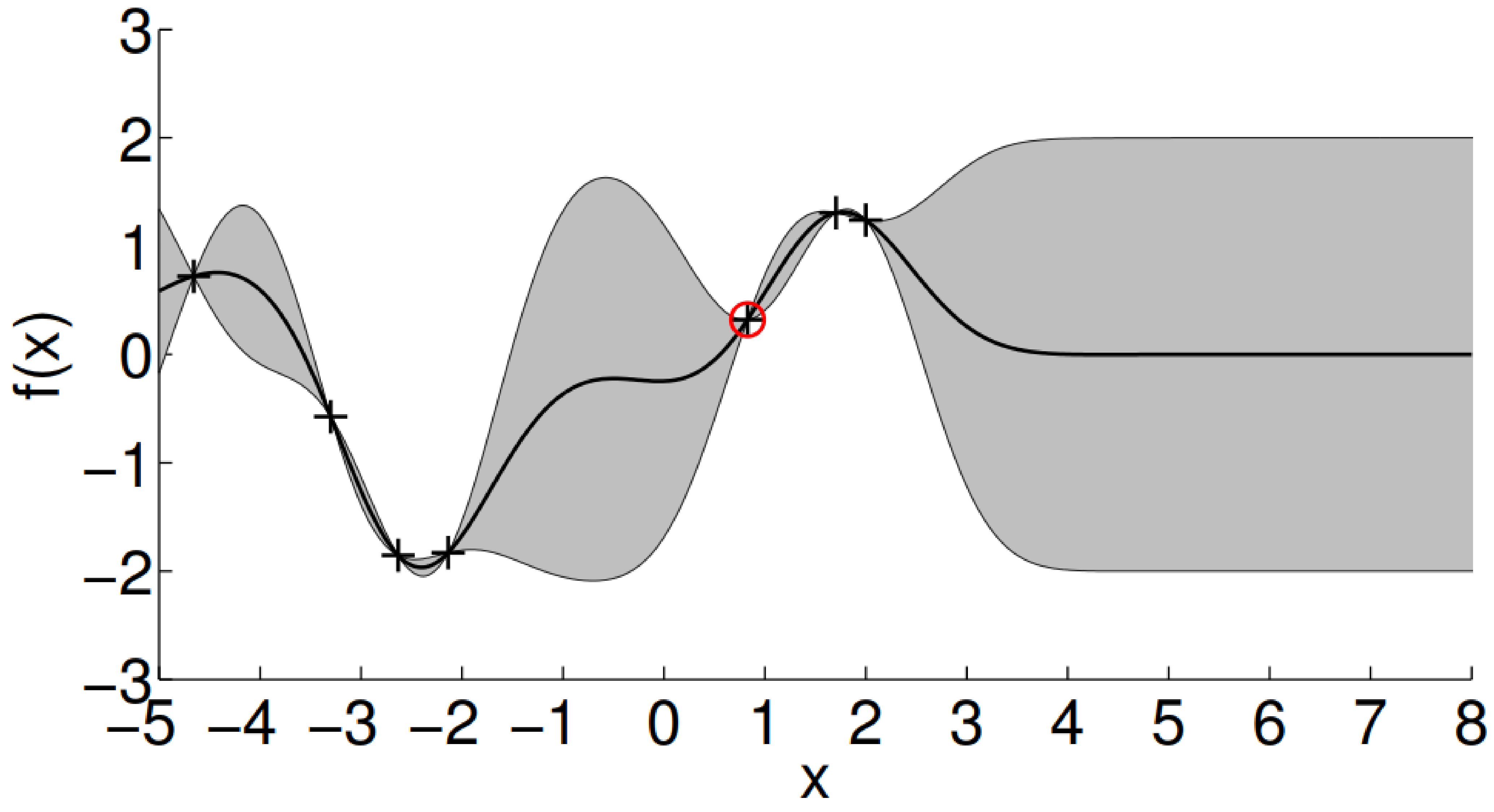
- Καθώς προσθέτουμε περισσότερα σημεία δεδομένων, προσαρμόζουμε τη διανομή



ΠΗΓΗ

Gaussian ανελίξεις κύρια ιδέα

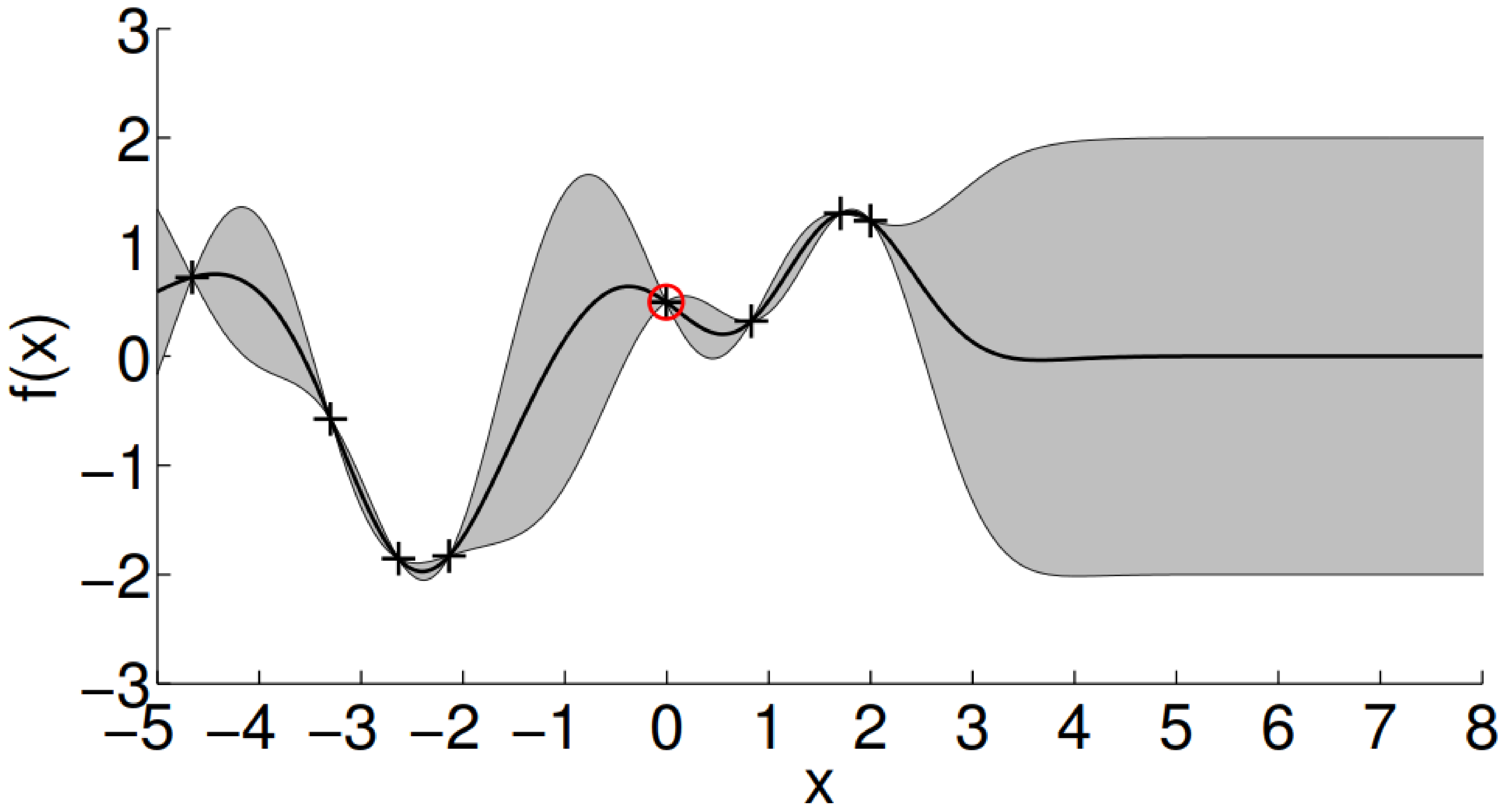
- Καθώς προσθέτουμε περισσότερα σημεία δεδομένων, προσαρμόζουμε τη διανομή



ΠΗΓΗ

Gaussian ανελίξεις κύρια ιδέα

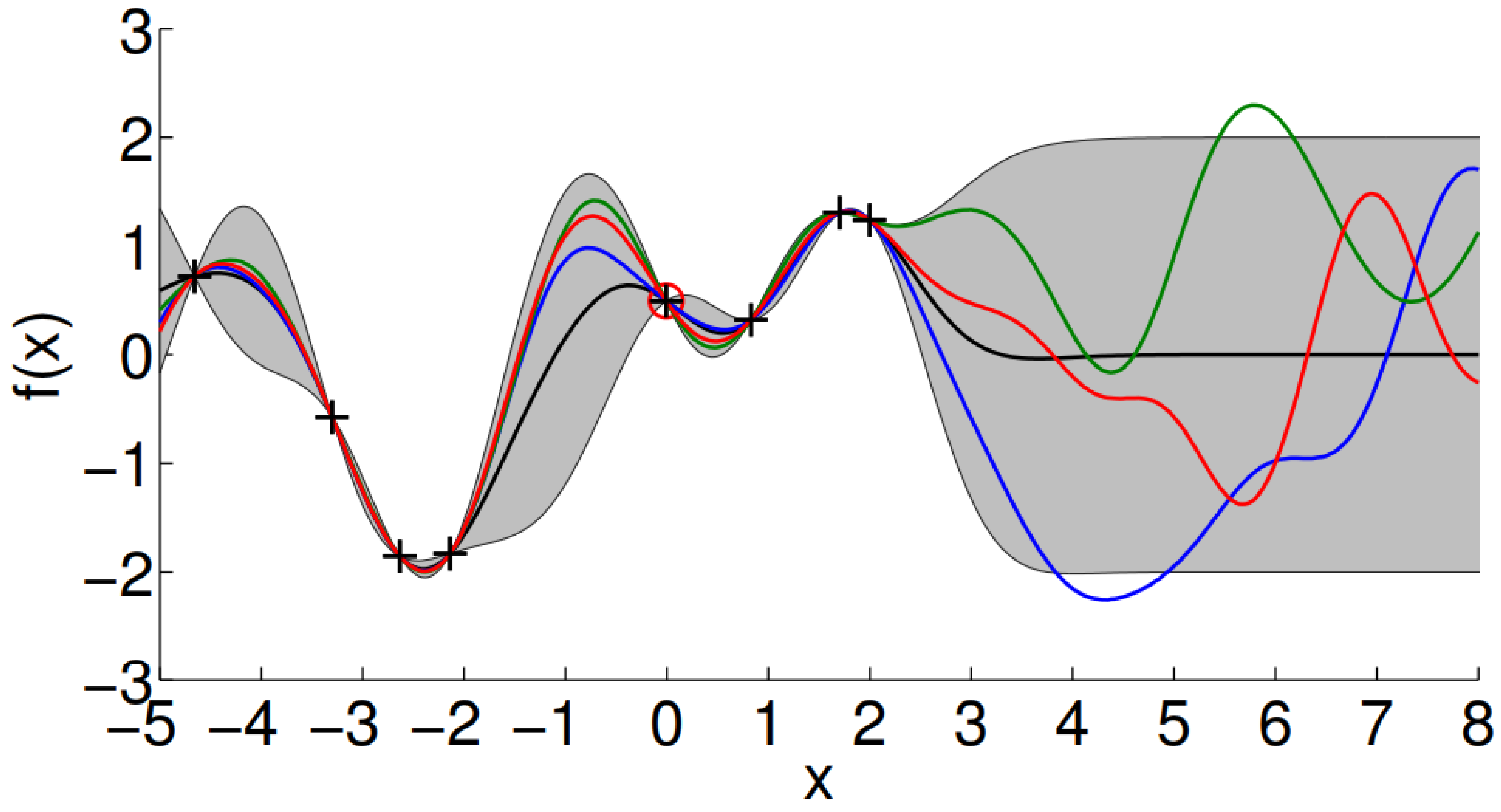
- Καθώς προσθέτουμε περισσότερα σημεία δεδομένων, προσαρμόζουμε τη διανομή



ΠΗΓΗ

Gaussian ανελίξεις κύρια ιδέα

- Δεδομένου ότι ένα GP είναι μια κατανομή πιθανοτήτων πάνω από συναρτήσεις, μπορούμε να δοκιμάσουμε συναρτήσεις από αυτό.



ΠΗΓΗ

Ορισμός Gaussian ανέλιξης

- Τοποθετούμε μια διανομή $p(f)$ στις λειτουργίες f
- Ανεπίσημα, μια συνάρτηση μπορεί να θεωρηθεί ένα απείρως μακρύ διάνυσμα των τιμών συναρτήσεων $f = [f_1, f_2, f_3, \dots]$
- Ένα GP είναι μια γενίκευση μιας πολυμεταβλητής κατανομής Gauss σε απείρως πολλές μεταβλητές.

Ορισμός (Rasmussen & Williams, 2006): Μια GΔ είναι μια συλλογή τυχαίων μεταβλητώνf_1, f_2, ..., κάθε πεπερασμένος αριθμός των οποίων είναι Gaussian κατανεμημένο.

- Μια κατανομή Gaussian καθορίζεται από ένα μέσο διάνυσμα $\boldsymbol{\mu}$ και ένας πίνακας συνδιακύμανσης $\boldsymbol{\Sigma}$
- Ένας GP καθορίζεται από μια **μέση συνάρτηση** $m(.)$ και μια **συνδιακύμανση (πυρήνας)** $k(.,.)$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Gaussian Process Regression as Bayesian Inference

Στόχος: Για ένα σύνολο παρατηρήσεων $y^{(i)} = f\left(x^{(i)}\right) + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)$ βρείτε μια (οπίσθια) κατανομή πάνω από συναρτήσεις $p(f(.)|\boldsymbol{X}, \boldsymbol{y})$ που εξηγεί τα δεδομένα. Εδώ είναι: $\boldsymbol{X}$ Είσοδος εκπαίδευσης, $\boldsymbol{y}$ στόχοι εκπαίδευσης.

Δεδομένα εκμάθησης: $\boldsymbol{X}$,$\boldsymbol{y}$. Το θεώρημα Bayes αποδίδει

$$p(f(.)|\boldsymbol{X}, \boldsymbol{y}) = \frac{p(\boldsymbol{y}|f(.), \boldsymbol{X})p(f(.))}{p(\boldsymbol{y}|\boldsymbol{X})}$$

Πριν από: $p\big(f(.)\big) = GP(m, k)$ → Καθορισμός μέσης συνάρτησης m και πυρήνας k

Πιθανότητα (μοντέλο θορύβου): $p(\boldsymbol{y}|f(.), \boldsymbol{X}) = N(f(\boldsymbol{X}), \sigma^2 \boldsymbol{I})$

Οριακή πιθανότητα (αποδεικτικά στοιχεία): $p(\boldsymbol{y}|\boldsymbol{X}) = \int p(\boldsymbol{y}|f(.), \boldsymbol{X})p(f(.)|\boldsymbol{X})\, df$

Οπίσθια όψη: $p(f(.)|\boldsymbol{X}, \boldsymbol{y}) = GP(m_{post}, k_{post})$

Πριν τη GΔ

Τι χαρακτηρίζει τη λειτουργία που θέλουμε να μοντελοποιήσουμε;

- Μέση λειτουργία
  - Μας επιτρέπει να μεροληπτούμε το μοντέλο (μπορεί να έχει νόημα στις ρυθμίσεις που αφορούν την εφαρμογή)
  - Μπορεί να είναι μια παραμετροποιημένη λειτουργία. Για παράδειγμα: $m_{\boldsymbol{\theta}}\,(x) = \boldsymbol{\theta}^T \boldsymbol{\varphi}(\boldsymbol{x})$
  - Μπορεί να ενσωματώσει προγενέστερες γνώσεις που αφορούν συγκεκριμένα προβλήματα (π.χ. στη ρομποτική, στις φυσικές επιστήμες)
  - Μπορεί να απλοποιήσει το πρόβλημα
  - Συχνά: «Αγνωστικιστική»: λειτουργία ελλείψει δεδομένων ή προηγούμενης γνώσης: $m(.) = 0$παντού
- Συνάρτηση συνδιακύμανσης (πυρήνας)
  - Κωδικοποιεί δομικές παραδοχές υψηλού επιπέδου (π.χ. ομαλότητα, περιοδικότητα) της συνάρτησης που θέλουμε να μοντελοποιήσουμε
  - Πυρήνας Gaussian: $k_{Gauss}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) = \sigma_f^2 \exp\left(-\left(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\right)^T \left(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\right)/\lambda^2\right)$
    - σ_f Πλάτος (κάθετο μέγεθος) της συνάρτησης
    - λ Κλίμακα μήκους (τυπική απόκλιση σε κατανομή Gaussian)

Gaussian Διαδικασίες Διαδραστικά διαγράμματα

- https://distill.pub/2019/visual-exploration-gaussian-processes/
- http://chifeng.scripts.mit.edu/stuff/gp-demo/
- http://www.infinitecuriosity.org/vizgp/
- https://edward-rees.com/gp/
- http://smlbook.org/GP/

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Εφαρμογή των ανελίξεων Gaussian: Αποκατάσταση ζημιών ρομπότ

[Βίντεο στο YouTube]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Επόμενη Διάλεξη

- Νευρωνικά δίκτυα

Σας ευχαριστούμε

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423