MAI4CAREU
Master programmes in Artificial Intelligence 4 Careers in Europe

Πανεπιστήμιο Κύπρου - Τεχνητή Νοημοσύνη

MAI612 - ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διάλεξη 7: Μέθοδοι με βάση τον πυρήνα 1

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

Επανάληψη

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

Αξιολόγηση και βελτίωση μοντέλου

- Θέλουμε τα μοντέλα μας να παρουσιάζουν δυνατότητες γενίκευσης αντί να απομνημονεύουν το σύνολο εκπαίδευσης.
- Γενίκευση: καλή απόδοση σε αόρατα δεδομένα, που προέρχονται από την ίδια κατανομή
- Η μοντελοποίηση απλουστευτικών παραδοχών σχετικά με τη σχέση των δεδομένων με τον στόχο (π.χ. με τη χρήση γραμμικού μοντέλου) μπορεί να οδηγήσει σε υποπροσαρμογή. Σε αυτή την περίπτωση λέμε ότι το μοντέλο πάσχει από υψηλή προκατάληψη.
- Από την άλλη πλευρά, αν το μοντέλο είναι πολύ περίπλοκο, ταιριάζει με το θόρυβο και δεν συλλαμβάνει την τάση στα δεδομένα, οδηγεί σε υπερβολική προσαρμογή. Σε αυτή την περίπτωση, λέμε ότι το μοντέλο πάσχει από μεγάλη διακύμανση.
- Ο συμβιβασμός προκατάληψης-διακύμανσης είναι η σύγκρουση της προσπάθειας ελαχιστοποίησης τόσο της προκατάληψης όσο και της διακύμανσης.
- Η βέλτιστη πολυπλοκότητα ενός μοντέλου είναι στο σημείο όπου το συνολικό σφάλμα (Προκατάληψη^2 + Διακύμανση + ε) είναι στο ελάχιστο
- Πρακτικά, το επιτυγχάνουμε αυτό διαιρώντας το σύνολο δεδομένων σε ένα σύνολο εκπαίδευσης, επικύρωσης και δοκιμών και επιλέγοντας το μοντέλο που έχει το χαμηλότερο σφάλμα επικύρωσης.

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

Αξιολόγηση και βελτίωση μοντέλου

- Επικύρωση k-fold cross:
  - διαιρεί το σύνολο δεδομένων εκμάθησης+επικύρωσης σε υποσύνολα k και εκπαιδεύει k ανεξάρτητα μοντέλα όπου χρησιμοποιείται υποσύνολο i για επικύρωση και το υπόλοιπο ως σύνολο εκπαίδευσης
  - η απόδοση του μοντέλου είναι το μέσο σφάλμα επικύρωσης σε όλα τα υποσύνολα k
  - χρησιμοποιείται όταν το σύνολο δεδομένων είναι μικρό, λόγω της πολυπλοκότητάς του στην εκπαίδευση μοντέλων k και όχι μόνο ένα.
- Οι καμπύλες μάθησης μπορούν να χρησιμοποιηθούν για να ελέγξουμε αν πρέπει να αποκτήσουμε περισσότερα δεδομένα
  - Περισσότερα δεδομένα ωφελούν τυπικά τα μοντέλα υψηλής διακύμανσης, αλλά όχι τα μοντέλα υψηλής προκατάληψης
- Η τακτοποίηση είναι μια ποινή που δίνεται στη λειτουργία απώλειας των μοντέλων υψηλής διακύμανσης για να μειωθεί το μέγεθος των παραμέτρων τους, και ως εκ τούτου, η πολυπλοκότητά τους.
  - Η κανονικοποίηση L1 χρησιμοποιεί τον κανόνα της απόλυτης τιμής και μπορεί μερικές φορές να χρησιμοποιηθεί για την επιλογή χαρακτηριστικών
  - Η κανονικοποίηση L2 χρησιμοποιεί το Euclidean norm και συνήθως παράγει μια καλύτερη εφαρμογή από το L1
- Υπερπαραμετρική ρύθμιση είναι η διαδικασία της διαφοροποίησης των υπερπαραμέτρων των μοντέλων και των αλγορίθμων μάθησης, και επιλέξτε το συνδυασμό που οδηγεί στο χαμηλότερο σφάλμα επικύρωσης.
- Η βελτίωση του μοντέλου μπορεί να επιτευχθεί με τη χρήση συνόλων: κατάρτιση πολλαπλών μοντέλων αντί για ένα και συνάθροιση των αποτελεσμάτων τους
- Βελτιώστε τα μοντέλα υψηλής προκατάληψης μέσω: πολυπλοκοποίηση, σύνολα
- Βελτιώστε τα πρότυπα υψηλής διακύμανσης μέσω: υπερπαραμετρική ρύθμιση, τακτοποίηση, σύνολα, περισσότερα δεδομένα

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

Δέντρα και δάση

- Τα δέντρα αποφάσεων είναι μοντέλα που έχουν μια φυσική αν-τότε-διαφορετικά δομή, έτσι ώστε να είναι ερμηνεύσιμα και γρήγορα.
- Για ένα δεδομένο σύνολο δεδομένων, μπορεί να υπάρχουν πολλά δέντρα αποφάσεων που ταξινομούν τα δεδομένα
- Ένας αλγόριθμος για την εκμάθηση δέντρων αποφάσεων πρέπει να επιλέξει έναν που γενικεύει καλά, αποφασίζοντας το χαρακτηριστικό που θα χρησιμοποιήσει για τον διαχωρισμό σε κάθε κόμβο και πότε να σταματήσει να χωρίζει
- Χρησιμοποιώντας άσχετα χαρακτηριστικά δημιουργεί μεγαλύτερα δέντρα αποφάσεων, έτσι προτιμάται η απλότητα
- Το καλύτερο χαρακτηριστικό που πρέπει να χρησιμοποιήσετε για τη διάσπαση είναι αυτό που είναι πιο ενημερωτικό, δηλαδή, αυτό που ελαχιστοποιεί τη διαταραχή γνωστή και ως εντροπία.
- Η εντροπία H λαμβάνει ως είσοδο το ποσοστό των θετικών παραδειγμάτων p_+, και όπως το p_+ πηγαίνει από 0 σε 0,5, η H αυξάνεται από 0 σε 1, και όπως το p_+ πηγαίνει από 0,5 σε 1, το H μειώνεται από 1 σε 0.
- Το κέρδος πληροφοριών μετρά την αναμενόμενη μείωση της εντροπίας λόγω της διάσπασης σε κάποιο χαρακτηριστικό A
  - Μετράται ως η διαφορά μεταξύ της εντροπίας του αρχικού συνόλου και του σταθμισμένου αθροίσματος των εντροπιών στα κλαδιά
  - Θέλουμε να μεγιστοποιήσουμε το κέρδος πληροφοριών, ή να ελαχιστοποιήσουμε ισοδύναμα το σταθμισμένο άθροισμα
- Για να χωρίσουμε σε μια συνεχή μεταβλητή, υπολογίζουμε πρώτα όλα τα πιθανά μοναδικά κατώφλια (χρησιμοποιώντας τα μέσα των ζευγών των ταξινομημένων σημείων) και επιλέγουμε αυτό με το υψηλότερο κέρδος πληροφοριών.

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

Δέντρα και δάση

- Δέντρα παλινδρόμησης:
    - πρόβλεψη του μέσου όρου των τιμών στους κόμβους των φύλλων τους
    - χρησιμοποιήστε τη διακύμανση αντί της εντροπίας, και τη μείωση διακύμανσης αντί για το κέρδος πληροφοριών
- Μέθοδοι συνόλων:
    - συνήθως έχουν χαμηλότερο σφάλμα γενίκευσης από τα μεμονωμένα μοντέλα
    - βασίζονται σε διαφορετικά μοντέλα που παράγουν διαφορετικά σφάλματα
    - λειτουργία συγκέντρωσης: μέσος όρος παλινδρόμησης, ψηφοφορία κατά πλειοψηφία για ταξινόμηση
- Το Bagging εκπαιδεύει μοντέλα παράλληλα, μεταβάλλοντας τα δεδομένα εκπαίδευσής τους με τη χρήση δειγματοληψίας με αντικατάσταση
- Τα τυχαία δάση χρησιμοποιούν bagging με το πρόσθετο βήμα της τυχαιοποίησης της επιλογής χαρακτηριστικών
- Σταδιακή ενίσχυση των μοντέλων εκμάθησης, εστιάζοντας σε προηγούμενα λανθασμένα παραδείγματα
- Το Stacking είναι μια μέθοδος που εκπαιδεύει μοντέλα τυπικά σε 2 επίπεδα, όπου οι προβλέψεις των μοντέλων στο επίπεδο 1 γίνονται δεδομένα κατάρτισης για ένα μοντέλο στο επίπεδο 2 το οποίο μαθαίνει πώς να τα συνδυάσει.

Διάλεξη 7: Μέθοδοι με βάση τον πυρήνα 1

Μαθησιακά αποτελέσματα

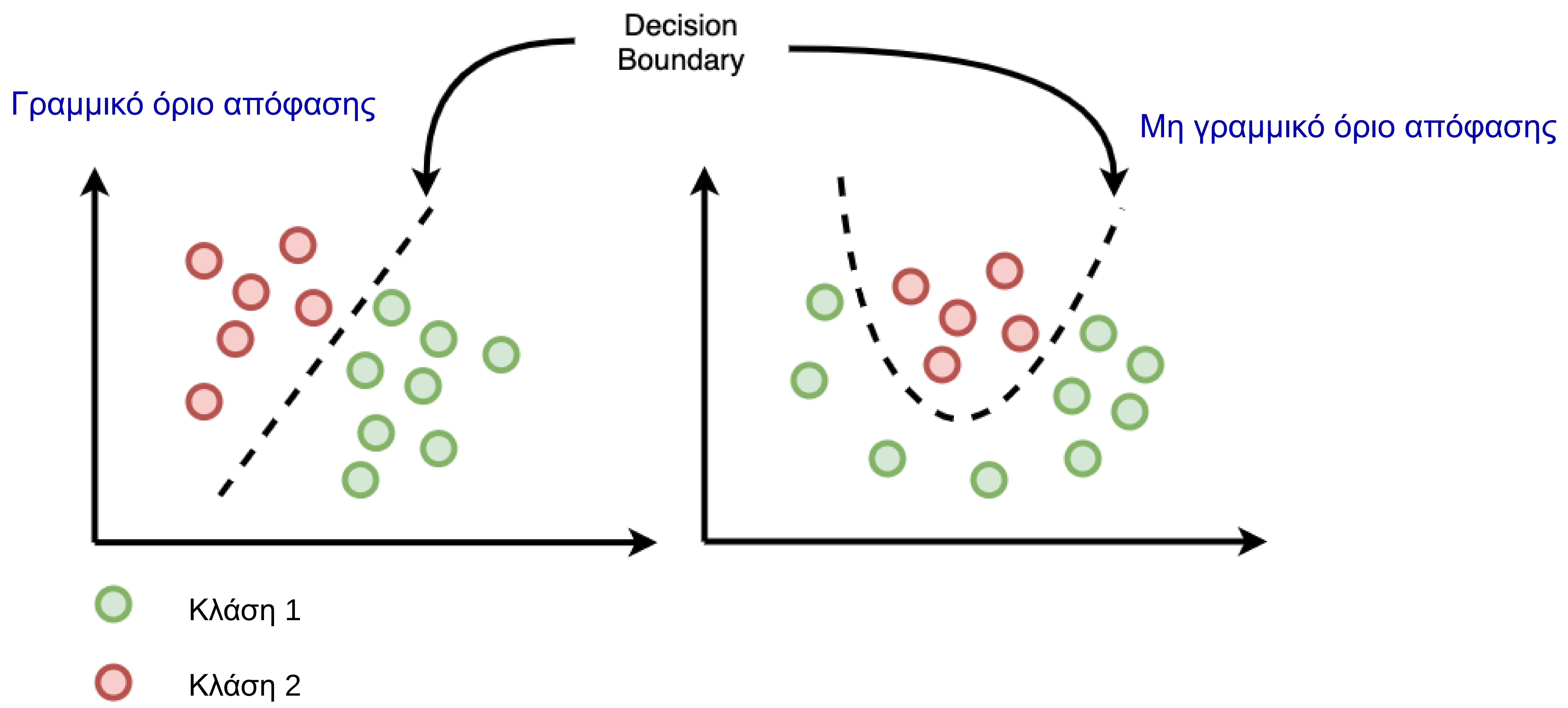
Θα καταλάβετε:

1. Η έννοια του γραμμικού και μη γραμμικού ορίου απόφασης.
2. Ποιες είναι οι μέθοδοι πυρήνων και το τέχνασμα του πυρήνα.
3. Διαφορετικοί πυρήνες και πώς να κατασκευάσετε έγκυρους πυρήνες.
4. Ο αλγόριθμος του Support Vector Machine (SVM).
5. Η έννοια του ταξινομητή μέγιστου περιθωρίου.
6. Η αντικειμενική λειτουργία των SVM και ο τρόπος με τον οποίο σχετίζεται με το logistic regression.
7. Πώς να δημιουργήσετε μη γραμμικά SVMs χρησιμοποιώντας πυρήνες.
8. Τα βασικά στοιχεία της διανυσματικής παλινδρόμησης υποστήριξης.

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

Τα όρια της απόφασης



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

Μαθαίνοντας μη γραμμικά όρια αποφάσεων

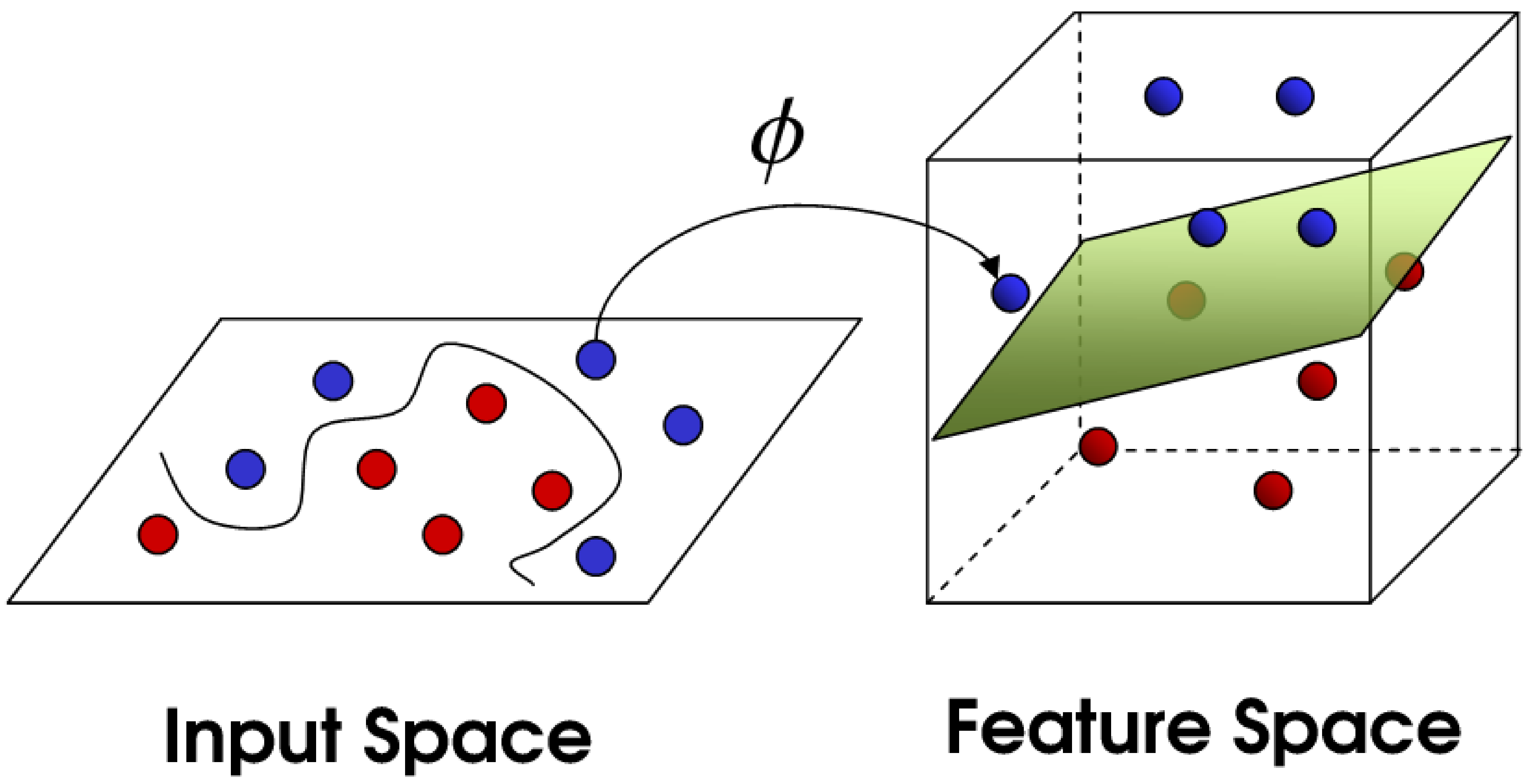
- Μπορούμε να χρησιμοποιήσουμε πολυωνυμικά χαρακτηριστικά για να μετατρέψουμε τα ανεπεξέργαστα δεδομένα σε διανύσματα που θα μας επιτρέψουν να μάθουμε μη γραμμικά όρια αποφάσεων.
- Για παράδειγμα:
  - 2 μεταβλητές (x,y), πολυωνυμικός βαθμός = 2: x, y, x^2, xy, y^2
  - 3 μεταβλητές (x,y,z), πολυωνυμικός βαθμός = 2: x, y, z, x^2, xy, xz, y^2, yz, z^2

- **Παρατήρηση**: Κάθε νέο χαρακτηριστικό αυξάνει τις διαστάσεις της εισόδου από κάποιον που θα μπορούσε να κάνει το πρόβλημα «ευκολότερο» να λυθεί στο νέο χώρο χαρακτηριστικών.

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

Γραμμική δυνατότητα διαχωρισμού σε υψηλότερες διαστάσεις



ΠΗΓΗ

Η προσθήκη μιας νέας διάστασης χαρακτηριστικών μπορεί να είναι αρκετή για να χωρίσει γραμμικά τις δύο κατηγορίες.

Το θεώρημα του εξωφύλλου για τη δυνατότητα διαχωρισμού των προτύπων:

«ένα σύνθετο πρόβλημα ταξινόμησης που ρίχνεται σε χώρο υψηλής διάστασης μη γραμμικά είναι πιο πιθανό να είναι *γραμμικά διαχωρίσιμο* από ό, τι σε ένα χώρο χαμηλής διάστασης» (Cover, 1965).

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

Μαθαίνοντας μη γραμμικά όρια αποφάσεων

- Μπορούμε να χρησιμοποιήσουμε πολυωνυμικά χαρακτηριστικά για να μετατρέψουμε τα ανεπεξέργαστα δεδομένα σε διανύσματα που θα μας επιτρέψουν να μάθουμε μη γραμμικά όρια αποφάσεων.
- Για παράδειγμα:
  - 2 μεταβλητές (x,y), πολυωνυμικός βαθμός = 2: x, y, x^2, xy, y^2
  - 3 μεταβλητές (x,y,z), πολυωνυμικός βαθμός = 2: $x, y, z, x^2, xy, xz, y^2, yz, z^2$

- Ωστόσο, έχουμε μια **συνδυαστική έκρηξη** στον αριθμό των παραμέτρων που πρέπει να μάθουμε.
- **Υπάρχει πιο αποτελεσματικός τρόπος για να μάθετε τα όρια των μη γραμμικών 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

Μέθοδοι πυρήνων

- Σε πολλούς αλγορίθμους ML, τα ανεπεξέργαστα δεδομένα πρέπει να μετατραπούν ρητά σε διανύσματα χαρακτηριστικών μέσω ενός **χάρτη** χαρακτηριστικών που προσδιορίζεται από τον χρήστη.
- Αντίθετα, οι μέθοδοι πυρήνα απαιτούν μόνο έναν πυρήνα που προσδιορίζεται **από** τον χρήστη, δηλαδή μια **συνάρτηση ομοιότητας μεταξύ ζευγών ακατέργαστων σημείων δεδομένων**.
- Αυτό τους επιτρέπει να λειτουργούν σε ένα υψηλής διάστασης, **υπονοούμενο χώρο χαρακτηριστικών** χωρίς να υπολογίζουν τις συντεταγμένες των δεδομένων σε αυτόν τον χώρο.
- Υπολογίζουν τα εσωτερικά προϊόντα μεταξύ των εικόνων όλων των ζευγών δεδομένων κατάρτισης στο χώρο χαρακτηριστικών: συχνά υπολογιστικά φθηνότερα από τον ρητό υπολογισμό των συντεταγμένων.
- Αυτό είναι γνωστό ως το «**κόλπο του πυρήνα**».

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

Το κόλπο του πυρήνα

- Επιτρέπει να λειτουργεί στον αρχικό χώρο χαρακτηριστικών χωρίς να υπολογίζει τις συντεταγμένες των δεδομένων στον υψηλότερο διαστατικό χώρο
- Για παράδειγμα: 2D διάστημα εισαγωγής, 3D διάστημα χαρακτηριστικών γνωρισμάτων

$$\begin{aligned} k(\mathbf{x}, \mathbf{z}) &= \left(\mathbf{x}^{\mathrm{T}}\mathbf{z}\right)^2 = (x_1 z_1 + x_2 z_2)^2 \\ &= x_1^2 z_1^2 + 2x_1 z_1 x_2 z_2 + x_2^2 z_2^2 \\ &= (x_1^2, \sqrt{2}x_1 x_2, x_2^2)(z_1^2, \sqrt{2}z_1 z_2, z_2^2)^{\mathrm{T}} \\ &= \phi(\mathbf{x})^{\mathrm{T}}\phi(\mathbf{z}). \end{aligned}$$

$$\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1 x_2, x_2^2)^T$$

$$\phi(\mathbf{z}) = (z_1^2, \sqrt{2}z_1 z_2, z_2^2)^T$$

Ο υπολογισμός των $\overbrace{\phi(\mathbf{x}) \;\; \phi(\mathbf{z}) \text{ και } \phi(\mathbf{x})^T \phi(\mathbf{z})}^{O(n^2)}$

είναι πιο ακριβά υπολογιστικά από

τον υπολογισμό του $\underbrace{(\mathbf{x}^T \mathbf{z})^2}_{O(n)}$

Πυρήνες

- Πολυωνυμικός πυρήνας

$$k(\boldsymbol{x}, \boldsymbol{z}) = (\boldsymbol{x}^T \boldsymbol{z} + c)^d$$

- Πυρήνας Gaussian ή Radial Basis Function (RBF)

$$k(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{z}\|^2}{2\sigma^2}\right)$$

$$k(\boldsymbol{x}, \boldsymbol{z}) = \exp(-\gamma\|\boldsymbol{x} - \boldsymbol{z}\|^2)$$

που $\gamma = \frac{1}{2\sigma^2}$

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

Φτιάχνοντας νέους πυρήνες

- Μπορούμε να φτιάξουμε νέους πυρήνες από έγκυρους πυρήνες με επιτρεπόμενες λειτουργίες
- Για παράδειγμα, η προσθήκη, ο πολλαπλασιασμός, η ανακλιμάκωση των πυρήνων δίνει ένα σωστό πυρήνα εφ' όσον ο πίνακας Gram που προκύπτει είναι συμμετρικός, θετικός ημιοριστικός

$$k(\boldsymbol{x},\boldsymbol{z}) = k_1(\boldsymbol{x},\boldsymbol{z}) + k_2(\boldsymbol{x},\boldsymbol{z})$$

$$k(\boldsymbol{x},\boldsymbol{z}) = k_1(\boldsymbol{x},\boldsymbol{z})k_2(\boldsymbol{x},\boldsymbol{z})$$

$$k(\boldsymbol{x},\boldsymbol{z}) = \lambda\, k_1(\boldsymbol{x},\boldsymbol{z}) \qquad , \qquad \lambda > 0$$

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

Κουίζ

Δεδομένου ότι k_1 και k_2 είναι κατάλληλοι πυρήνες, αποφασίστε ποιος από τους ακόλουθους τύπους ορίζει τους κατάλληλους πυρήνες:

$$k_3(\boldsymbol{x},\boldsymbol{z}) = 5k_1(\boldsymbol{x},\boldsymbol{z}) + 3k_2(\boldsymbol{x},\boldsymbol{z})$$

Κατάλληλος

$$k_4(\boldsymbol{x},\boldsymbol{z}) = -k_1(\boldsymbol{x},\boldsymbol{z})k_2(\boldsymbol{x},\boldsymbol{z})$$

Μη κατάλληλος

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 πυρήνα (ή μηχανές πυρήνα)

- Πρόκειται για **μοντέλα που βασίζονται σε** περιπτώσεις:
  - μαθαίνουν ένα βάρος, w_i για κάθε παράδειγμα προπόνησης $(\boldsymbol{x}_i, y_i)$
  - η πρόβλεψη για εισόδους χωρίς ετικέτες γίνεται με την εφαρμογή της συνάρτησης ομοιότητας k (δηλαδή, ο πυρήνας) μεταξύ της εισόδου χωρίς ετικέτα $\boldsymbol{z}$ και καθεμίας από τις εισόδους εκπαίδευσης $\boldsymbol{x}_i$
  - για παράδειγμα, ένας δυαδικός ταξινομητής πυρήνα υπολογίζει συνήθως ένα σταθμισμένο άθροισμα ομοιοτήτων:

$$\hat{y} = \operatorname{sgn} \sum_{i=1}^{m} w_i y_i k(\boldsymbol{x}_i, \boldsymbol{z})$$

όπου $\hat{y} \in \{-1, +1\}$, $k: X \times X \to \mathbb{R}$, $\operatorname{sgn} x := \begin{cases} -1 & if\ x < 0, \\ +1 & if\ x \geq 0. \end{cases}$

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

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Support Vector Machines

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

Support Vector Machines

- Ισχυρά εποπτευόμενα μοντέλα μάθησης για την ταξινόμηση

- Δύο βασικές ιδέες
  - Υποθέτοντας γραμμικά διαχωριζόμενες τάξεις, μάθετε να διαχωρίζετε το υπερπλάνο με το μέγιστο περιθώριο
    - Μέγιστο περιθώριο: ελάχιστο σφάλμα γενίκευσης
  - Χρησιμοποιήστε πυρήνες για να επεκτείνετε την είσοδο σε χώρο υψηλής διάστασης για την αντιμετώπιση γραμμικά μη διαχωρίσιμων περιπτώσεων

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

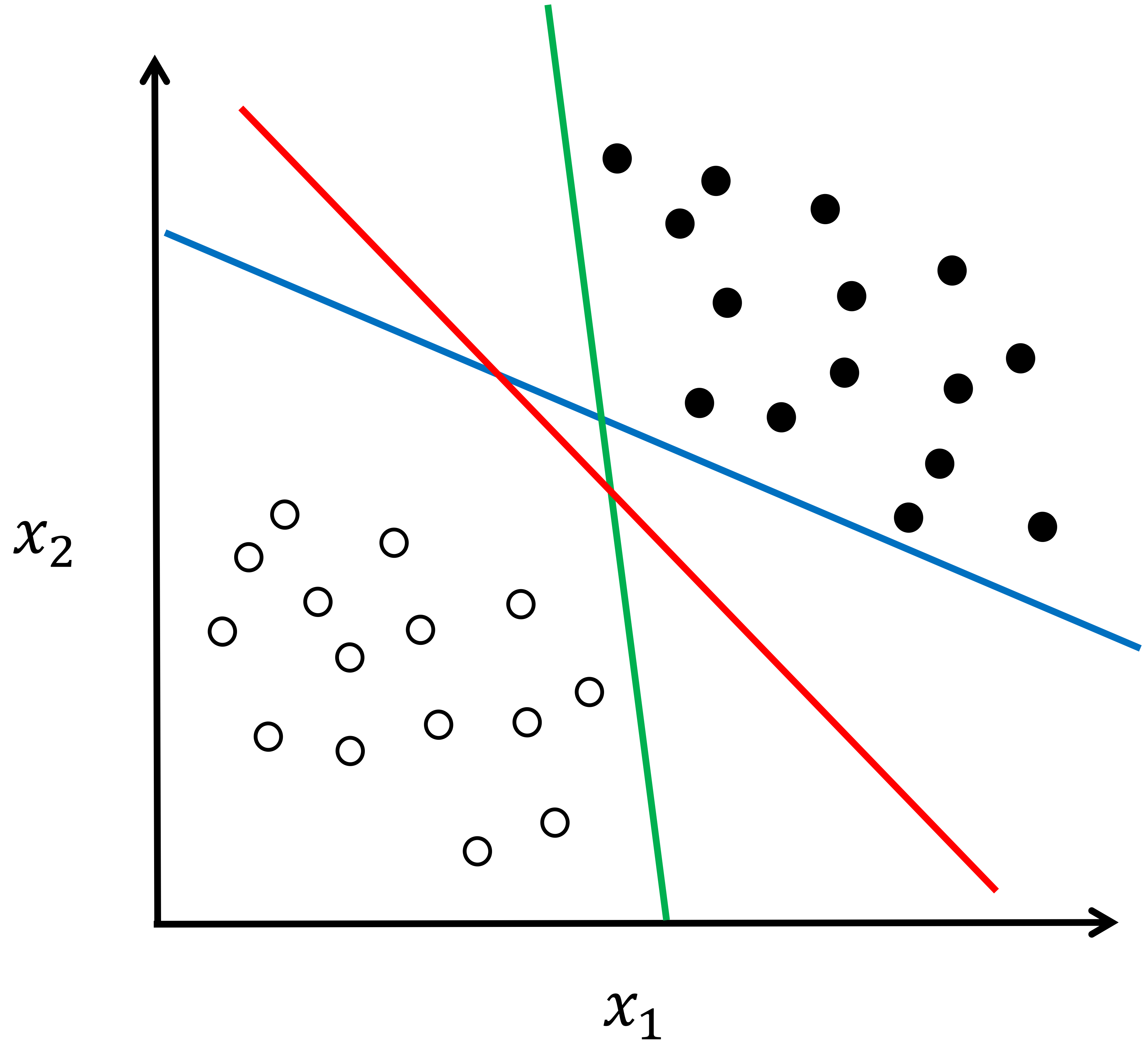
Διαχωρισμός υπερπλάνων

- Δεδομένα εκμάθησης: $\left(\boldsymbol{x}^{(i)}, y^{(i)}\right)$,

$i = 1,2, \dots, m;\ y^{(i)} \in \{+1, -1\}$

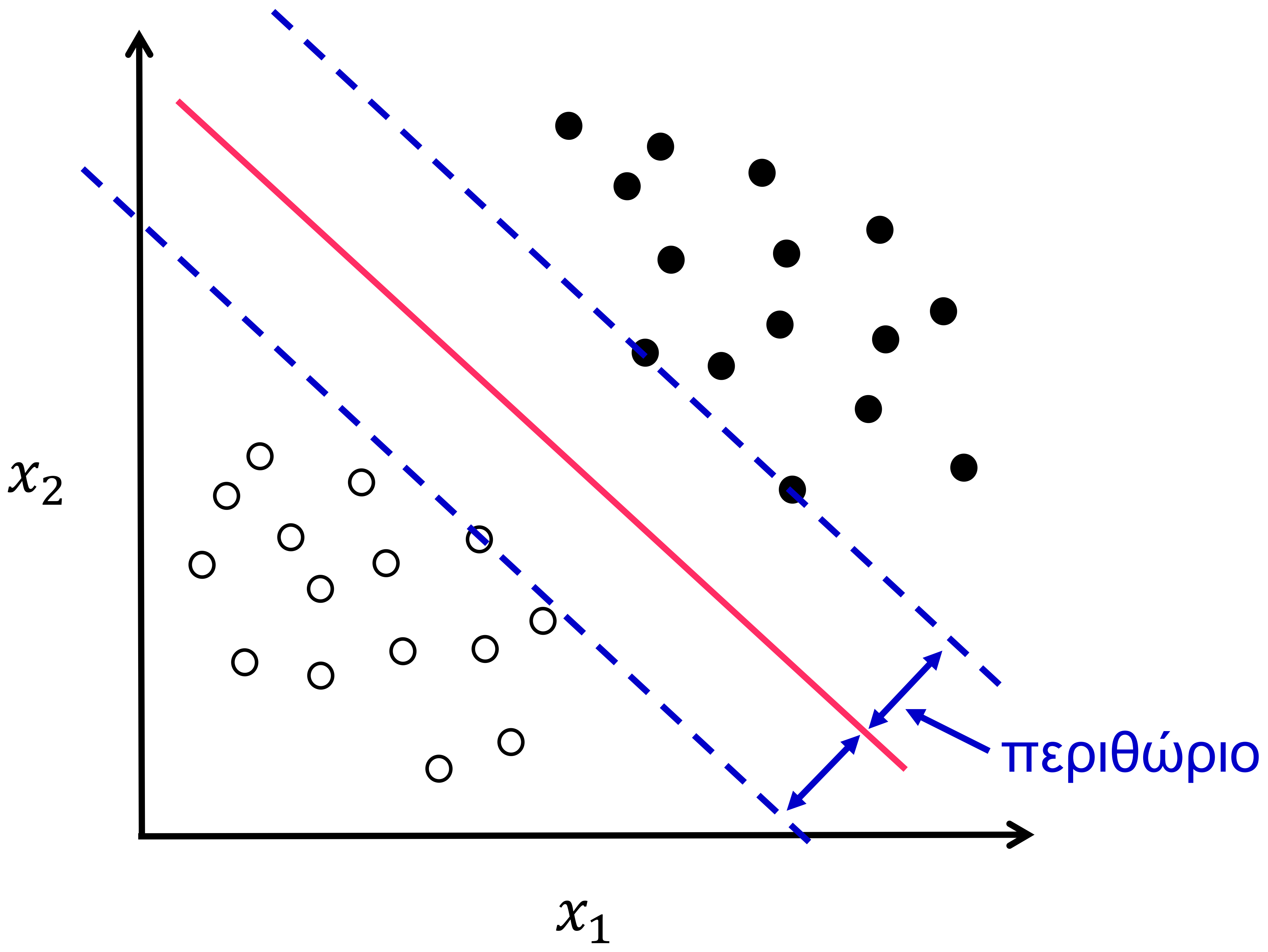
- Υπερπλάνο: $\boldsymbol{\theta}^T \boldsymbol{x} = 0$

$[\theta_0, \theta_1, \cdots, \theta_n]^T$ $\qquad$ $[1, x_1, \cdots, x_n]^T$



Ποια γραμμή θα επέλεγες;

Μέγιστο περιθώριο



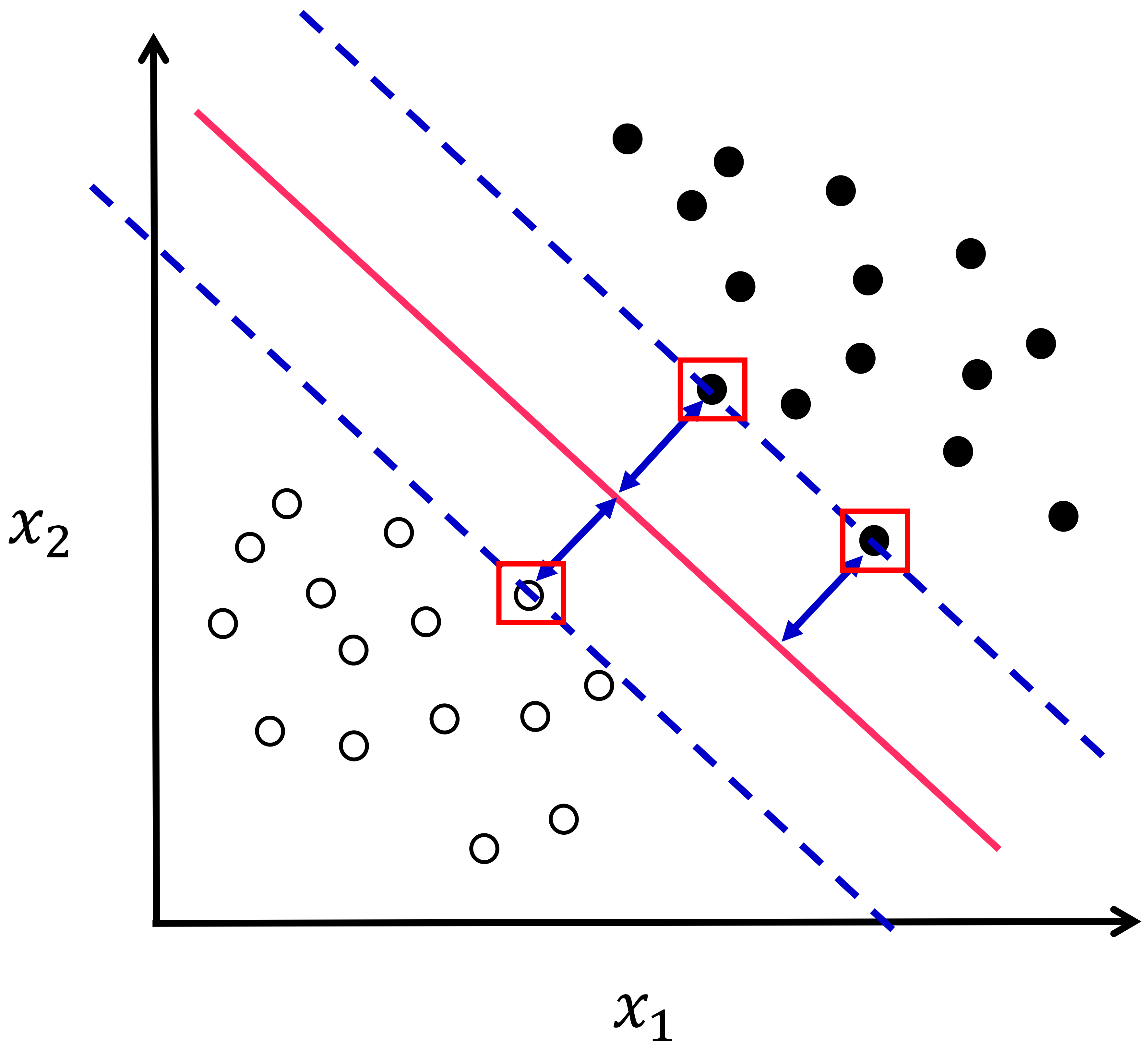
- Πρέπει να είμαστε σε θέση να διαχωρίσουμε σωστά τα νέα δεδομένα

- Σύμφωνα με το θεώρημα της Θεωρίας της Μάθησης, από όλα τα πιθανά γραμμικά όρια απόφασης, αυτό που μεγιστοποιεί το περιθώριο του συνόλου εκμάθησης θα ελαχιστοποιήσει το σφάλμα γενίκευσης.

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

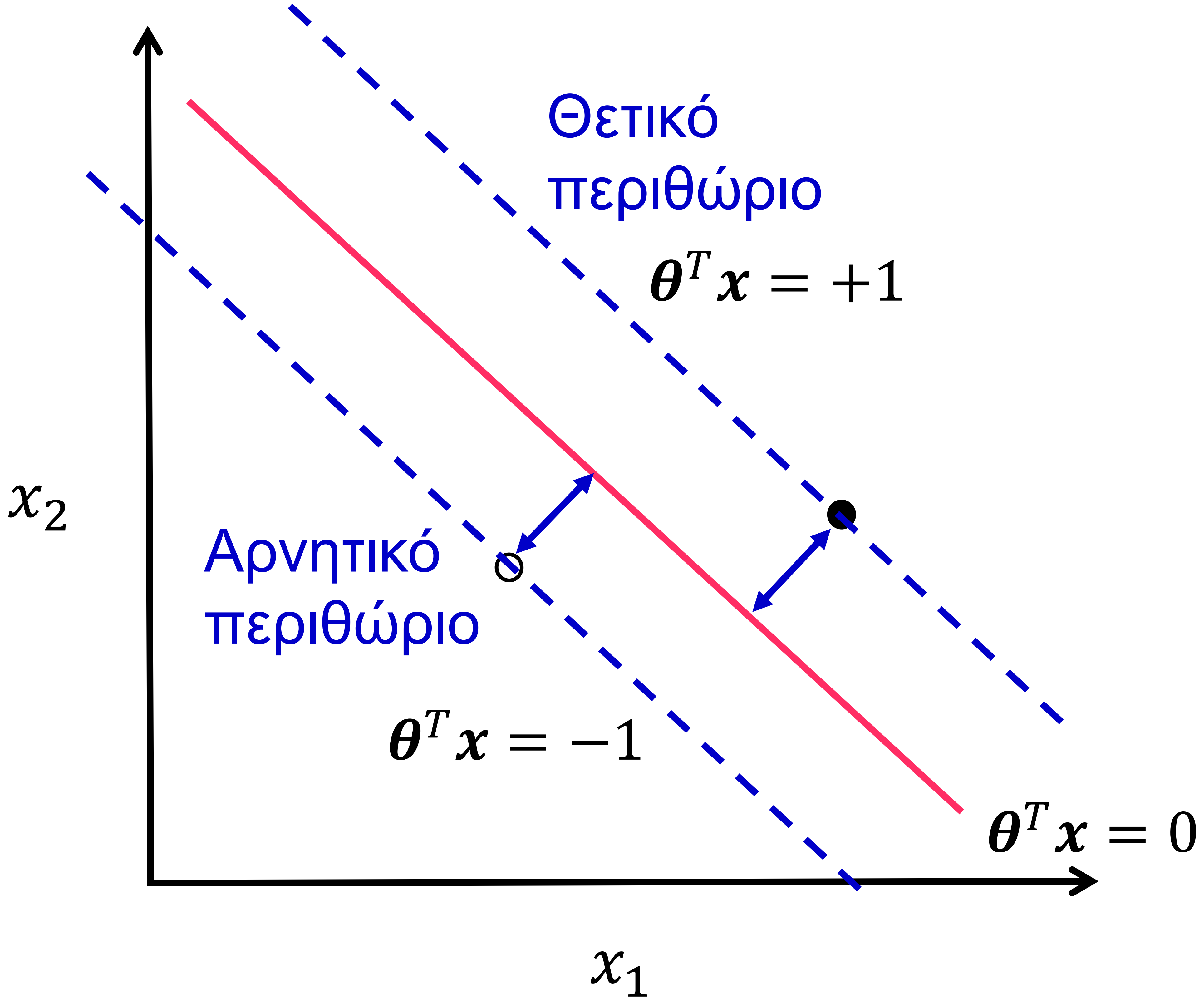
Υποστήριξη διανυσμάτων



- Τα σημεία εκπαίδευσης που βρίσκονται πλησιέστερα στο διαχωριζόμενο υπερεπίπεδο ονομάζονται φορείς υποστήριξης.

- Αυτά είναι τα πιο σημαντικά σημεία, καθώς καθορίζουν ποιο υπερεπίπεδο θα επιλεγεί με τη διαδικασία βελτιστοποίησης.

Μέγιστο περιθώριο



Αν $y = 1$ θέλουμε $\boldsymbol{\theta}^T\boldsymbol{x} \geq 1$) όχι μόνο$\geq 0$)

Αν $y = -1$ θέλουμε $\boldsymbol{\theta}^T\boldsymbol{x} \leq -1$) όχι μόνο$< 0$)

Ορίζουμε το περιθώριο σε μια σταθερή τιμή και θέλουμε να περιορίσουμε τα σημεία δεδομένων να βρίσκονται πέρα από αυτό το περιθώριο.

Στόχος βελτιστοποίησης

$$\min_{\theta} \; C \sum_{i=1}^{m} loss\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)}, y^{(i)}\right) + \frac{1}{2} \| \boldsymbol{\theta} \|_2^2$$

που:

$$loss\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)}, y^{(i)}\right) := \max\left(0, 1 - y^{(i)} \boldsymbol{\theta}^T \boldsymbol{x}^{(i)}\right)$$

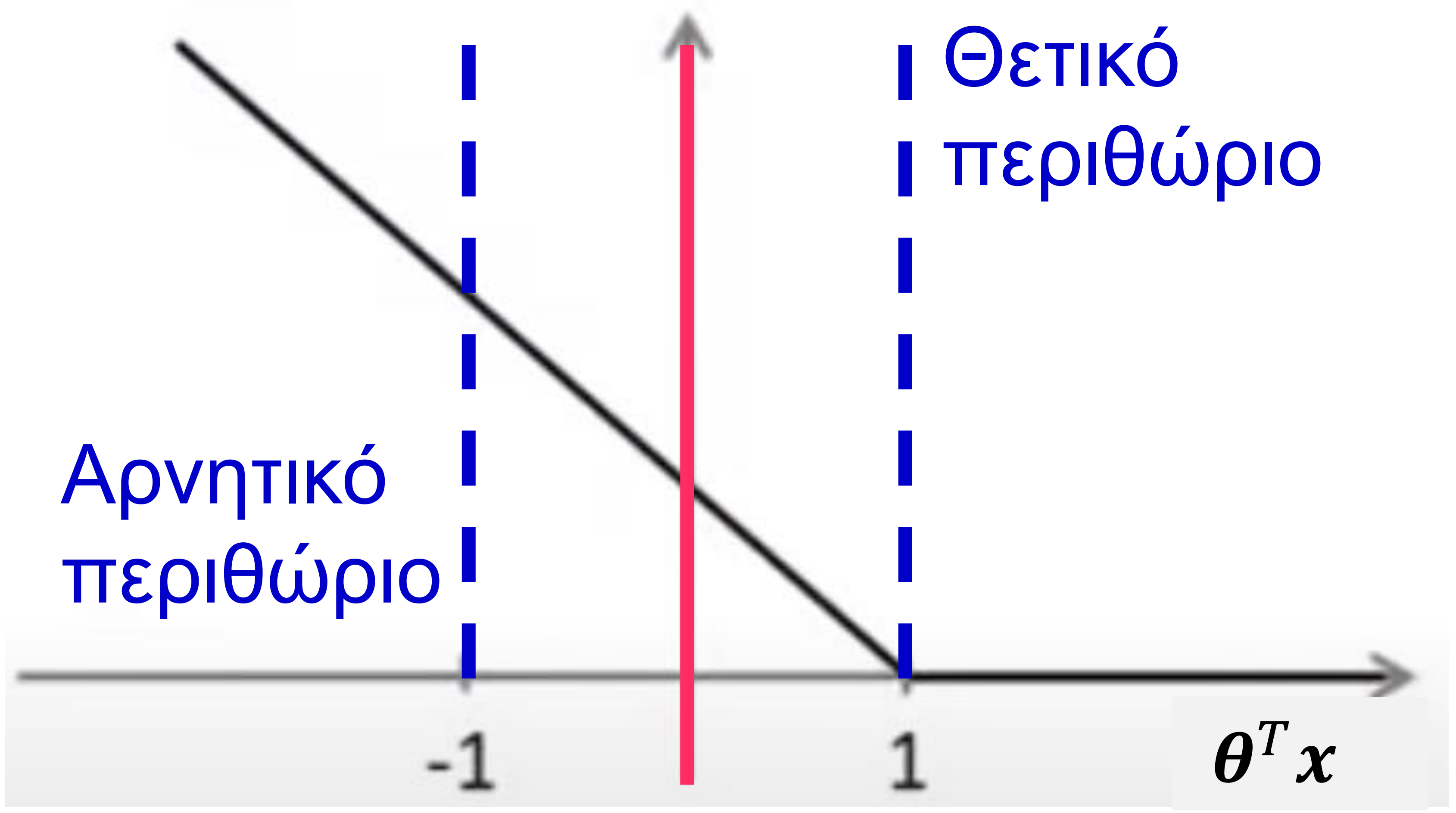
Hinge Loss
(μη διαφοροποιήσιμο)

$$f_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) = \operatorname{sgn}\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)}\right) = \begin{cases} +1 & if \quad \boldsymbol{\theta}^T \boldsymbol{x}^{(i)} \geq 0 \\ -1 & otherwise \end{cases}$$

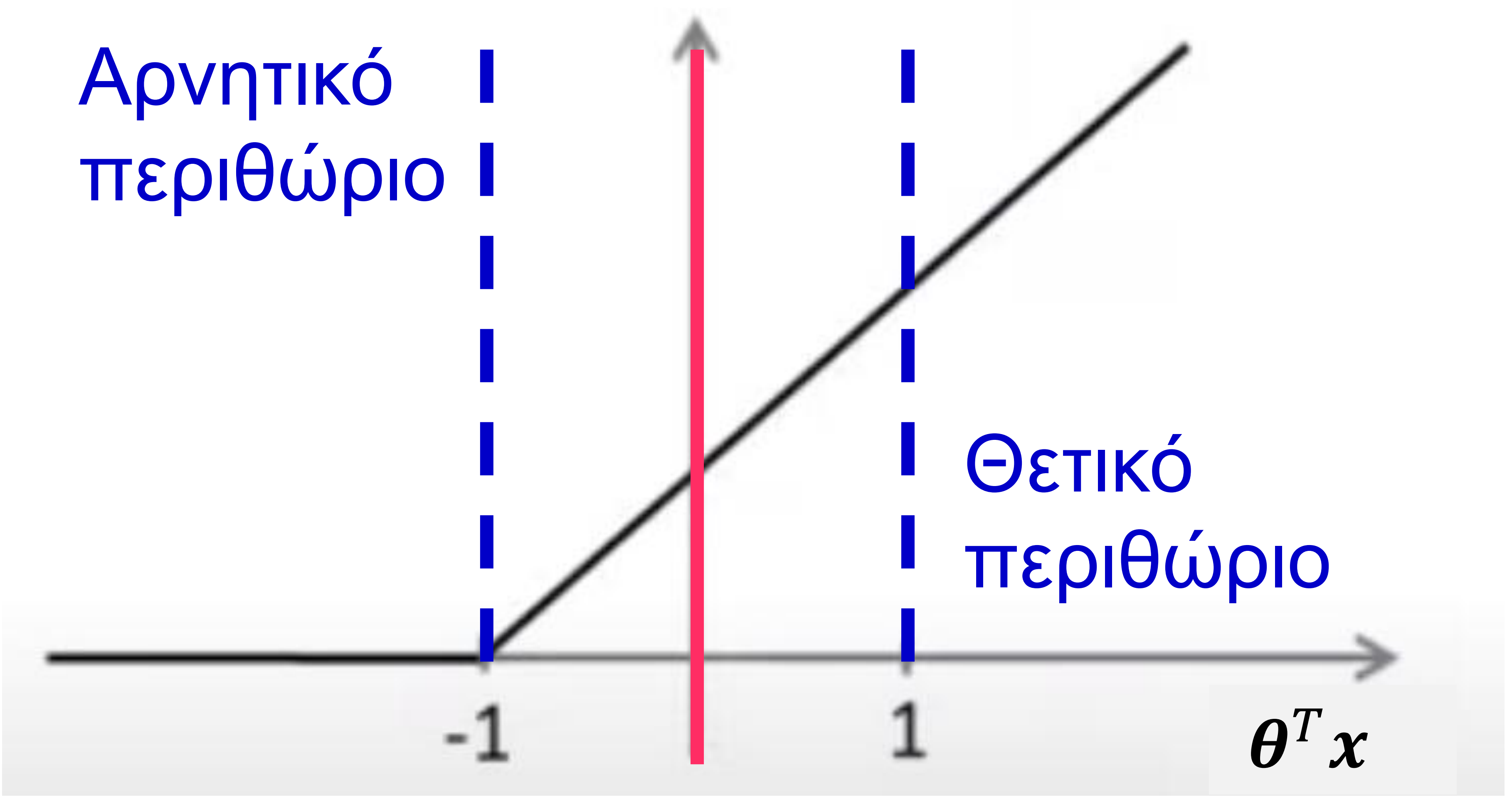
Hinge Loss

$$\max(0, 1 - y\boldsymbol{\theta}^T\boldsymbol{x})$$

Εάν $y = 1$



Εάν $y = -1$



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

Στόχος βελτιστοποίησης σε σύγκριση με το Logistic Regression

Regularized Logistic Regression

$$\min_{\theta} \underbrace{\frac{1}{m}\sum_{i=1}^{m} loss_{LR}\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)}, y^{(i)}\right)}_{A} + \underbrace{\frac{\lambda}{2m} || \boldsymbol{\theta} ||_2^2}_{B}$$

Support Vector Machine

$$\min_{\theta} C\sum_{i=1}^{m} loss_{SVM}\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)}, y^{(i)}\right) + \frac{1}{2} || \boldsymbol{\theta} ||_2^2$$

$$A + \lambda B$$

$$CA + B$$

$$C = \frac{1}{\lambda}$$

έτσι, ένα μικρό C αντιστοιχεί στο να δώσει B ένα μεγαλύτερο βάρος

Το μεγάλο γ: Χαμηλότερη προκατάληψη, υψηλή διακύμανση

Το μικρό γ: Υψηλότερη προκατάληψη, χαμηλή διακύμανση

Στόχος βελτιστοποίησης

$$\min_{\theta} \frac{1}{2} || \boldsymbol{\theta} ||_2^2$$

υ Σ.Τ.

$$y^{(i)}\boldsymbol{\theta}^T\boldsymbol{x}^{(i)} - 1 \geq 0, \qquad i = 1, \dots, m$$

Αυτό είναι ένα **τετραγωνικό πρόβλημα προγραμματισμού** με γραμμικούς περιορισμούς ανισότητας

Υπάρχουν γνωστές διαδικασίες για την επίλυσή της, όπου τα δεδομένα εισάγονται μόνο με τη μορφή προϊόντων κουκίδων.

Η βέλτιστη παράμετρος διάνυσμα μπορεί να γραφτεί με την ακόλουθη μορφή:

$$\boldsymbol{\theta}^* = \sum_{i=1}^{m} a_i^* y^{(i)} \boldsymbol{x}^{(i)}$$

πού $a_i^* \geq 0$ είναι οι συντελεστές Langrangian και δεν είναι μηδενικοί μόνο για τους φορείς υποστήριξης.

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

Ταξινόμηση νέων σημείων δεδομένων

$$f_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(new)}\right) = \operatorname{sgn}\left(\boldsymbol{\theta}^T \boldsymbol{x}^{(new)}\right) = \begin{cases} +1 & if \quad \boldsymbol{\theta}^T \boldsymbol{x}^{(new)} \geq 0 \\ -1 & otherwise \end{cases}$$

$$\boldsymbol{\theta}^T \boldsymbol{x}^{(new)} = \sum_{i=1}^{m} a_i^* y^{(i)} \boldsymbol{x}^{(i)} \boldsymbol{x}^{(new)} = \sum_{i \in SV} a_i^* y^{(i)} \boxed{\boldsymbol{x}^{(i)} \boldsymbol{x}^{(new)}}$$

Μόλις βρεθούν οι παράμετροι με την επίλυση του απαιτούμενου προβλήματος QP στο σύνολο εκπαίδευσης, το SVM μπορεί να χρησιμοποιηθεί για την ταξινόμηση νέων σημείων χρησιμοποιώντας **μόνο τα διανύσματα υποστήριξης**, την ετικέτα τους και τους σχετικούς συντελεστές.

Τα δεδομένα εισάγονται με τη μορφή εσωτερικών γινομένων

Τα SVMs λειτουργούν καλά για γραμμικά διαχωριζόμενα δεδομένα.

Τι γίνεται αν τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα;

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

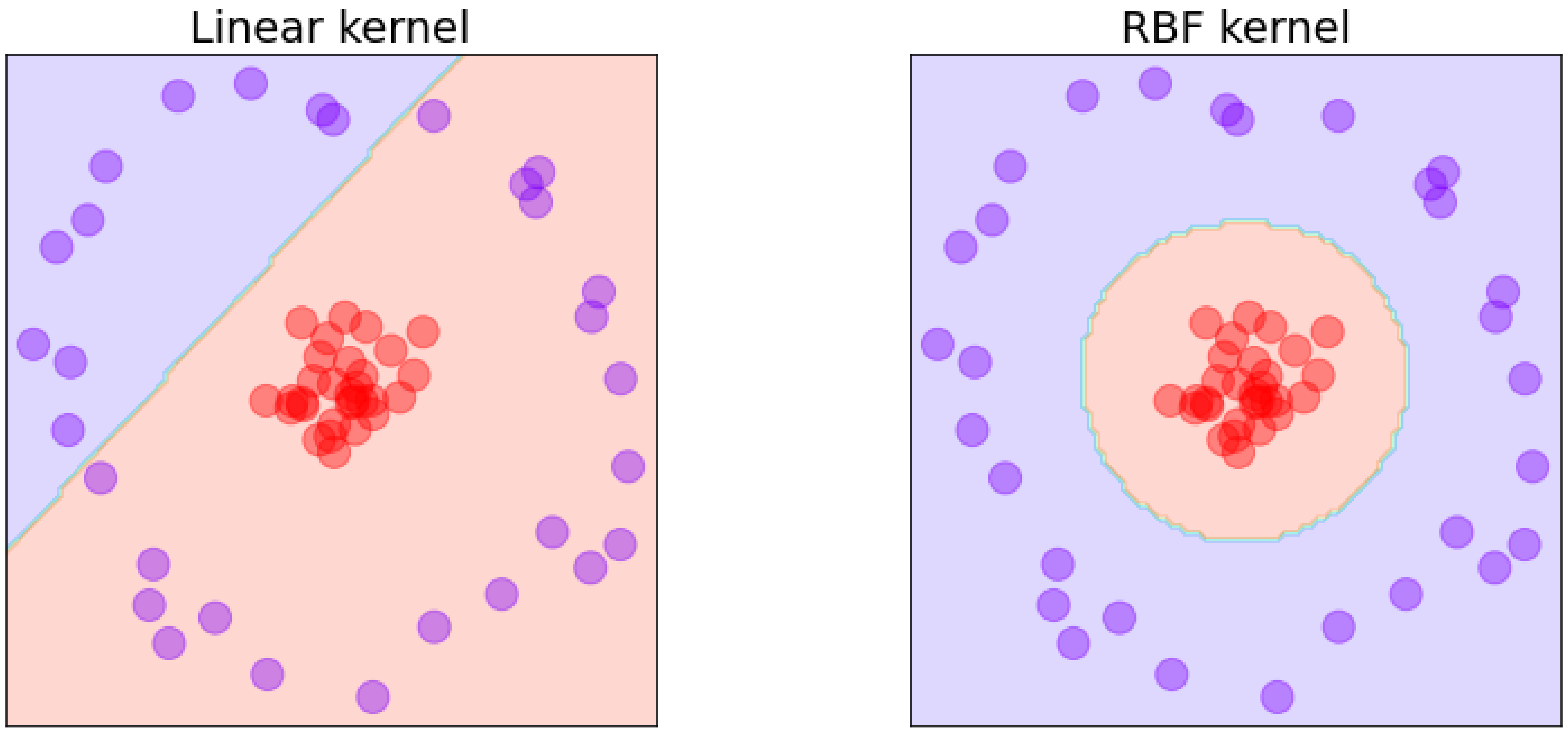
Μη γραμμικά SVMs

- Η λύση του SVM έχει την ωραία ιδιότητα που τα δεδομένα εισάγουν μόνο με τη μορφή εσωτερικών γινομένων.
- Αυτό μας επιτρέπει να κάνουμε τα SVMs μη γραμμικά χωρίς να περιπλέκουμε τον αλγόριθμο, χρησιμοποιώντας πυρήνες.
- Το γραμμικό SVM εξαρτάται από $\boldsymbol{x}^{(i)}\boldsymbol{x}^{(j)}$
- Το μη γραμμικό SVM αντικαθιστά $\boldsymbol{x}^{(i)}\boldsymbol{x}^{(j)}$ με $k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$

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

Μη γραμμικά SVMs

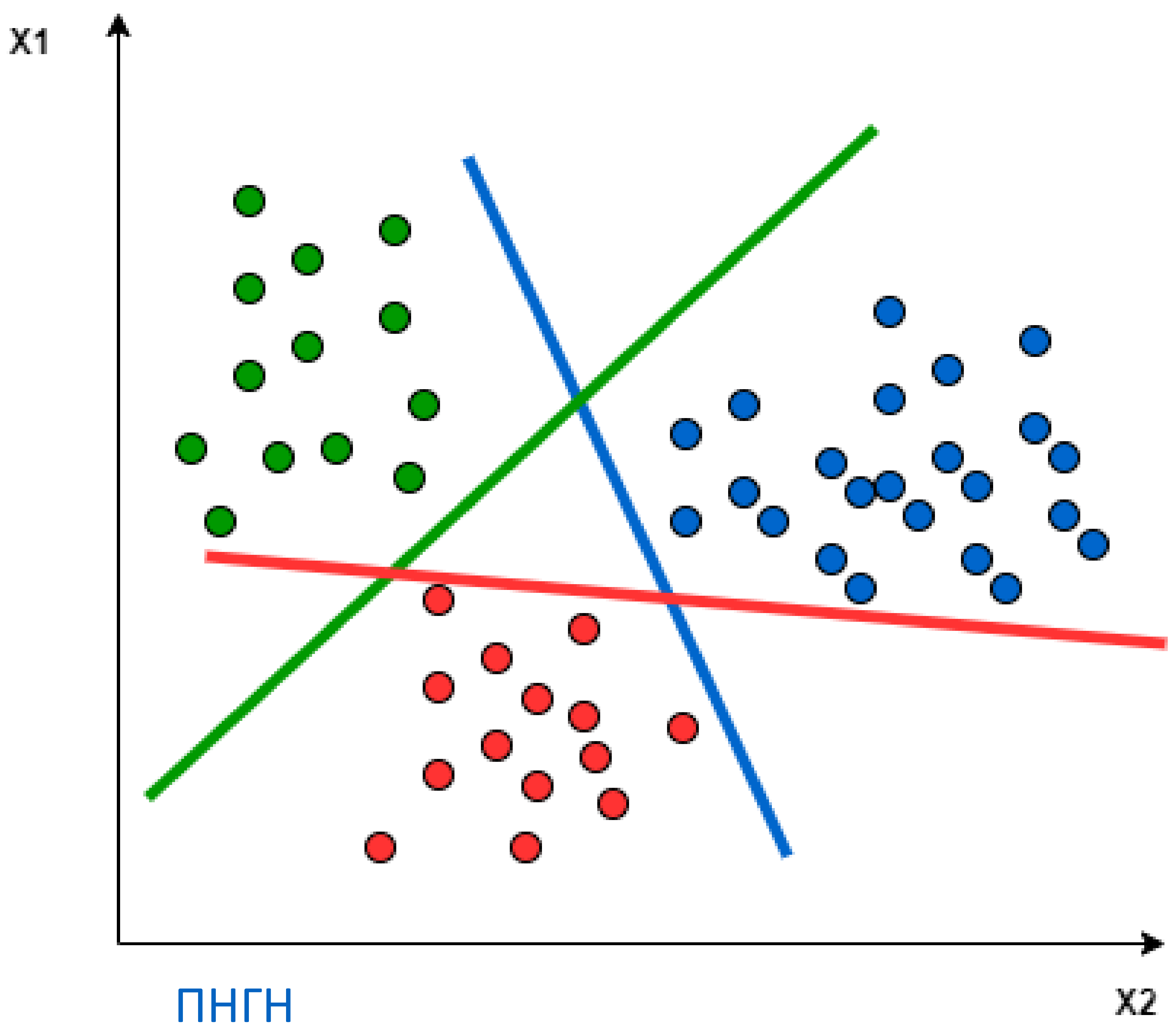


ΠΗΓΗ

SVM με πολυωνυμική απεικόνιση πυρήνων

[Βίντεο στο YouTube]

Ταξινόμηση πολλαπλών κλάσεων

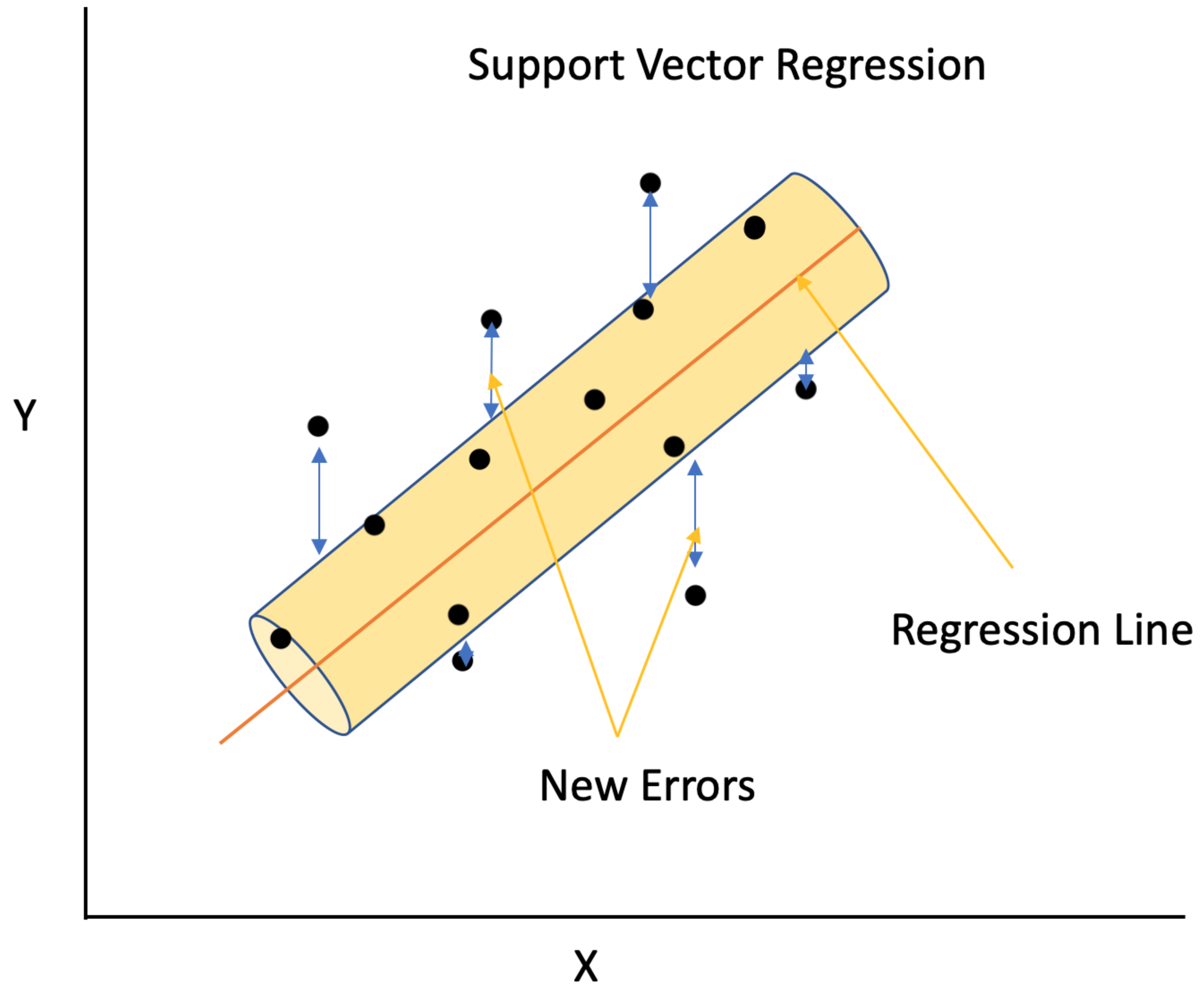


ΠΗΓΗ

- Πολλά πακέτα SVM έχουν ενσωματωμένη λειτουργικότητα για ταξινόμηση πολλαπλών κλάσεων
- Χρησιμοποιήστε τη μέθοδο one-vs-all
  - Εκμάθηση k SVM (k=αριθμός κλάσεων)
  - Πάρτε τα διανύσματα παραμέτρων $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}$
  - Διαλέξτε τη κλάση με το μεγαλύτερο $\left(\boldsymbol{\theta}^{(i)}\right)^T \boldsymbol{x}$

Διανυσματική παλινδρόμηση υποστήριξης

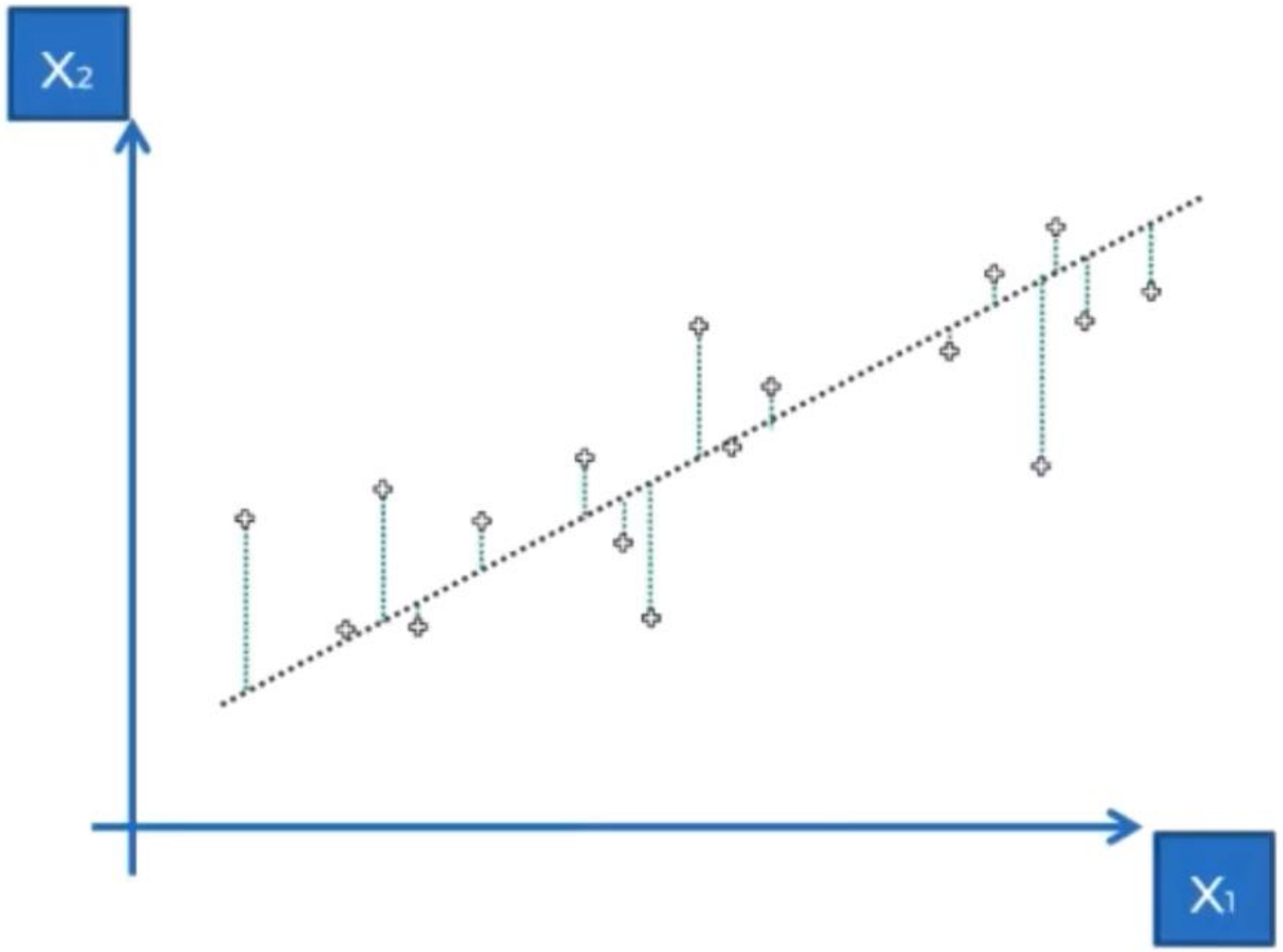
- Επέκταση του SVM σε προβλήματα παλινδρόμησης
- ε-insensitive tube: περιθώριο λάθους που επιτρέπουμε στο μοντέλο μας να έχει
  - Μην νοιάζεστε για το σφάλμα στο εσωτερικό του σωλήνα
  - Αγνοήστε δείγματα των οποίων η πρόβλεψη είναι κοντά στο στόχο τους
- Χαλαρές μεταβλητές: σημεία έξω από το σωλήνα
  - Νοιάζονται μόνο για τα λάθη τους



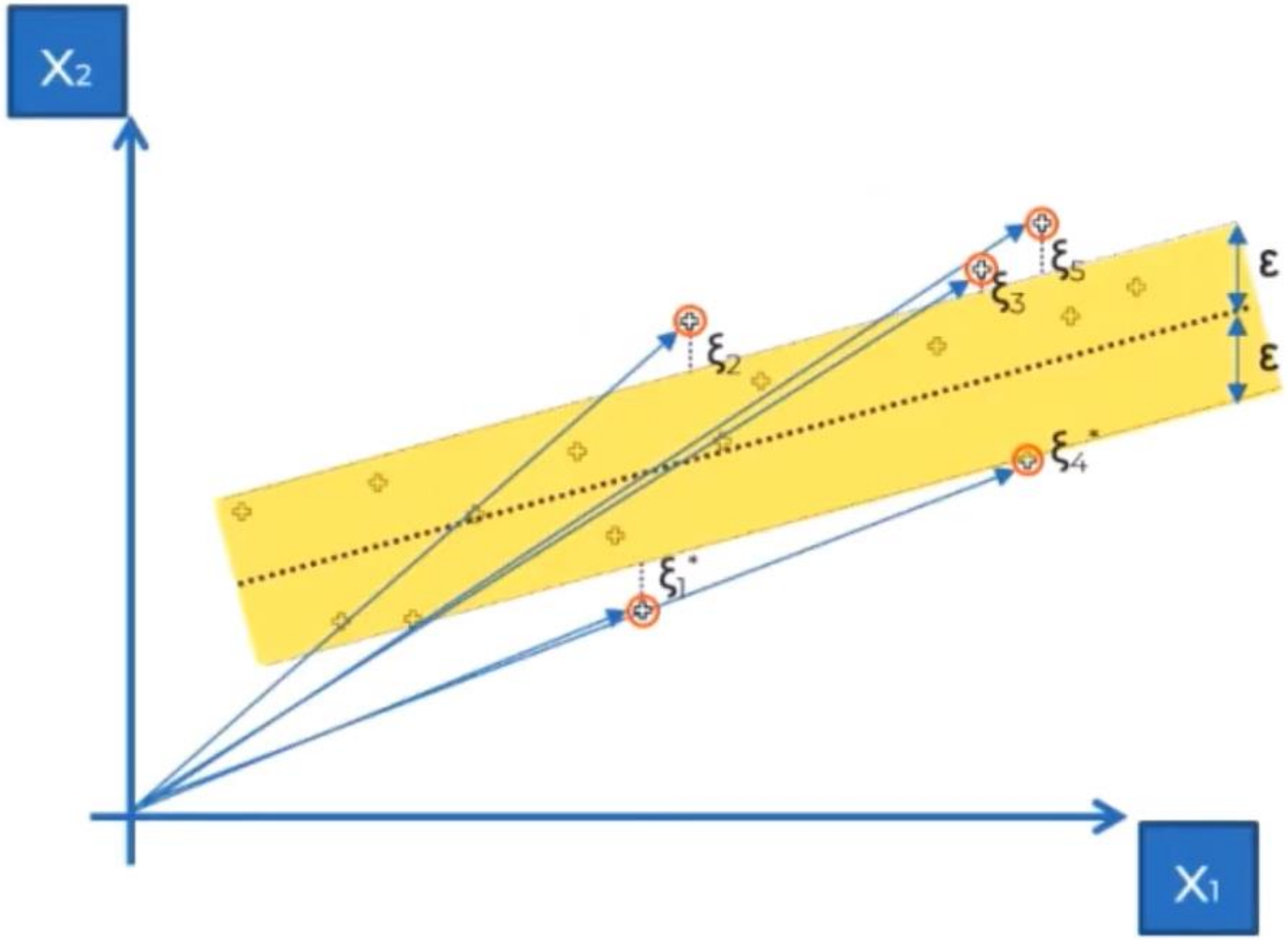
ΠΗΓΗ

Διανυσματική παλινδρόμηση υποστήριξης

Simple Linear Regression



Support Vector Regression



ΠΗΓΗ

Logistic regression vs SVMs

- Εάν n είναι μεγάλο σε σχέση με m, π.χ., $n = 10{,}000, m = 10, \ldots 1000$
  - Χρησιμοποιήστε Logistic regression ή SVM χωρίς πυρήνα (γραμμικός πυρήνας)

- Εάν n είναι μικρό, m είναι ενδιάμεσο, π.χ., $n = 1 - 1000, m = 100 - 10{,}000$
  - Χρήση SVM με πυρήνα Gaussian ή Polynomial

- Αν n είναι μικρό, m είναι μεγάλο, π.χ., $n = 1 - 1000, m \geq 50{,}000$
  - Δημιουργία/προσθήκη περισσότερων χαρακτηριστικών και, στη συνέχεια, χρήση Logistic regression ή SVM χωρίς πυρήνα

Διαφάνεια προσαρμοσμένη από το μάθημα Machine Learning του Andrew Ng, Coursera

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

Περίληψη

- Ένα μη γραμμικά διαχωρίσιμο πρόβλημα μπορεί να γίνει γραμμικά διαχωρίσιμο σε υψηλότερες διαστάσεις
- Το κόλπο του πυρήνα μας επιτρέπει να λειτουργούμε σε ένα χώρο υψηλής διάστασης χωρίς να υπολογίζουμε ρητά τις συντεταγμένες των δεδομένων σε αυτόν τον χώρο.
- Το Support Vector Machine χρησιμοποιεί την έννοια του μέγιστου περιθωρίου για να βελτιώσει το σφάλμα γενίκευσης
- Τα SVMs μπορούν να χρησιμοποιήσουν το τέχνασμα του πυρήνα για να μάθουν μη γραμμικά όρια αποφάσεων
- Τα SVMs μπορούν να χρησιμοποιηθούν για προβλήματα ταξινόμησης πολλαπλών κλάσεων και έχουν επεκταθεί σε προβλήματα παλινδρόμησης

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

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

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