MAI4CAREU
Master programmes in Artificial Intelligence 4 Careers in Europe

Πανεπιστήμιο Κύπρου - Τεχνητή Νοημοσύνη

MAI612 - ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διάλεξη 3: Παλινδρόμηση

Βασίλης Βασιλειάδης, PhD
Χειμερινό Εξάμηνο 2022-2023

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

Διάλεξη 3: Παλινδρόμηση

Μαθησιακά αποτελέσματα

Θα καταλάβετε:

1. Το πρόβλημα της «παλινδρόμησης» στην επιβλεπόμενη μάθηση
2. Το μοντέλο παλινδρόμησης των k- κοντινών γειτόνων
3. Το μοντέλο γραμμικής παλινδρόμησης
4. Κάθοδος κλίσης για την εύρεση των παραμέτρων του μοντέλου γραμμικής παλινδρόμησης
5. Η αναλυτική λύση στο μοντέλο γραμμικής 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

Επιβλεπόμενη Μάθηση

Παρουσιάζουμε μια συμβολή στο σύστημα: ο «δάσκαλος» παρέχει την έξοδο «στόχο»

Το σύστημα προσπαθεί να ταιριάξει τις προβλέψεις του με τις προβλεπόμενες προβλέψεις

- Για παράδειγμα: εικόνες από γάτες και σκύλους με την αντίστοιχη ετικέτα τους

Στόχος: Χρησιμοποιήστε δεδομένα για να δημιουργήσετε ένα μοντέλο που γενικεύει σε άγνωστα ζεύγη εισόδου-εξόδου

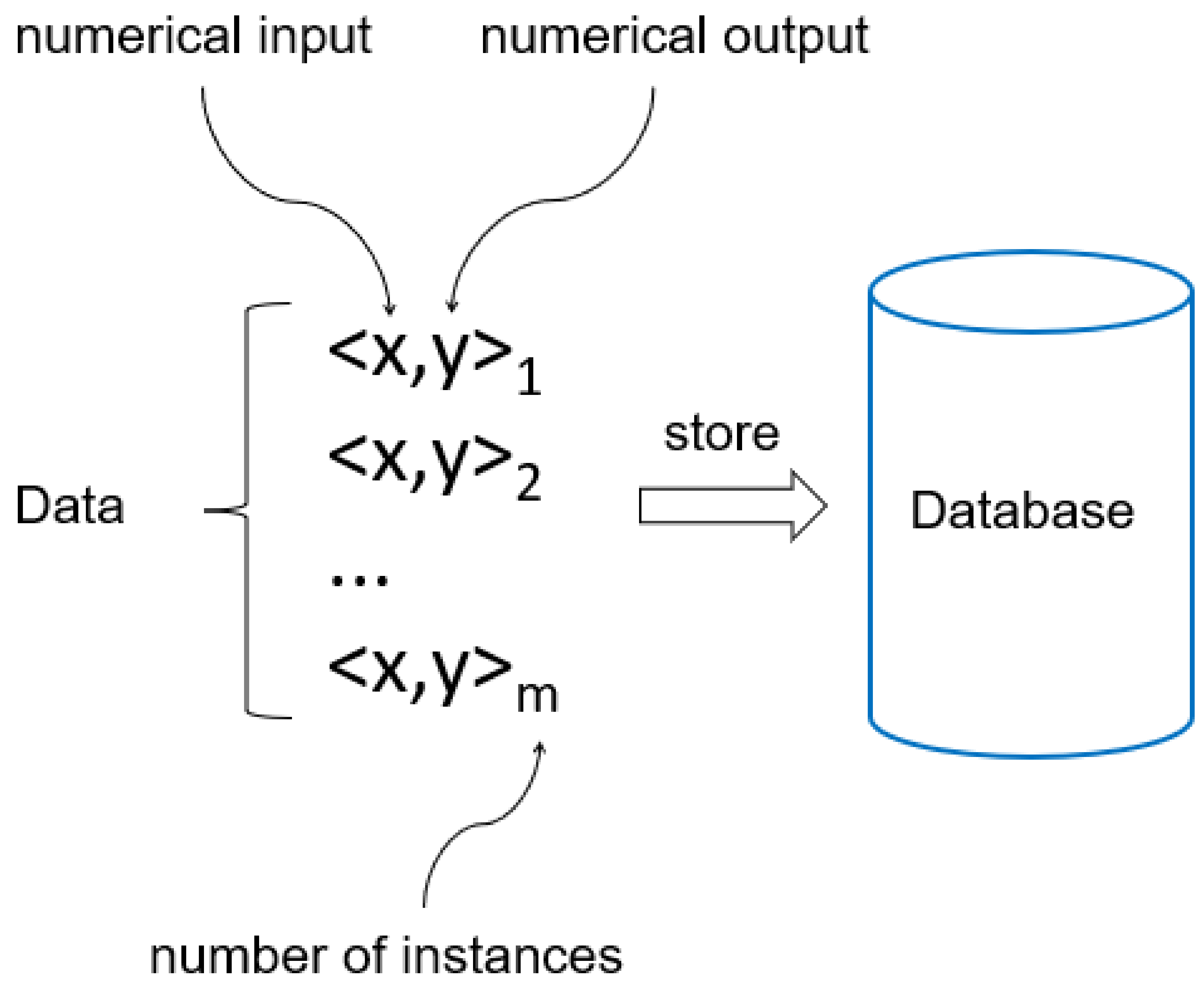
Δύο κοινά προβλήματα:

1. Παλινδρόμηση - Έξοδος: συνεχής αξία
2. Ταξινόμηση - Έξοδος: διακριτή τιμή

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

Παλινδρόμηση



Ποια είναι η τιμή του y δοθέντος x;

~~Y = ανα η(x)~~

Y = f [γείτονες(x)]

Χαρακτηριστικά γνωρίσματα:

+ απλό

+ γρήγορο

+ θυμάται

Μειονεκτήματα:

- Δεν γενικεύει

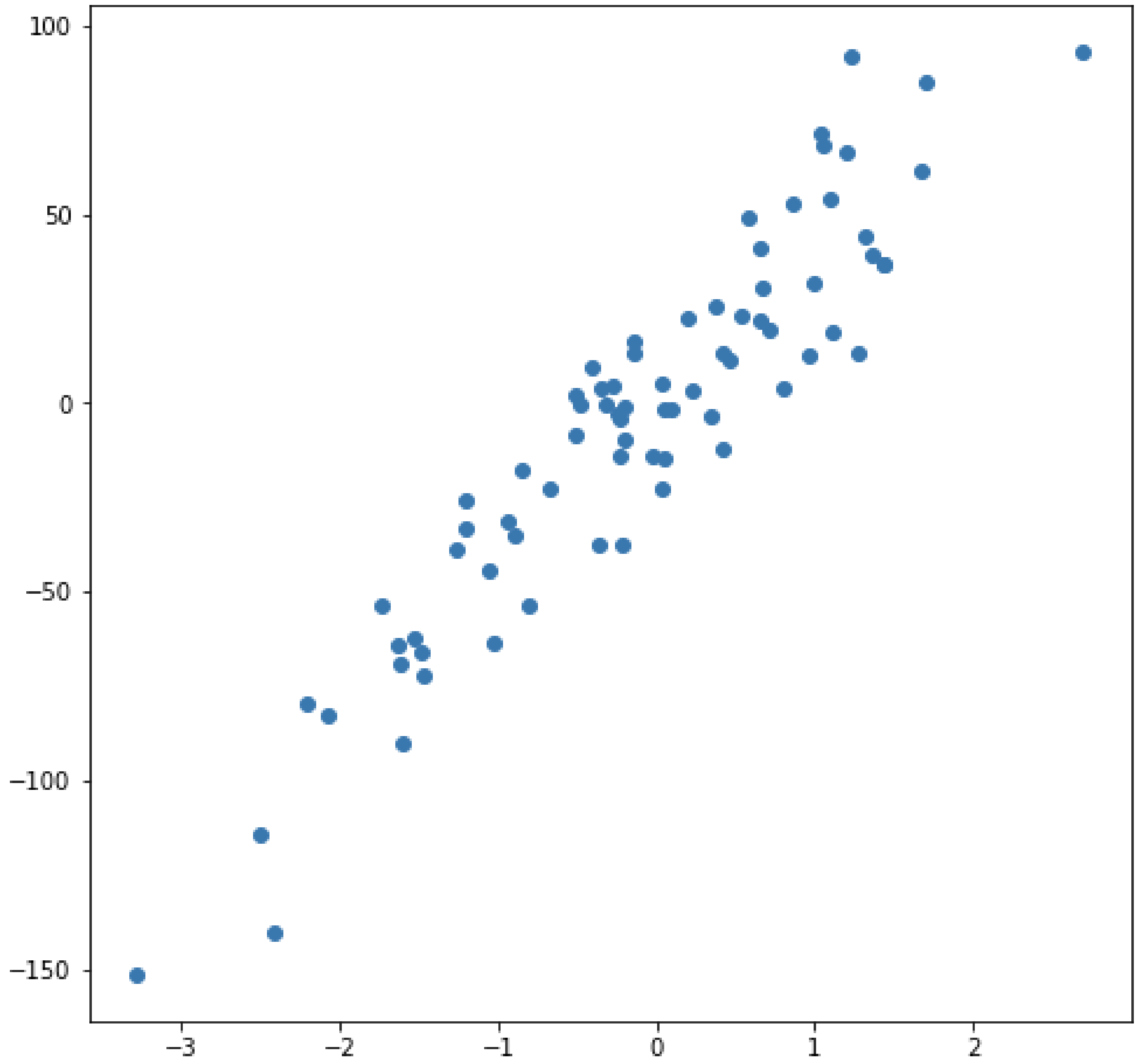
Παλινδρόμηση με Κ-Κοντινότερο γείτονα

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

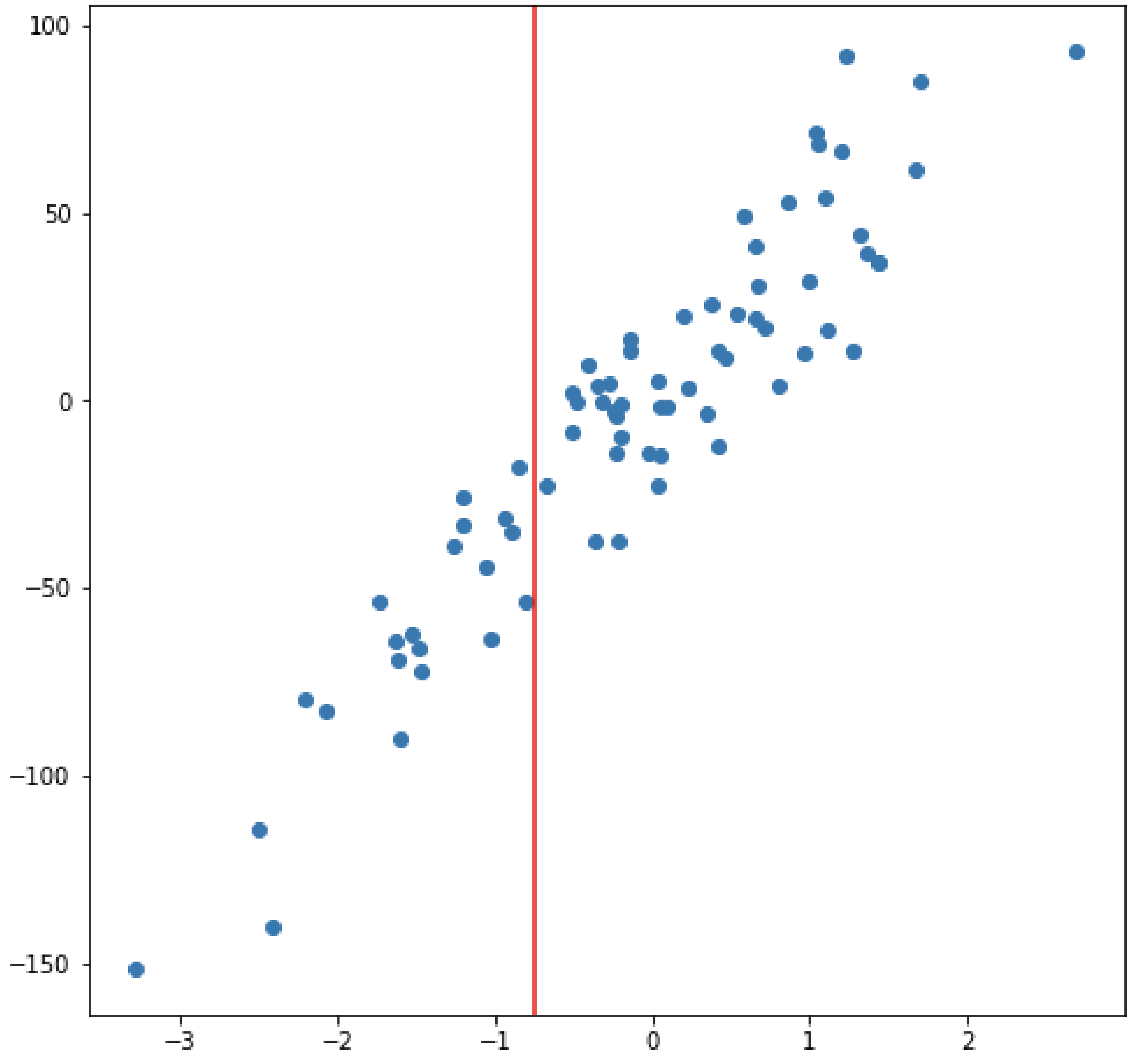
Παλινδρόμηση με Κ-Κοντινότερο γείτονα

ΠΗΓΗ



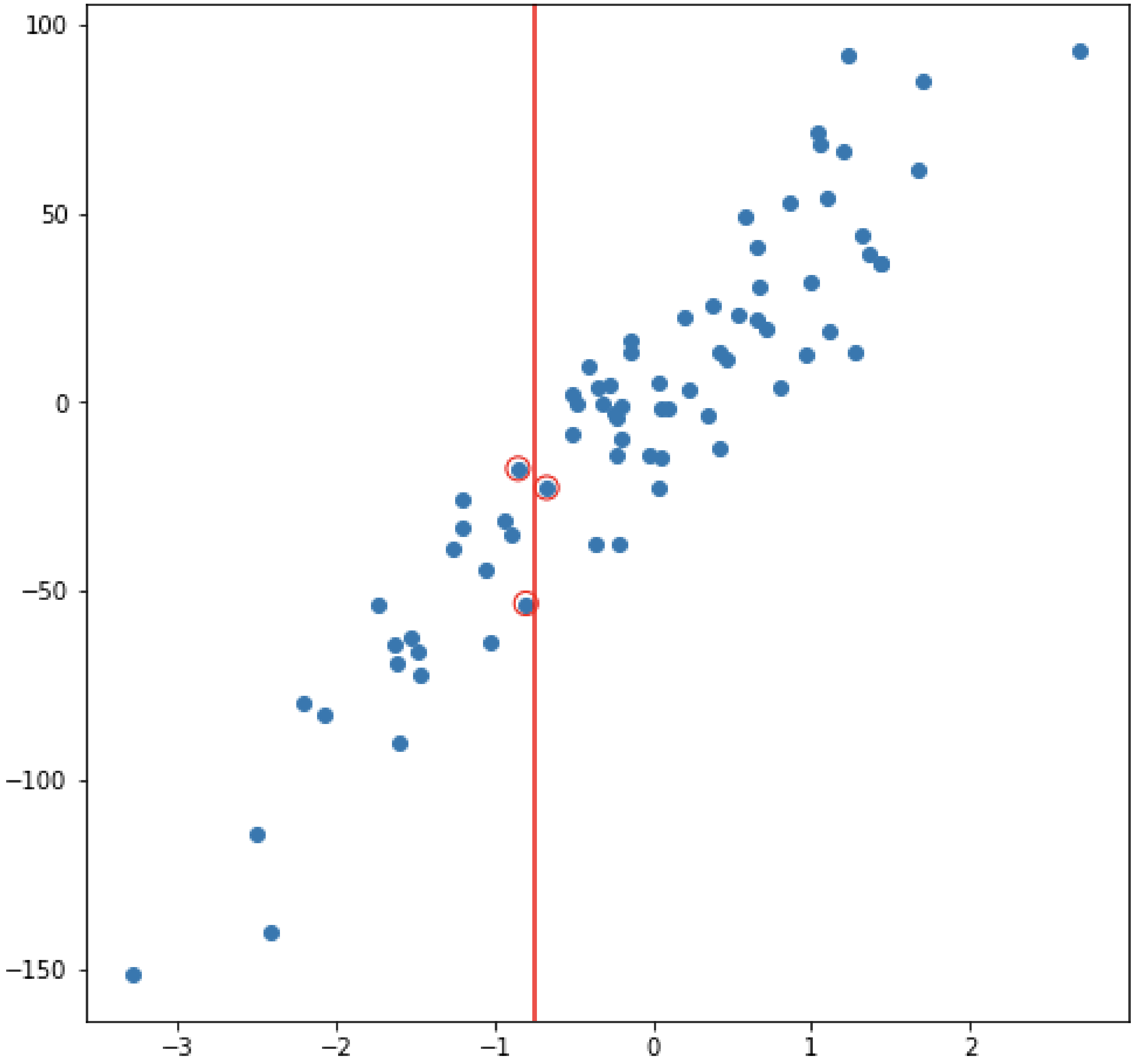
Παλινδρόμηση με Κ-Κοντινότερο γείτονα

ΠΗΓΗ



Παλινδρόμηση με Κ-Κοντινότερο γείτονα

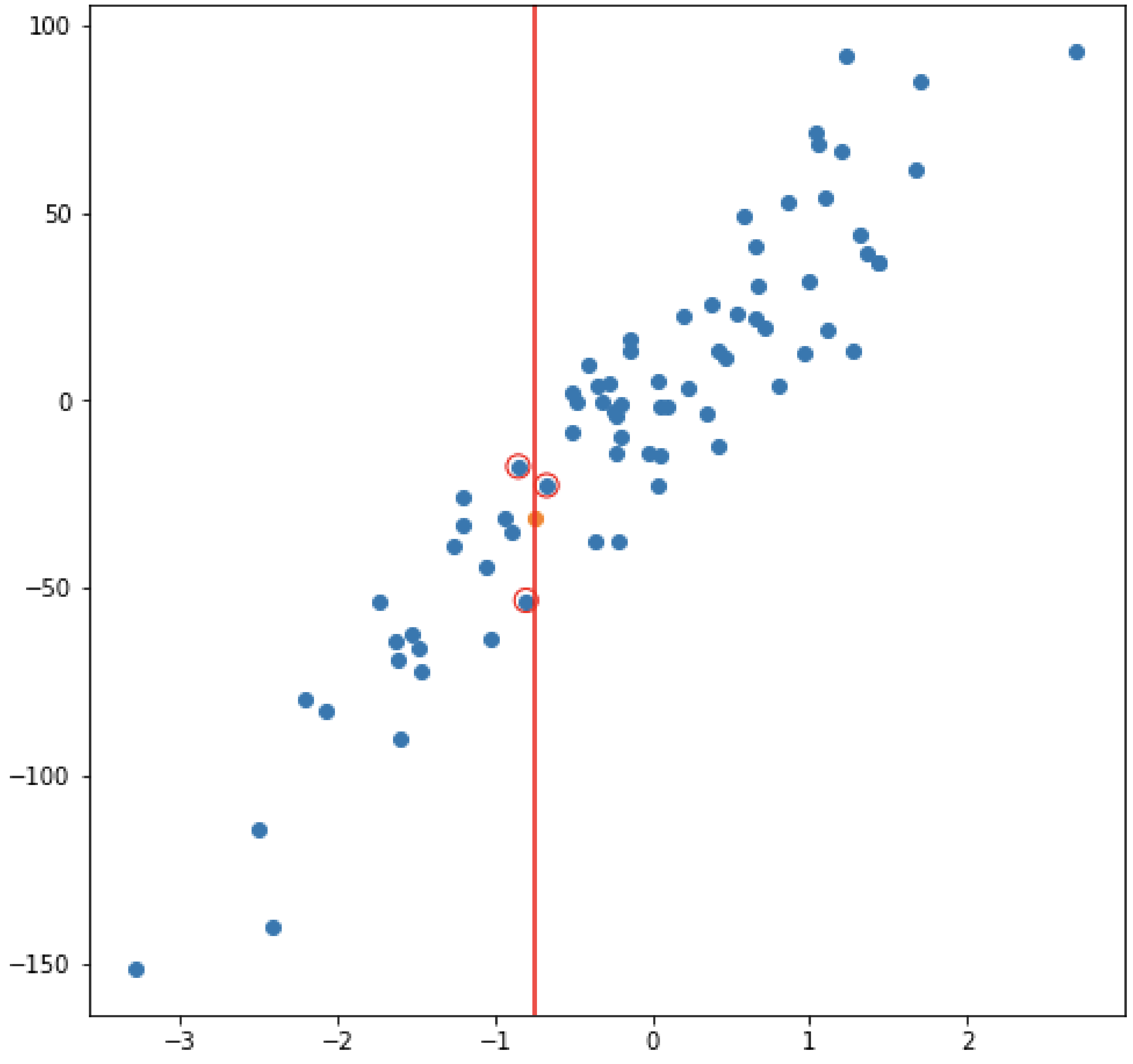
[ΠΗΓΗ](#)



K = 3

Παλινδρόμηση με Κ-Κοντινότερο γείτονα

ΠΗΓΗ



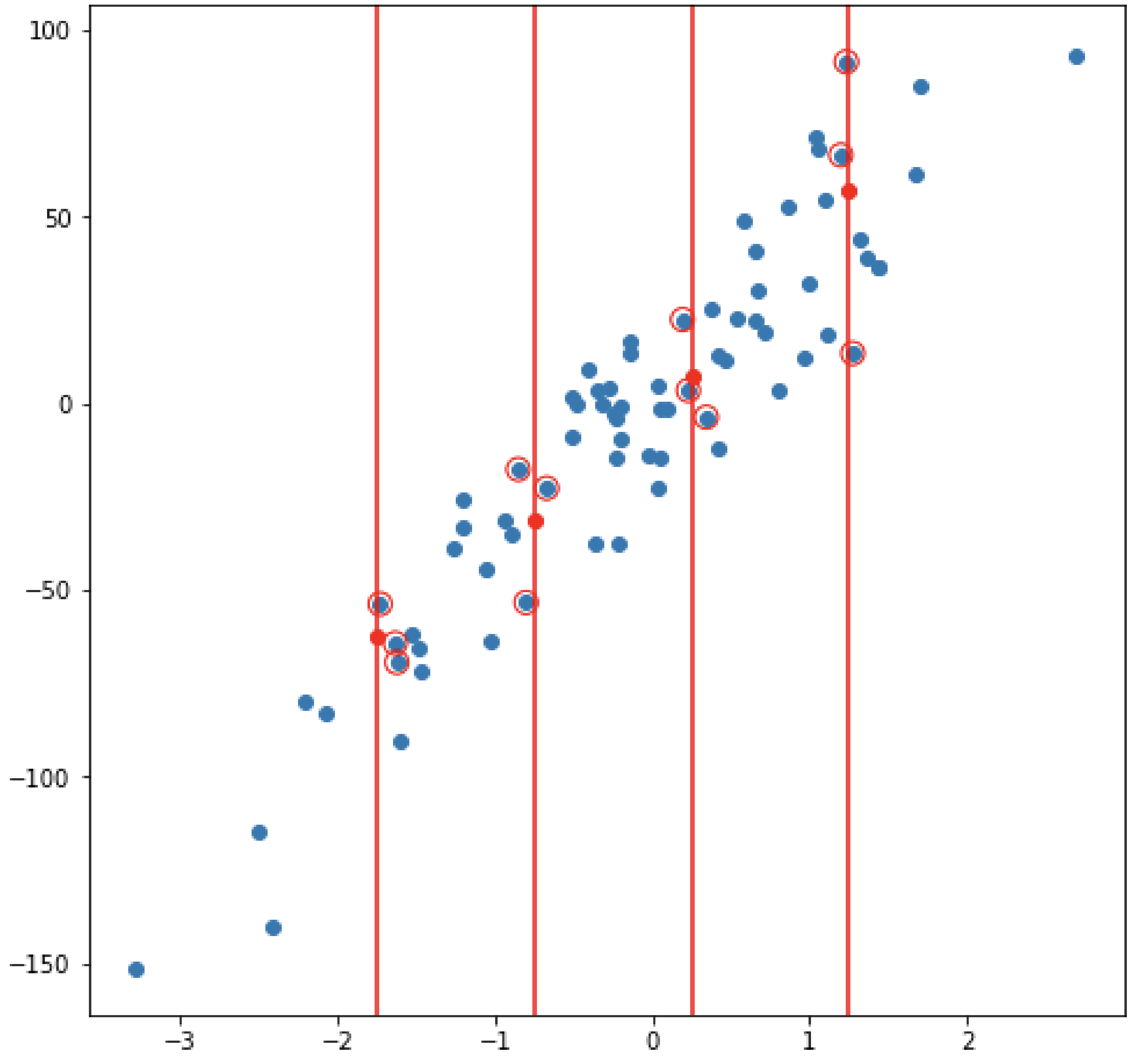
K = 3

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

Παλινδρόμηση με K-Κοντινότερο γείτονα

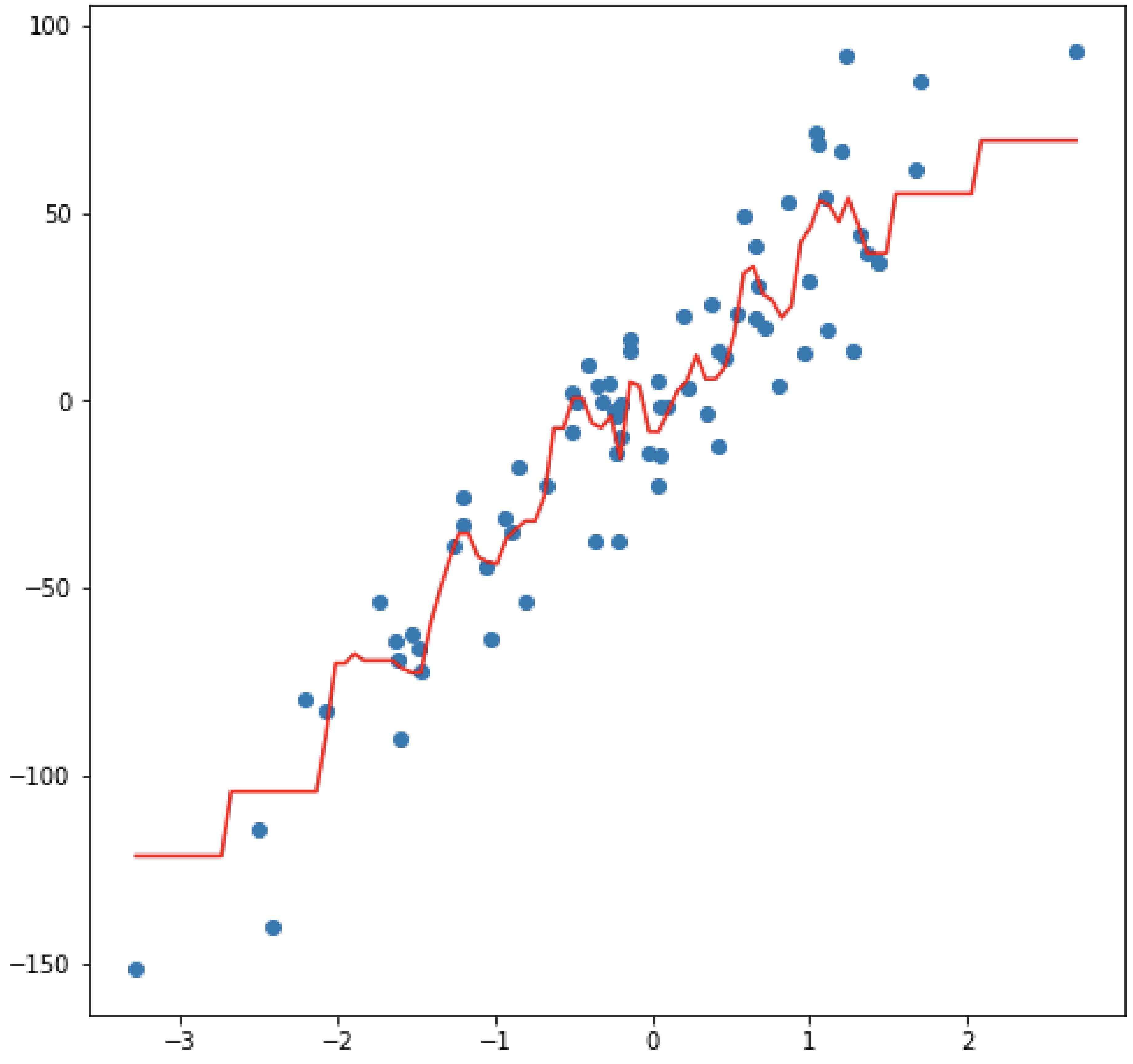
ΠΗΓΗ



K = 3

Παλινδρόμηση με K-Κοντινότερο γείτονα

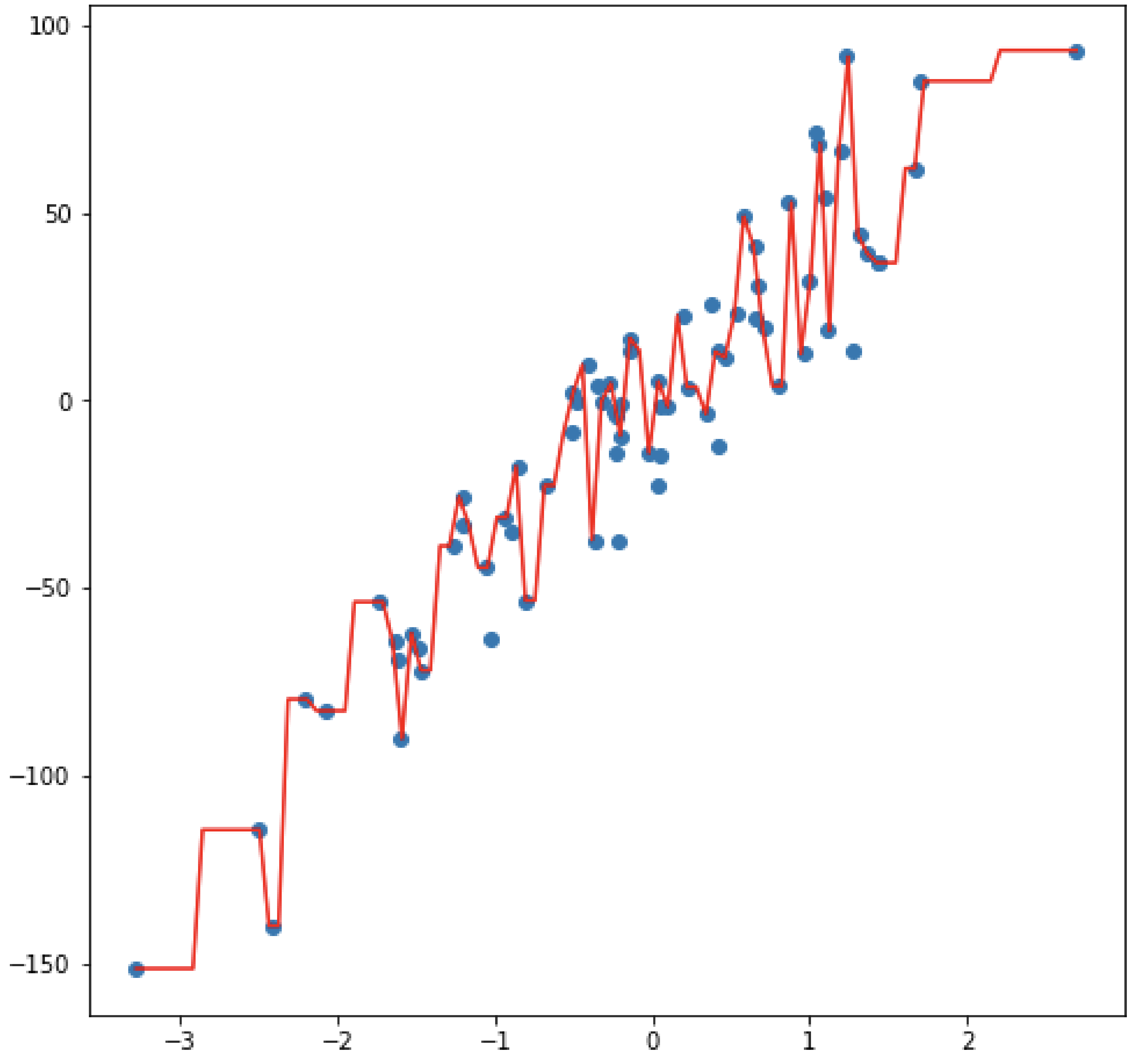
ΠΗΓΗ



K = 3

Παλινδρόμηση με Κ-Κοντινότερο γείτονα

ΠΗΓΗ



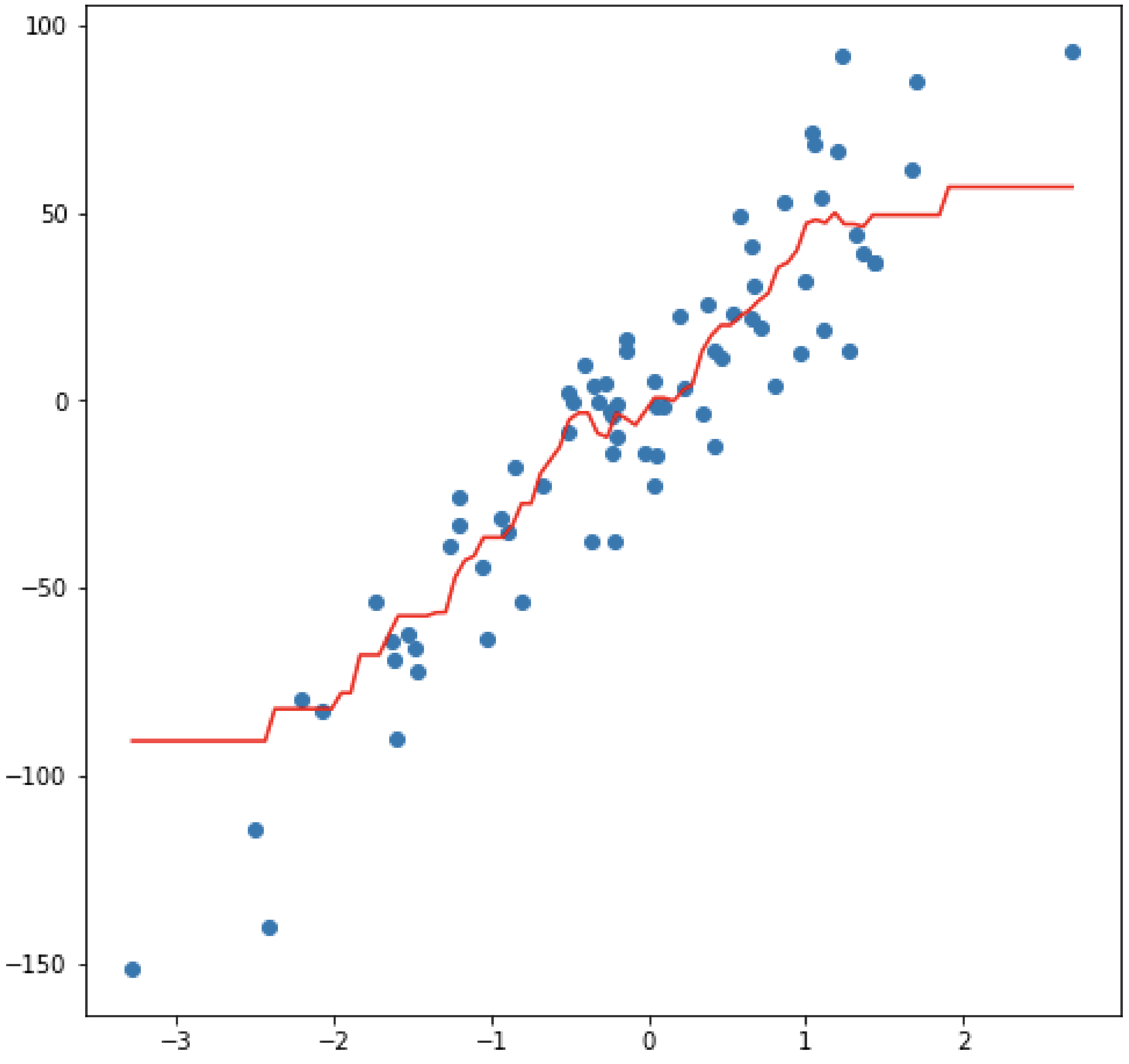
Κ = 1

Το μοντέλο καταγράφει τις μικρές διαφορές και το θόρυβο αντί για την τάση

Αυτό είναι γνωστό ως υπερπροσαρμογή

Παλινδρόμηση με Κ-Κοντινότερο γείτονα

ΠΗΓΗ



K = 10

Ομαλότερη πρόβλεψη, αλλά πρόβλημα στις άκρες

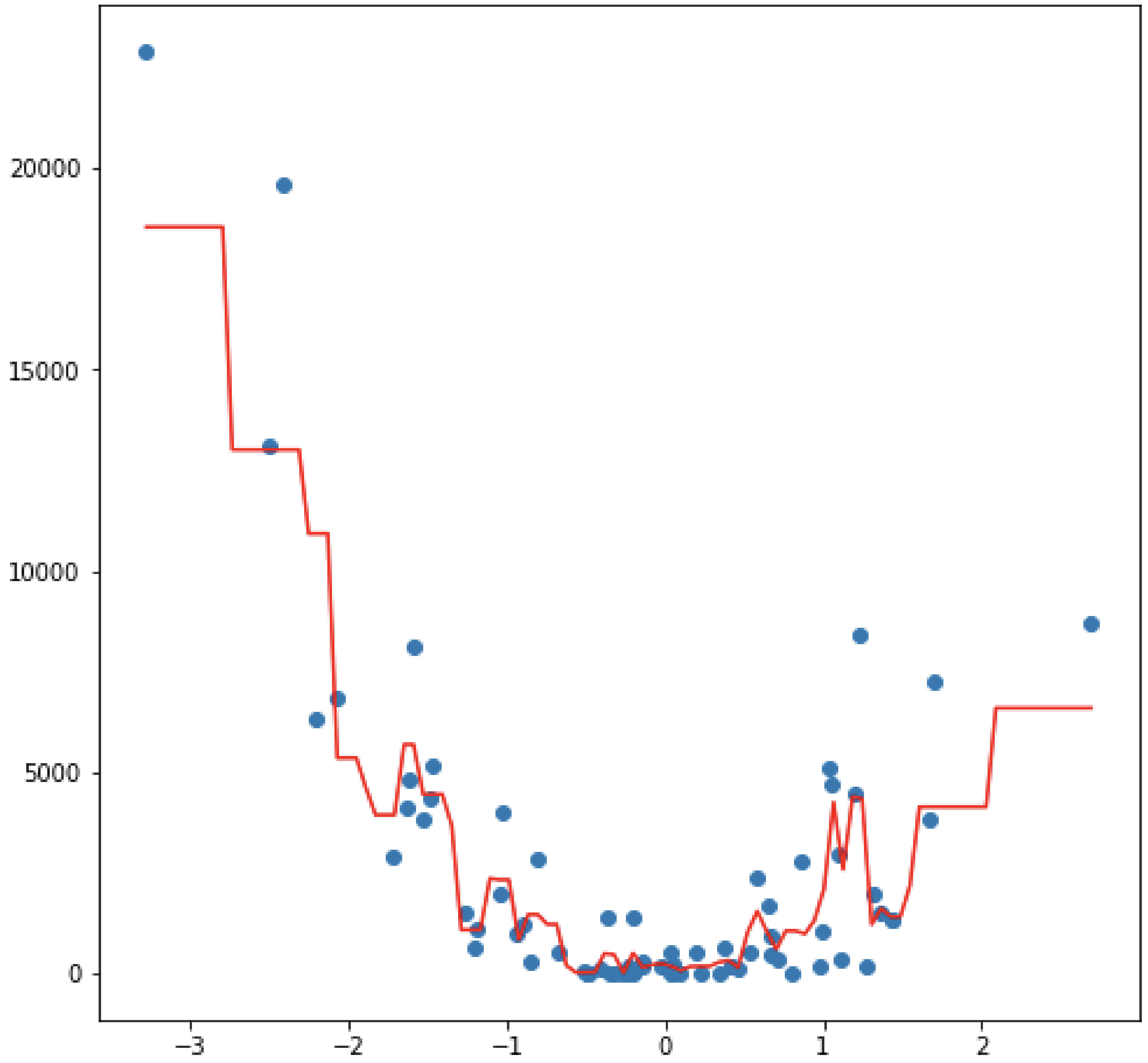
Κι αν K=m;

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

Παλινδρόμηση με Κ-Κοντινότερο γείτονα

ΠΗΓΗ



K = 3

Παλινδρόμηση με Κ-Κοντινότερο γείτονα

Προτερήματα
- εύκολη στην υλοποίηση
- δεν υπάρχει χρόνος εκμάθησης
- Μπορεί να χειριστεί μη γραμμικότητα

Αδυναμίες
- η πρόβλεψη γίνεται πιο αργή καθώς τα δεδομένα γίνονται περισσότερα
- δεν λειτουργεί καλά σε μεγάλες διαστάσεις
- χρειάζεται κανικοποίηση χαρακτηριστικών
- ευαίσθητα σε θορυβώδη δεδομένα ή ακραίες τιμές
- έλλειψη ερμηνευσιμότητας

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

Παράδειγμα

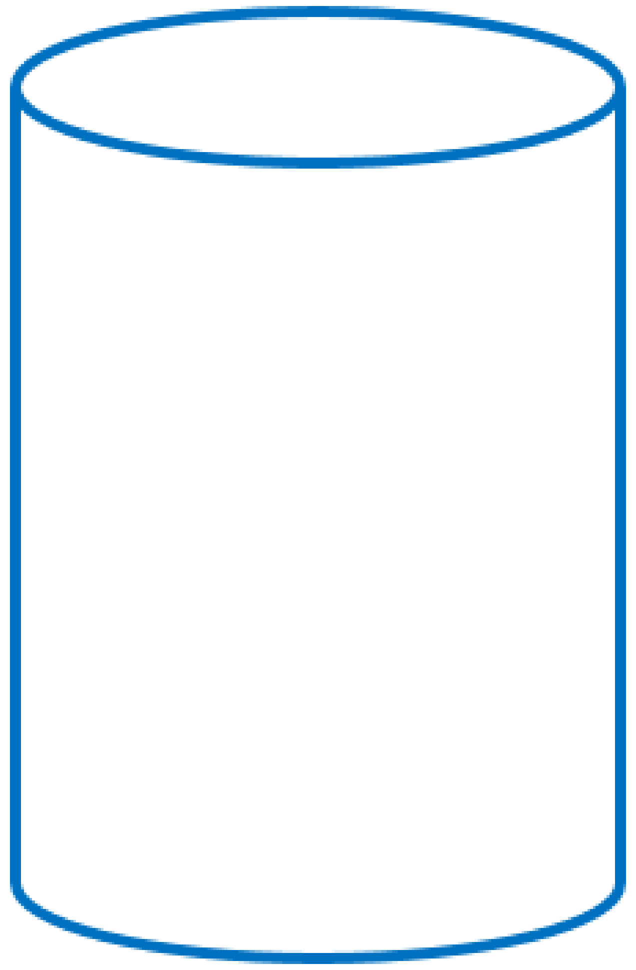
```
In [1]: X = [[0], [1], [2], [3]]
   ...: y = [0, 0, 1, 1]

In [2]: from sklearn.neighbors import KNeighborsRegressor
   ...: neigh = KNeighborsRegressor(n_neighbors=2)
   ...: neigh.fit(X, y)
Out[2]:
KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
          metric_params=None, n_jobs=None, n_neighbors=2, p=2,
          weights='uniform')

In [3]: print(neigh.predict([[1.5]]))
[0.5]
```


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

Παλινδρόμηση

$\langle x,y\rangle_1$
$\langle x,y\rangle_2$
…
$\langle x,y\rangle_m$



y = knn(x)

Μάθηση με βάση την περίπτωση/μάθηση με βάση τη μνήμη
ή
Μη παραμετρικά μοντέλα /παραδειγματικά μοντέλα

Εναλλακτική προσέγγιση

$\langle x,y\rangle_1$
$\langle x,y\rangle_2$
…
$\langle x,y\rangle_m$

y = f(x)

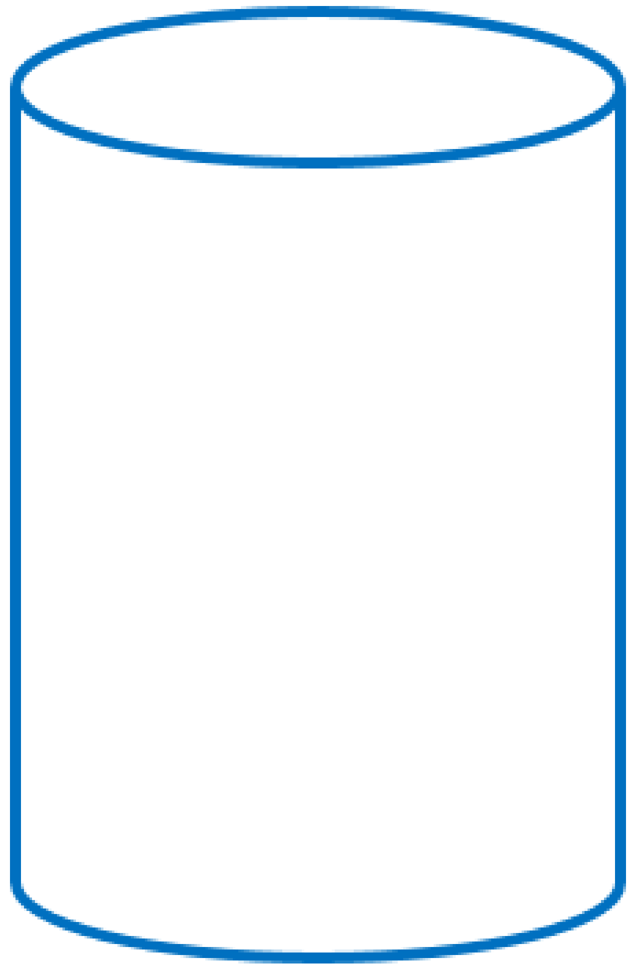
Ποια είναι η απλούστερη συνάρτηση που μπορούμε να χρησιμοποιήσουμε;

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

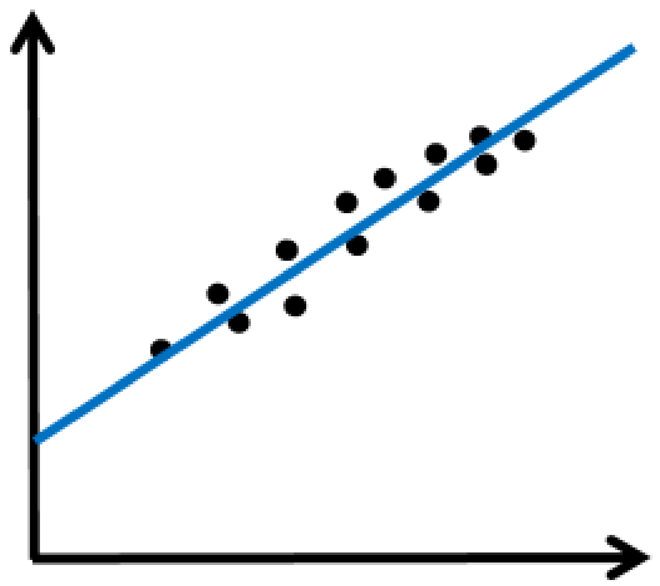
Παλινδρόμηση

$\langle x,y\rangle_1$
$\langle x,y\rangle_2$
…
$\langle x,y\rangle_m$



$y = knn(x)$

$\langle x,y\rangle_1$
$\langle x,y\rangle_2$
…
$\langle x,y\rangle_m$



$y = \alpha x + \beta$

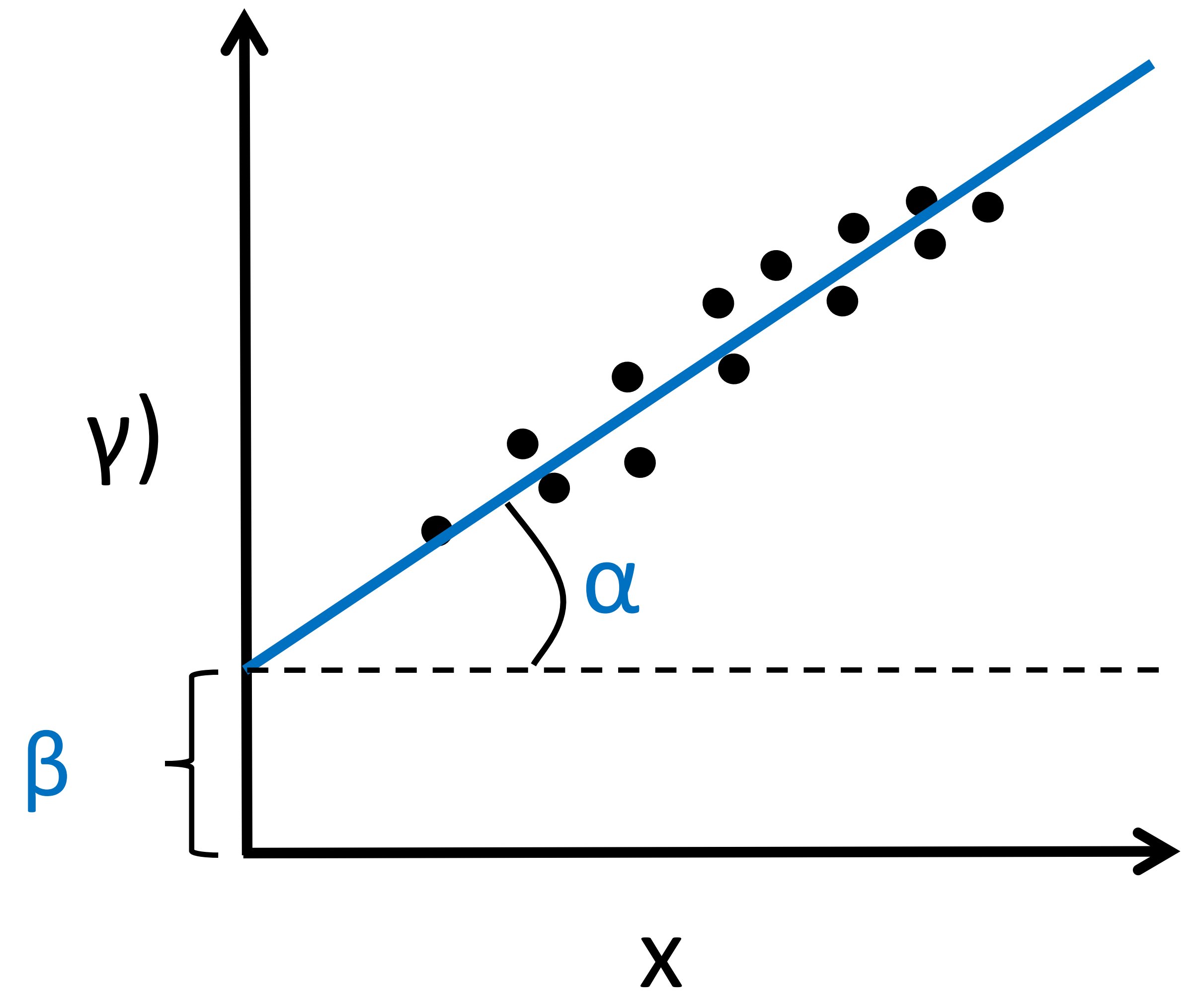
Μάθηση με βάση την περίπτωση/μάθηση με βάση τη μνήμη
ή
Μη παραμετρικά μοντέλα /υποδειγματικά μοντέλα

Μάθηση με βάση το μοντέλο
ή
Παραμετρικά μοντέλα

Co-financed by the European Union
Connecting Europe Facility

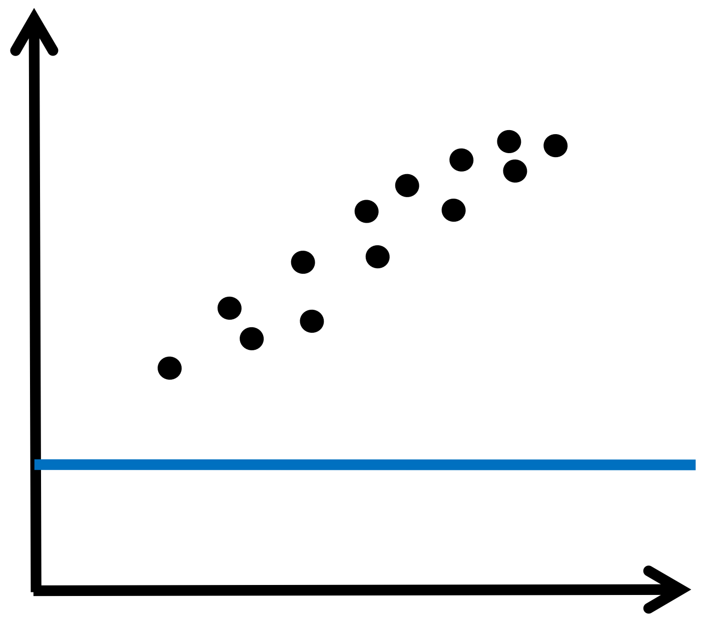
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Γραμμική παλινδρόμηση

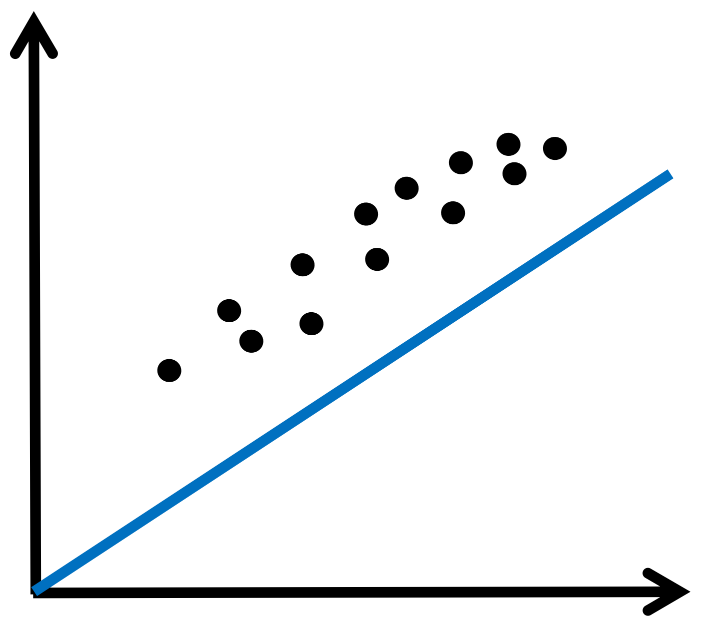


$y = \alpha x + \beta$

$\alpha=0$
$y=\beta$



$\beta=0$
$y=\alpha x$



Αναπαράσταση Μοντέλου

Training Set

$\langle x,y \rangle_1$
$\langle x,y \rangle_2$
...
$\langle x,y \rangle_m$

$\boldsymbol{x} \in \mathbb{R}^n$ → dimensionality of the input

$y \in \mathbb{R}$

Training Set → **Learning Algorithm** → **Model**

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

$\boldsymbol{\theta} \in \mathbb{R}^{n+1}$

intercept term (β)

estimated value of y

slopes (α) in every dimension

n=1 Simple Linear Regression
n>1 Multiple Linear Regression

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

Αναπαράσταση Μοντέλου

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$ όπου $x_0 = 1$, έτσι τώρα $\boldsymbol{x} \in \mathbb{R}^{n+1}$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}^T \boldsymbol{x}$$

$$\begin{bmatrix} \theta_0 & \theta_1 & \ldots & \theta_n \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = f_{\boldsymbol{\theta}}(\boldsymbol{x})$$

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

Επιλέξτε $\theta_0, \ldots, \theta_n$ έτσι ώστε το $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ να είναι κοντά το y σε όλα τα εκπαιδευτικά παραδείγματα μας $< \boldsymbol{x}, y >_{1:m}$

Πώς;

Μια κοινή προσέγγιση είναι η χρήση της μεθόδου των **ελάχιστων τετραγώνων**

- Ελαχιστοποιήστε το άθροισμα των τετραγώνων των υπολείπων
- **Υπόλοιπα**: διαφορά μεταξύ της πραγματικής τιμής (y) και της εκτιμώμενης τιμής (που $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ τοποθετείται από το μοντέλο)

$$(f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)})^2$$

Στόχος και συνάρτηση σφάλματος

Επιλέξτε $\theta_0, \ldots, \theta_n$ έτσι ώστε το $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ να είναι κοντά το y σε όλα τα εκπαιδευτικά παραδείγματα μας $< \boldsymbol{x}, y >_{1:m}$

$$\underset{\boldsymbol{\theta}}{minimize}\ L(\boldsymbol{\theta})$$

$$L(\boldsymbol{\theta}) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}\right)^2$$

Συνάρτηση μέσου τετραγωνικού σφάλματος

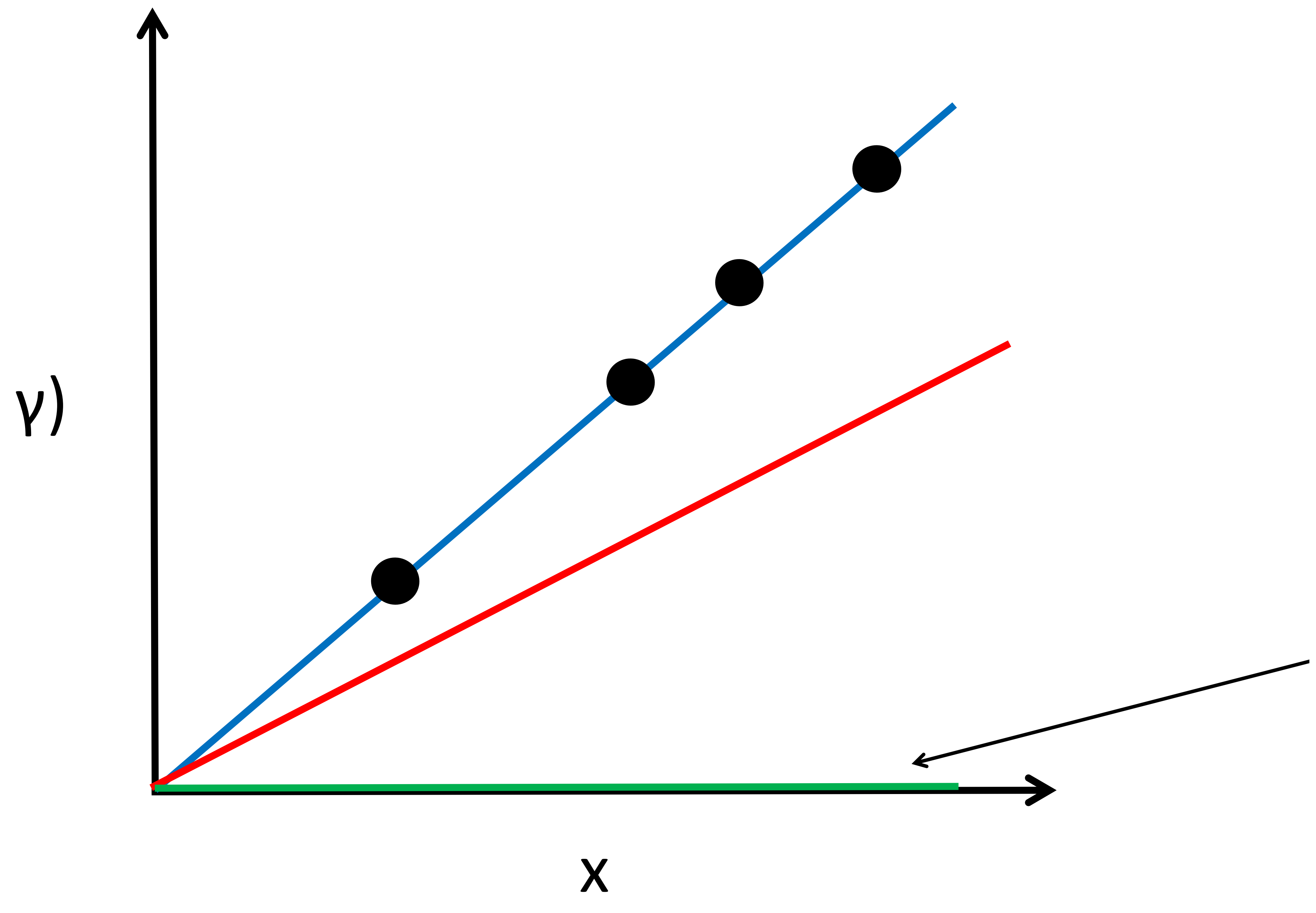
Μπορεί το σφάλμα να είναι < 0;

Co-financed by the European Union
Connecting Europe Facility

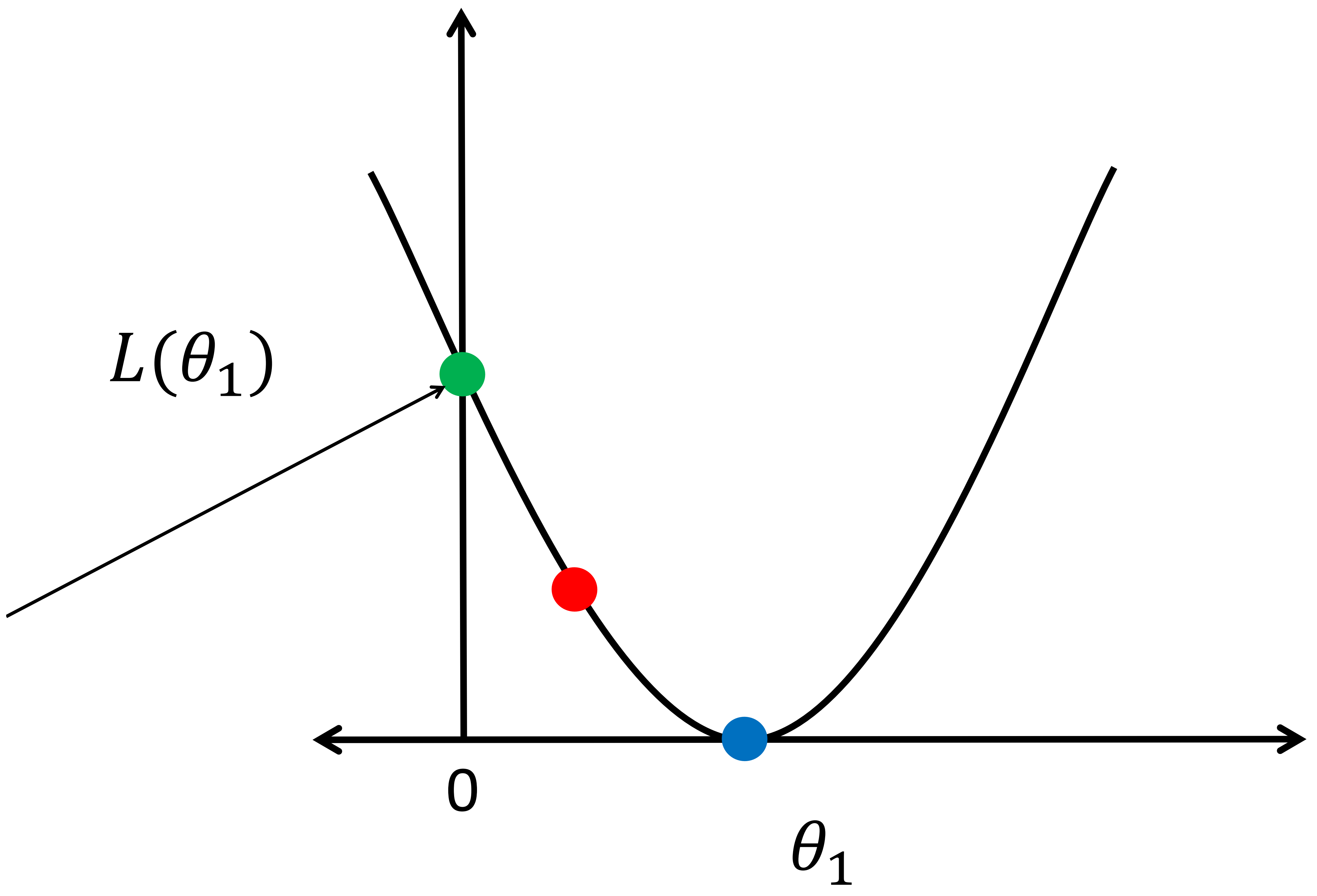
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Συνάρτηση σφάλματος

$f_{\theta_1}(x)$



$L(\theta_1)$



MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

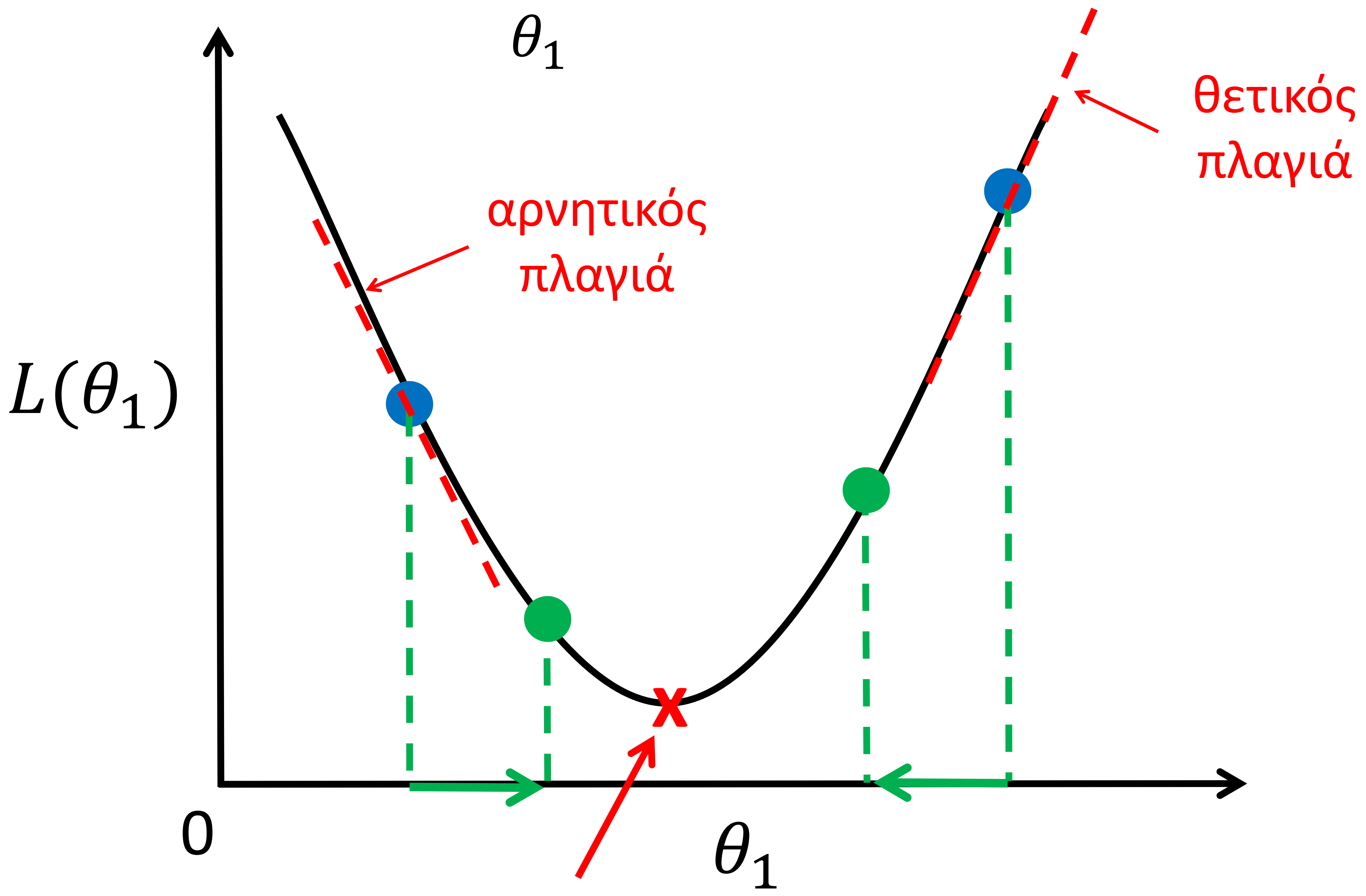
Gradient Descent

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

Βελτιστοποίηση: Gradient Descent

$$\underset{\theta_1}{minimize}\ L(\theta_1)$$



Εδώ θέλουμε να καταλήξουμε
(τοπικό ελάχιστο)

1. Ξεκινήστε από τυχαίο θ_1
2. Επαναλάβετε μέχρι τη σύγκλιση

Αλλαγή θ_1 έτσι ώστε να κινηθεί προς τα κάτω

$$\theta_1 = \overbrace{\theta_1}\ \overbrace{- \alpha \frac{dL(\theta_1)}{d\theta_1}} \qquad \alpha \in [0,1]$$

$\frac{dL(\theta_1)}{d\theta_1} > 0$ μειώσεις θ_1

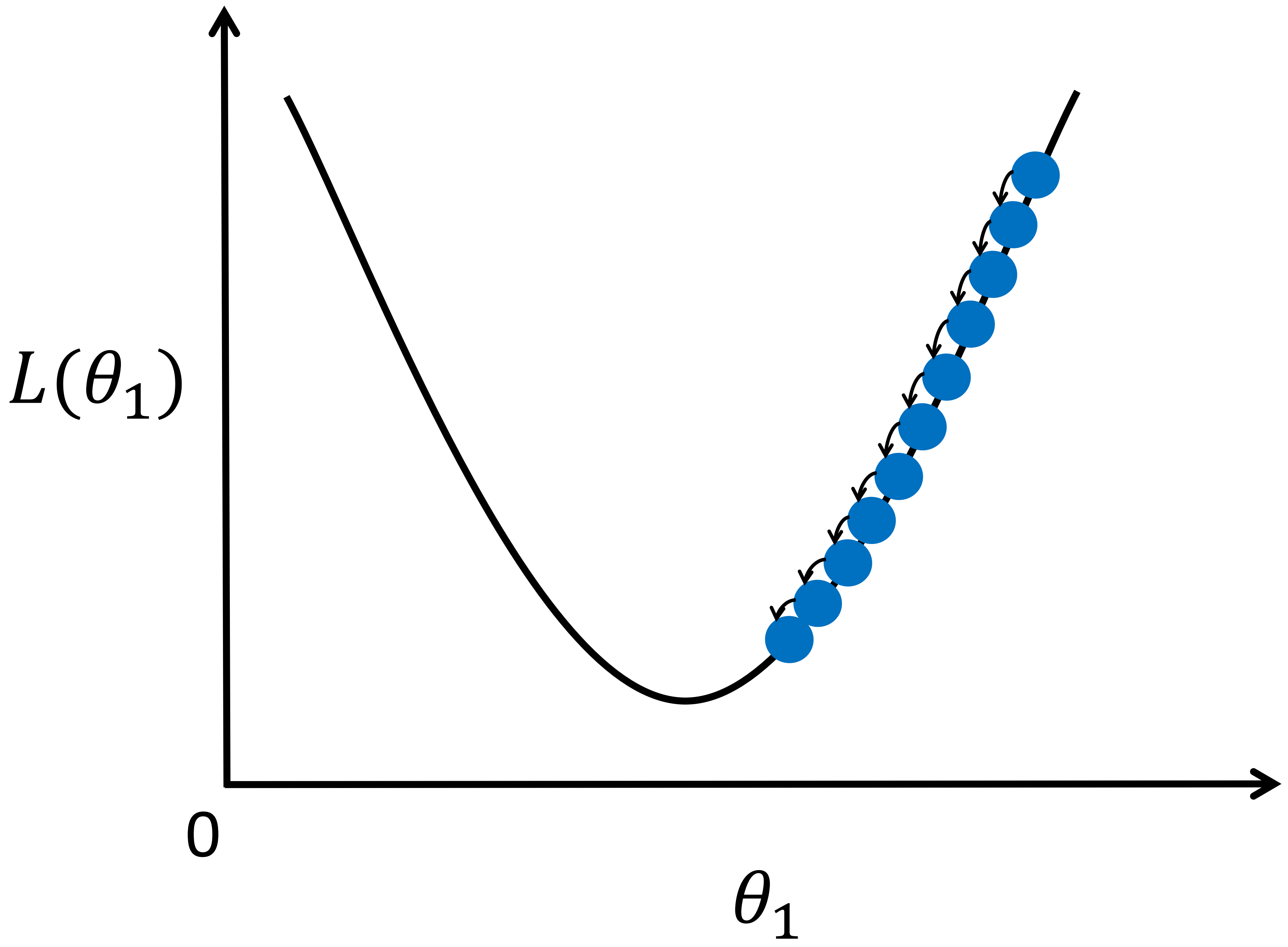
$\frac{dL(\theta_1)}{d\theta_1} < 0$ αυξήσεις θ_1

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

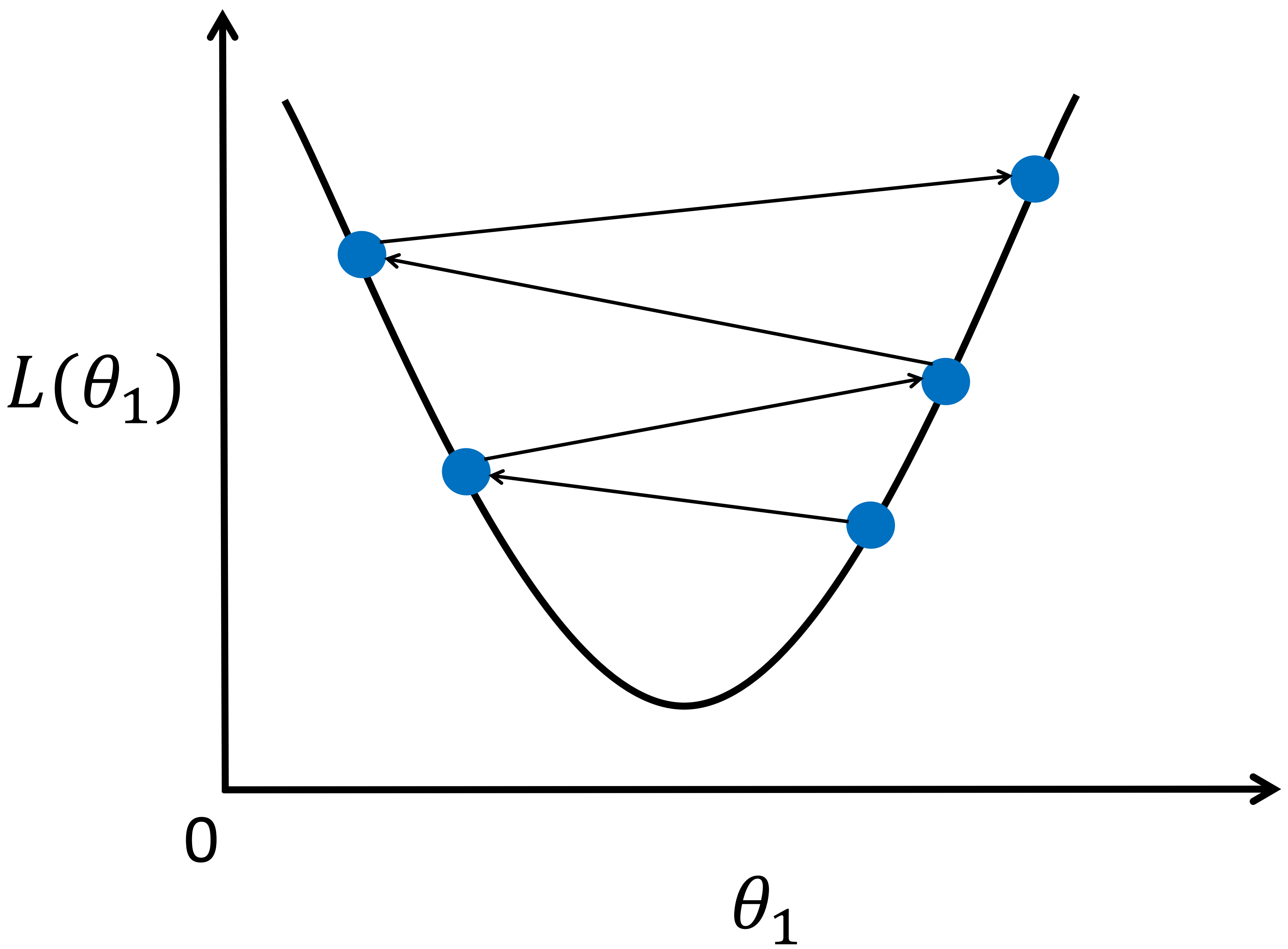
Gradient Descent: ρυθμός εκμάθησης α

πολύ μικρό



αργή σύγκλιση

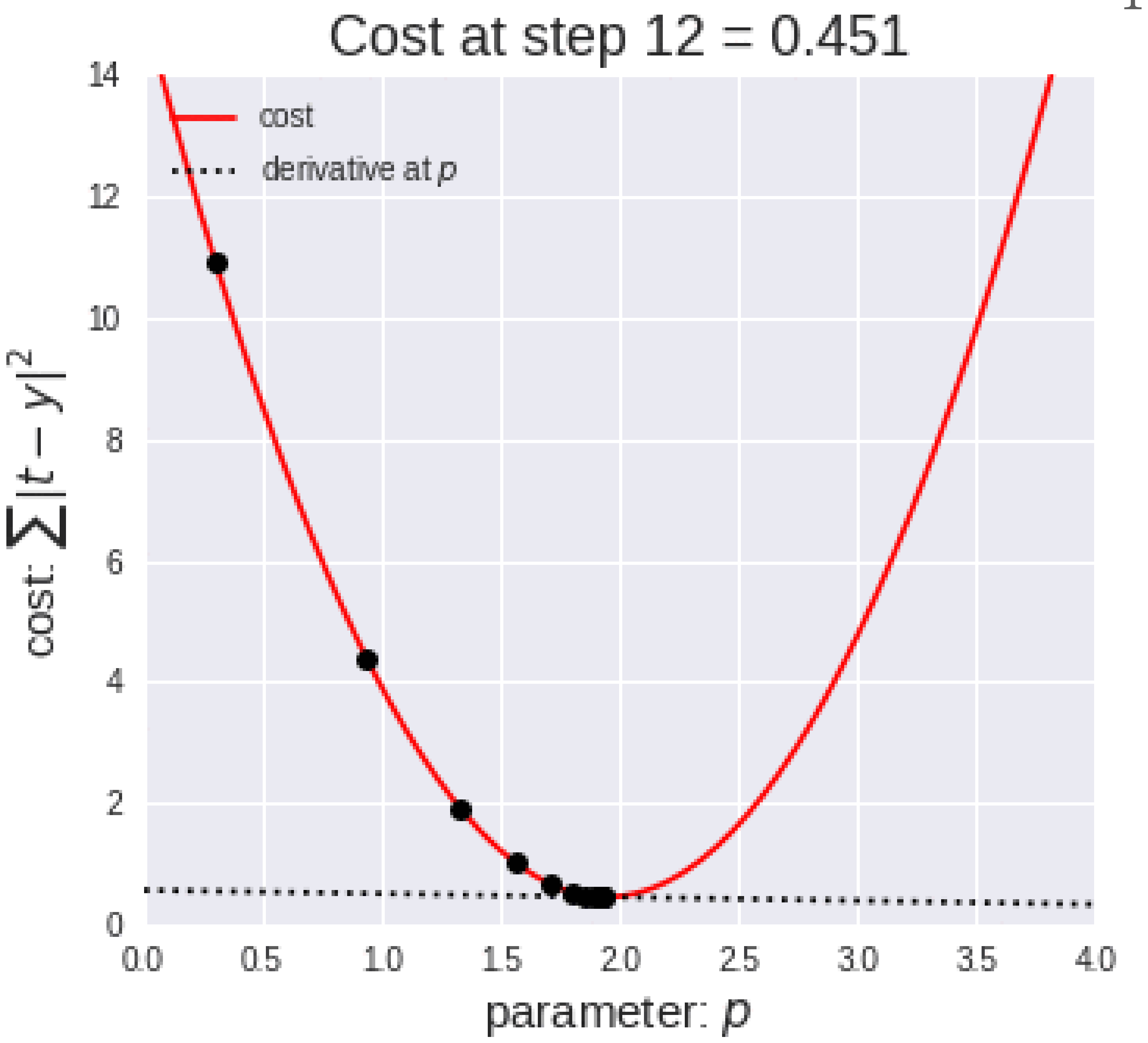
πολύ μεγάλο

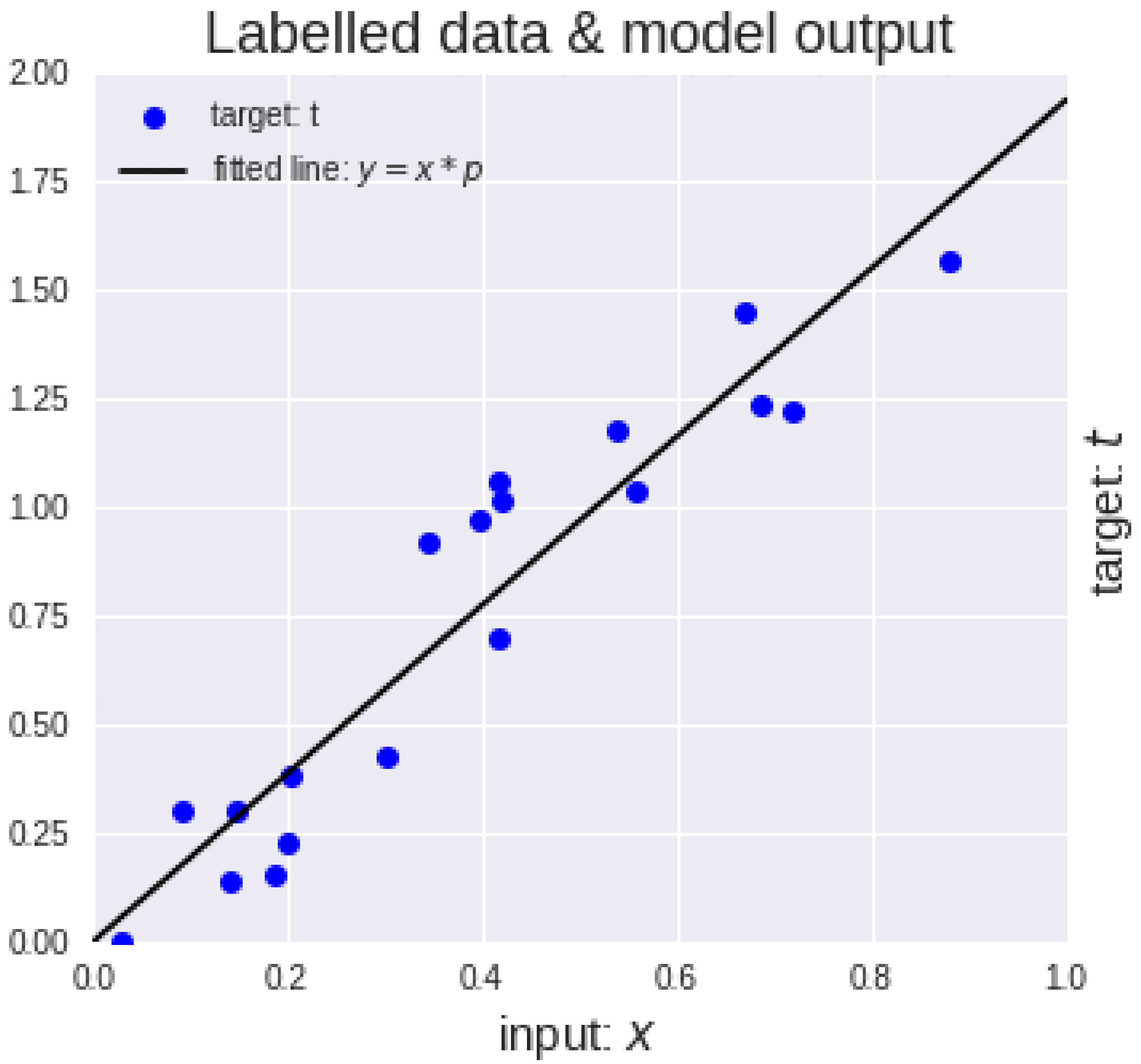


πιθανή απόκλιση

Gradient Descent

$$\underset{\theta_1}{minimize}\ L(\theta_1)$$

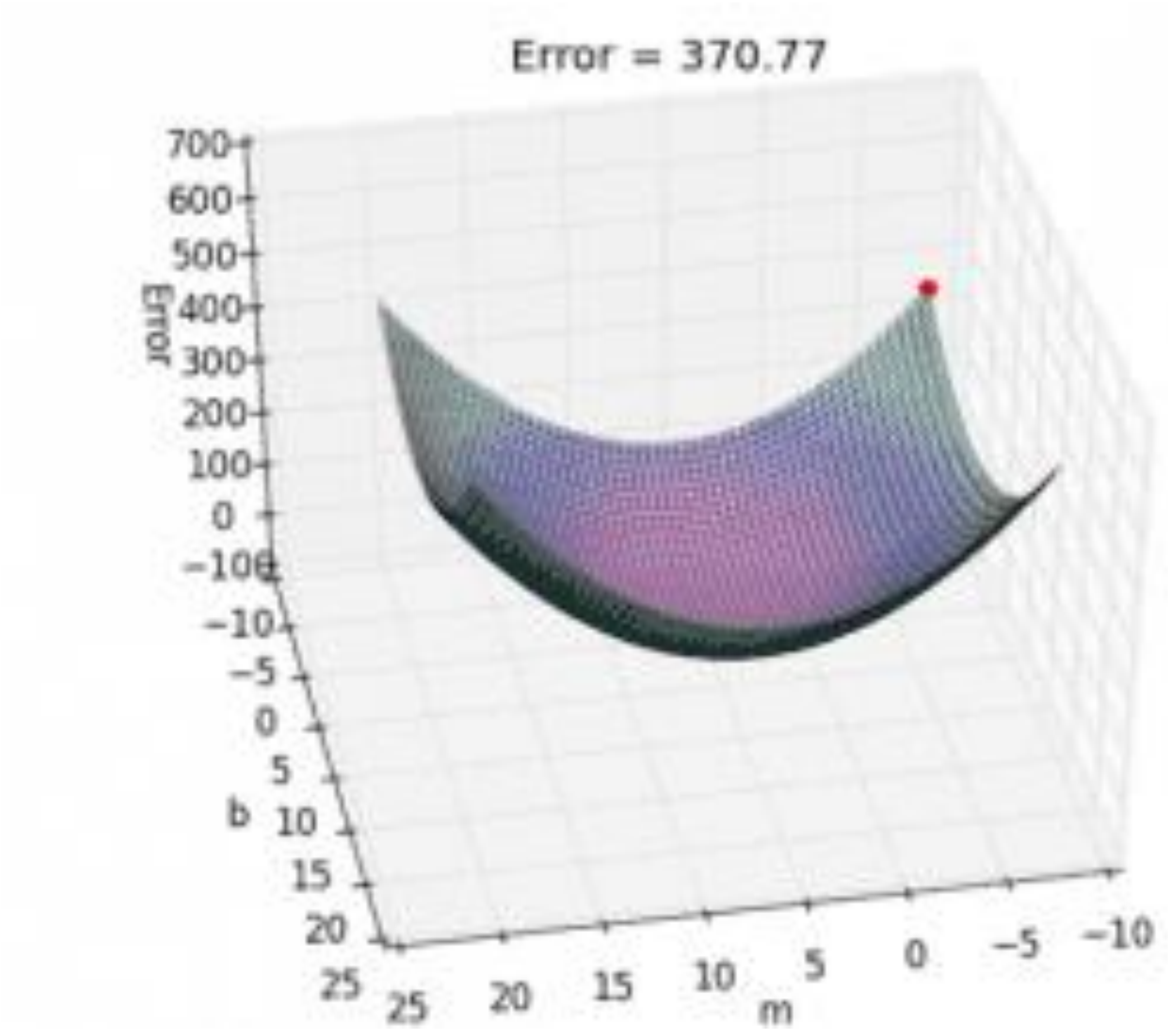


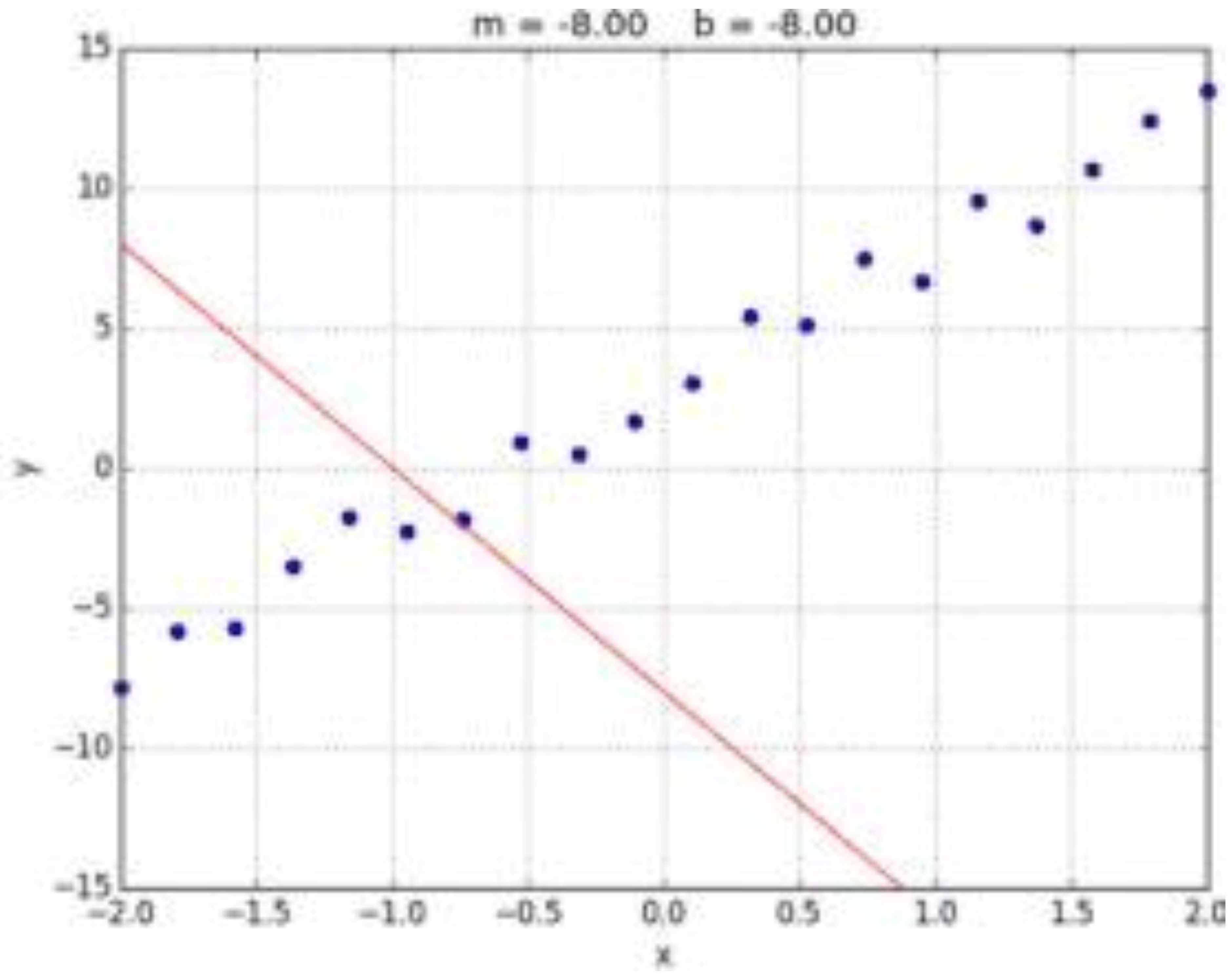


ПНГН

Gradient Descent

$$\underset{\theta_0, \theta_1}{minimize}\ L(\theta_0, \theta_1)$$





ΠΗΓΗ

Gradient Descent στην πράξη

$$f_{\theta_0,\theta_1}(x) = \theta_0 + \theta_1 x$$

Υλοποίηση:

$$||\nabla L|| < \varepsilon$$

Επαναλάβετε μέχρι τη σύγκλιση:

$$L(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(f_{\theta_0,\theta_1}(x^{(i)}) - y^{(i)})^2$$

1. Υπολογίστε το σφάλμα $L(\theta_0, \theta_1)$
2. Υπολογισμός διανύσματος κλίσης $\begin{bmatrix} \frac{\partial L(\theta_0,\theta_1)}{\partial \theta_0} \\ \frac{\partial L(\theta_0,\theta_1)}{\partial \theta_1} \end{bmatrix}$

$$\frac{1}{m}\sum_{i=1}^{m}(f_{\theta_0,\theta_1}(x^{(i)}) - y^{(i)})$$

$$\frac{1}{m}\sum_{i=1}^{m}(f_{\theta_0,\theta_1}(x^{(i)}) - y^{(i)})x^{(i)}$$

3. Ενημέρωση διανύσματος παραμέτρων $\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} \theta_0 - \alpha\frac{\partial L(\theta_0,\theta_1)}{\partial \theta_0} \\ \theta_1 - \alpha\frac{\partial L(\theta_0,\theta_1)}{\partial \theta_1} \end{bmatrix}$

Gradient Descent στην πράξη

Gradient Descent «παρτίδας»:

- Χρησιμοποιεί όλα τα παραδείγματα κατάρτισης (m)
- Χρησιμοποιείται όταν το μέγεθος του συνόλου εκπαίδευσης είναι μικρό (π.χ. m = 10^1 — 10^3)

Άλλες παραλλαγές του Gradient Descent:

- Στοχαστικό Gradient Descent
- Gradient Descent μίνι παρτίδων
- θα τα δούμε σε μεταγενέστερες διαλέξεις

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

Gradient Descent στην πράξη

Χρήση κανονικοποίησης χαρακτηριστικών

Χωρίς κανονικοποίηση χαρακτηριστικών

Με κανονικοποίηση χαρακτηριστικών

Βεβαιωθείτε ότι τα χαρακτηριστικά είναι σε παρόμοια κλίμακα

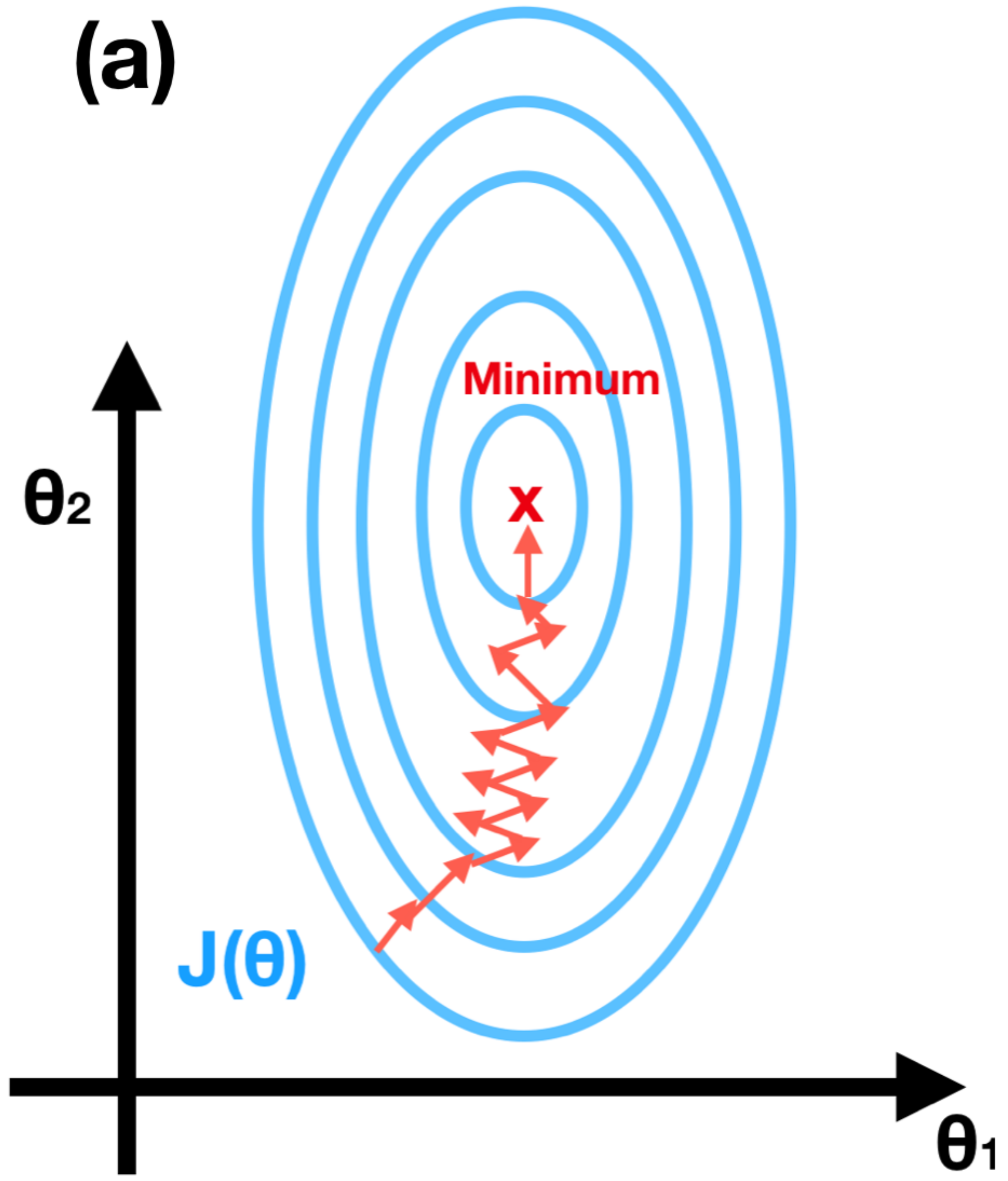
Παράδειγμα:

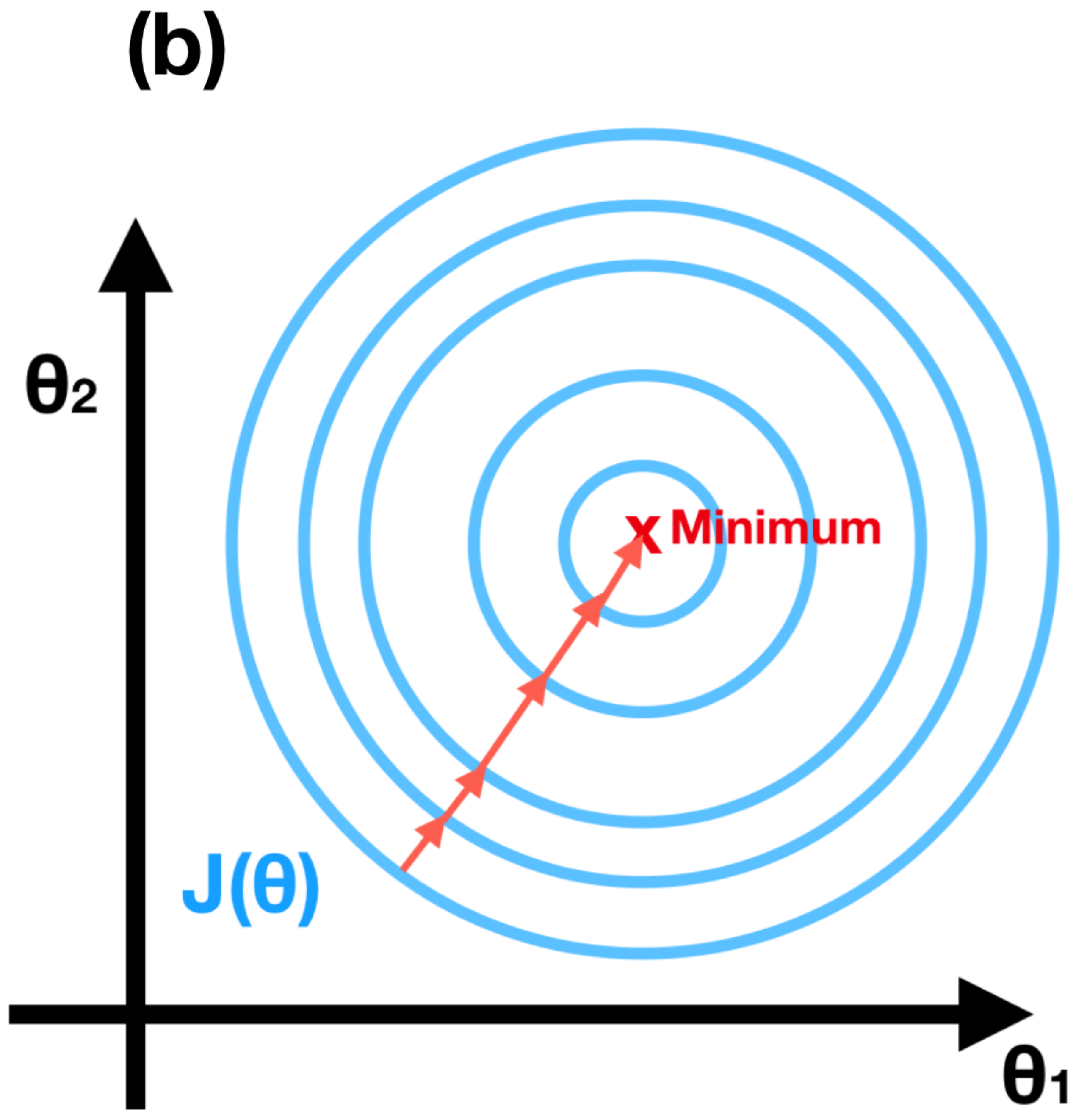
x1 = μέγεθος (0-2000 πόδια2)

x2 = αριθμός υπνοδωματίων (1-5)

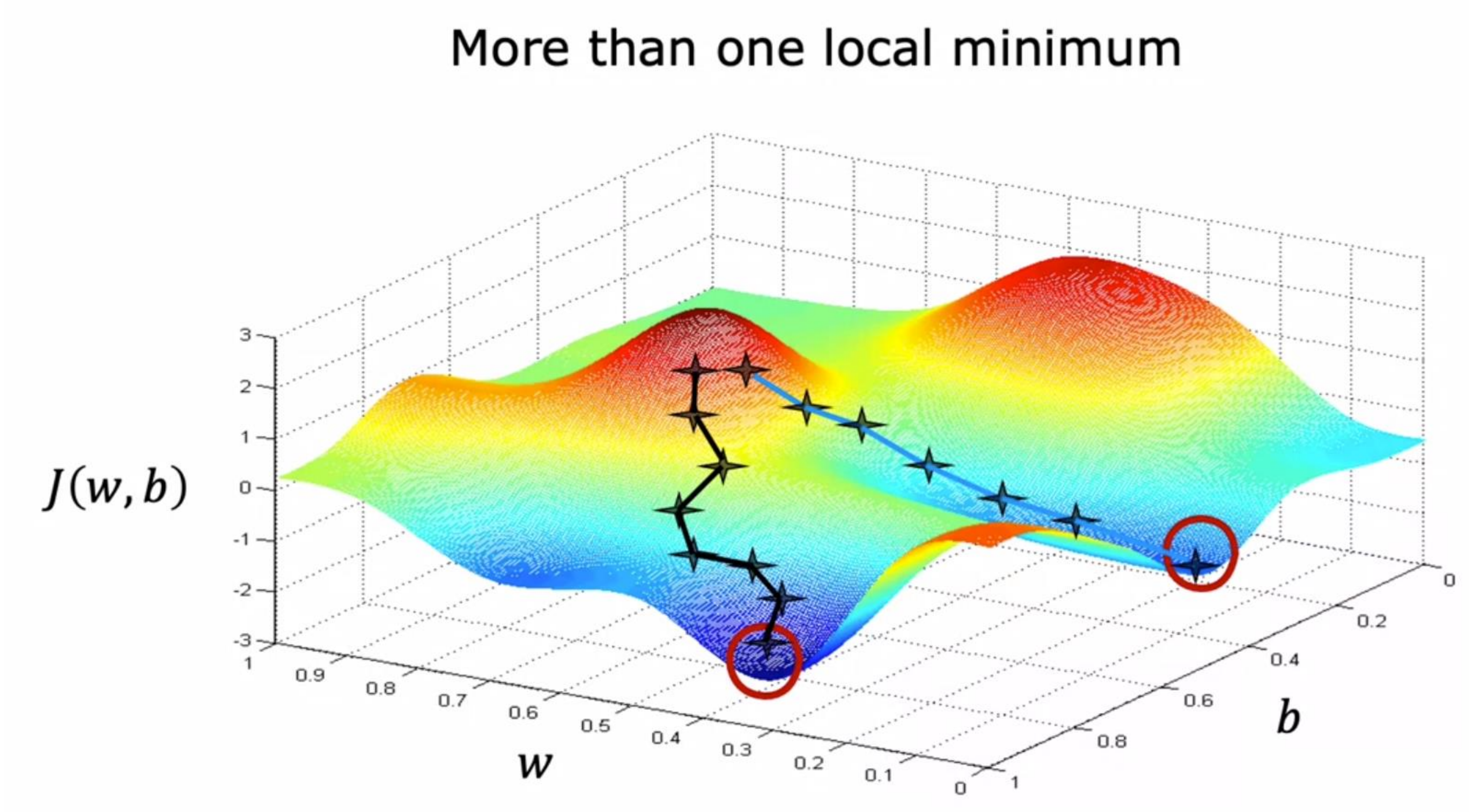
new_x1 = x1 /2000 new_x1 ∈ [0,1]

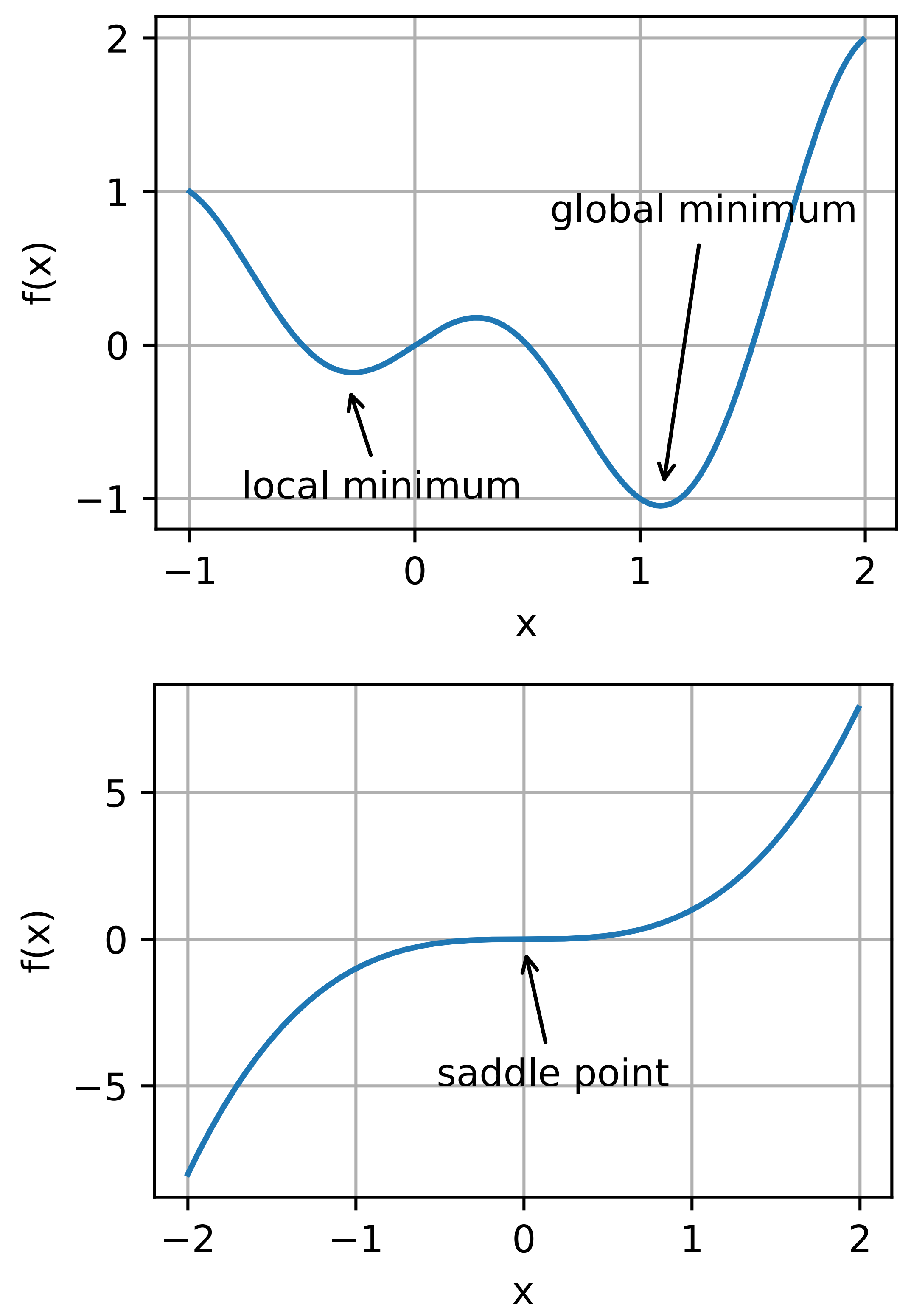
new_x2 = x2 /5 new_x2 ∈ [0,1]





Gradient Descent σε μια μη κυρτή 2D συνάρτηση





ΠΗΓΗ

Gradient Descent στο μέσο τετραγωνικό σφάλμα

Το ΜΤΣ όταν χρησιμοποιείται το **γραμμικό μοντέλο παλινδρόμησης** έχει περισσότερα από ένα ελάχιστα;

- Όχι — ένα ενιαίο παγκόσμιο ελάχιστο

Για τη συγκεκριμένη περίπτωση, υπάρχει μια αναλυτική (κλειστή μορφή/μη-δραστική) λύση.

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

Αναλυτική λύση: Κανονικές εξισώσεις

$$\theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n \approx y$$

x	y)
$x^{(1)}$	$y^{(1)}$
$x^{(2)}$	$y^{(2)}$
…	…
$x^{(m)}$	$y^{(m)}$

$$\underbrace{\begin{bmatrix} 1 & x_1^{(1)} & \ldots & x_n^{(1)} \\ 1 & x_1^{(2)} & \ldots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(m)} & \ldots & x_n^{(m)} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}}_{\theta} \approx \underbrace{\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}}_{y}$$

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

Σημαντικό: $m > n$

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

Παράδειγμα

```
In [1]: import numpy as np
   ...: from sklearn.linear_model import LinearRegression
   ...: X = np.array([[1, 1], [1, 2], [2, 2], [2, 3]])

In [2]: # y = 1 * x_0 + 2 * x_1 + 3
   ...: y = np.dot(X, np.array([1, 2])) + 3
   ...: reg = LinearRegression().fit(X, y)

In [3]: reg.score(X, y)
Out[3]: 1.0

In [4]: reg.coef_
Out[4]: array([1., 2.])

In [5]: reg.intercept_
Out[5]: 3.0000000000000018

In [6]: reg.predict(np.array([[3, 5]]))
Out[6]: array([16.])
```


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 παλινδρόμηση

Θετικά

- εύκολο στην υλοποίηση (αναλυτική λύση)
- μικρός χρόνος υπολογισμού για την προσαρμογή (αναλυτικό διάλυμα)
- σταθερός χρόνος πρόβλεψης (δεν αποθηκεύει σημεία δεδομένων)
- η αναλυτική λύση δεν χρειάζεται κανονικοποίηση χαρακτηριστικών
- λιγότερο ευαίσθητα σε θορυβώδη δεδομένα
- πιο ερμηνευσιμα

Αρνητικά

- Δεν μπορεί να χειρίστεί μη γραμμικές σχέσεις στα δεδομένα
- για μεγάλα σύνολα δεδομένων απαιτεί επαναληπτικές μεθόδους, π.χ. gradient descent
- Το gradient descent χρειάζεται κανονικοποίηση 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 παλινδρόμηση

Θετικά

- εύκολο στην υλοποίηση (αναλυτική λύση)
- μικρός χρόνος υπολογισμού για την προσαρμογή (αναλυτικό διάλυμα)
- σταθερός χρόνος πρόβλεψης (δεν αποθηκεύει σημεία δεδομένων)
- η αναλυτική λύση δεν χρειάζεται κανονικοποίηση χαρακτηριστικών
- λιγότερο ευαίσθητα σε θορυβώδη δεδομένα
- πιο ερμηνευσιμα

Αρνητικά

- Δεν μπορεί να χειρίστεί μη γραμμικές σχέσεις στα δεδομένα
- για μεγάλα σύνολα δεδομένων απαιτεί επαναληπτικές μεθόδους, π.χ. gradient descent
- Το gradient descent χρειάζεται κανονικοποίηση χαρακτηριστικών

Πώς μπορούμε να το διορθώσουμε αυτό;
... χωρίς να αλλάξει ο αλγόριθμος;

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

Πολυωνυμική παλινδρόμηση

1. Χρήση της μηχανικής χαρακτηριστικών για την κατασκευή πολυωνυμικών χαρακτηριστικών

π.χ. για 2 μεταβλητές
και πολυωνυμικός βαθμός = 2

$$x_1, x_2, x_1^2, x_2^2, x_1 x_2$$

2. Βρείτε τη λύση στο νέο πρόβλημα γραμμικής παλινδρόμησης

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2 \approx y$$

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

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