Μηχανική Μάθηση: Μαθηματικό Υπόβαθρο

Κωνσταντίνος Καραμανής

The University of Texas at Austin & Archimedes/Athena RC

constantine@utexas.edu

https://caramanis.github.io/

Ας θυμηθούμε τα προηγούμενα…

Εκπαίδευση Αλγορίθμου

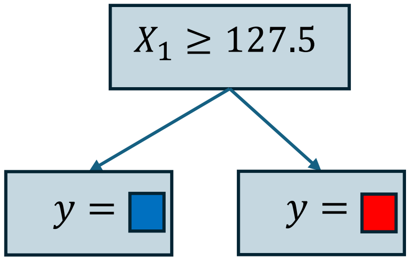
1. Επιλέγουμε την οικογένεια αλγορίθμων: παράδειγμα – δέντρα απόφασης βάθους 1

```
mymodel = DecisionTrees(max_depth=1)
```

2. Βρίσκουμε τις παραμέτρους που ελαχιστοποιούν τα λάθη στα δεδομένα μας

```
mymodel.fit(X,y)
mymodel.predict(x)
```

Δέντρο απόφασης με βάθος = 1



Εκπα
Αλγο

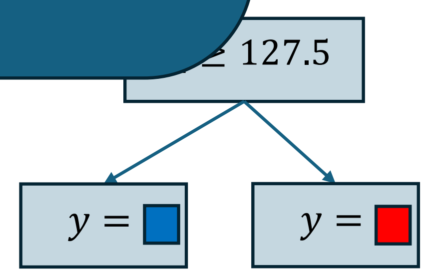
1. Επιλέγουμε την οικογένεια αλγορίθμων: παράδειγμα – δέντρα απόφασης βάθους 1

h=1)

Empirical Risk Minimization

να μας

σης με βάθος = 1



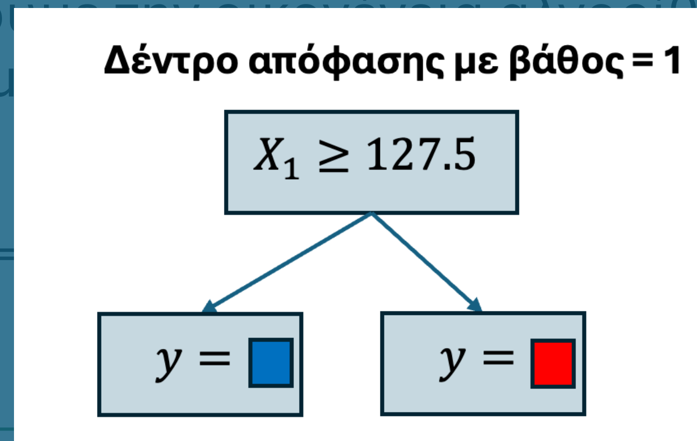
mymodel.predict(x)

Εκπαίδευση Αλγορίθμου

1. Επιλέγου... ...μων: παράδειγμ... ...υς 1

mymodel = ... =1)

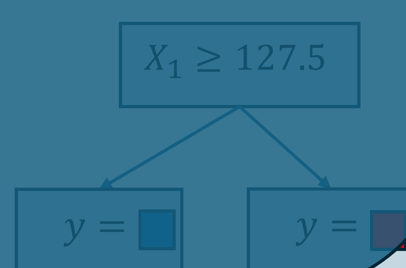
mymodel.fit(X,y)



2. Βρίσκουμε τις παραμέτρους που ελαχιστοποιούν τα λάθη στα δεδομένα μας

mymodel.fit(X,y)

mymodel.predict(x)

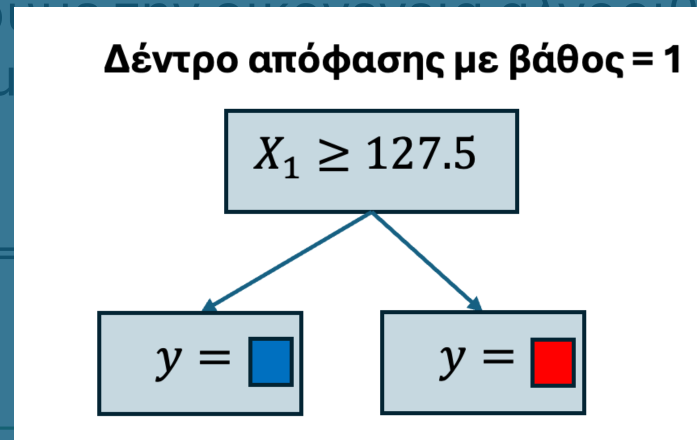


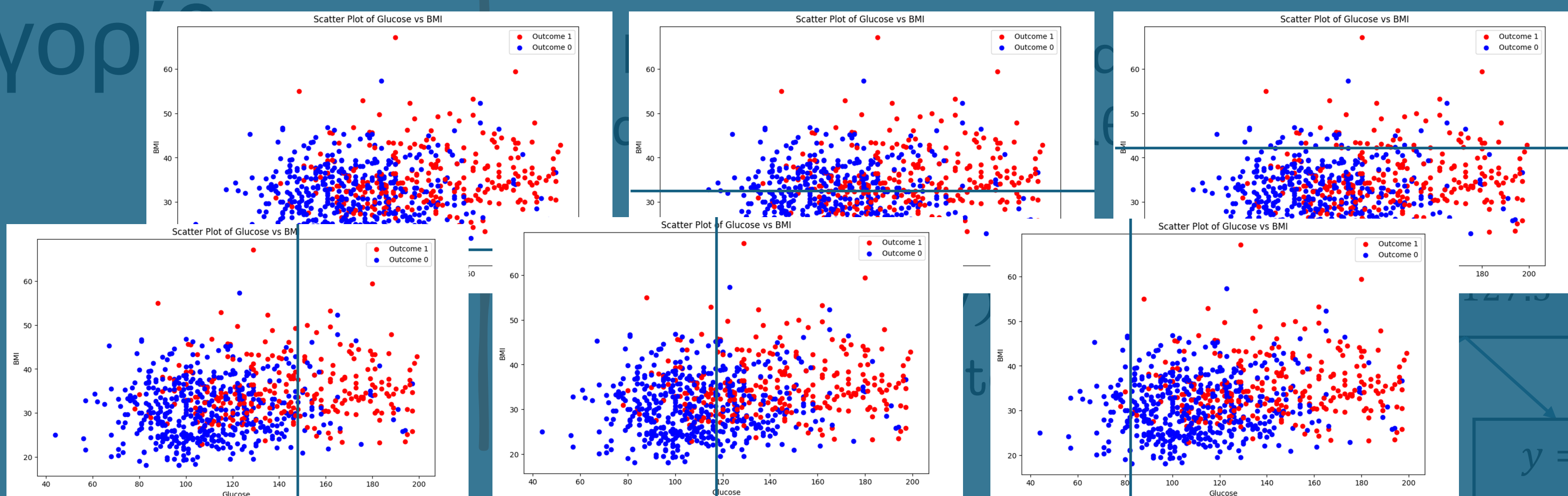
Εκπαίδευση
Αλγορ

1. Επιλέγο... παράδειγμ... υς 1

mymodel.fit(X,y)

mymodel = ... =1)





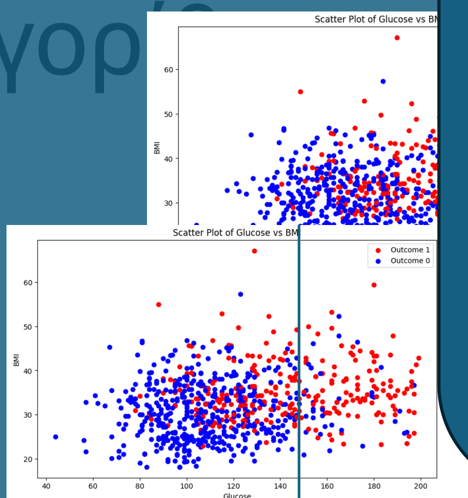
βάθος = 1

Εκπαίδευση

mymodel.fit(X,y)

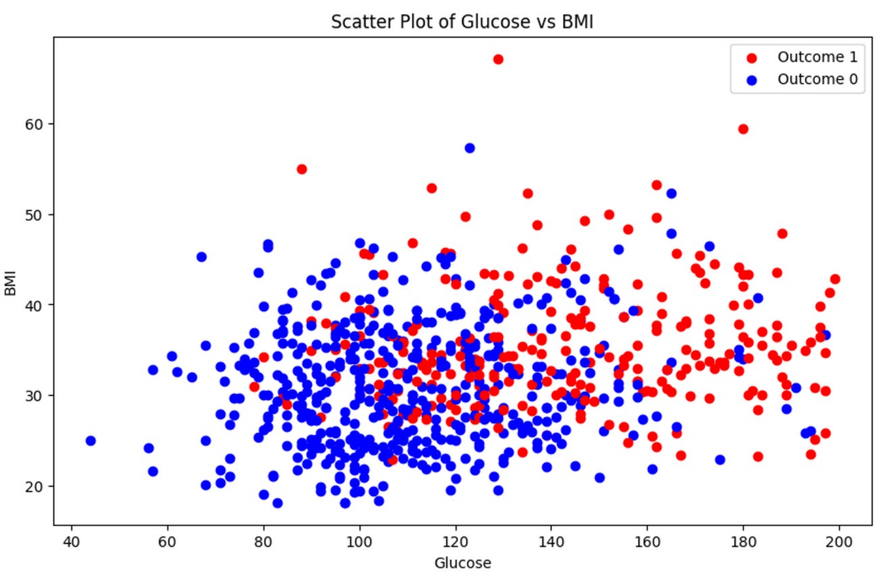
Δέντρο απόφασης με βάθος = 1

$X_1 \geq 127.5$



Το mymodel.fit(X,y) προσπαθεί να ελαχιστοποιήσει την απώλεια

Ελαχιστοποιώντας την Απώλεια



Άθροισμα

Σωστή ετικέτα (label)

Τι λέει το μοντέλο

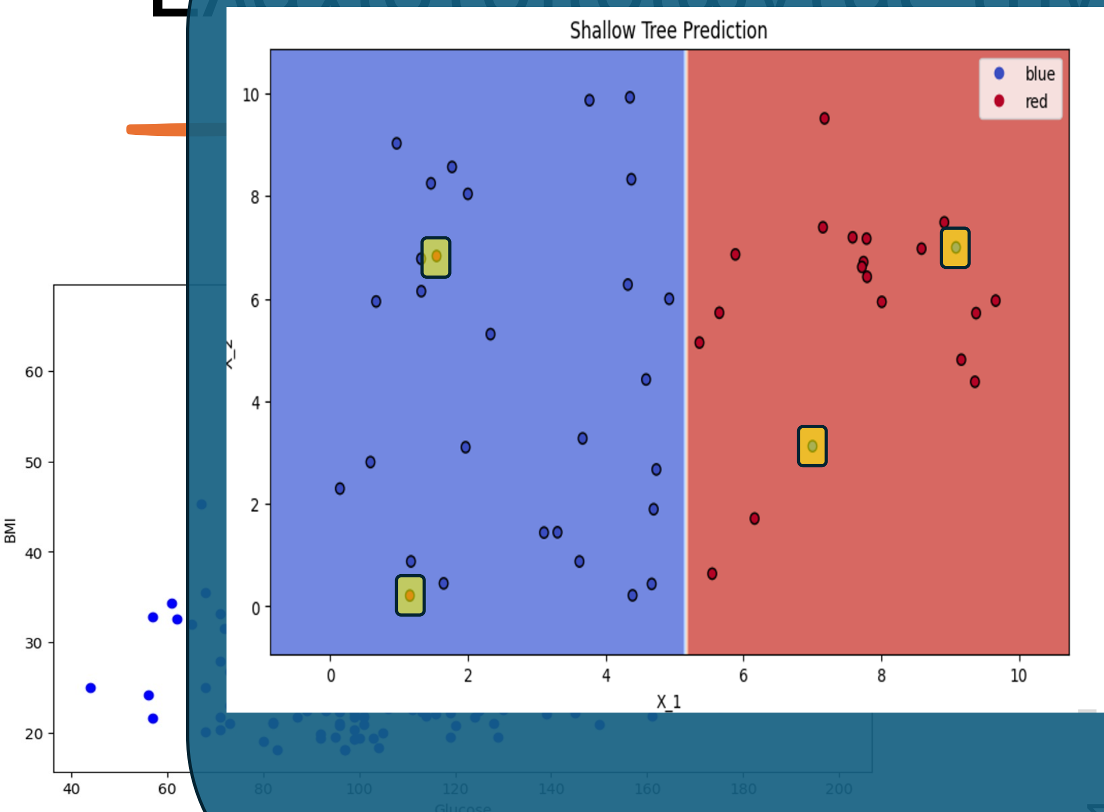
$$\sum \mathrm{Loss}(y_i, \widehat{y_i})$$

Συνάρτηση απώλειας

Ελαχιστοποιώντας την Απώλεια

$$\sum \mathrm{Loss}(\mathrm{y_i}, \widehat{y_i})$$

1. 0/1 – μετράμε λάθη



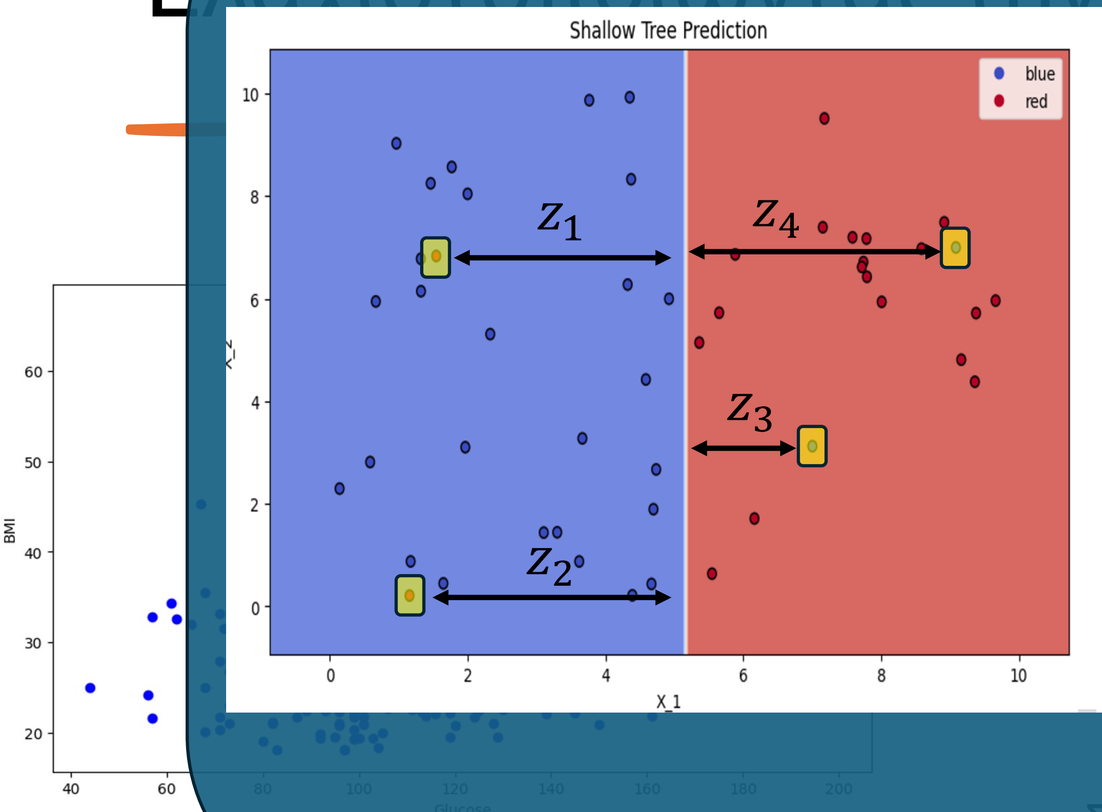
Σωστή ετικέτα (label)

Τι λέει το μοντέλο

$$\mathrm{Loss}(y_i, \widehat{y_i})$$

Συνάρτηση απώλειας

Ελαχιστοποιώντας την Απώλεια



$$\sum \text{Loss}(y_i, \hat{y}_i)$$

1. 0/1 – μετράμε λάθη
2. Απόσταση από το σωστό: $\sum z_i$

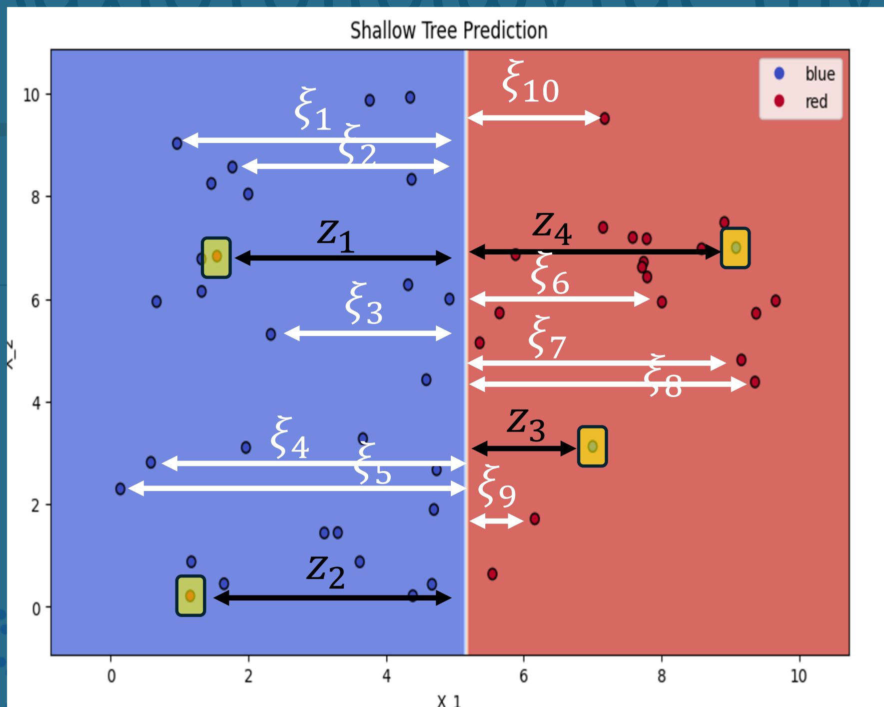
Σωστή ετικέτα (label)

...μοντέλο

$$\text{Loss}(y_i, \hat{y}_i)$$

Συνάρτηση απώλειας

$$\sum \mathrm{Loss}(\mathrm{y_i}, \hat{y}_i)$$



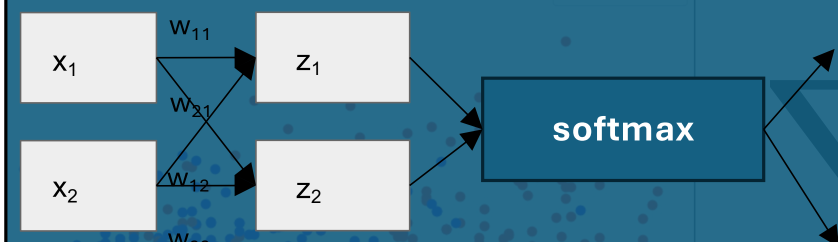
1. 0/1 – μετράμε λάθη
2. Απόσταση από το σωστό: $\sum \boldsymbol{z_i}$
3. Πόσο σίγουροι είμαστε για το σωστό ή το λάθος:

$$\sum e^{z_i} + \sum e^{-\xi_j}$$

4. Συνάρτηση Απώλειας της Λογιστικής Παλινδρόμησης (cross-entropy)

$y = (y_1, y_2)$ όπου $y_1 = 1, y_2 = 0$ για κατηγορία (ετικέτα) 1

$y = (y_1, y_2)$ όπου $y_1 = 0, y_2 = 1$ για κατηγορία (ετικέτα) 2



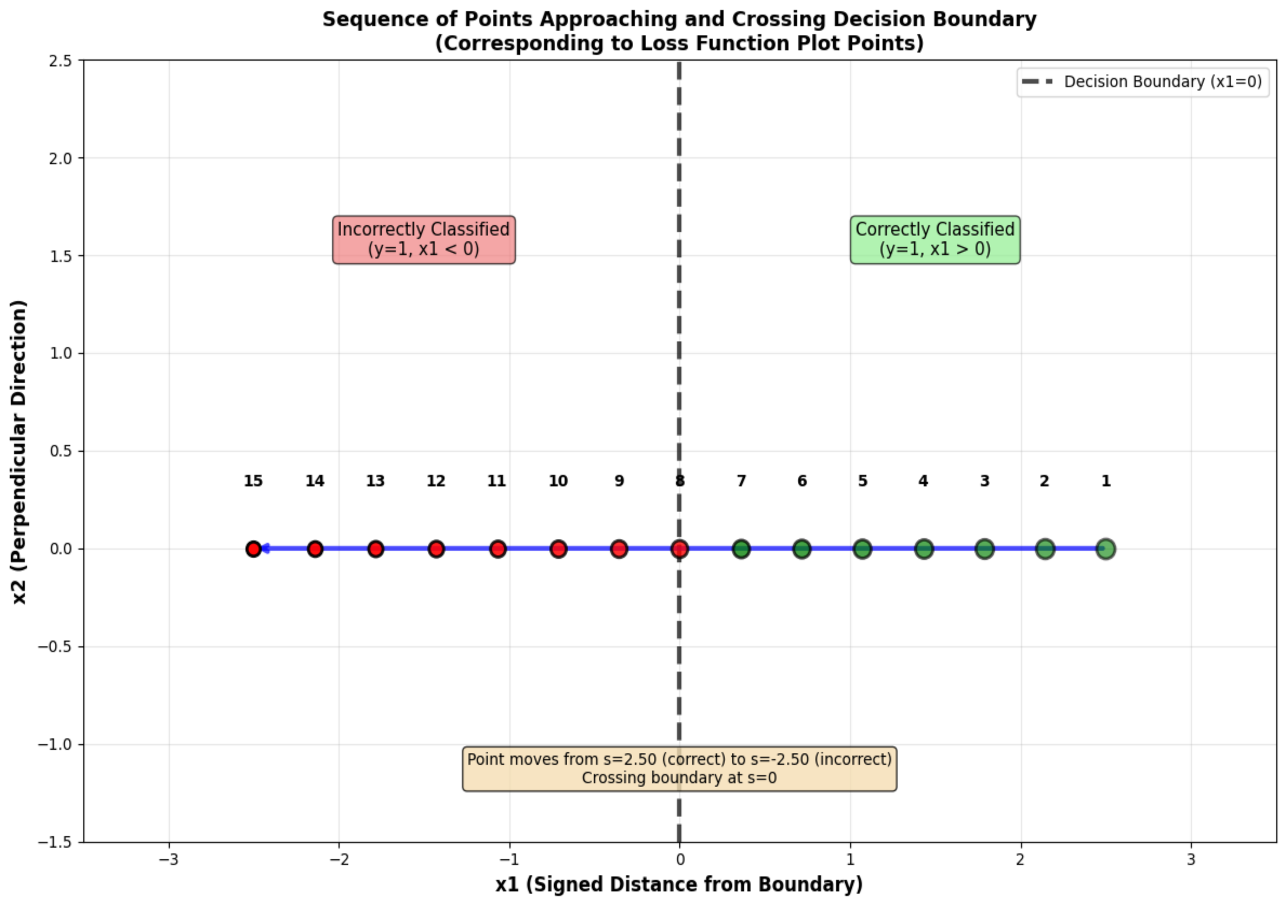
$$\hat{y}_1 = \frac{\exp(z_1)}{\exp(z_1)+\exp(z_2)} = \frac{\exp(w_1^{\top} x)}{\exp(w_1^{\top} x)+\exp(w_2^{\top} x)}$$

$$\hat{y}_2 = \frac{\exp(z_2)}{\exp(z_1)+\exp(z_2)} = \frac{\exp(w_2^{\top} x)}{\exp(w_1^{\top} x)+\exp(w_2^{\top} x)}$$

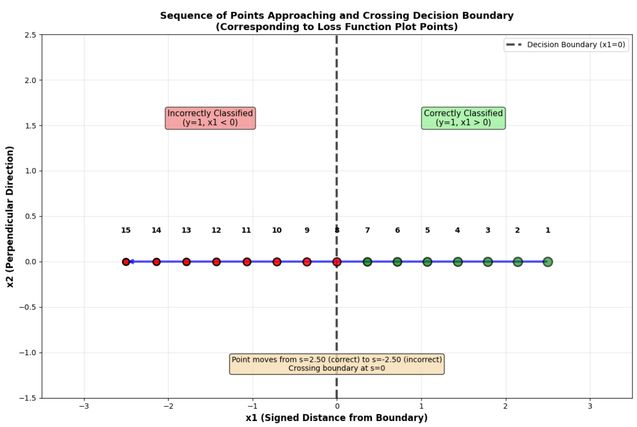
$$Loss\big((y_1, y_2), (\hat{y}_1, \hat{y}_2)\big) = -(y_1 \log(\hat{y}_1) + y_2 \log(\hat{y}_2))$$

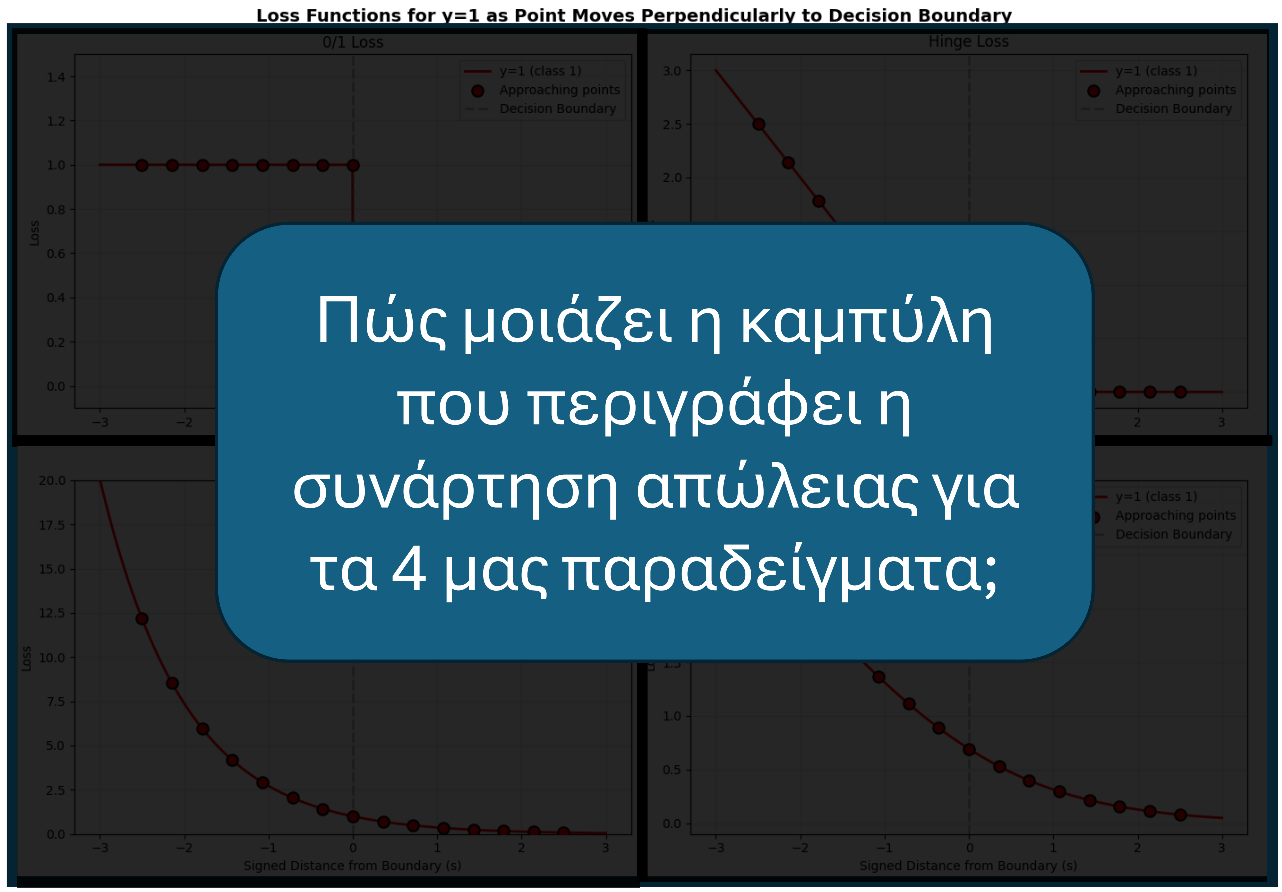
Πώς διαφέρουν οι 4 συναρτήσεις απώλειας

Τι απώλεια αντιστοιχεί σε σημεία που έχουν ταξινομηθεί σωστά ή λανθασμένα;



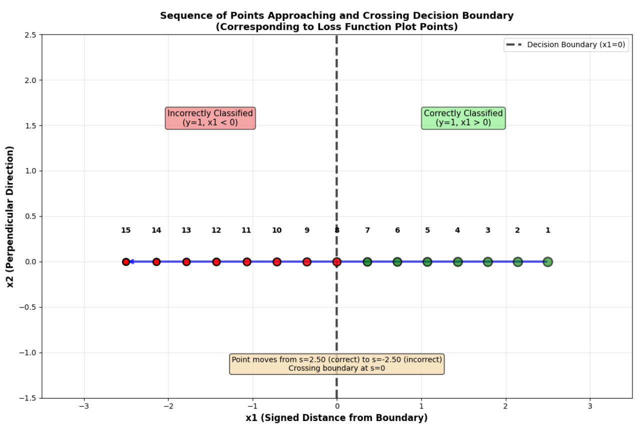
Πώς διαφέρουν οι 4 συναρτήσεις απώλειας

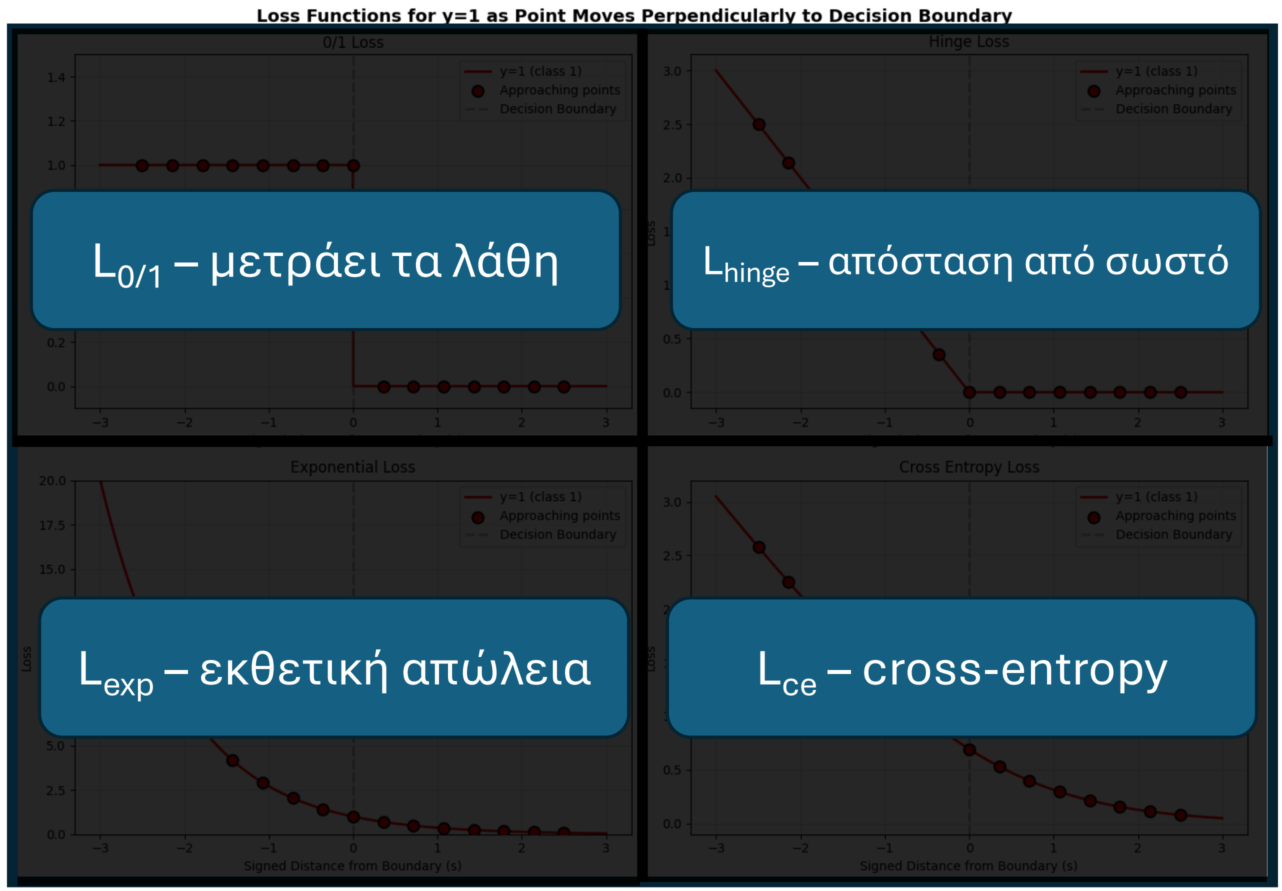




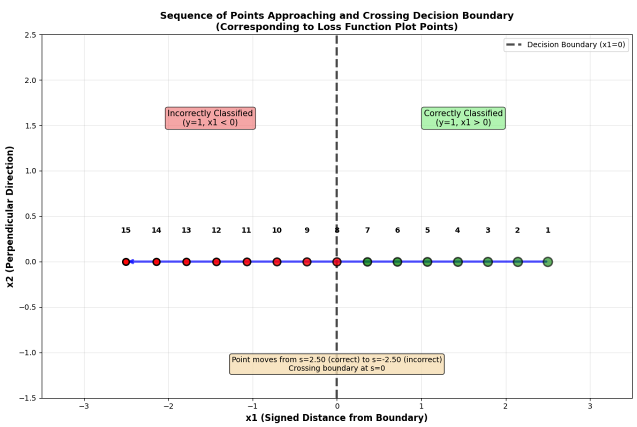
Πώς μοιάζει η καμπύλη που περιγράφει η συνάρτηση απώλειας για τα 4 μας παραδείγματα;

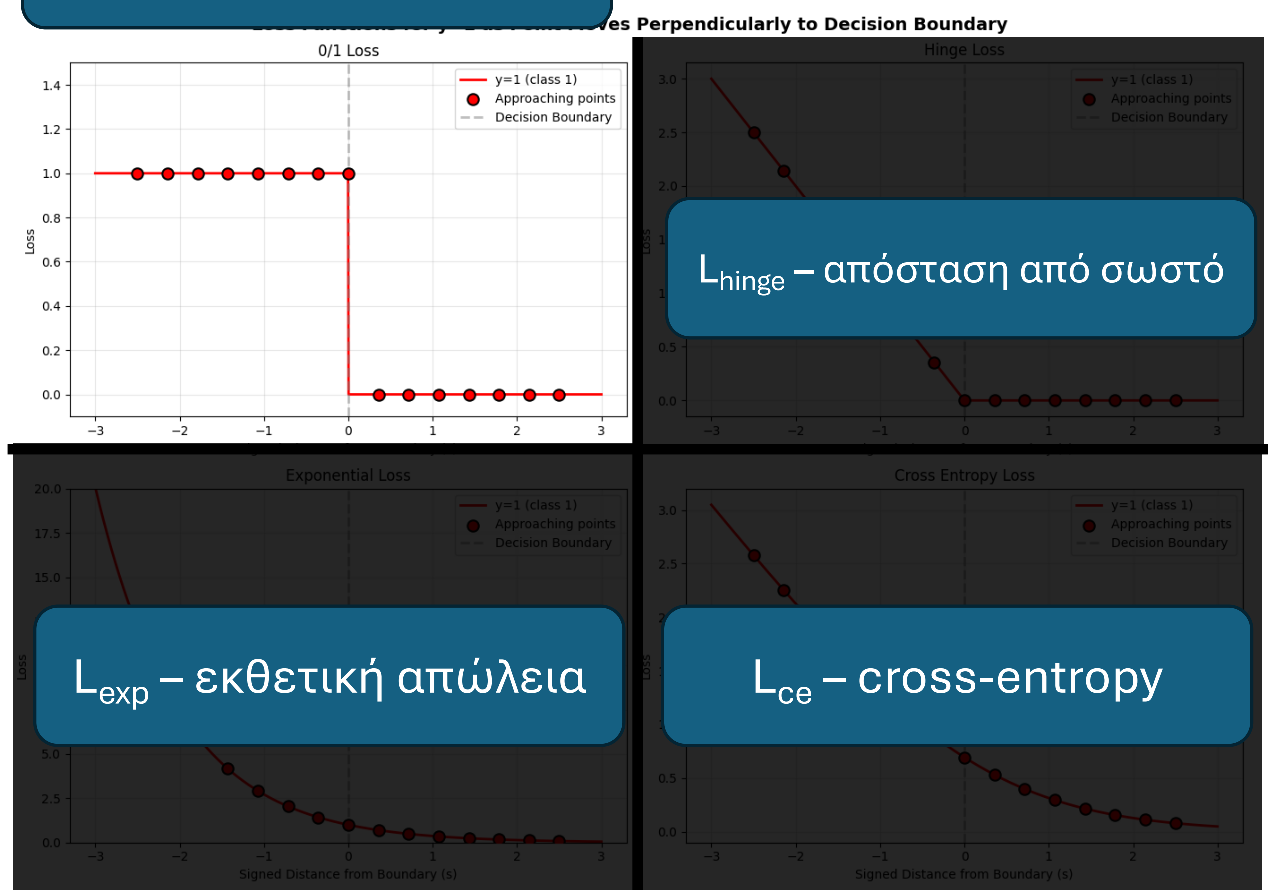
Πώς διαφέρουν οι 4 συναρτήσεις απώλειας





Πώς διαφέρουν οι 4 συναρτήσεις απώλειας





$L_{0/1}$ – μετράει τα λάθη

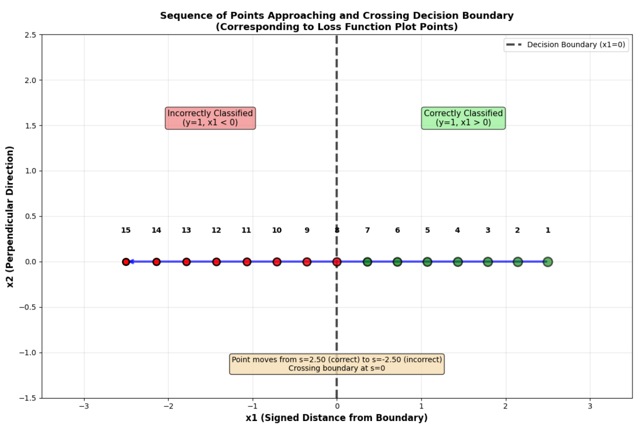
L_{hinge} – απόσταση από σωστό

L_{exp} – εκθετική απώλεια

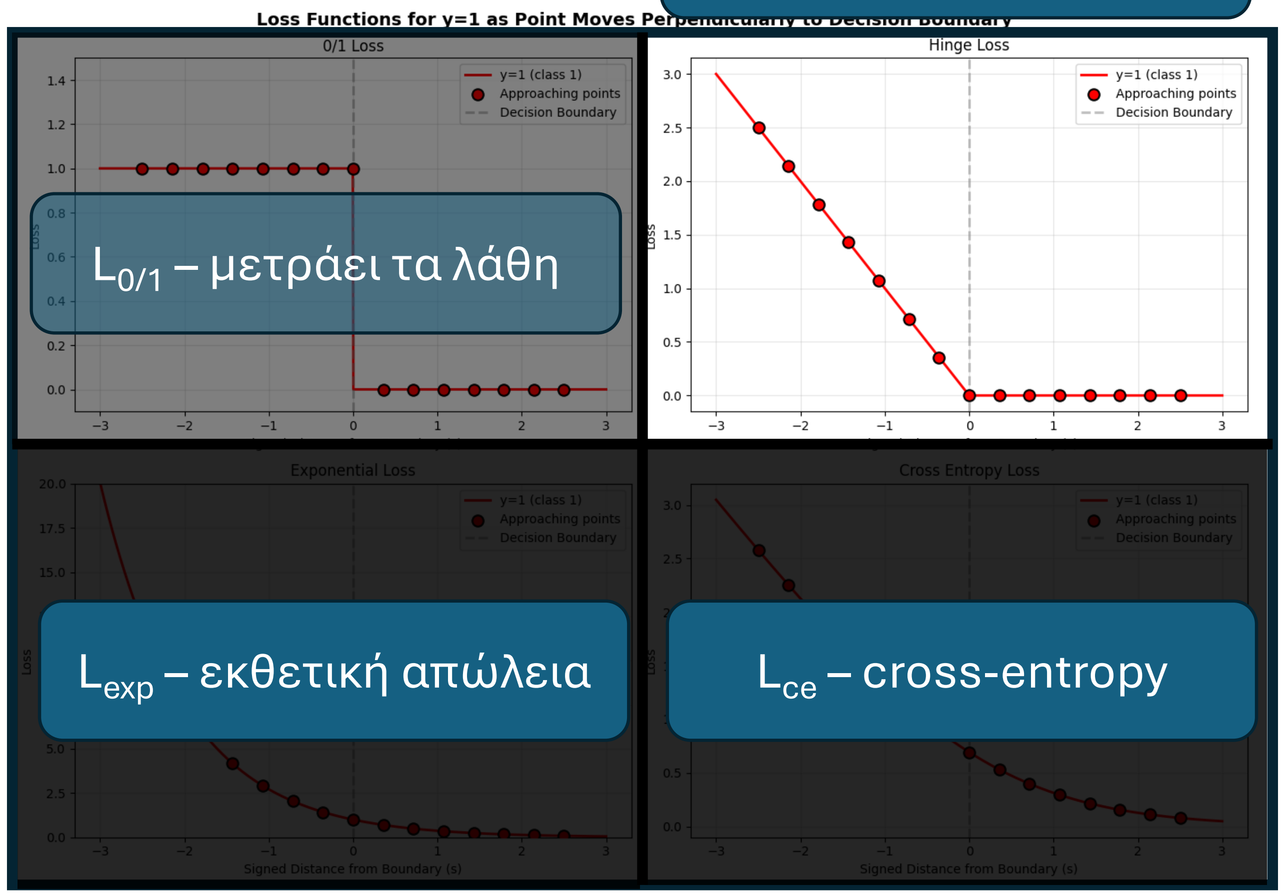
L_{ce} – cross-entropy

Κωνσταντίνος Καραμανής

Πώς διαφέρουν οι 4 συναρτήσεις απώλειας



L_{hinge} – απόσταση από σωστό

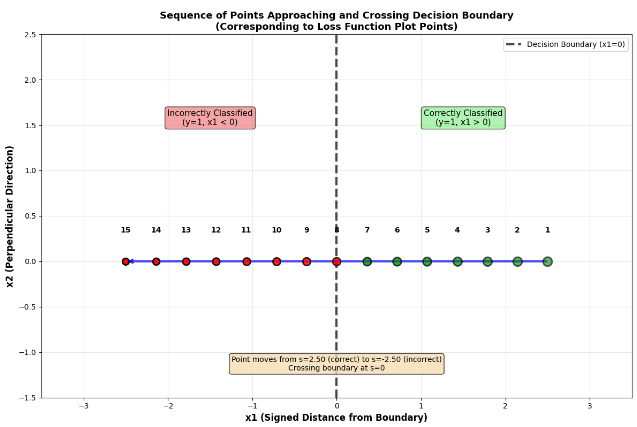


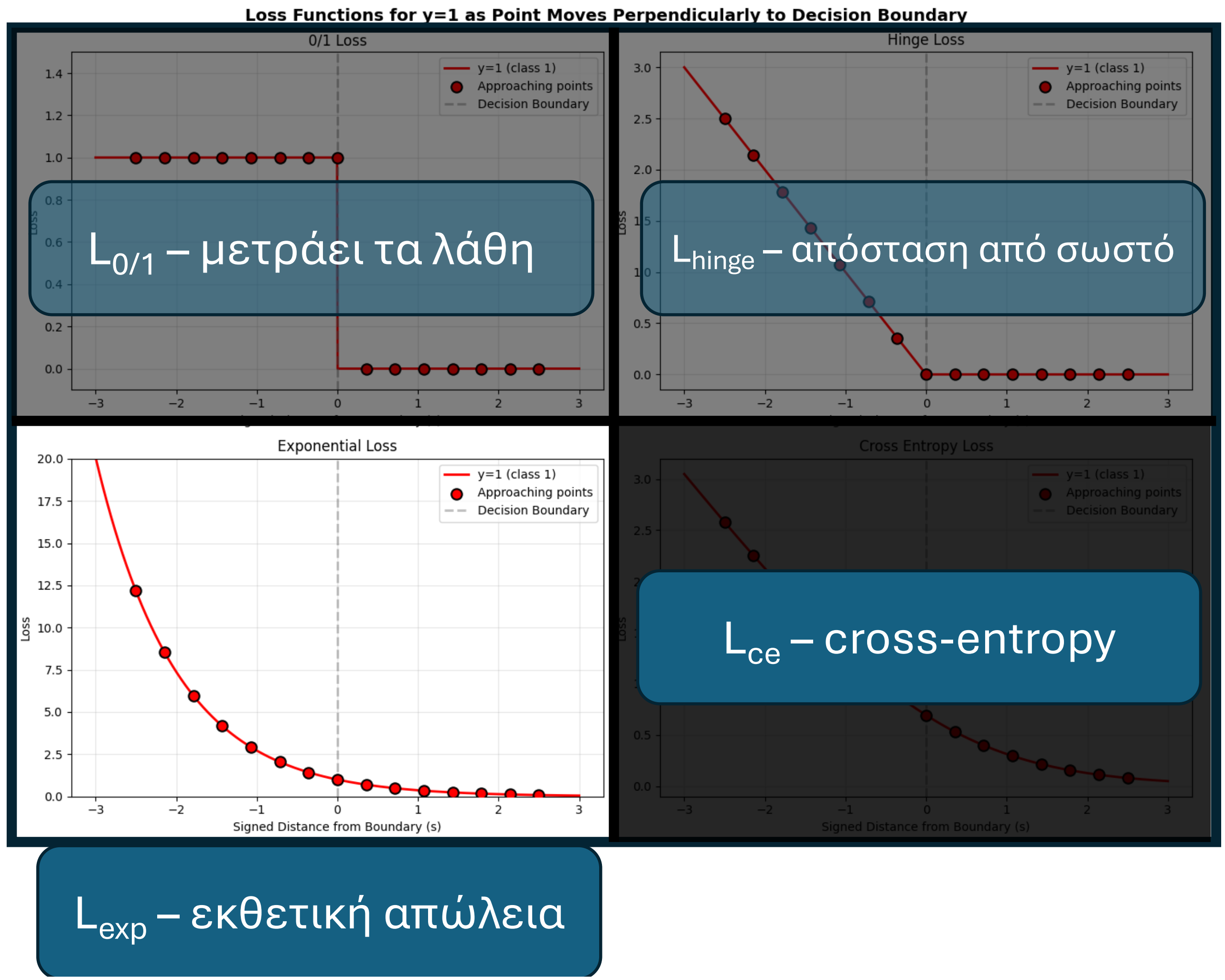
$L_{0/1}$ – μετράει τα λάθη

L_{exp} – εκθετική απώλεια

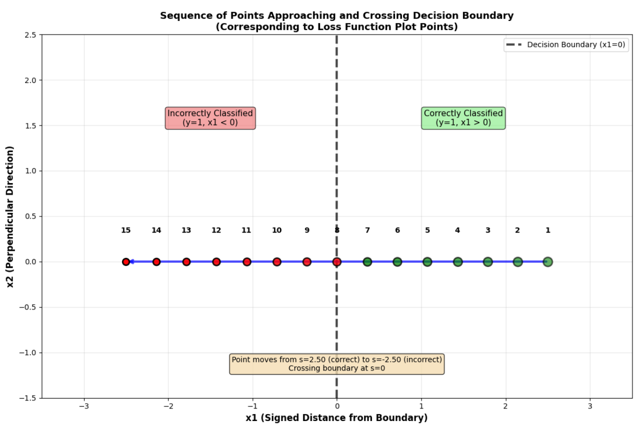
L_{ce} – cross-entropy

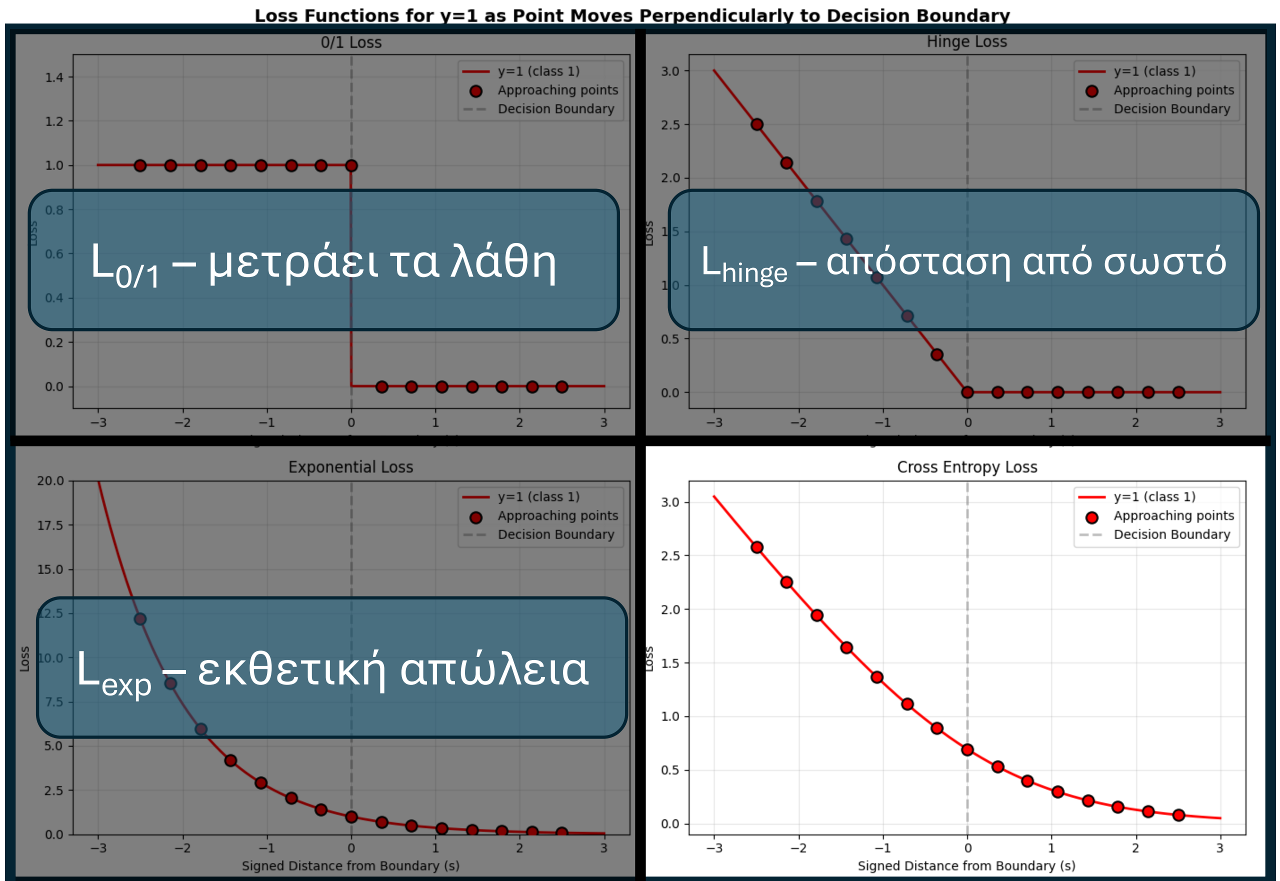
Πώς διαφέρουν οι 4 συναρτήσεις απώλειας





Πώς διαφέρουν οι 4 συναρτήσεις απώλειας





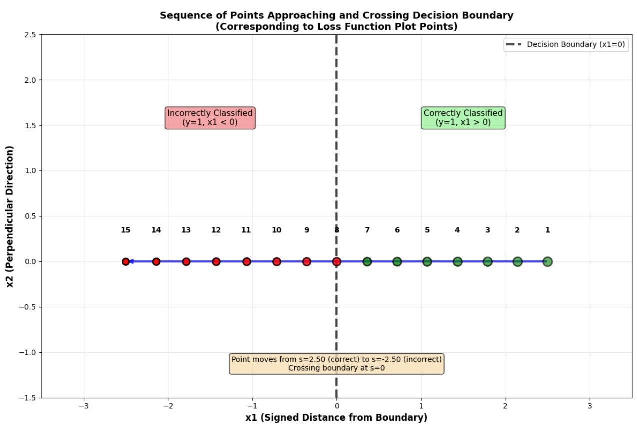
$L_{0/1}$ – μετράει τα λάθη

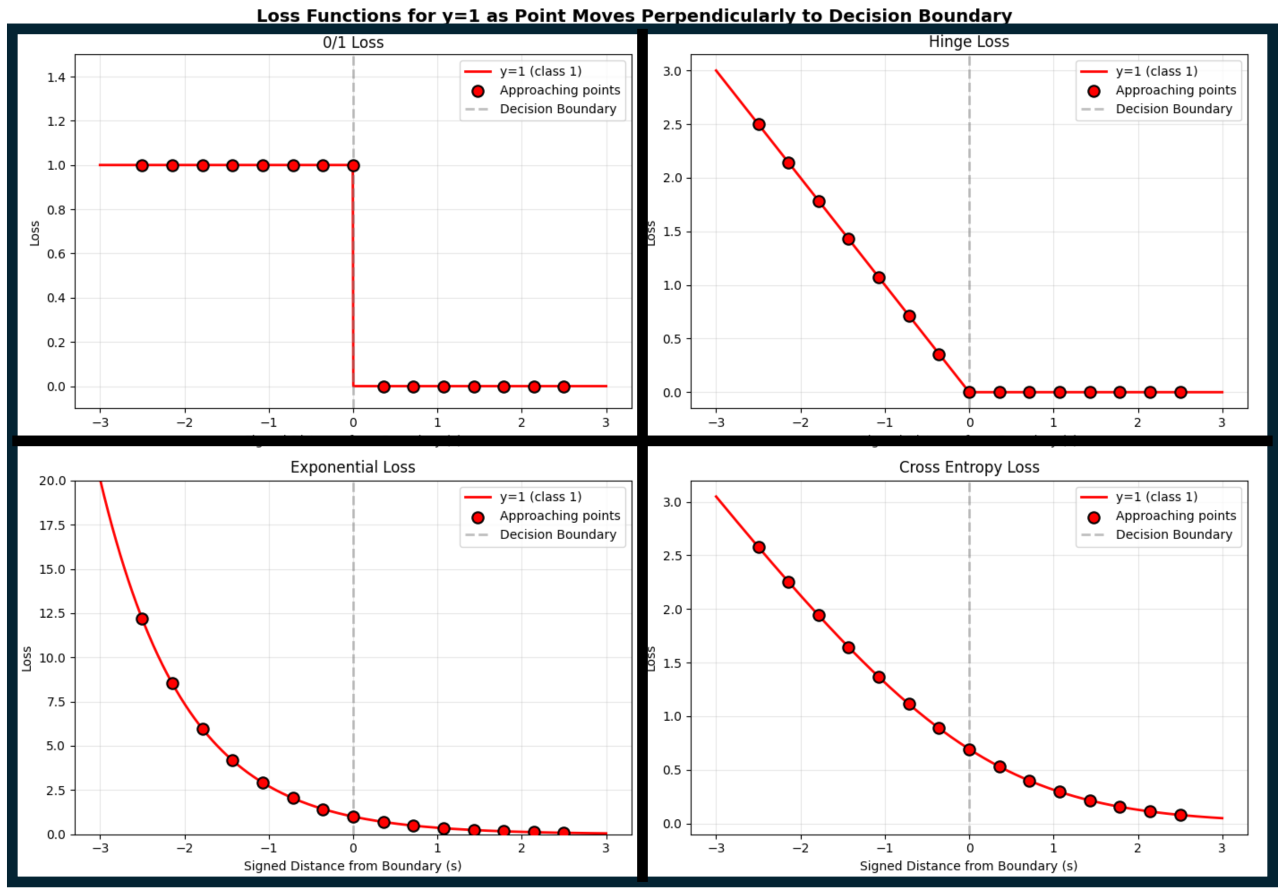
L_{hinge} – απόσταση από σωστό

L_{exp} – εκθετική απώλεια

L_{ce} – cross-entropy

Πώς διαφέρουν οι 4 συναρτήσεις απώλειας

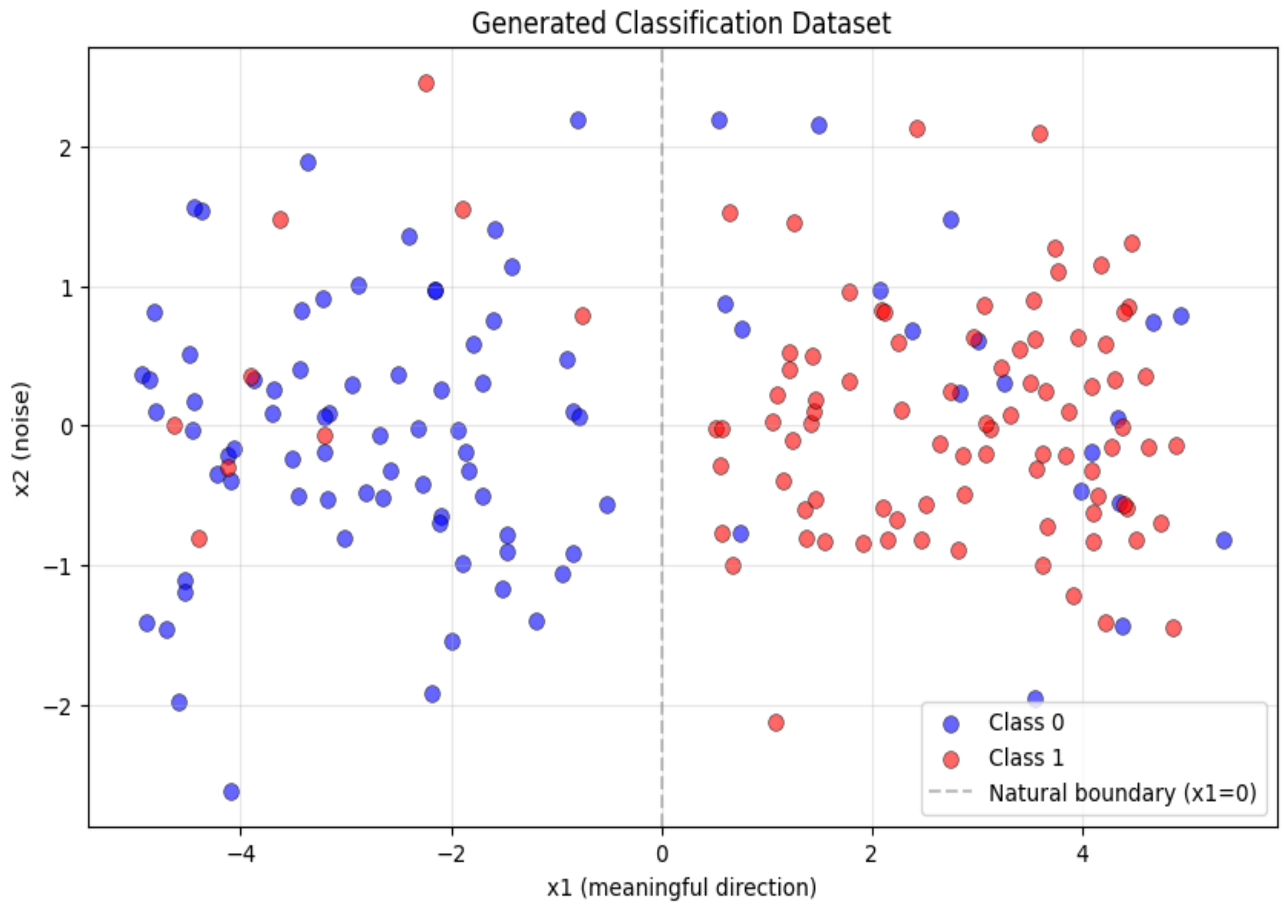




Κωνσταντίνος Καραμανής

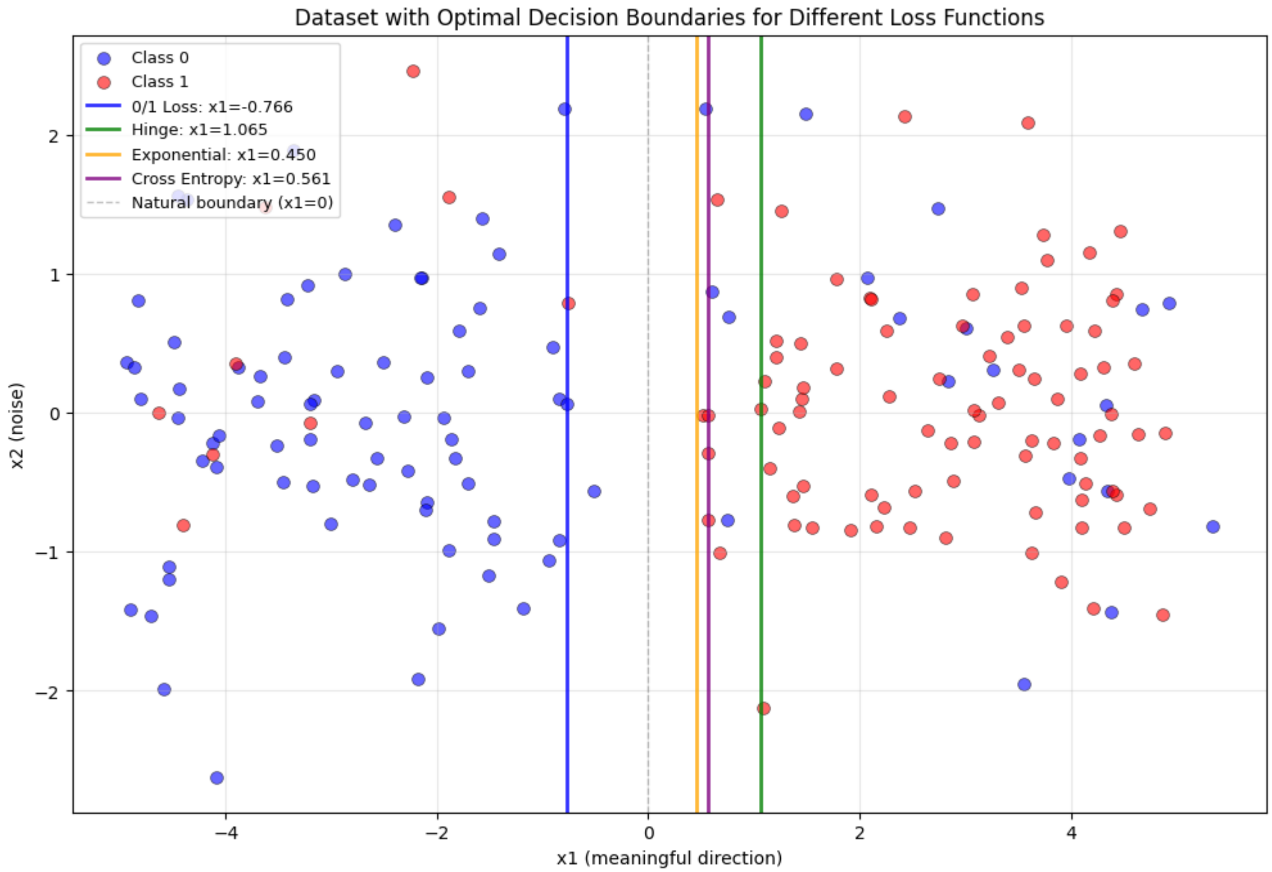
Πώς διαφέρουν οι 4 συναρτήσεις απώλειας

Όταν αλλάζουμε την συνάρτηση απώλειας, αλλάζουμε και την βέλτιστη λύση



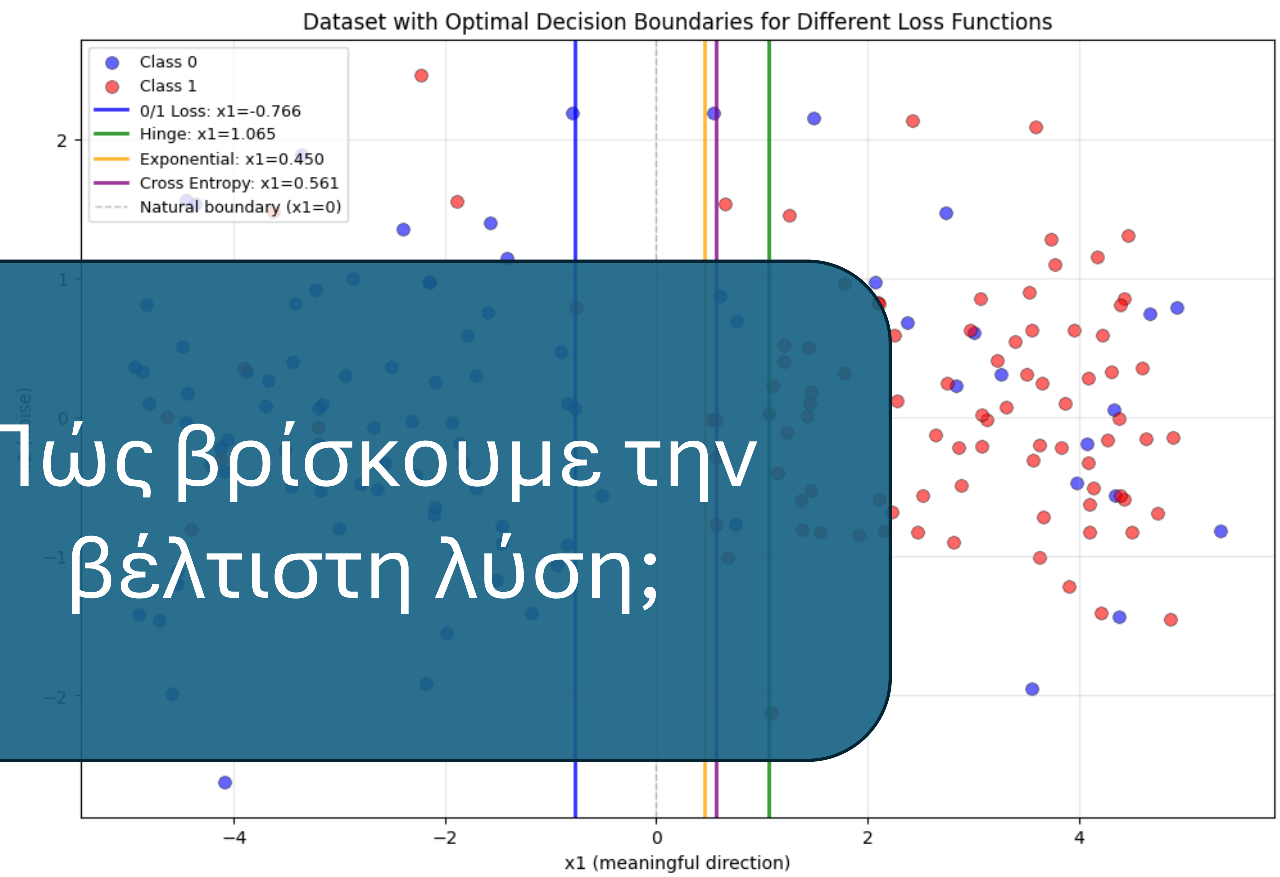
Πώς διαφέρουν οι 4 συναρτήσεις απώλειας

Όταν αλλάζουμε την συνάρτηση απώλειας, αλλάζουμε και την βέλτιστη λύση



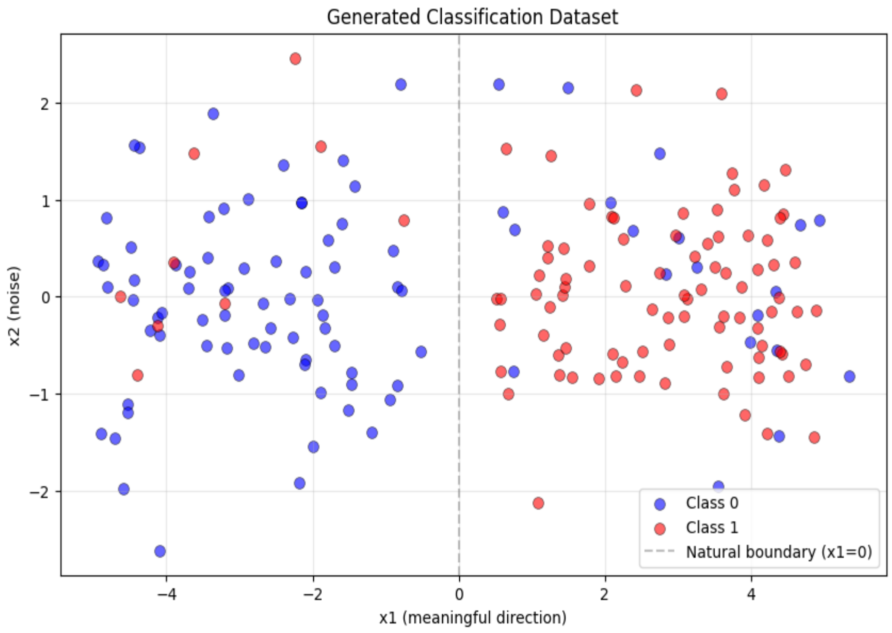
Πώς διαφέρουν οι 4 συναρτήσεις απώλειας

Όταν αλλάζουμε την συνάρτηση απώλειας, αλλάζουμε και την βέλτιστη λύση

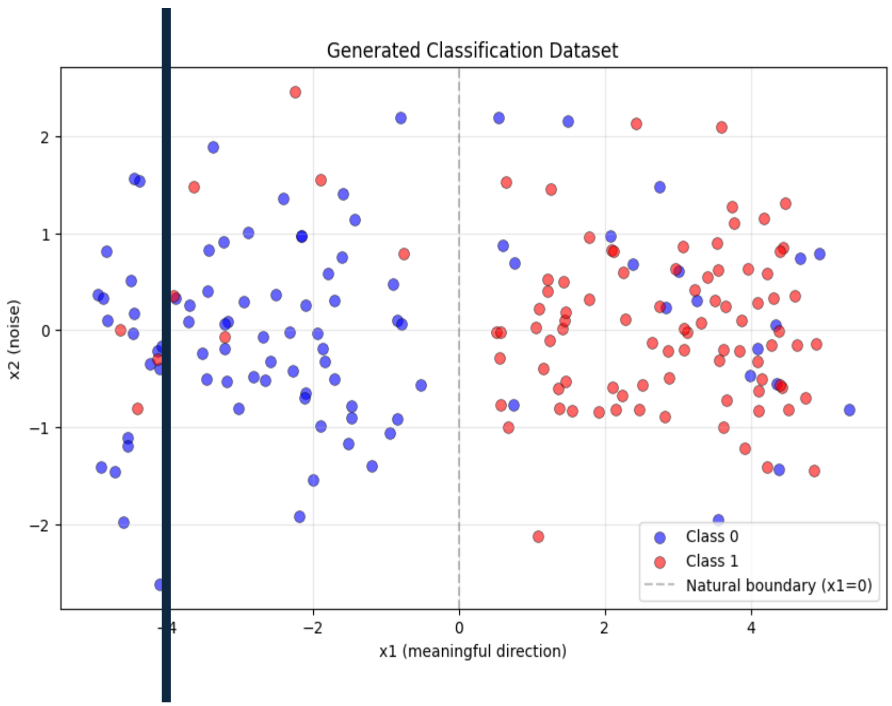


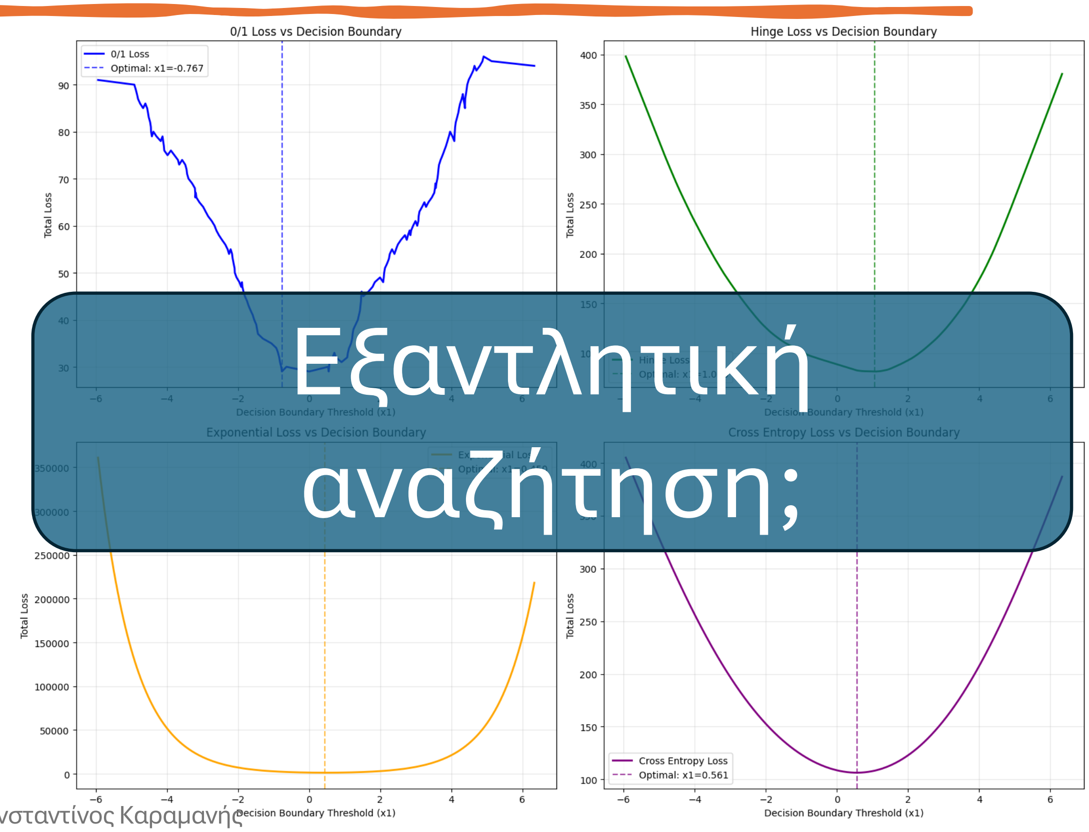
Πώς βρίσκουμε την βέλτιστη λύση;

Πώς βρίσκουμε την βέλτιστη λύση;



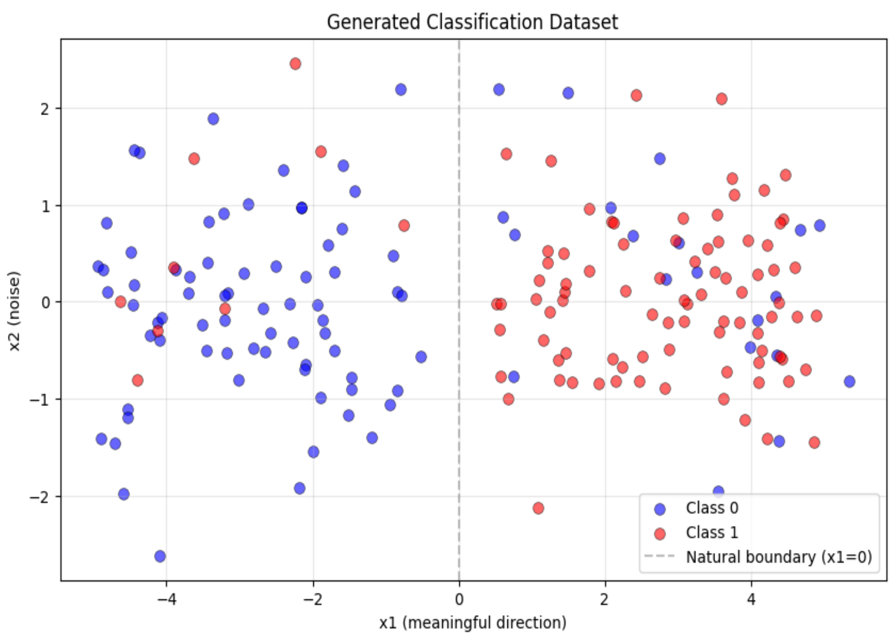
Πώς βρίσκουμε την βέλτιστη λύση;

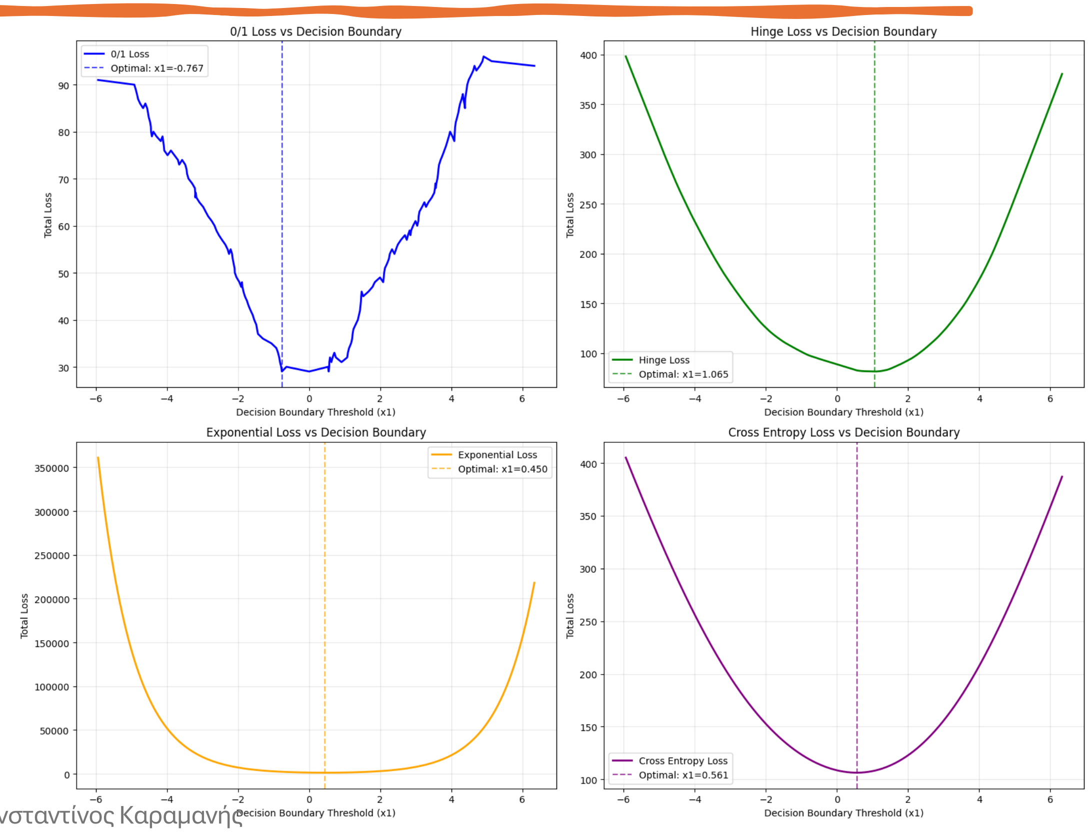




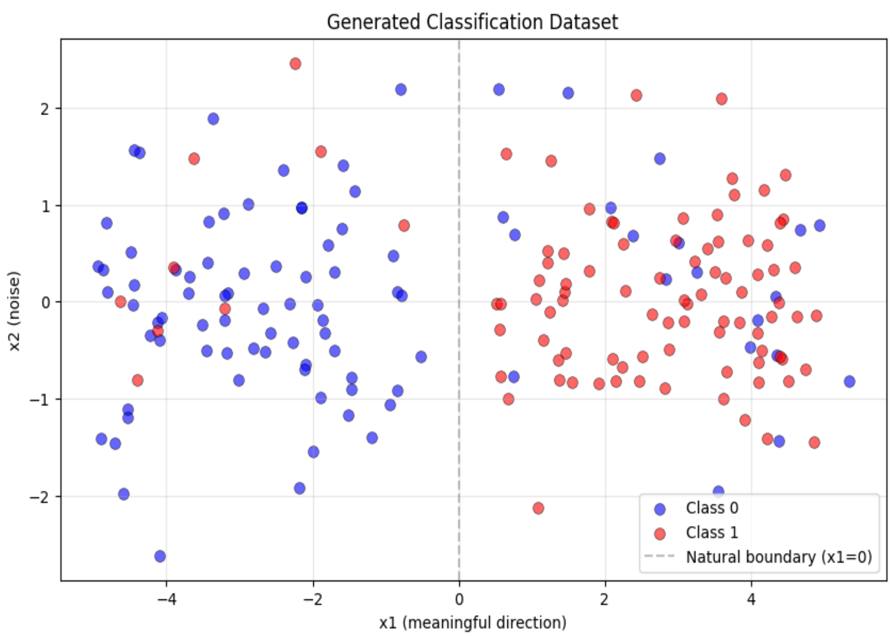
Εξαντλητική αναζήτηση;

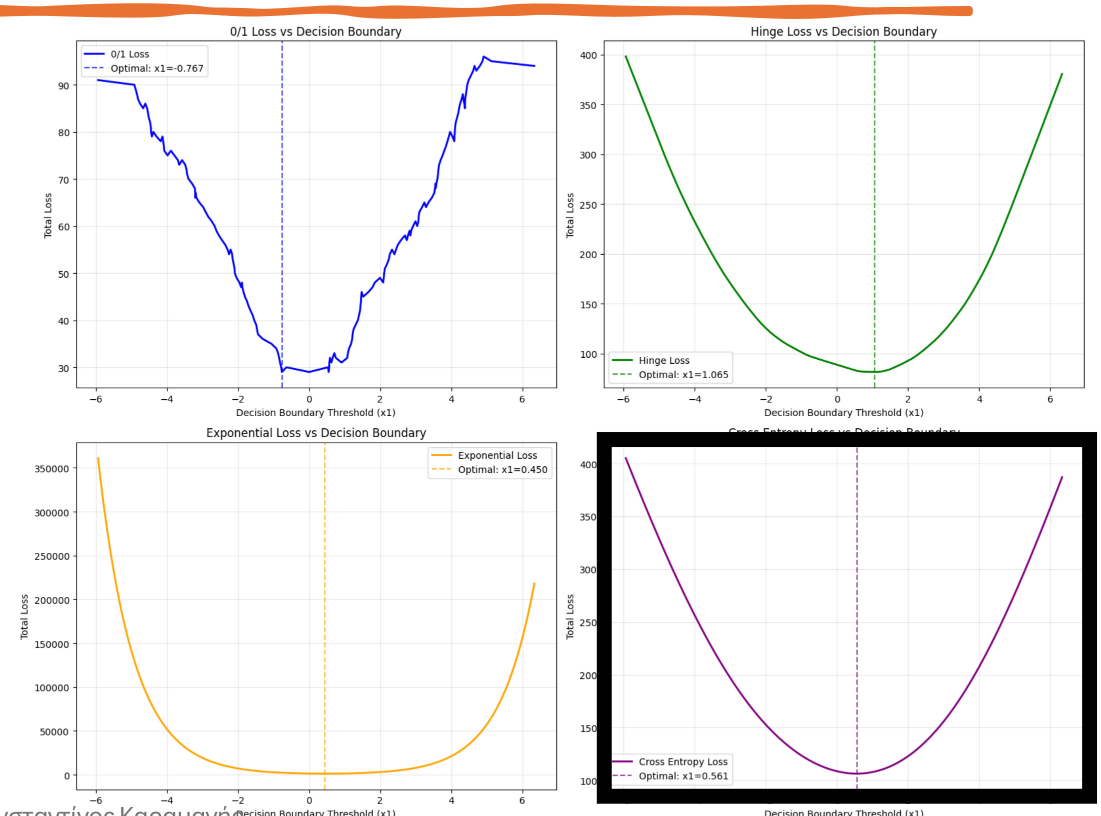
Πώς βρίσκουμε την βέλτιστη λύση;



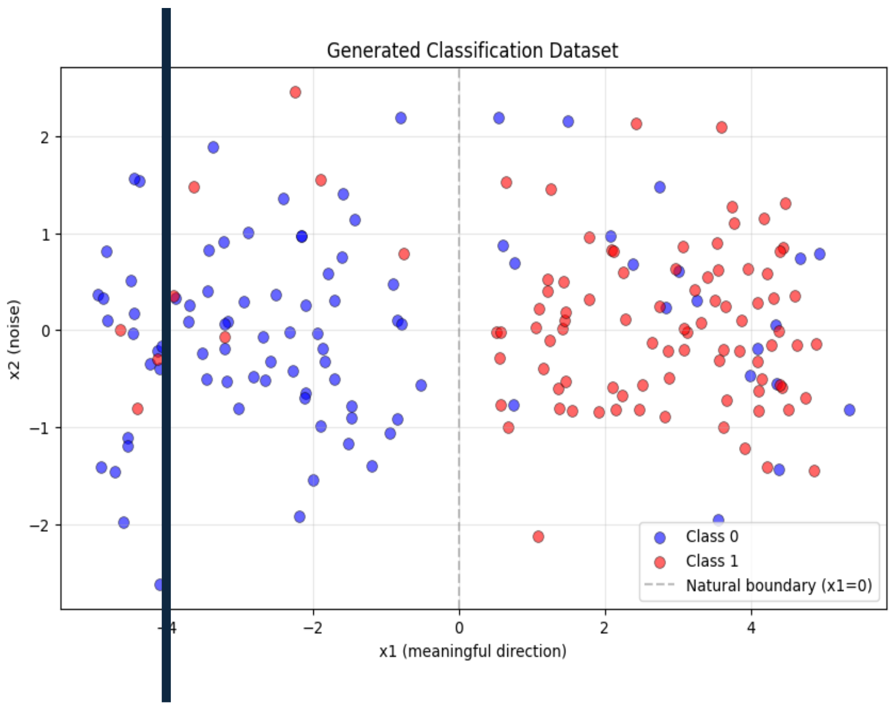


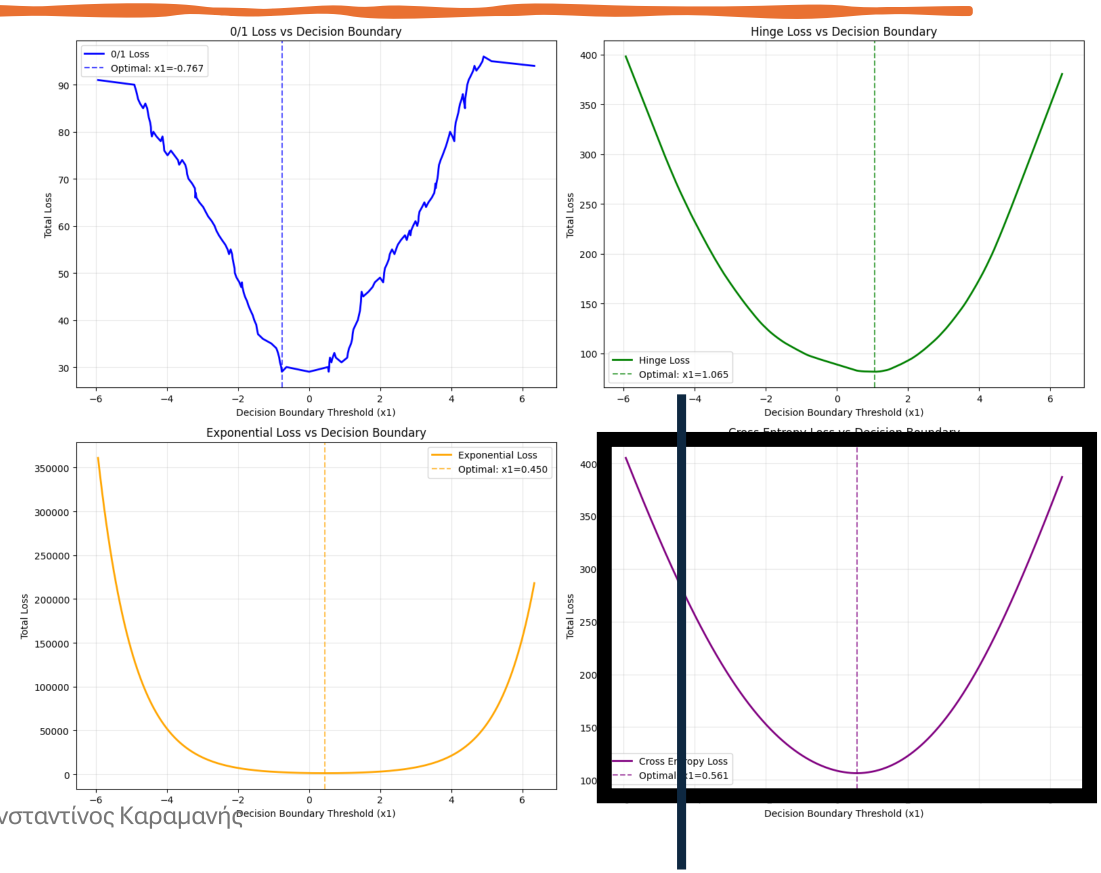
Πώς βρίσκουμε την βέλτιστη λύση;





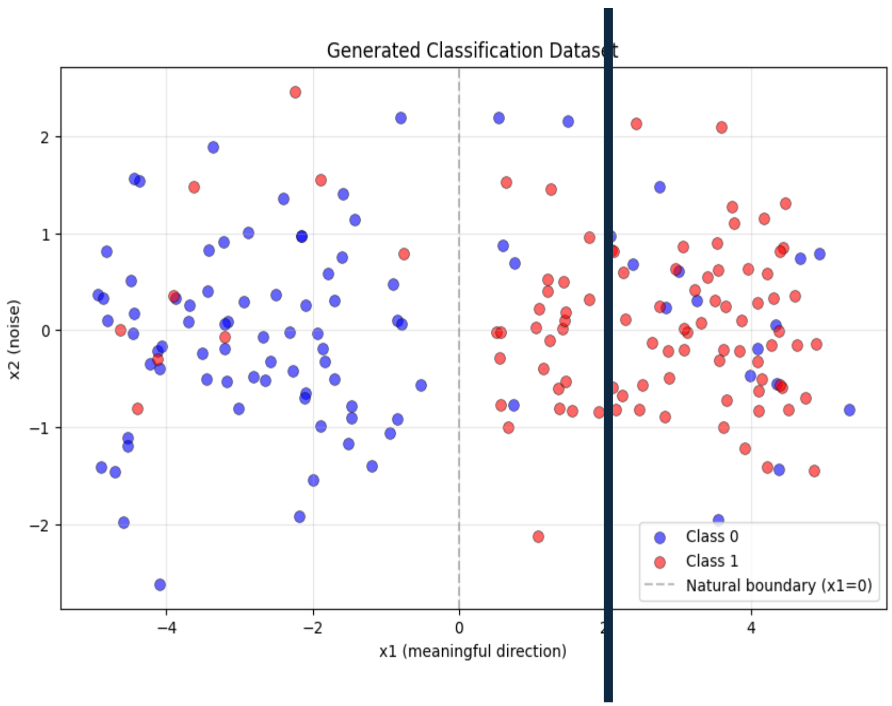
Πώς βρίσκουμε την βέλτιστη λύση;

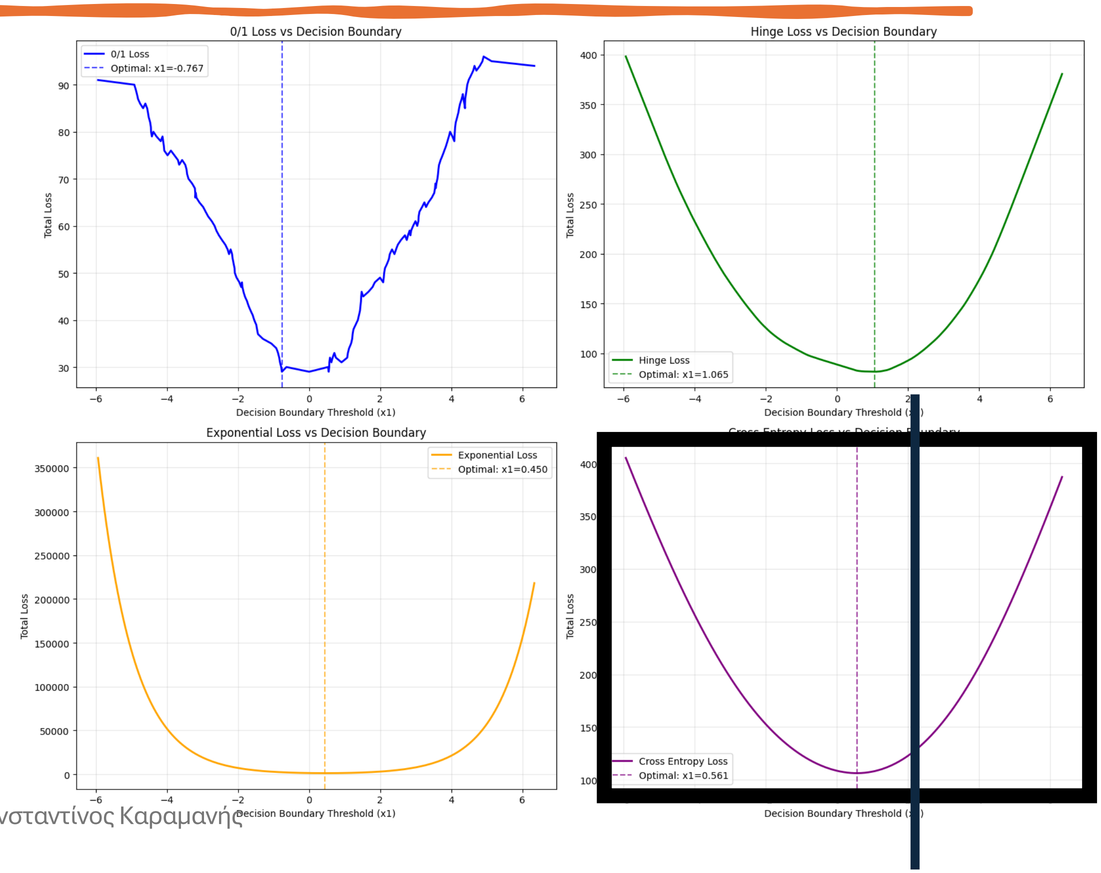




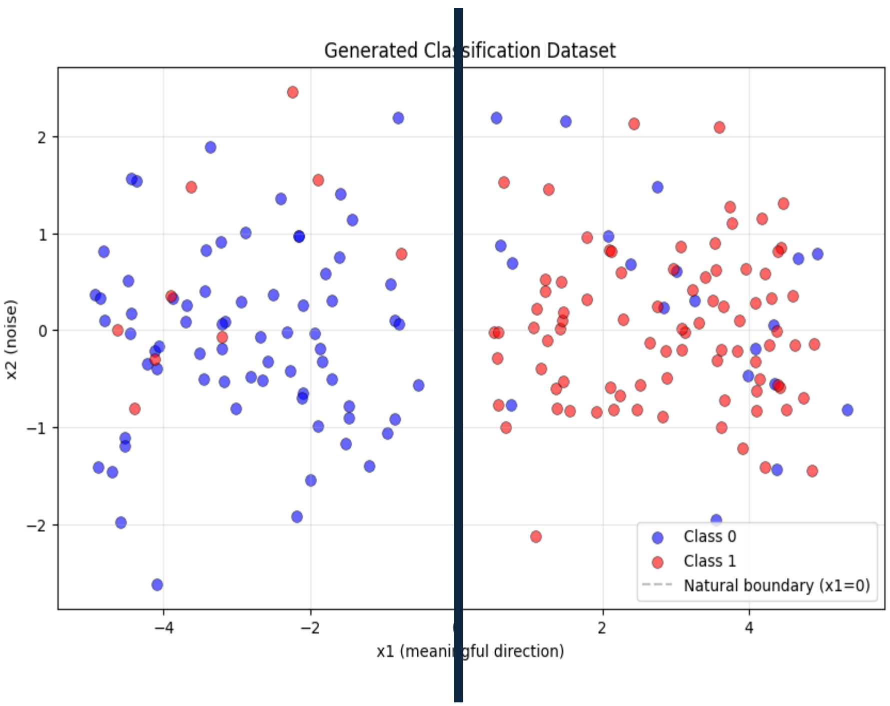
Κωνσταντίνος Καραμανής

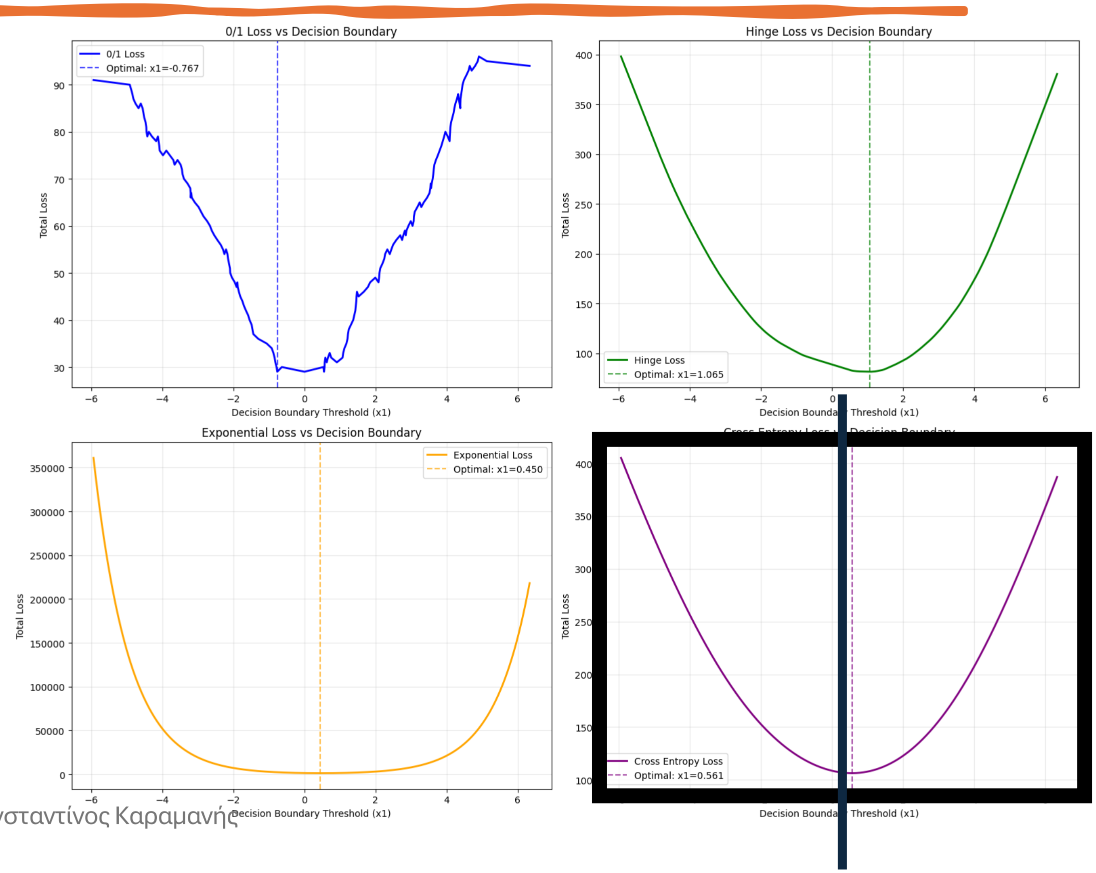
Πώς βρίσκουμε την βέλτιστη λύση;



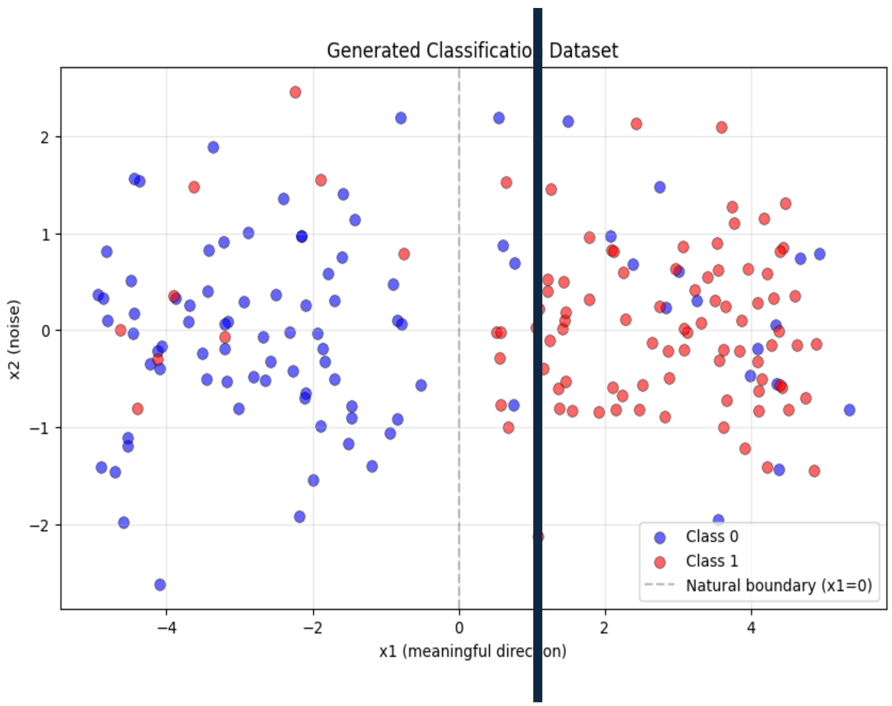


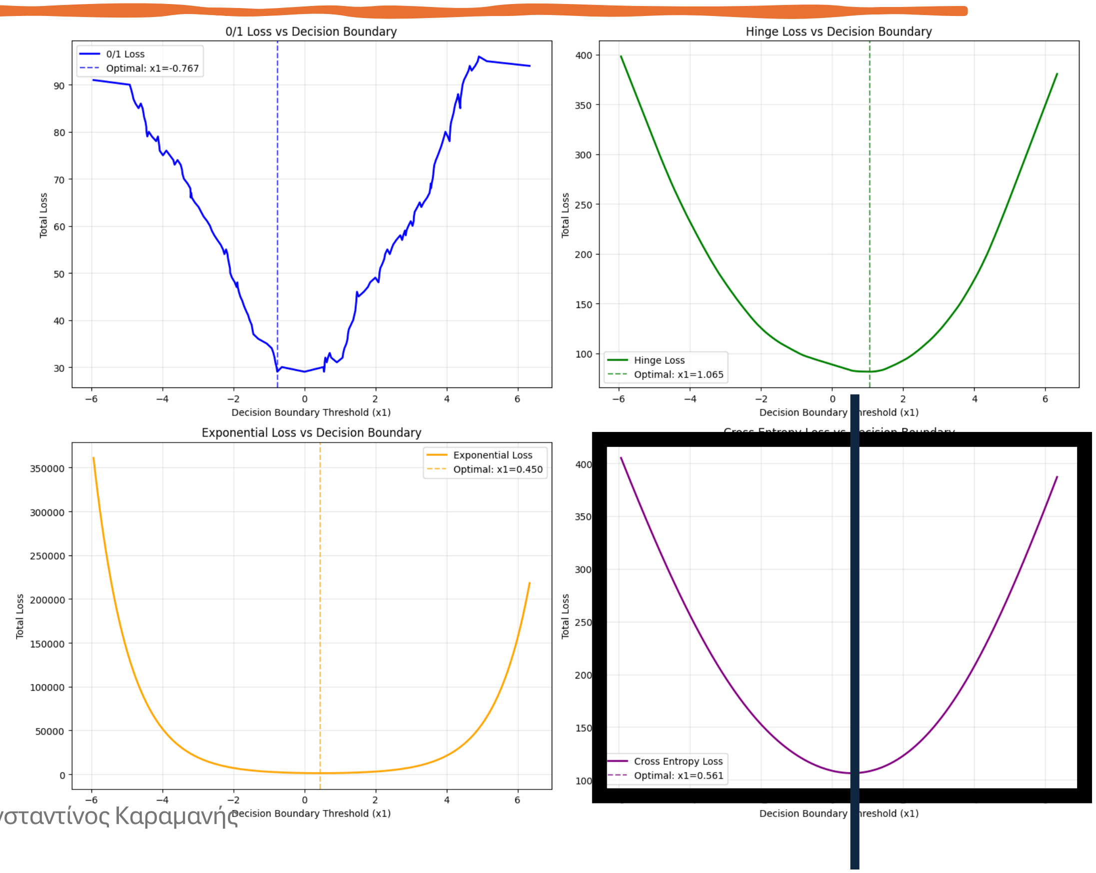
Πώς βρίσκουμε την βέλτιστη λύση;



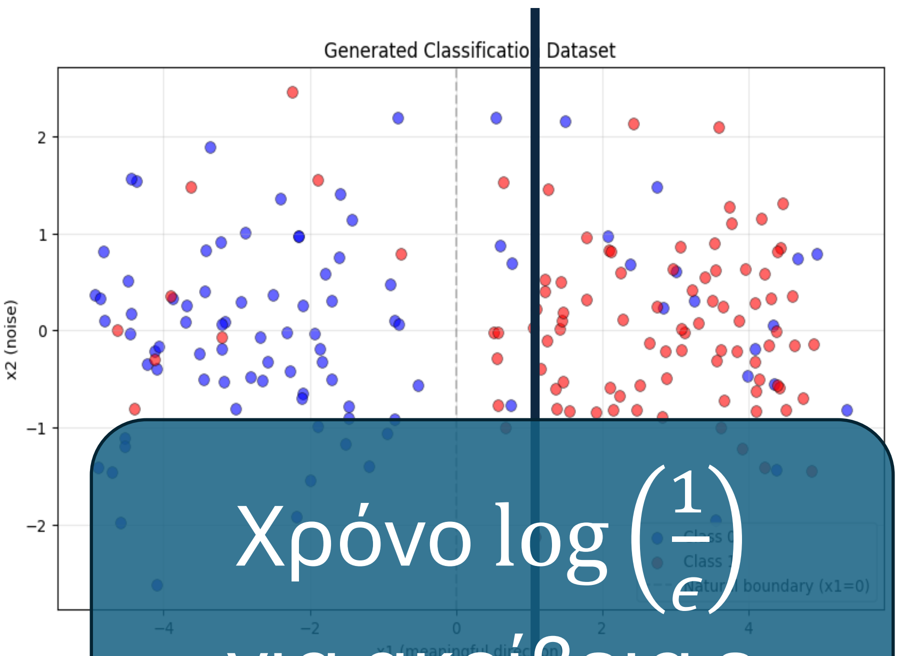


Πώς βρίσκουμε την βέλτιστη λύση;

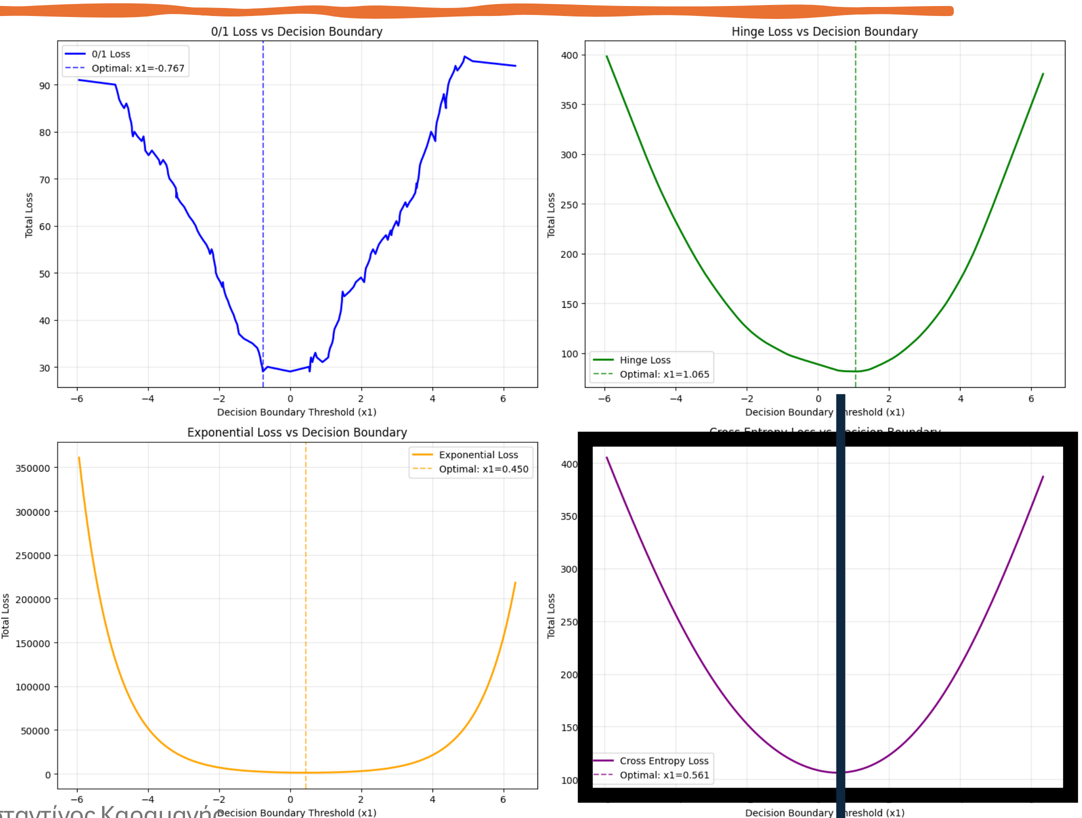




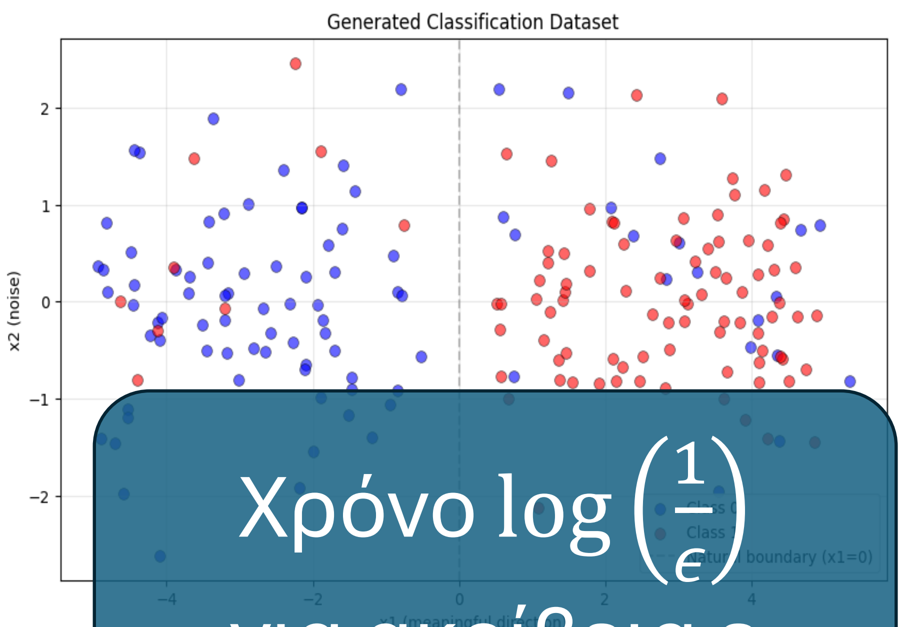
Πώς βρίσκουμε την βέλτιστη λύση;

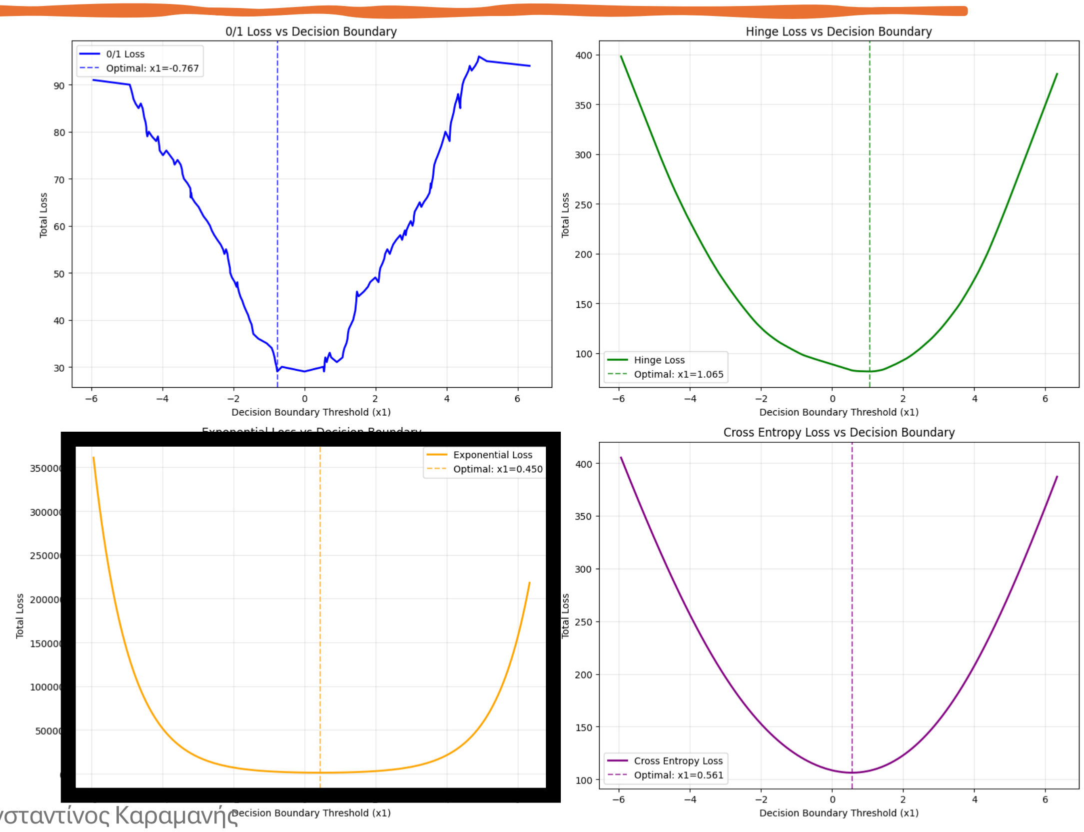


Χρόνο $\log\left(\frac{1}{\epsilon}\right)$
για ακρίβεια ε



Πώς βρίσκουμε την βέλτιστη λύση;

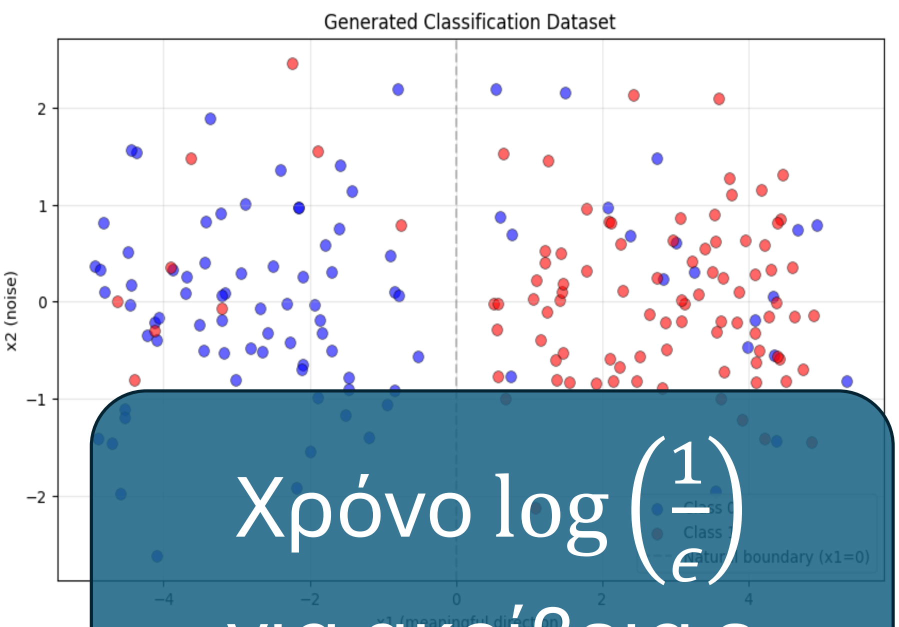




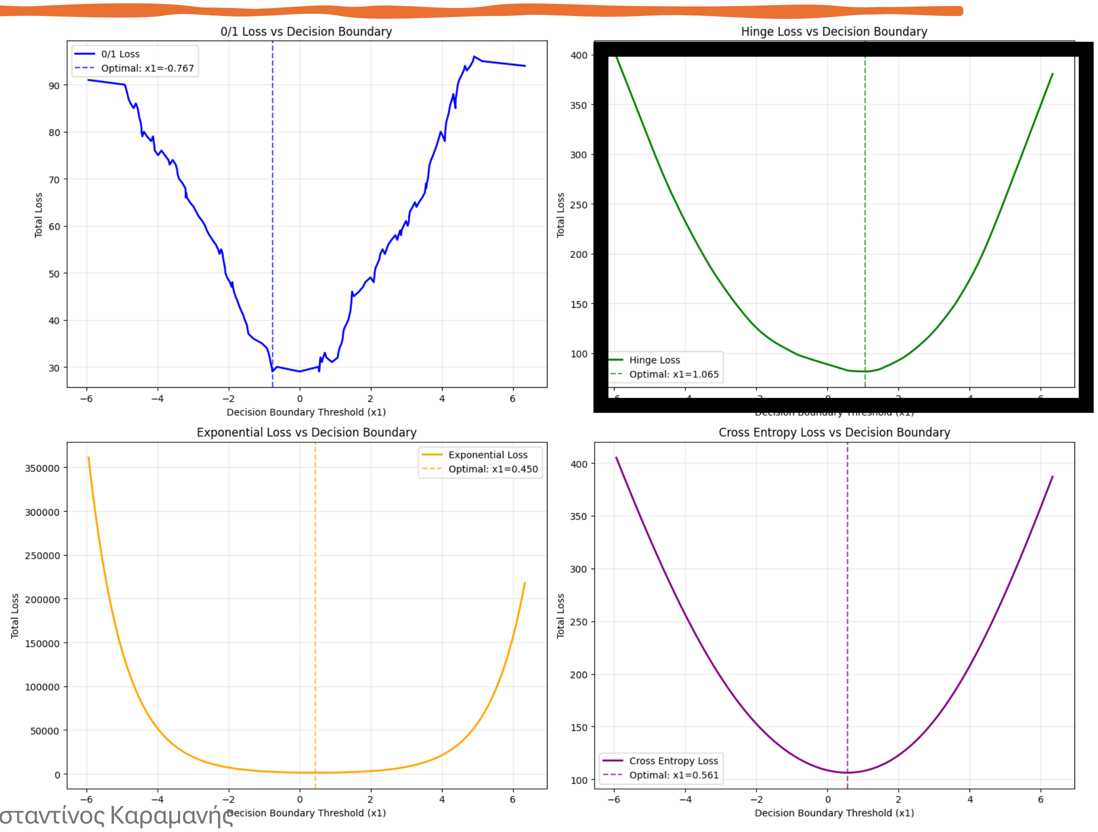
Χρόνο $\log\left(\frac{1}{\epsilon}\right)$

για ακρίβεια ε

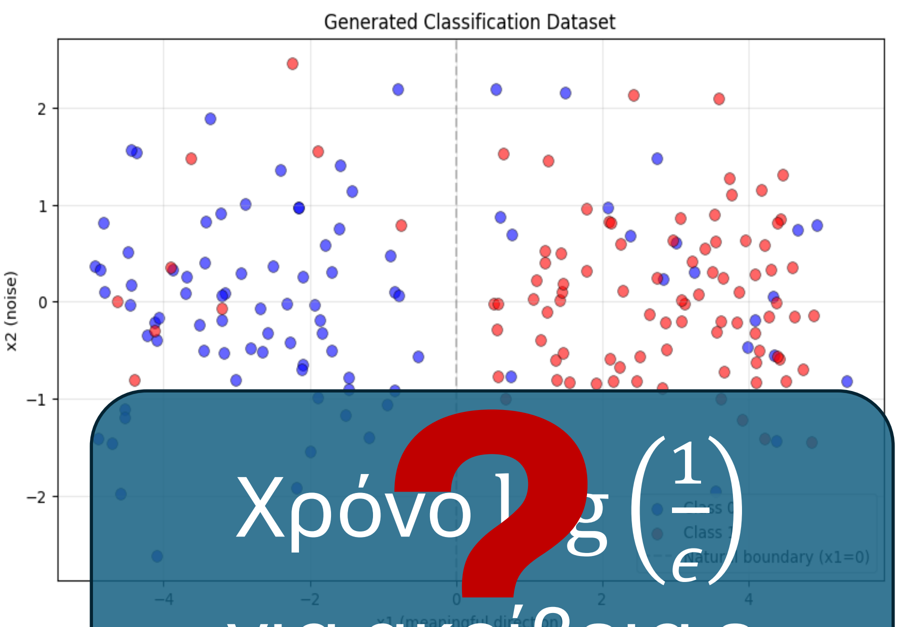
Πώς βρίσκουμε την βέλτιστη λύση;

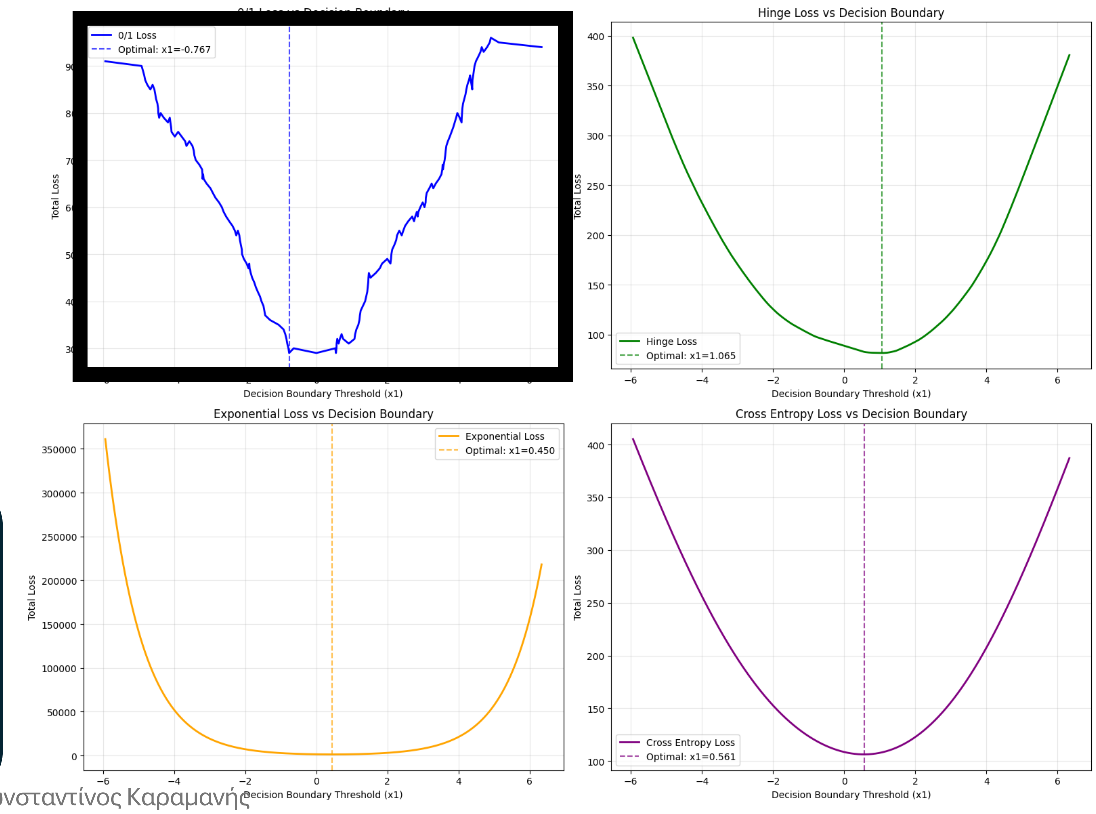


Χρόνο $\log\left(\frac{1}{\epsilon}\right)$
για ακρίβεια ε



Πώς βρίσκουμε την βέλτιστη λύση;

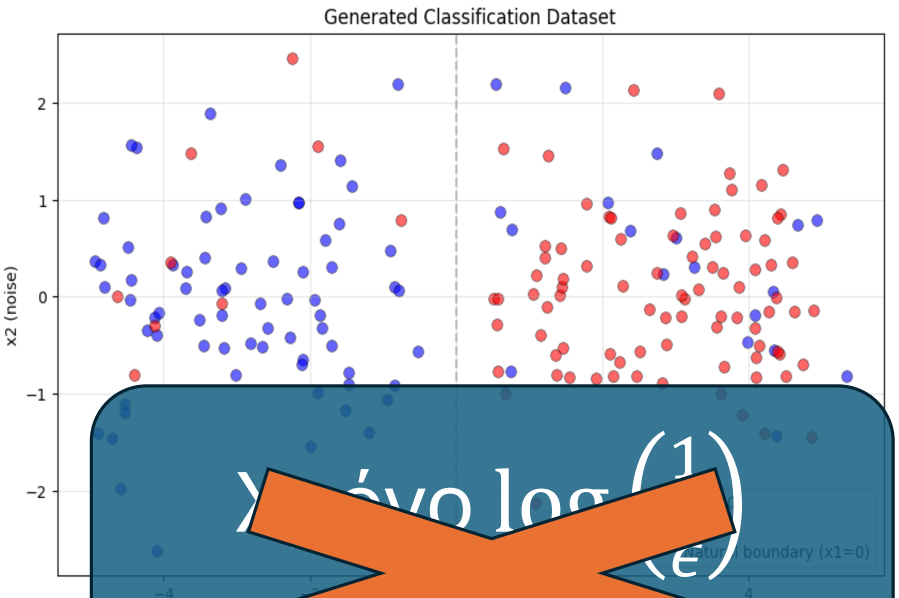


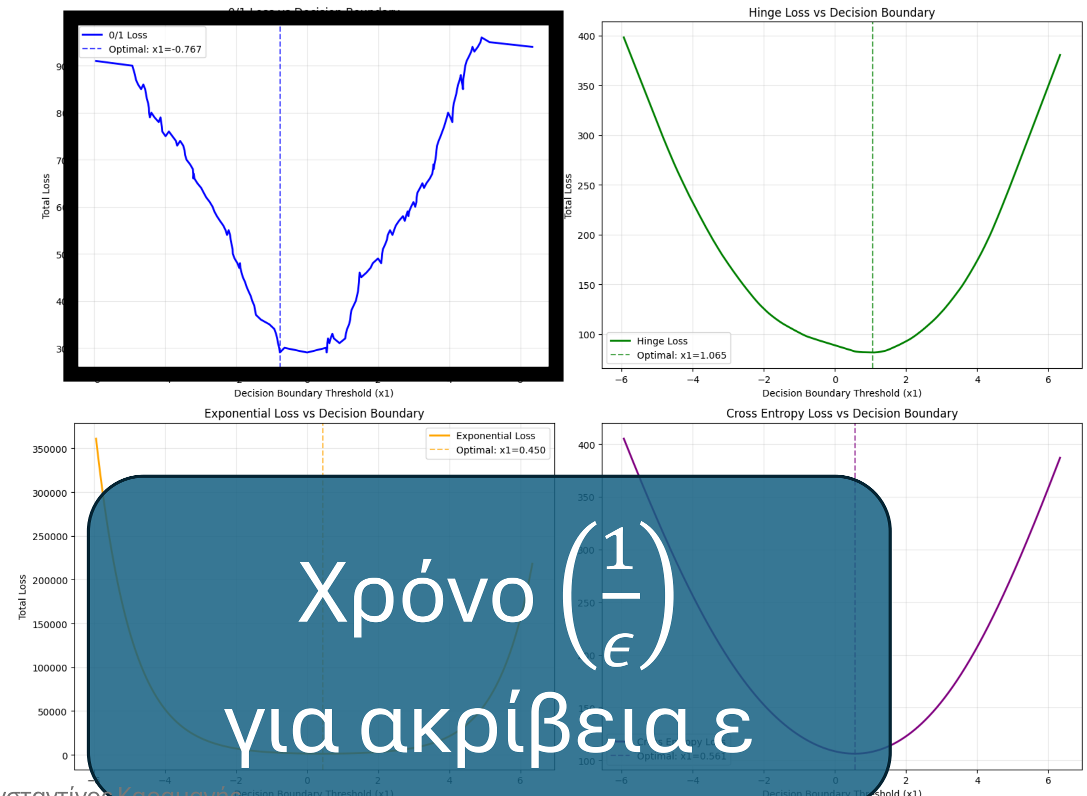


Χρόνο $\log\left(\frac{1}{\epsilon}\right)$ για ακρίβεια ε

?

Πώς βρίσκουμε την βέλτιστη λύση;





~~Χρόνο $\log\left(\frac{1}{\epsilon}\right)$ για ακρίβεια ε~~

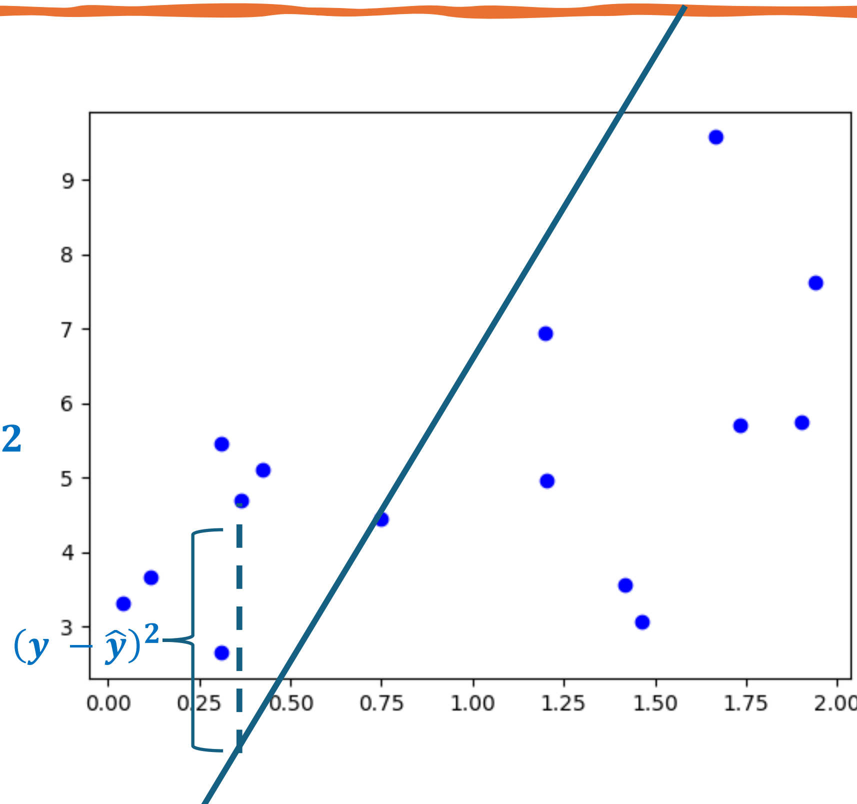
Χρόνο $\left(\frac{1}{\epsilon}\right)$ για ακρίβεια ε

Παλινδρόμιση (Linear Regression)

Ποια είναι η βέλτιστη σχέση μεταξύ X και Y;

Πώς μετράμε την απώλεια;

MSE Loss: $\boldsymbol{L(y, \hat{y}) = (y - \hat{y})^2}$



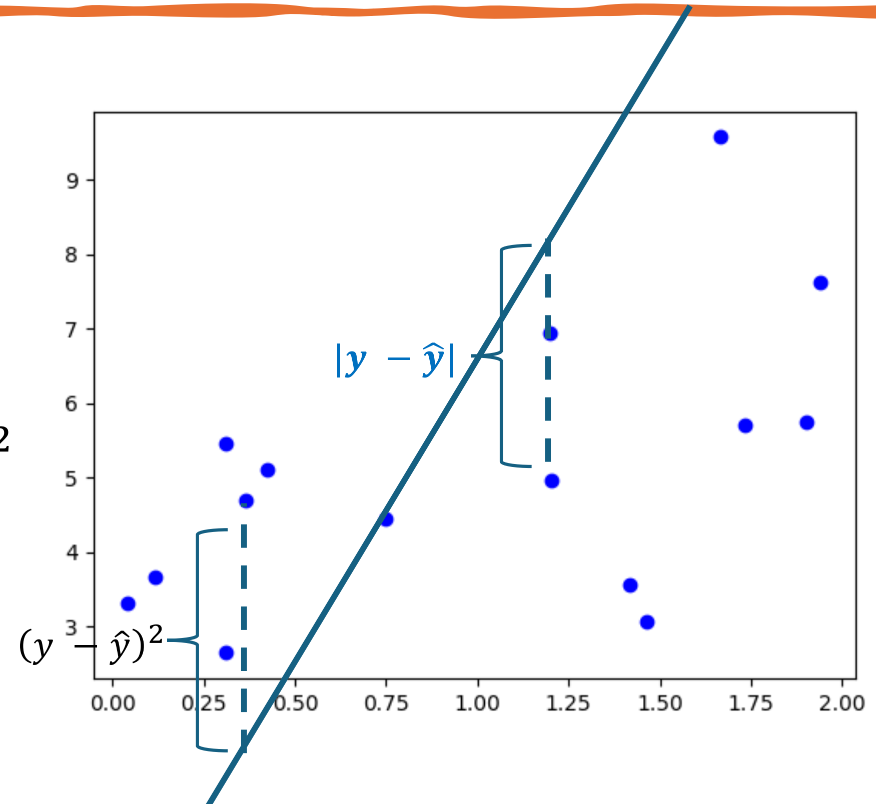
Παλινδρόμιση (Linear Regression)

Ποια είναι η βέλτιστη σχέση μεταξύ X και Y;

Πώς μετράμε την απώλεια;

MSE Loss: $L(y, \hat{y}) = (y - \hat{y})^2$

L₁ Loss: $L(y, \hat{y}) = |y - \hat{y}|$



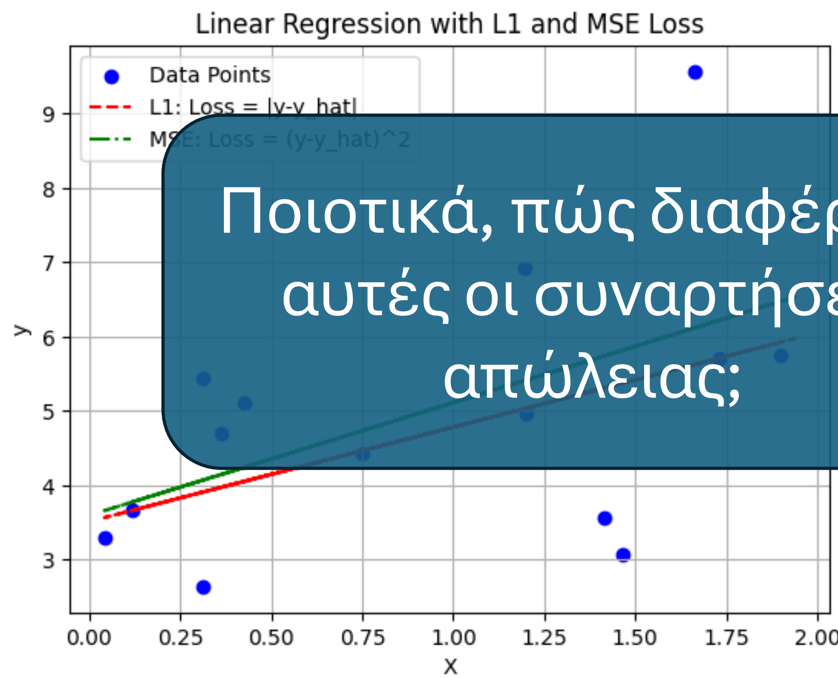
Παλινδρόμιση (Linear Regression)

Ποια είναι η βέλτιστη σχέση μεταξύ X και Y;

Πώς μετράμε την απώλεια;

MSE Loss: $L(y, \hat{y}) = (y - \hat{y})^2$

L_1 Loss: $L(y, \hat{y}) = |y - \hat{y}|$



Ποιοτικά, πώς διαφέρουν αυτές οι συναρτήσεις απώλειας;

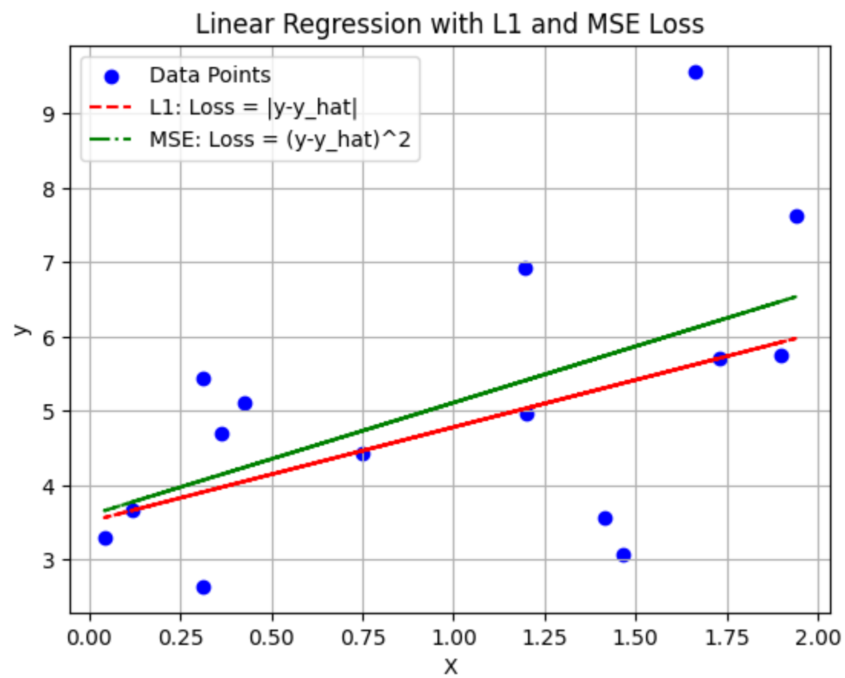
Παλινδρόμιση (Linear Regression)

Ποια είναι η βέλτιστη σχέση μεταξύ X και Y;

Πώς μετράμε την απώλεια;

MSE Loss: $L(y, \hat{y}) = (y - \hat{y})^2$

L_1 Loss: $L(y, \hat{y}) = |y - \hat{y}|$



Παλινδρόμιση (Linear Regression)

Ποιοτικά, πώς διαφέρουν αυτές οι συναρτήσεις απώλειας;

Ένα απλό παράδειγμα: $y = h(X) = \beta_0 + \beta_1 X$

MSE Loss: $L(y, \hat{y}) = (y - \hat{y})^2$

Με δεδομένα $\{(X_i, y_i)\}$ το απλό μοντέλο μας ζητάει να λύσουμε: $\min_{\beta_0} \sum_i (y_i - \beta_0)^2$

$$\frac{d}{d\beta_0} \sum_i (y - \beta_0)^2 = -2 \sum_i (y_i - \beta_0) = 0$$

$$\hat{\beta}_0 = \frac{1}{n} \sum y_i$$

Ο μέσος όρος!

L1 Loss: $L(y, \hat{y}) = |y - \hat{y}|$

Παλινδρόμιση (Linear Regression)

Ποιοτικά, πώς διαφέρουν αυτές οι συναρτήσεις απώλειας;

Ένα απλό παράδειγμα: $y = h(X) = \beta_0$

$\boxed{\text{MSE Loss: } L(y, \hat{y}) = (y - \hat{y})^2}$

Με δεδομένα $\{(X_i, y_i)\}$ το απλό μοντέλο μας ζητάει να λύσουμε: $\min_{\beta_0} \sum_i (y_i - \beta_0)^2$

$$\frac{d}{d\beta_0} \sum_i (y - \beta_0)^2 = -2 \sum_i (y_i - \beta_0) = 0$$

$$\boxed{\hat{\beta}_0 = \frac{1}{n} \sum y_i}$$

Ο μέσος όρος!

$\boxed{\text{L}_1 \text{ Loss: } L(y, \hat{y}) = |y - \hat{y}|}$

Με δεδομένα $\{(X_i, y_i)\}$ το απλό μοντέλο μας ζητάει να λύσουμε: $\min_{\beta_0} \sum_i |y_i - \beta_0|$

Πρόκληση: πως θα το λύσουμε ως προς β_0;

$$\boxed{\hat{\beta}_0 = median(\{y_i\})}$$

Η διάμεσος!

Παλινδρόμιση: δεύτερο παράδειγμα

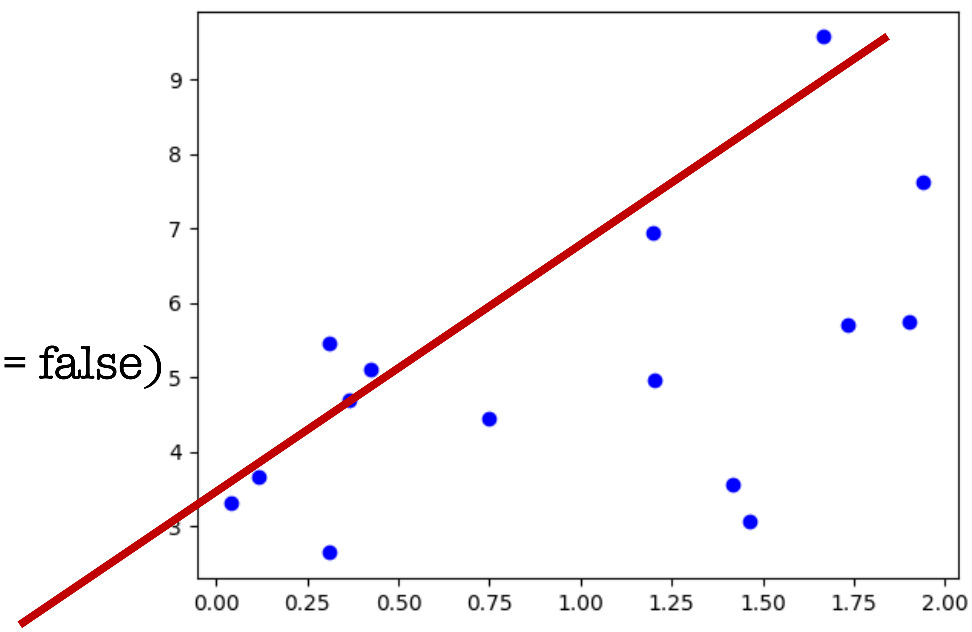
Συναρτήσεις που περνάνε από το 0

μοντέλο: $y = h(X) = \beta_0 + \beta_1 X$

Python:

```
model = LinearRegression(fit_intercept = false)
model.fit(X,y)
```

β_1 = model.coef_



Εμείς;

Παλινδρόμιση: δεύτερο παράδειγμα

Συναρτήσεις που

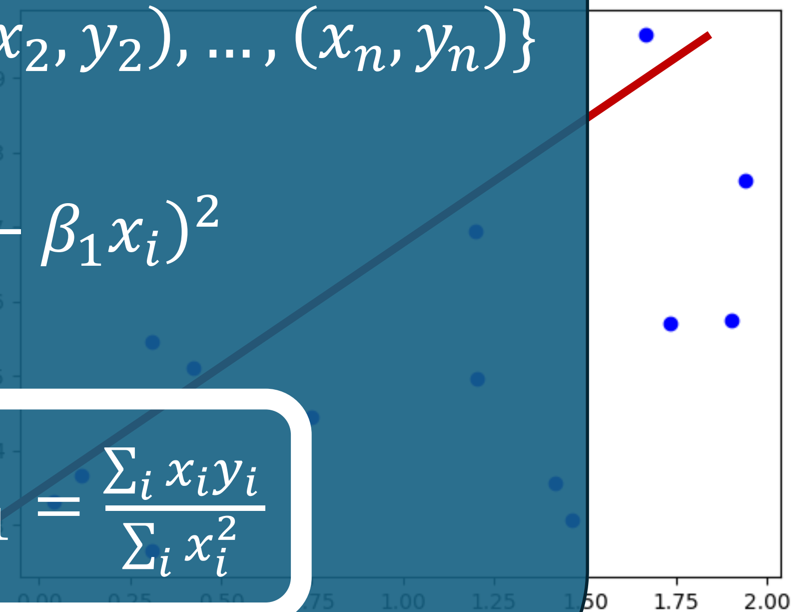
μοντέλο: $y = h(X) = \beta_0 X + \beta_1 X$

Python:

```
model = LinearRegression(fit_intercept = false)
model.fit(X,y)
```

β_1 = model.coef_

Εμείς:



Δεδομένα: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

$$\min_{\beta_1} \sum_i (y_i - \beta_1 x_i)^2$$

$$\boxed{\text{Απάντηση: } \hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}}$$

Παλινδρόμιση: δεύτερο παράδειγμα

Συναρτήσ

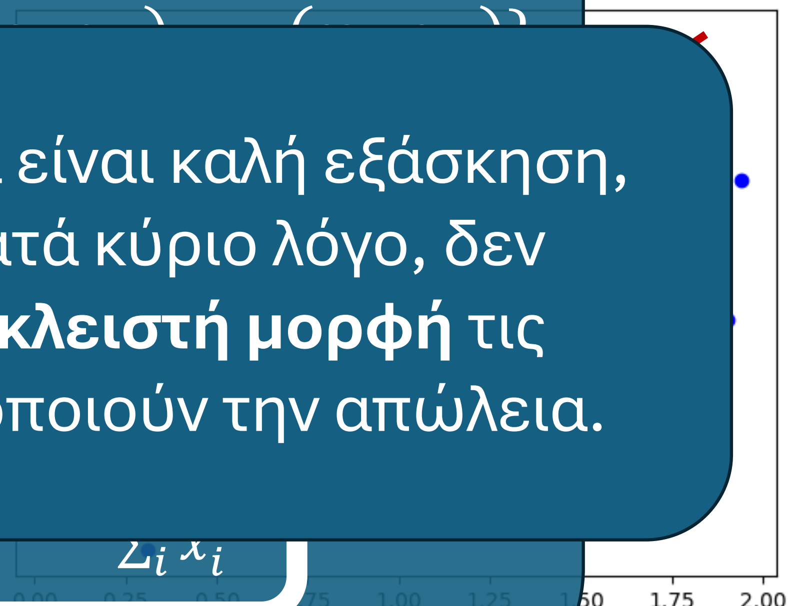
μοντέλο:

Python:

model =

model.

$\beta_1 =$ mod



$\sum_i x_i$

Αυτά τα απλά παραδείγματα είναι καλή εξάσκηση, αλλά δεν γενικεύονται: κατά κύριο λόγο, δεν μπορούμε να βρούμε **σε κλειστή μορφή** τις παραμέτρους που ελαχιστοποιούν την απώλεια.

Εμείς:

Εκπαίδευση Αλγορίθμου

1. Επιλέγουμε την οικογένεια αλγορίθμων: παράδειγμα – λογιστική παλινδρόμιση

2. Επιλέγουμε την συνάρτηση απώλειας

```
mymodel = LogisticRegression()
```

3. Βρίσκουμε τις παραμέτρους που ~~ελαχιστοποιούν τα λάθη στα δεδομένα~~ ελαχιστοποιούν το άθροισμα της απώλειας στα δεδομένα (εκπαίδευσης)

```
mymodel.fit(X,y)
```

Εκπαίδευση Αλγορίθμου

Αλλάζοντας την συνάρτηση απώλειας, αλλάζουμε και την βέλτιστη λύση

1. Επιλέγουμε την οικογένεια αλγορίθμων: παράδειγμα – λογιστική παλινδρόμιση
2. Επιλέγουμε την συνάρτηση απώλειας

mymodel = LogisticRegression()

3. Βρίσκουμε τις παραμέτρους που ~~ελαχιστοποιούν τα λάθη στα δεδομένα~~ ελαχιστοποιούν το άθροισμα της απώλειας στα δεδομένα (εκπαίδευσης)

```
mymodel.fit(X,y)
```

Εκπαίδευση Αλγορίθμου

Αλλάζοντας την συνάρτηση απώλειας, αλλάζουμε και την πολυπλοκότητα επίλυσης

1. Επιλέγουμε την οικογένεια αλγορίθμων: παράδειγμα – λογιστική παλινδρόμιση

2. Επιλέγουμε την συνάρτηση απώλειας

   mymodel = LogisticRegression()

3. Βρίσκουμε τις παραμέτρους που ~~ελαχιστοποιούν τα λάθη στα δεδομένα~~ ελαχιστοποιούν το άθροισμα της απώλειας στα δεδομένα (εκπαίδευσης)

   mymodel.fit(X,y)

Εκπαίδευση Αλγορίθμου

1. Επιλέγουμε την οικογένεια αλγορίθμων: παράδειγμα – λογιστική παλινδρόμιση

2. Επιλέγουμε την συνάρτηση απώλειας

   mymodel = LogisticRegression()

3. Βρίσκουμε τις παραμέτρους που ~~ελαχιστοποιούν τα λάθη στα δεδομένα~~ ελαχιστοποιούν το άθροισμα της απώλειας στα δεδομένα (εκπαίδευσης)

   mymodel.fit(X,y)

Αλλάζοντας την συνάρτηση απώλειας, αλλάζουμε και την πολυπλοκότητα επίλησης

Πώς βρίσκουμε τις παραμέτρους που ελαχιστοποιούν την απώλεια;