



# Μηχανική Μάθηση: Μαθηματικό Υπόβαθρο

---

Κωνσταντίνος Καραμανής

The University of Texas at Austin & Archimedes/Athena RC

[constantine@utexas.edu](mailto:constantine@utexas.edu)

<https://caramanis.github.io/>





Ας θυμηθούμε τα  
προηγούμενα...



# Λογιστική Παλινδρόμηση

---

Στόχος: εκτίμηση της πιθανότητας  $Y=1$  ή  $Y=0$

Πρώτο βήμα: υπολογίζουμε το λεγόμενο **score** ή **logit**:

$$Z_X = \beta_0 + \beta^T X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d$$

Χρησιμοποιούμε την sigmoid (softmax) συνάρτηση για να μετατρέψουμε το «score» σε πιθανότητα:

$$P(Y = 1|X) = \hat{p}_X = \text{sigmoid}(z_X) = \frac{\exp(z_X)}{1 + \exp(z_X)} = \frac{1}{1 + \exp(-z_X)}$$

# Λογιστική Παλινδρόμηση

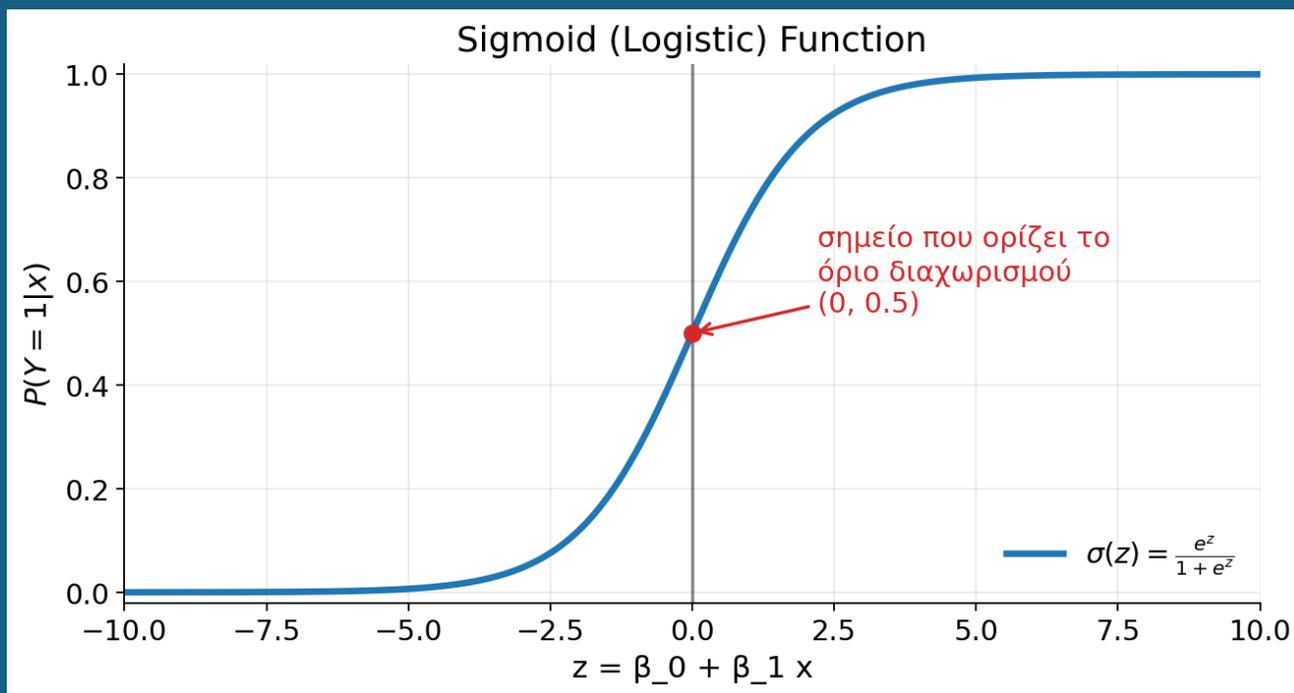
Στόχο

Πρώτ

Χρησι

το «SC

P(



ουμε

x)



# Λογιστική Παλινδρόμηση: πώς λειτουργεί

$$Z_X = \beta_0 + \beta^T X = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Περιοχή με  $Y=1$ :  $\beta_0 + \beta^T X > 0$

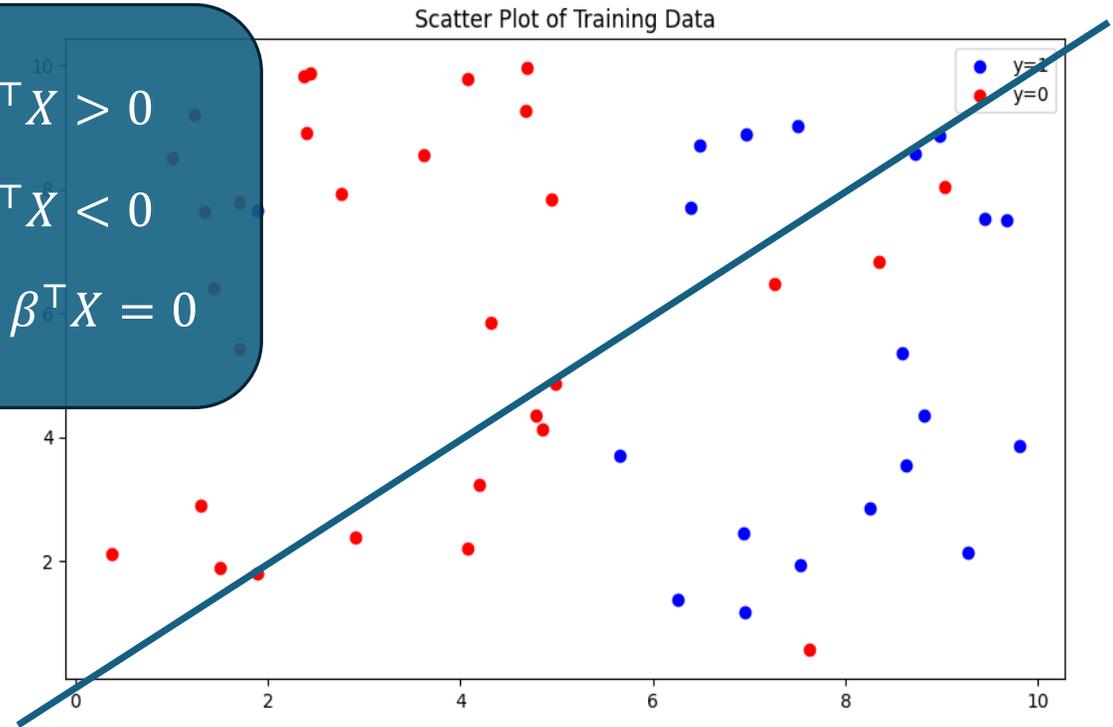
$$P(Y = 1|X) = \frac{\exp(z_X)}{1 + \exp(z_X)}$$

Περιοχή με  $Y=0$ :  $\beta_0 + \beta^T X < 0$

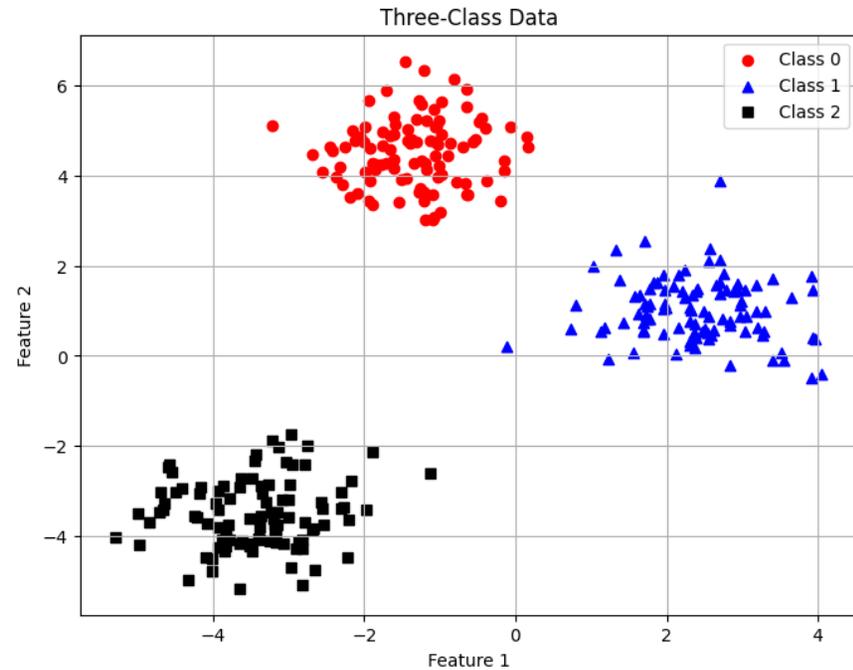
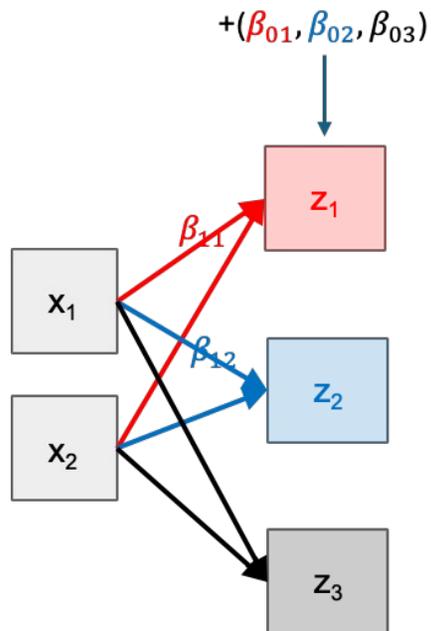
Περιοχή με  $Y=1/2$ :  $\beta_0 + \beta^T X = 0$

Με δύο κατηγορίες, ο διαχωρισμός του χώρου είναι γραμμικός:

$$Z_X = \beta_0 + \beta^T X = 0$$



# Λογιστική Παλινδρόμηση με K κατηγορίες

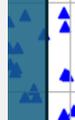


# Λογιστική Παλινδρόμηση με K κατηγορίες

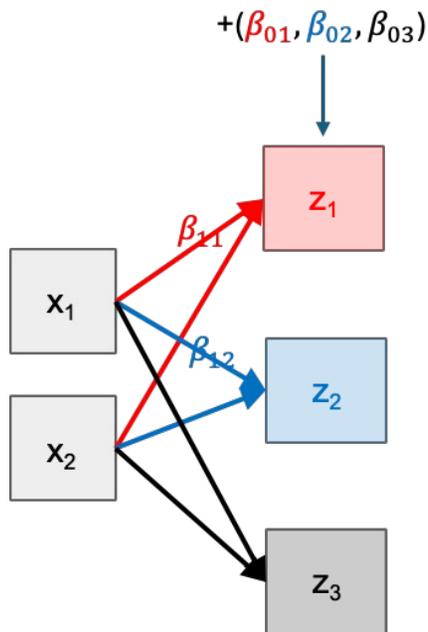
## Νευρωνικό Δίκτυο για 3 κατηγορίες

```
class ThreeClassLogisticNN(nn.Module):  
    def __init__(self):  
        super(ThreeClassLogisticNN, self).__init__()  
        self.linear = nn.Linear(2, 3)  
  
    def forward(self, x):  
        x = self.linear(x)  
        x = nn.functional.softmax(x, dim=1)  
        return x
```

Class 0  
Class 1  
Class 2



# Λογιστική Παλινδρόμηση με K κατηγορίες



$$Z_R = \beta_0^R + \beta^{R\top} X = \beta_0^R + \beta_1^R X_1 + \beta_2^R X_2$$

$$Z_B = \beta_0^B + \beta^{B\top} X = \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

$$Z_{Bl} = \beta_0^{Bl} + \beta^{Bl\top} X = \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

# Λογιστική Παλινδρόμηση με K κατηγορίες

$$y_R = \frac{\exp(Z_R)}{\exp(Z_R) + \exp(Z_B) + \exp(Z_{Bl})}$$
$$= \frac{\exp(\beta_0^R + \beta^{RT} X)}{\exp(\beta_0^R + \beta^{RT} X) + \exp(\beta_0^B + \beta^{BT} X) + \exp(\beta_0^{Bl} + \beta^{BlT} X)}$$

# Λογιστική Παλινδρόμηση με K κατηγορίες

$$y_B = \frac{\exp(Z_B)}{\exp(Z_R) + \exp(Z_B) + \exp(Z_{Bl})}$$
$$= \frac{\exp(\beta_0^B + \beta^{B\top} X)}{\exp(\beta_0^R + \beta^{R\top} X) + \exp(\beta_0^B + \beta^{B\top} X) + \exp(\beta_0^{Bl} + \beta^{Bl\top} X)}$$

# Λογιστική Παλινδρόμηση με K κατηγορίες

$$\begin{aligned} y_{Bl} &= \frac{\exp(Z_{Bl})}{\exp(Z_R) + \exp(Z_B) + \exp(Z_{Bl})} \\ &= \frac{\exp(\beta_0^{Bl} + \beta^{BlT} X)}{\exp(\beta_0^R + \beta^{RT} X) + \exp(\beta_0^B + \beta^{BT} X) + \exp(\beta_0^{Bl} + \beta^{BlT} X)} \end{aligned}$$

# Λογιστική Παλινδρόμηση με K κατηγορίες

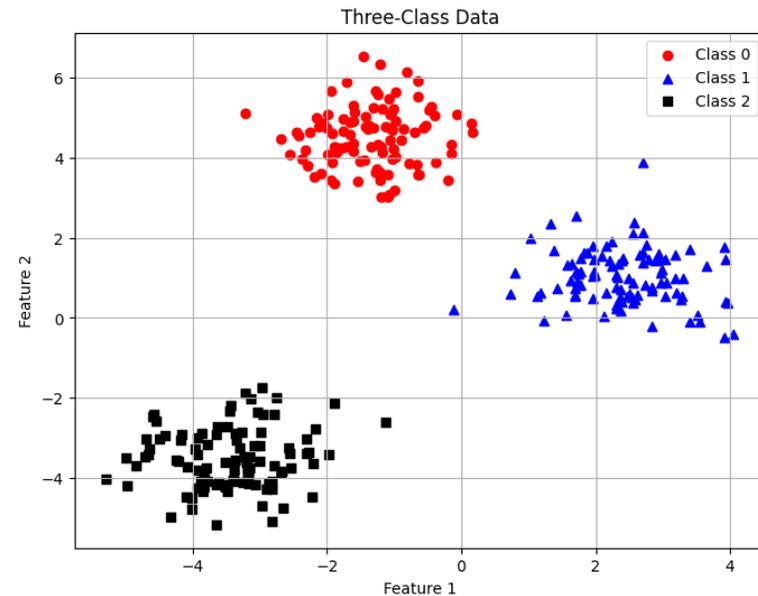
$$Z_R = \beta_0^R + \beta^{R\top} X = \beta_0^R + \beta_1^R X_1 + \beta_2^R X_2$$

$$Z_B = \beta_0^B + \beta^{B\top} X = \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

$$Z_{Bl} = \beta_0^{Bl} + \beta^{Bl\top} X = \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$y_R = \frac{\exp(Z_R)}{\exp(Z_R) + \exp(Z_B) + \exp(Z_{Bl})}$$

$$= \frac{\exp(\beta_0^R + \beta^{R\top} X)}{\exp(\beta_0^R + \beta^{R\top} X) + \exp(\beta_0^B + \beta^{B\top} X) + \exp(\beta_0^{Bl} + \beta^{Bl\top} X)}$$



# Λογιστική Παλινδρόμηση με K κατηγορίες

$$Z_R = \beta_0^R + \beta^{R\top} X = \beta_0^R + \beta_1^R X_1 + \beta_2^R X_2$$

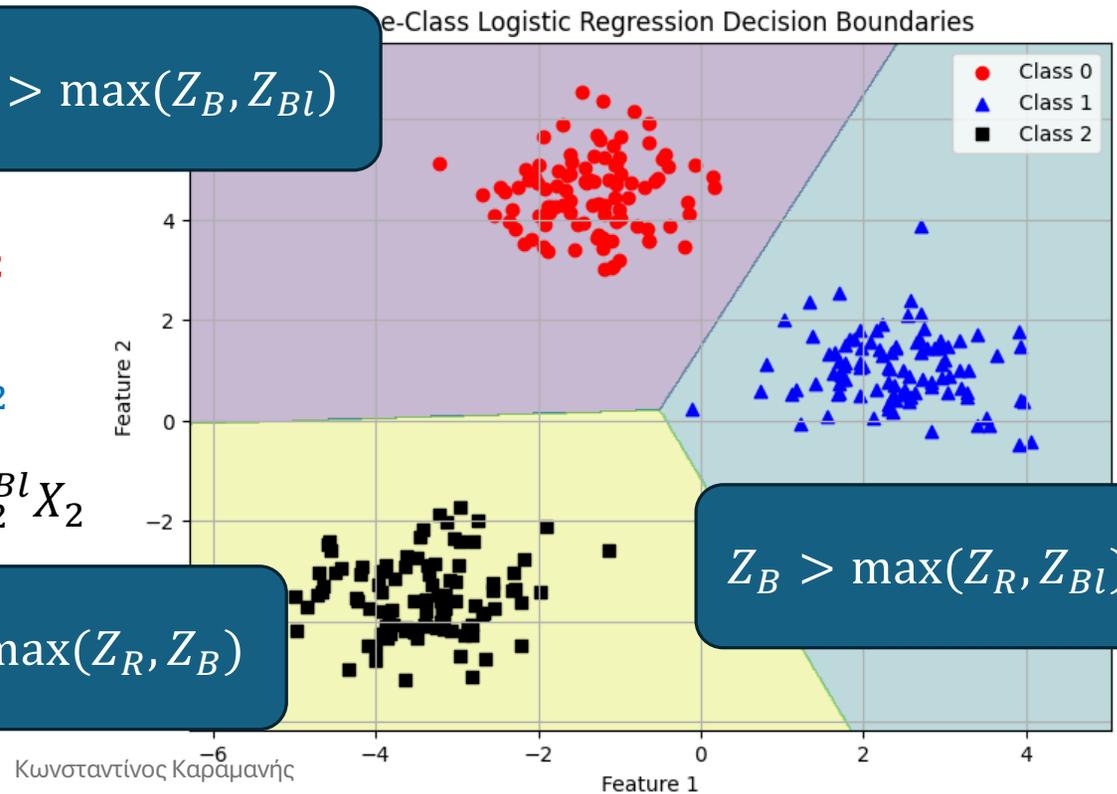
$$Z_B = \beta_0^B + \beta^{B\top} X = \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

$$Z_{Bl} = \beta_0^{Bl} + \beta^{Bl\top} X = \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$Z_R > \max(Z_B, Z_{Bl})$$

$$Z_{Bl} > \max(Z_R, Z_B)$$

$$Z_B > \max(Z_R, Z_{Bl})$$



# Λογιστική Παλινδρόμηση με K κατηγορίες

---

$$Z_R > \max(Z_B, Z_{Bl})$$

$$Z_B > \max(Z_R, Z_{Bl})$$

$$Z_{Bl} > \max(Z_R, Z_B)$$

Πώς μοιάζει (γενικά) ο διαχωρισμός του χώρου που μπορεί να προκύψει από την λογιστική παλινδρόμηση με K κατηγορίες;

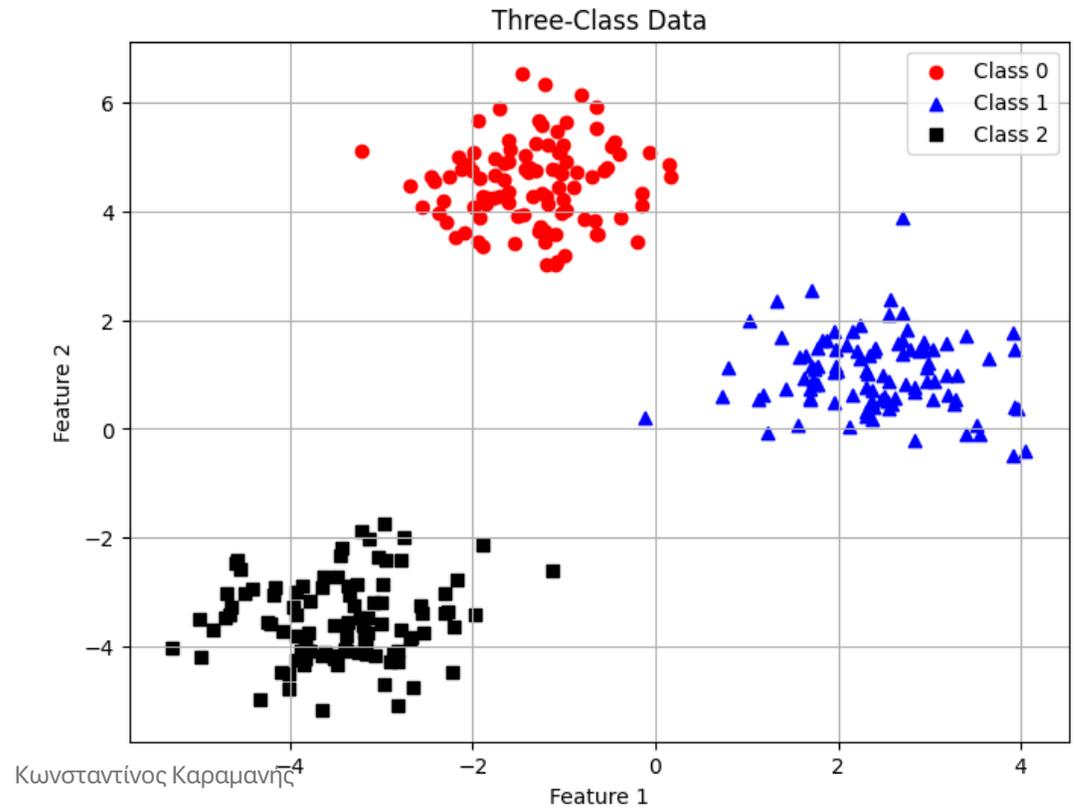
# Λογιστική Παλινδρόμηση με Κ κατηγορίες

$$Z_R = \beta_0^R + \beta_1^R X_1 + \beta_2^R X_2$$

$$Z_B = \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

$$Z_{Bl} = \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$Z_R > \max(Z_B, Z_{Bl})$$



# Λογιστική Παλινδρόμηση με Κ κατηγορίες

$$Z_R = \beta_0^R + \beta_1^R X_1 + \beta_2^R X_2$$

$$Z_B = \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

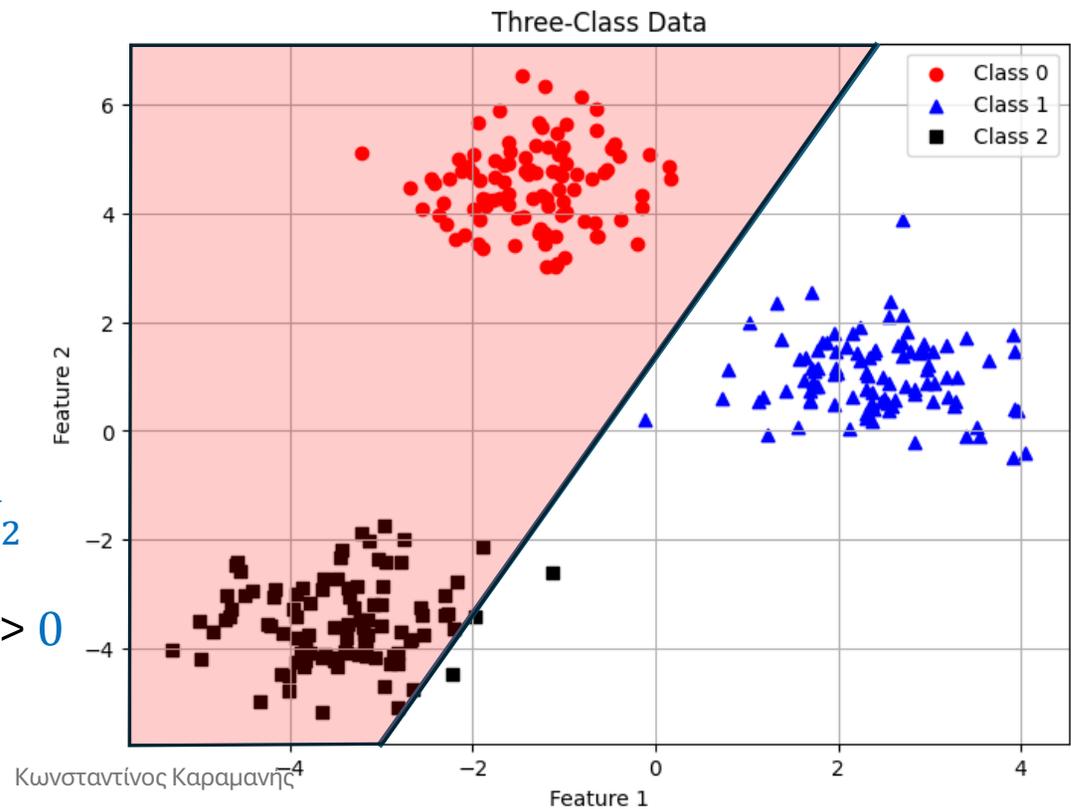
$$Z_{Bl} = \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$Z_R > Z_B \text{ \& } Z_R > Z_{Bl}$$

$$\beta_0^R + \beta_1^R X_1 + \beta_2^R X_2 > \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

$$(\beta_0^R - \beta_0^B) + (\beta_1^R - \beta_1^B) X_1 + (\beta_2^R - \beta_2^B) X_2 > 0$$

$$A + B X_1 + C X_2 > 0$$



# Λογιστική Παλινδρόμηση με Κ κατηγορίες

$$Z_R = \beta_0^R + \beta_1^R X_1 + \beta_2^R X_2$$

$$Z_B = \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

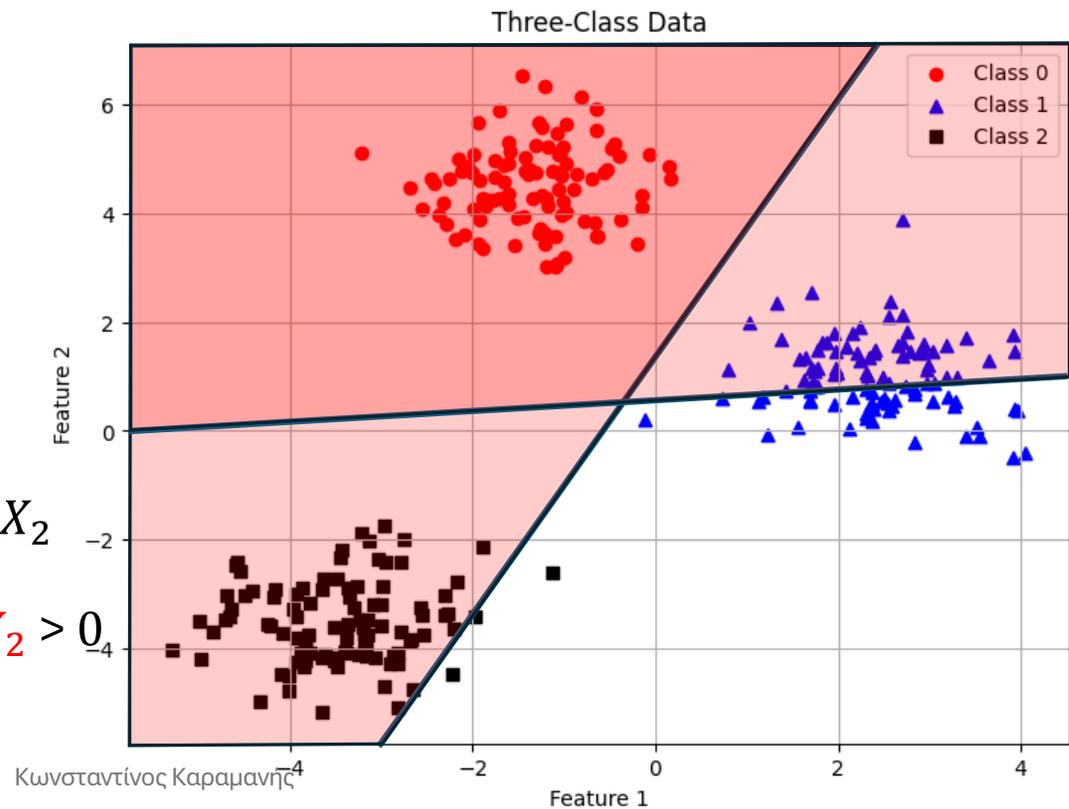
$$Z_{Bl} = \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$Z_R > Z_B \ \& \ Z_R > Z_{Bl}$$

$$\beta_0^R + \beta_1^R X_1 + \beta_2^R X_2 > \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$(\beta_0^R - \beta_0^{Bl}) + (\beta_1^R - \beta_1^{Bl}) X_1 + (\beta_2^R - \beta_2^{Bl}) X_2 > 0$$

$$A' + B' X_1 + C' X_2 > 0$$



# Λογιστική Παλινδρόμηση με Κ κατηγορίες

$$Z_R = \beta_0^R + \beta_1^R X_1 + \beta_2^R X_2$$

$$Z_B = \beta_0^B + \beta_1^B X_1 + \beta_2^B X_2$$

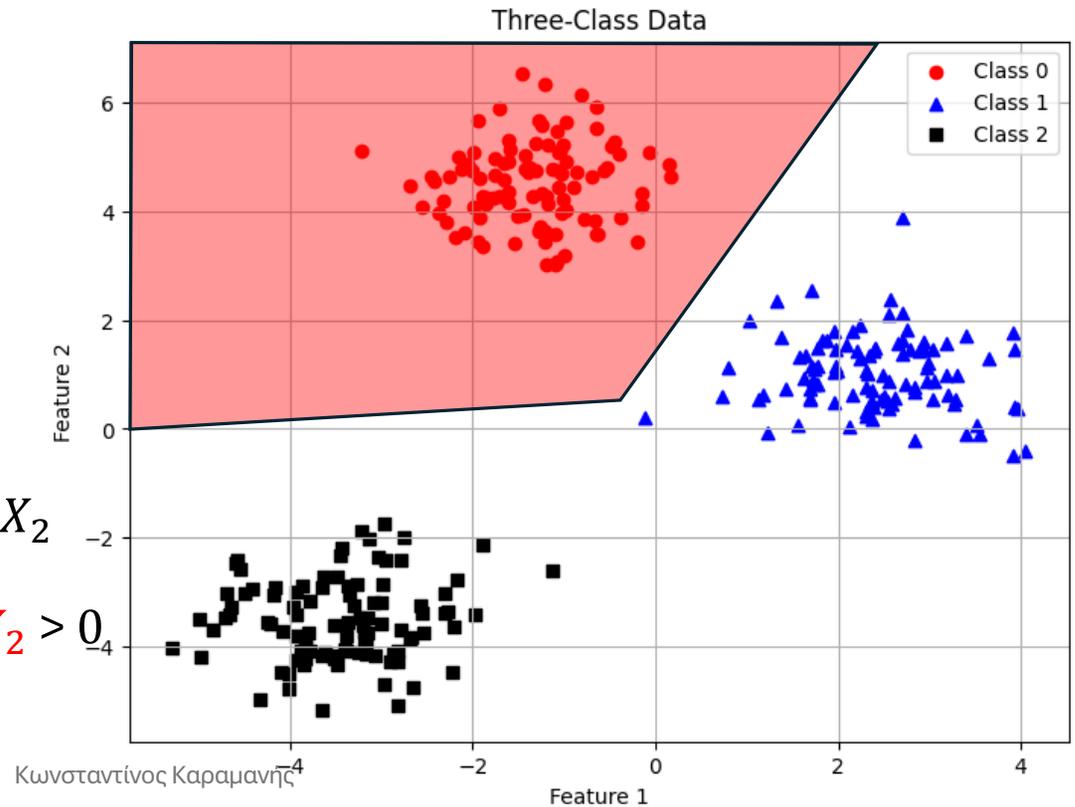
$$Z_{Bl} = \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$Z_R > Z_B \ \& \ Z_R > Z_{Bl}$$

$$\beta_0^R + \beta_1^R X_1 + \beta_2^R X_2 > \beta_0^{Bl} + \beta_1^{Bl} X_1 + \beta_2^{Bl} X_2$$

$$(\beta_0^R - \beta_0^{Bl}) + (\beta_1^R - \beta_1^{Bl}) X_1 + (\beta_2^R - \beta_1^{Bl}) X_2 > 0$$

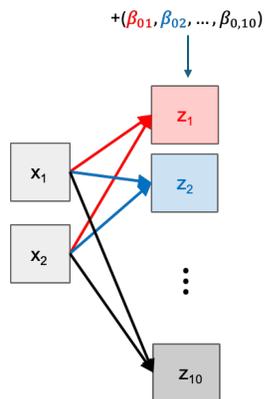
$$A' + B' X_1 + C' X_2 > 0$$



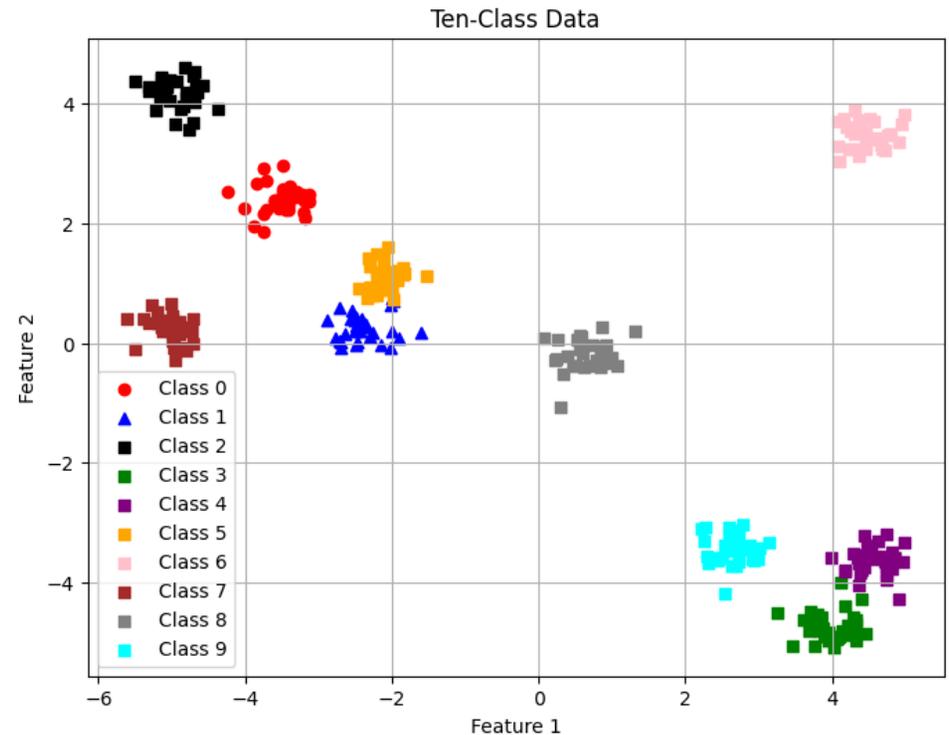
# Λογιστική Παλινδρόμηση με K κατηγορίες

10 Κατηγορίες

1. Πόσες παραμέτρους θα έχει η λογιστική παλινδρόμηση;



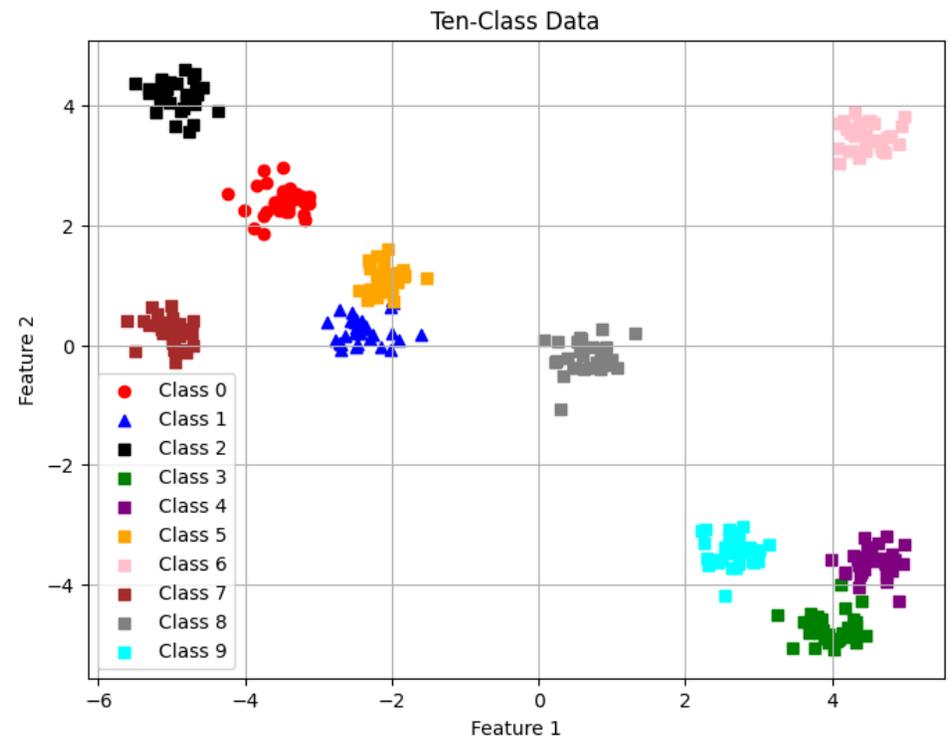
$$2 \times 10 + 10 = 30$$



# Λογιστική Παλινδρόμηση με $K$ κατηγορίες

10 Κατηγορίες

2. Πώς θα μοιάζει ο διαχωρισμός του χώρου;



# Λογιστική Παλινδρόμηση με K κατηγορίες

$$Z_0 > \max(Z_1, Z_2, \dots, Z_{10})$$

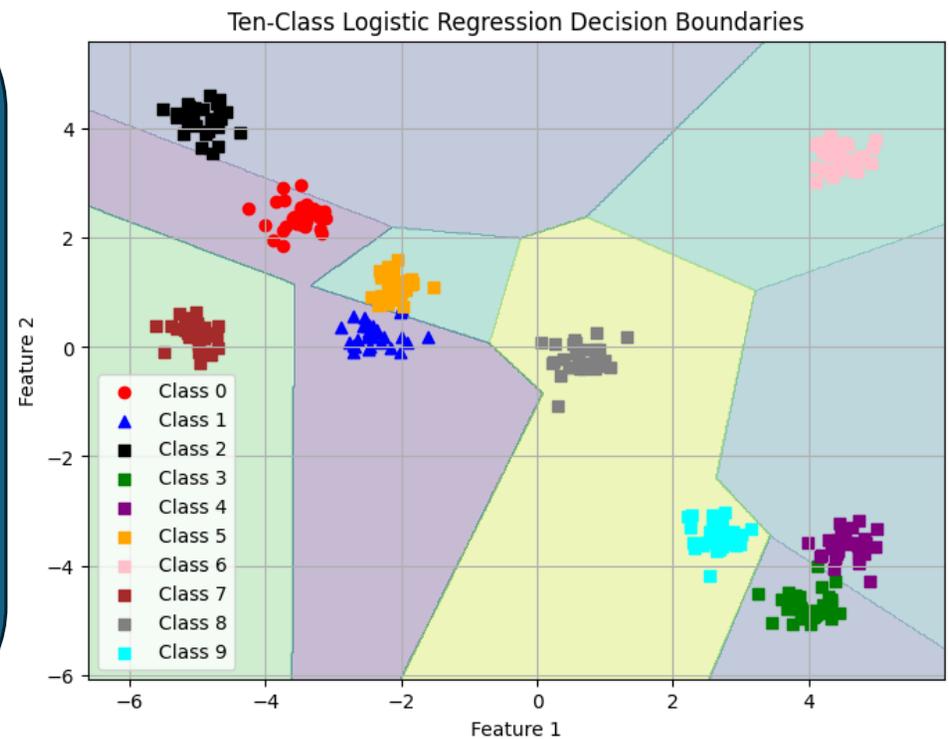
$$Z_1 > \max(Z_0, Z_2, \dots, Z_{10})$$

⋮

$$Z_{10} > \max(Z_0, Z_1, \dots, Z_9)$$

$Z_i > Z_j$ : περιγράφει ημιχώρο

Οπότε;



# Λογιστική Παλινδρόμηση με K κατηγορίες

$$Z_0 > \max(Z_1, Z_2, \dots, Z_{10})$$

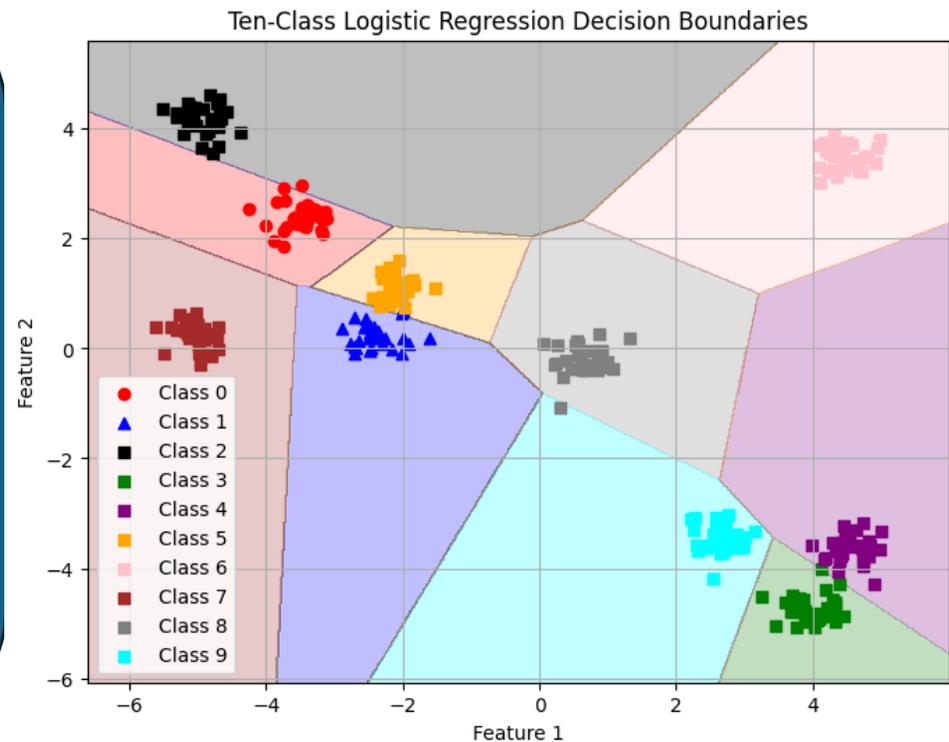
$$Z_1 > \max(Z_0, Z_2, \dots, Z_{10})$$

⋮

$$Z_{10} > \max(Z_0, Z_1, \dots, Z_9)$$

$Z_i > Z_j$ : περιγράφει ημιχώρο

Οπότε;



# Λογιστική Παλινδρόμηση με K κατηγορίες

$$Z_0 > \max(Z_1, Z_2, \dots, Z_{10})$$

$$Z_1 > \max(Z_0, Z_2, \dots, Z_{10})$$

$$Z_{10} > \max(Z_0, Z_1, \dots, Z_9)$$

$$Z_i > Z_j: \text{ περιγράφει ημιχώρο}$$

Οπότε,

Ο διαχωρισμός του χώρου ορίζεται από την τομή ημιχώρων. Οπότε η λογιστική παλινδρόμηση χωρίζει τον χώρο σε (κυρτά) πολύεδρα.

Ten-Class Logistic Regression Decision Boundaries

