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…

Συναρτήσεις απώλειας

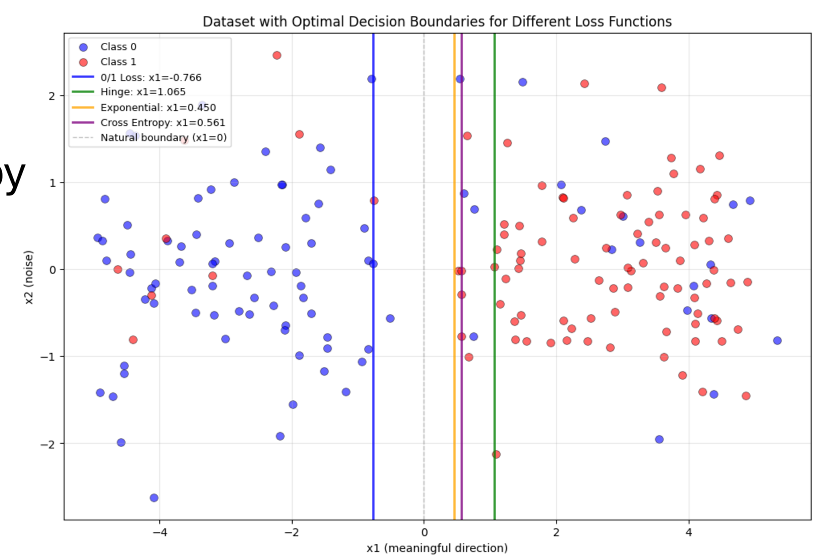
Ταξινόμηση

Errors

Cross Entropy

Hinge Loss

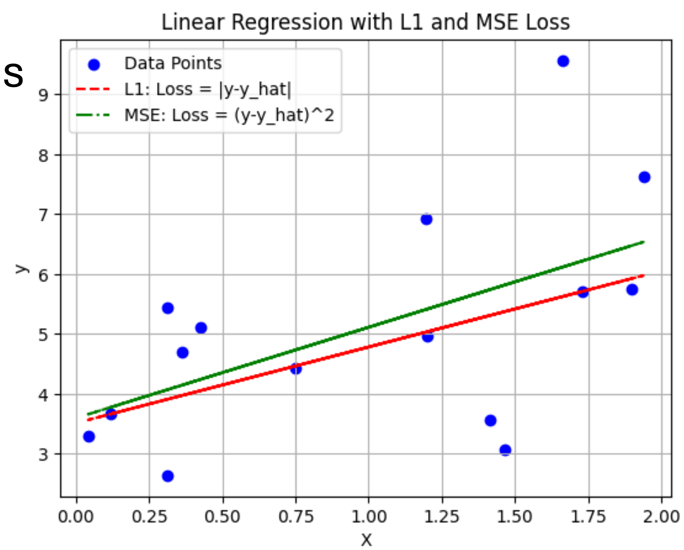
Exp Loss



Παλινδρόμηση

MSE loss

L1 loss



Συναρτήσεις απώλειας

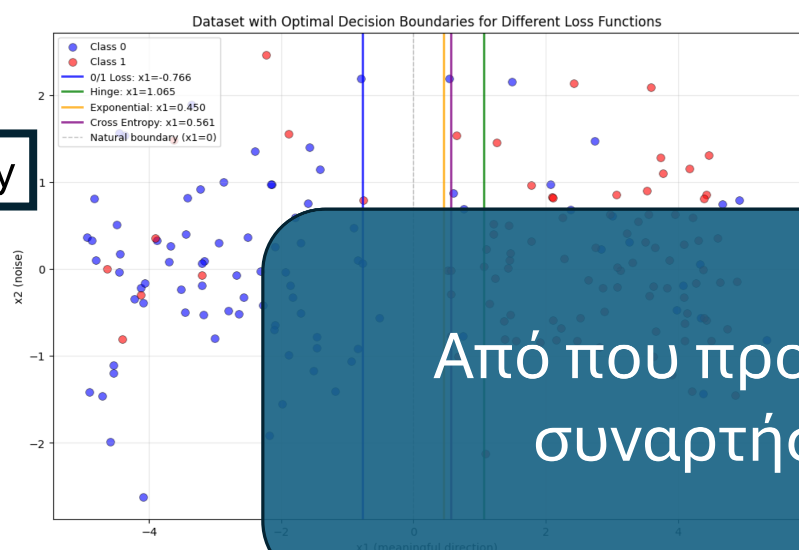
Ταξινόμηση

Errors

Cross Entropy

Hinge Loss

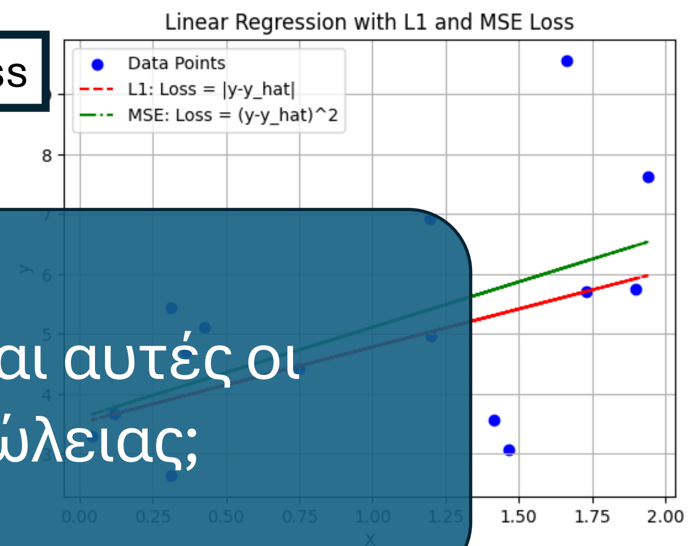
Exp Loss



Παλινδρόμηση

MSE loss

L1 loss



Από που προέρχονται αυτές οι συναρτήσεις απώλειας;

Ένα απλό πείραμα σκέψης

Διαλέγουμε ένα κέρμα, και το ρίχνουμε 100 φορές. Παρατηρούμε 37 φορές «κορόνα» και 63 φορές «γράμματα». Θέλουμε να εκτιμήσουμε την **πιθανότητα** πως θα ξαναβγεί κορόνα:

$$\mathrm{p} = \mathbb{P}(\text{κορόνα})$$

Δύο κοσμοθεωρίες: (1) Ξέρουμε πώς είναι κατασκευασμένα τα κέρματα, και αυτό θα μας δώσει την απάντηση (Bayesian). (2) Ξέρουμε μόνο αυτά που έχουμε παρατηρήσει (Frequentist).

Ένα απλό πείραμα σκέψης

Διαλέγουμε ένα κέρμα, και το ρίχνουμε 100 φορές. Παρατηρούμε 37 φορές «κορόνα» και 63 φορές «γράμματα». Θέλουμε να εκτιμήσουμε την **πιθανότητα** πως θα ξαναβγεί κορόνα:

$$\mathrm{p} = \mathbb{P}(\text{κορόνα})$$

Ιδέα 1: Τα κέρματα είναι ίδια, με $\mathrm{p} = \frac{1}{2}$ και απλά έτυχε να βγουν 37Κ/63Γ.

Ιδέα 2: Το μόνο που ξέρουμε είναι ότι βγήκαν 37 Κ / 63 Γ

Ένα απλό πείραμα σκέψης

Διαλέγουμε ένα κέρμα, και το ρίχνουμε 100 φορές. Παρατηρούμε 37 φορές «κορόνα» και 63 φορές «γράμματα». Θέλουμε να εκτιμήσουμε την **πιθανότητα** πως θα ξαναβγεί κορόνα:

$$\mathrm{p} = \mathbb{P}(\text{κορόνα})$$

Ιδέα 1: Τα κέρματα είναι ίδια, με $\mathrm{p} = \frac{1}{2}$ και απλά έτυχε να βγουν 37Κ/63Γ.

Ιδέα 2: Το μόνο που ξέρουμε είναι ότι βγήκαν 37 Κ / 63 Γ

Αρχή Μέγιστης Πιθανοφάνειας (Maximum Likelihood)

Εάν η πραγματική πιθανότητα να βγεί «κορόνα» είναι $\mathrm{p} = \mathbb{P}(\kappa o \rho \acute{o} \nu \alpha)$, με ποια πιθανότητα θα βλέπαμε 37Κ/63Γ;

Πιθανότητα Κ: p

Πιθανότητα (Κ,Κ): $p \cdot p$

Πιθανότητα (Γ,Κ): $(1 - p) \cdot p$

Πιθανότητα (Κ,Γ): $p \cdot (1 - p)$

Πιθανότητα 1 Κ και 1 Γ: $(1 - p) \cdot p + p \cdot (1 - p)$

Πιθανότητα 2 Κ και 1 Γ: $\underbrace{p^2 \cdot (1 - p)}_{\text{ΚΚΓ}} + \underbrace{p \cdot (1 - p) \cdot p}_{\text{ΚΓΚ}} + \underbrace{(1 - p) \cdot p^2}_{\text{ΓΚΚ}}$

Αρχή Μέγιστης Πιθανοφάνειας (Maximum Likelihood)

Εάν η πραγματική πιθανότητα να βγεί «κορόνα» είναι $\mathrm{p} = \mathbb{P}(κορόνα)$, με

Εάν η πραγματική πιθανότητα να βγει «κορόνα» είναι $\mathrm{p} = \mathbb{P}(κορόνα)$, με ποια πιθανότητα θα βλέπαμε 37Κ/63Γ;

$$\frac{100!}{37!63!} p^{37} \cdot (1-p)^{63} = \binom{100}{37} p^{37} \cdot (1-p)^{63}$$

ΚΚΓ ΚΓΚ ΓΚΚ

Αρχή Μέγιστης Πιθανοφάνειας (Maximum Likelihood)

Εάν η πραγματική πιθανότητα να βγεί «κορόνα» είναι $\mathrm{p} = \mathbb{P}(κορόνα)$, η πιθανότητα να βλέπαμε 37Κ/63Γ είναι:

$$\frac{100!}{37!63!}p^{37} \cdot (1-p)^{63} = \binom{100}{37}p^{37} \cdot (1-p)^{63}$$

Αρχή Μέγιστης Πιθανοφάνειας (Maximum Likelihood)

Εάν η πραγματική πιθανότητα να βγεί «κορόνα» είναι $\mathrm{p} = \mathbb{P}(\kappa o\rho \acute{o} \nu \alpha)$, η πιθανότητα να βλέπαμε 37Κ/63Γ είναι:

$$\frac{\mathbf{100!}}{\mathbf{37!63!}}\boldsymbol{p}^{\mathbf{37}} \cdot (\mathbf{1}-\boldsymbol{p})^{\mathbf{63}} = \binom{\mathbf{100}}{\mathbf{37}}\boldsymbol{p}^{\mathbf{37}} \cdot (\mathbf{1}-\boldsymbol{p})^{\mathbf{63}}$$

Max Likelihood: εάν δούμε 37Κ/63Γ, η αρχή της Μέγιστης Πιθανοφάνειας εκτιμάει πως το πραγματικό $\hat{p}$ είναι αυτό που μεγιστοποιεί την πιθανότητα.

Αρχή Μέγιστης Πιθανοφάνειας (Maximum Likelihood)

$$f(p) = \binom{100}{37} p^{37} \cdot (1-p)^{63}$$

$$argmax:\ f(p) \Rightarrow f'(p) = 0$$

$$f'(p) = \binom{100}{37} [37p^{36} \cdot (1-p)^{63} - 63p^{37}(1-p)^{62}]$$

$$[37p^{36} \cdot (1-p)^{63} - 63p^{37}(1-p)^{62}] = 0$$

$$p^{36} \cdot (1-p)^{62}[37(1-p)^{1} - 63p^{1}] = 0$$

$$[37(1-p)^{1} - 63p^{1}] = 0$$

$$\boldsymbol{\hat{p} = \frac{37}{100}}$$

Maximum Likelihood και Λογιστική Παλινδρόμηση

Πως εφαρμόζουμε την ιδέα της μέγιστης πιθανοφάνειας στη λογιστική παλινδρόμηση, ώστε να βρούμε κατάλληλη συνάρτηση απώλειας;

Λογιστική Παλινδρόμηση: Με δεδομένα $\{(X_i, y_i)\}$, όπου έχουμε $y \in \{0,1\}$, θέλουμε να εκτιμήσουμε την *πιθανότητα* πως το $y = 1$:

$$P(Y = 1|X = x)$$

Maximum Likelihood και Λογιστική Παλινδρόμηση

Πως εφαρμόζουμε την ιδέα της
παλινδρόμηση, ώστε να βρούμε

Λογιστική Παλινδρόμηση: Με δε
$y \in \{0,1\}$, θέλουμε να εκτιμήσου

$$P(Y = 1$$

Όπως έχουμε ήδη δει, αυτές οι ιδέες επεκτείνονται σε γενικότερα προβλήματα με Κ κατηγορίες (π.χ., CIFAR-10), όπου ο στόχος μας είναι να προβλέψουμε την πιθανότητα Y = κ, για κάθε κατηγορία κ.

$$\hat{p}_\kappa = P(Y = \kappa | X = x)$$

$$0 \leq \hat{p}_\kappa \leq 1, \sum_{\kappa \in \mathrm{K}} \hat{p}_\kappa = 1$$

Λογιστική Παλινδρόμηση

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

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

$$Z_X = \beta_0 + \beta^\top X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_d X_d$$

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

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

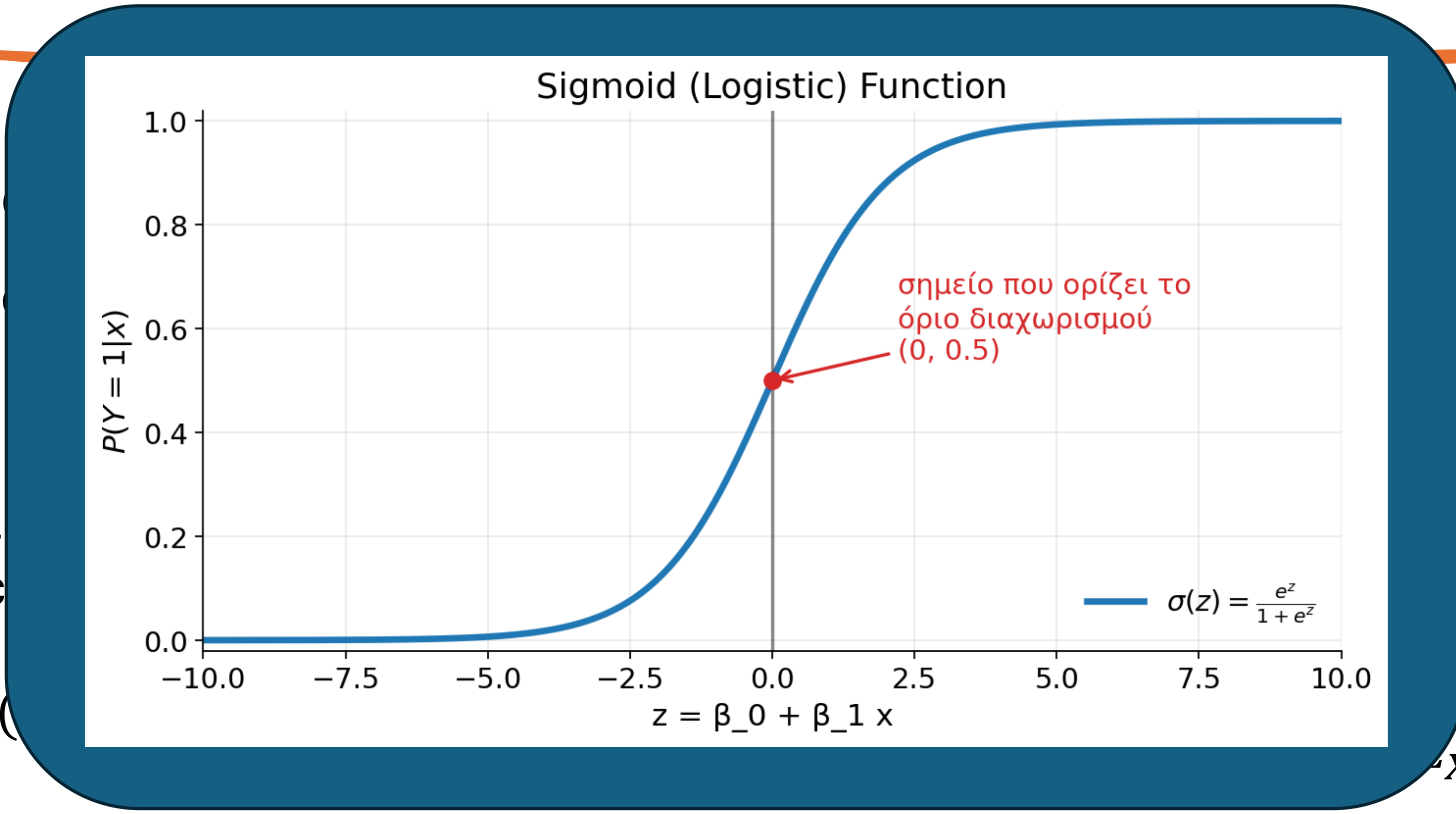
Λογιστική Παλινδρόμηση

Στόχος

Πρώτ

Χρησι ουμε
το «sc

$P($ $_X)$



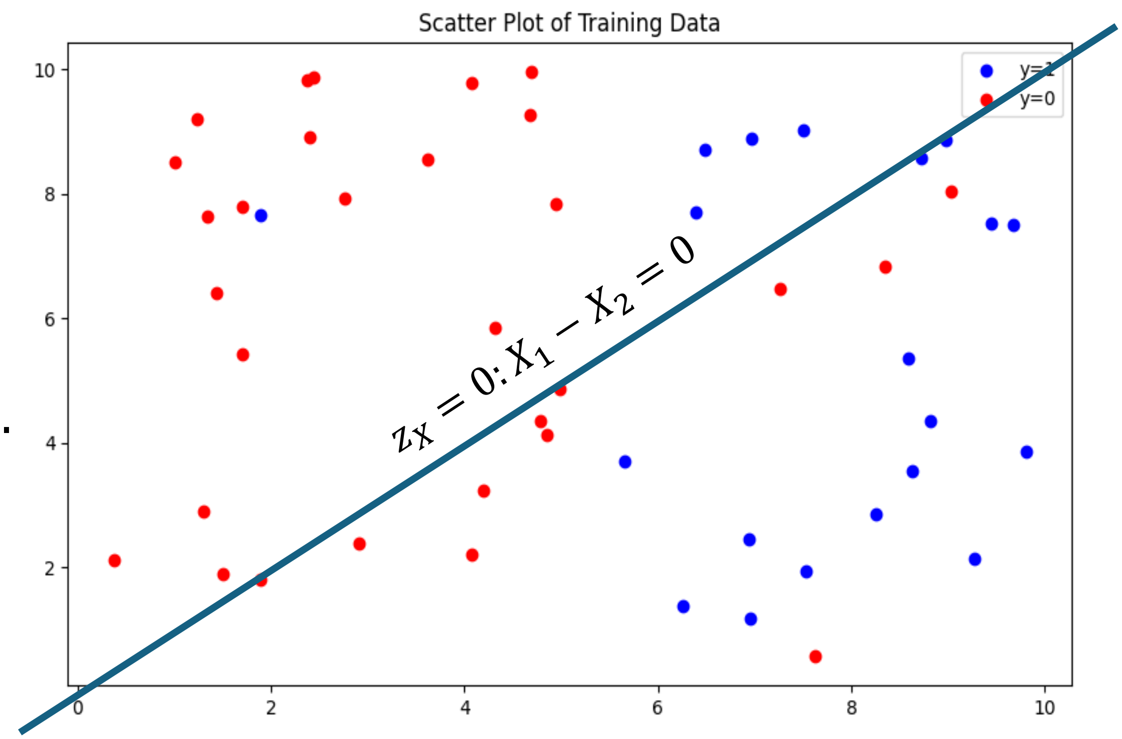
Λογιστική Παλινδρόμηση: πώς λειτουργεί

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

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

Παράδειγμα Α: $\beta_0 = 0, \beta_1 = 2, \ \beta_2 = -2.$

Ποιος διαχωρισμός του χώρου προκύπτει;

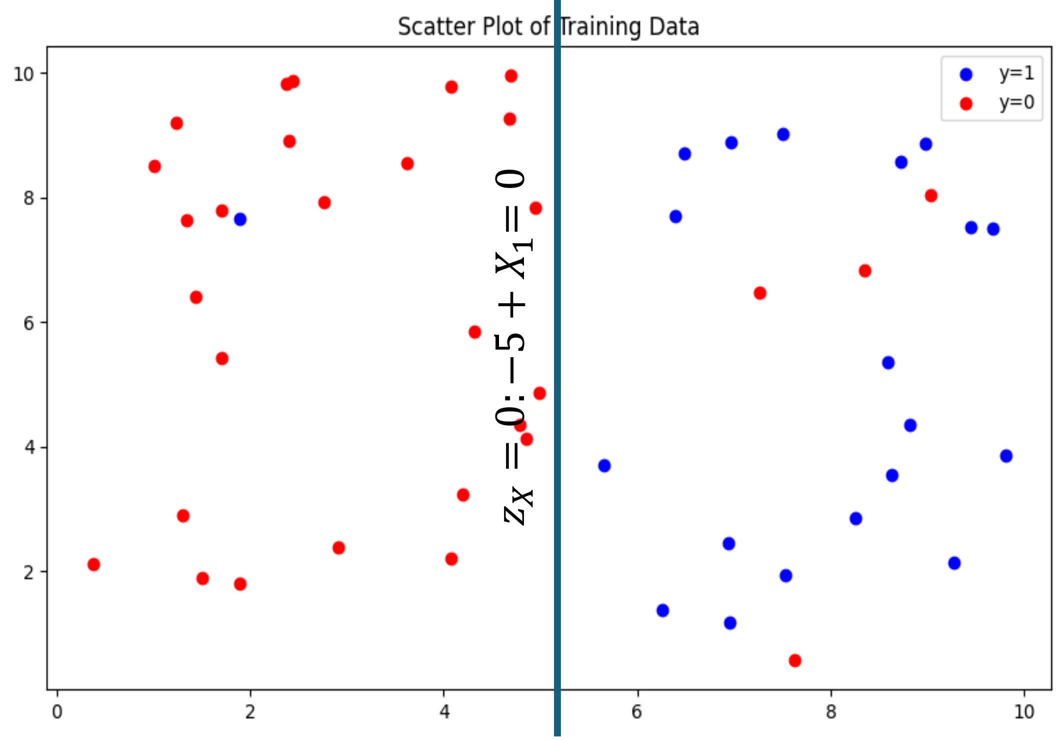


Λογιστική Παλινδρόμηση: πώς λειτουργεί

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

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

Παράδειγμα Β : $\beta_0 = -5, \beta_1 = 1, \ \beta_2 = 0.$



Ποιος διαχωρισμός του χώρου προκύπτει;

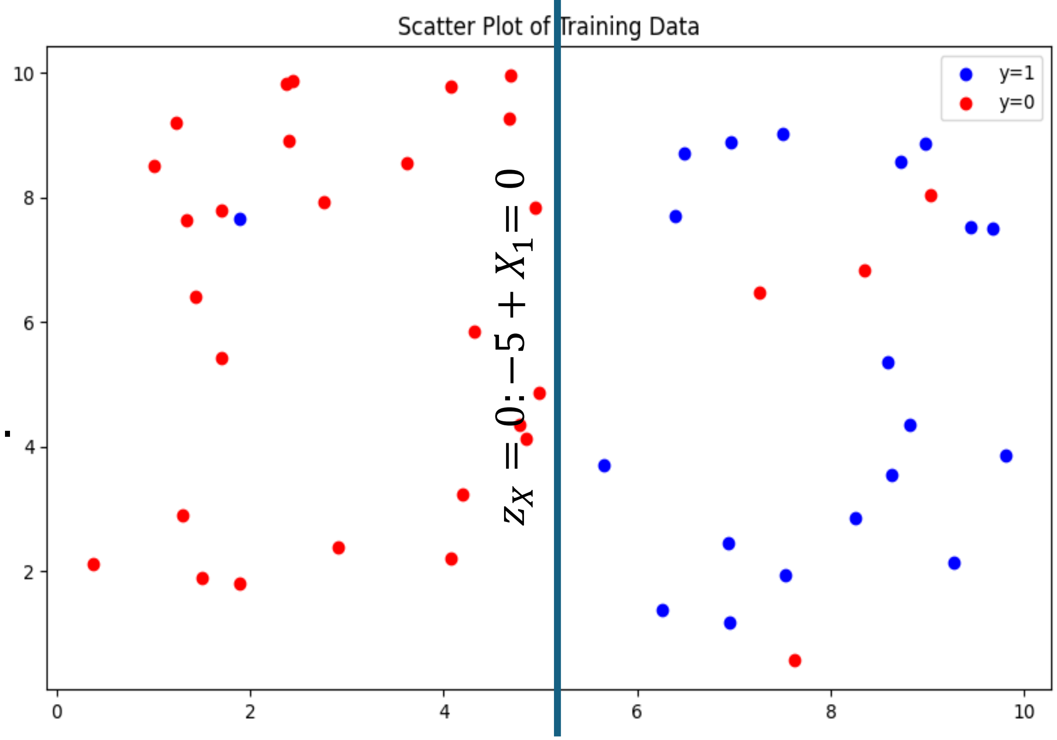
Λογιστική Παλινδρόμηση: πώς λειτουργεί

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

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

Παράδειγμα Γ: $\beta_0 = -10, \beta_1 = 2, \ \beta_2 = 0.$

Ποιος διαχωρισμός του χώρου προκύπτει;



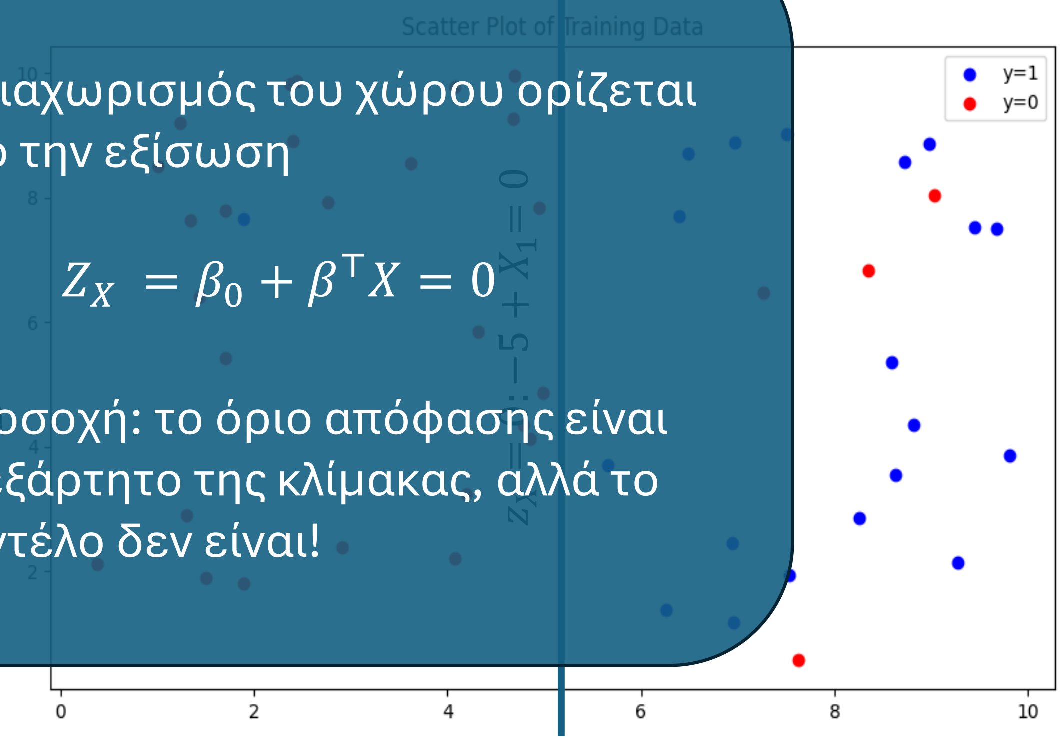
Λογιστική Παλινδρόμηση: πώς λειτουργεί

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

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

Παράδειγμα Γ: $\beta_0 = -15, \beta_1 = 3, \ \beta_2$

Ποιος διαχωρισμός του χώρου προκύπτει;



Ο διαχωρισμός του χώρου ορίζεται από την εξίσωση

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

Προσοχή: το όριο απόφασης είναι ανεξάρτητο της κλίμακας, αλλά το μοντέλο δεν είναι!

Λογιστική Παλινδρόμηση: πώς λειτουργεί

$Z_X = \beta_0 +$

$P(Y = 1|X$

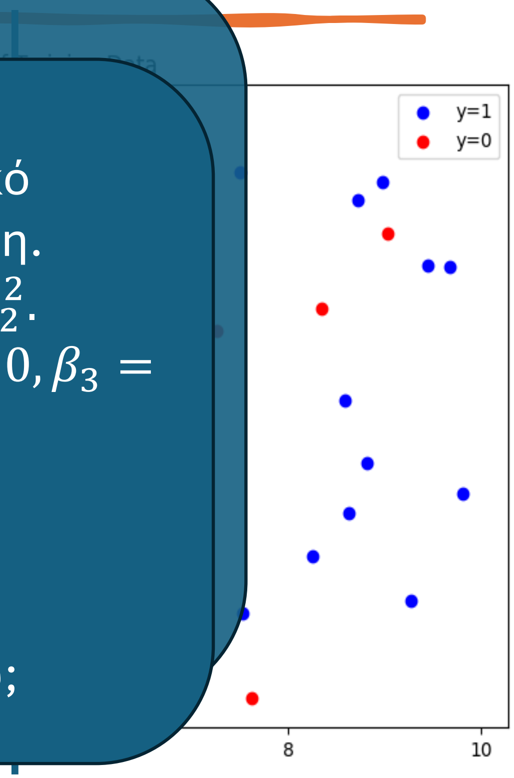
Παράδειγ

Ποιος δια
προκύπτει;

Παρατηρήστε πως μπορούμε να ορίσουμε μη-γραμμικό διαχωρισμό του χώρου για την λογιστική παλινδρόμηση. Παράδειγμα: προσθέτουμε τις στήλες $X_3 = X_1^2, X_4 = X_2^2$. Θεωρήστε το μοντέλο που έχει: $\beta_0 = -4, \beta_1 = 0, \beta_2 = 0, \beta_3 = 1, \beta_4 = 1$. Για ένα σημείο $(\mathrm{X}_1, \mathrm{X}_2)$, έχουμε:

$$Z_X = -4 + 0X_1 + 0X_2 + 1X_1^2 + 1X_2^2$$

Ποια είναι η μορφή του όριου απόφασης του μοντέλου;

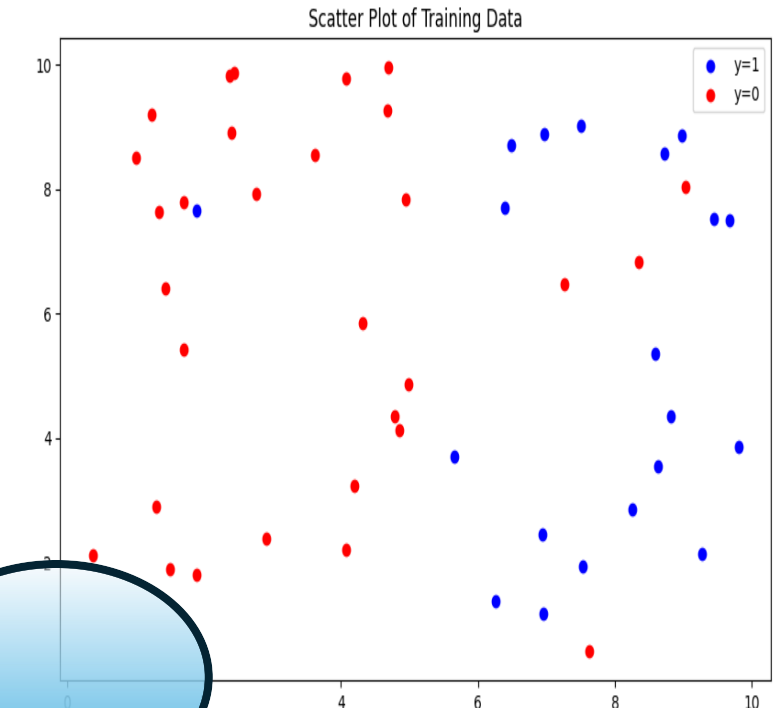


Λογιστική Παλινδρόμηση: πώς λειτουργεί

Παρατηρήστε πως μπορούμε να ορίσουμε μη-γραμμικό διαχωρισμό του χώρου για την λογιστική παλινδρόμηση. Παράδειγμα: προσθέτουμε τις στήλες $X_3 = X_1^2, X_4 = X_2^2$. Θεωρήστε το μοντέλο που έχει: $\beta_0 = -4, \beta_1 = 0, \beta_2 = 0, \beta_3 = 1, \beta_4 = 1.$ Για ένα σημείο $(\mathrm{X}_1, \mathrm{X}_2)$, έχουμε:

$$Z_X = -4 + 0X_1 + 0X_2 + 1X_1^2 + 1X_2^2$$

Ποια είναι η μορφή του όριου απόφασης του μοντέλου;



Λογιστική Παλινδρόμηση: πώς λειτουργεί

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

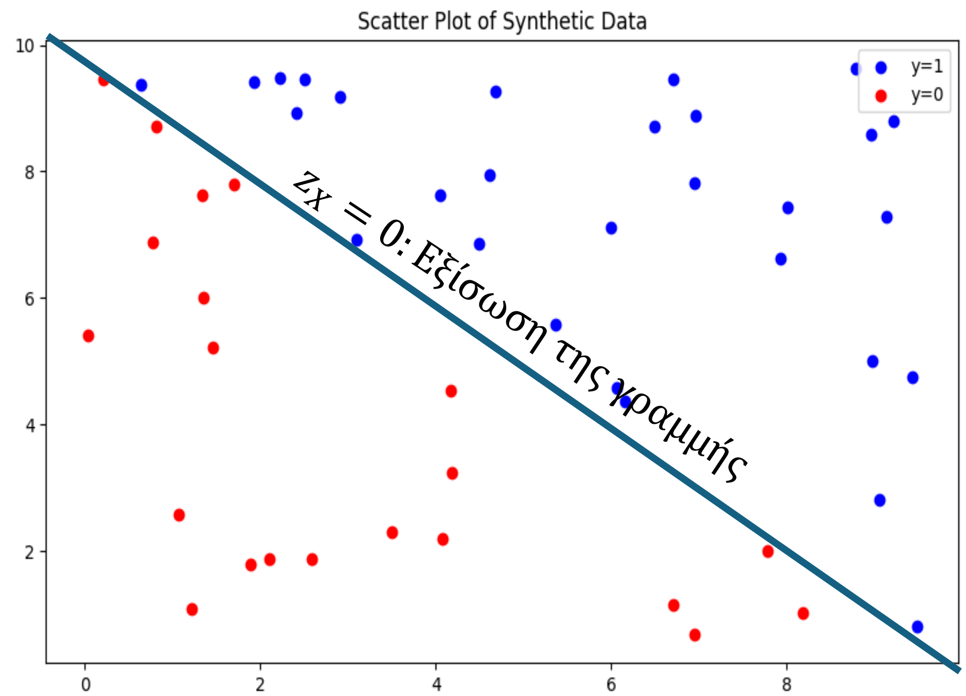
$$P(Y=1|X) = \frac{\exp(z_X)}{1+\exp(z_X)}$$

Ένα καλό μοντέλο;

Ποιο μοντέλο είναι καλύτερο;

$$Z_X = -10 + X_1 + X_2$$

$$\boxed{Z_X = -50 + 5 \cdot X_1 + 5 \cdot X_2}$$



Λογιστική Παλινδρόμηση: πώς λειτουργεί

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

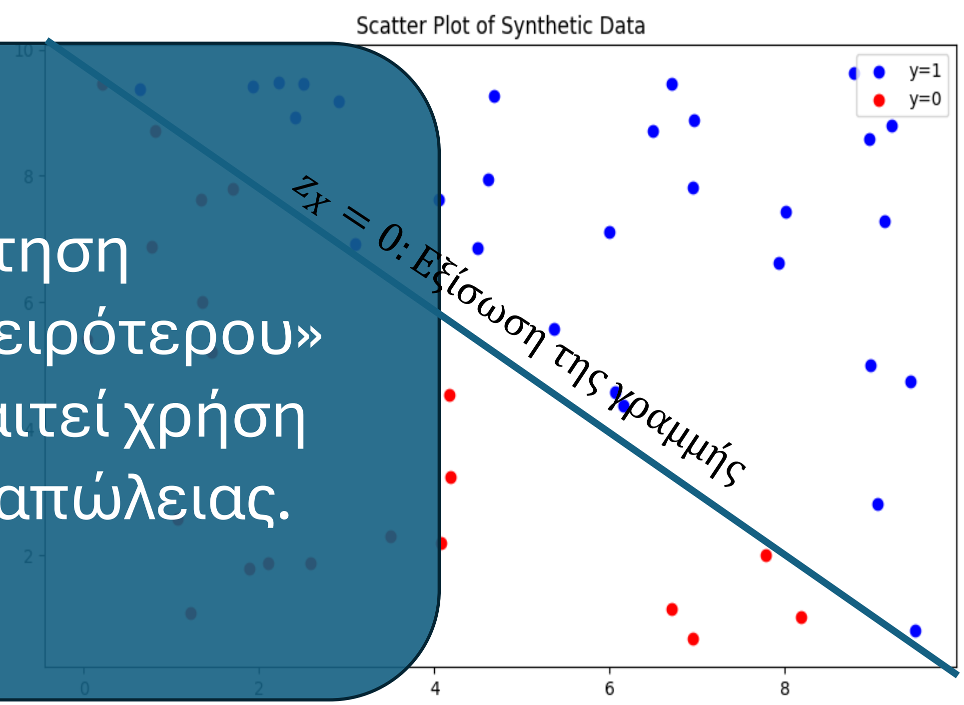
$$P(Y = 1|X) = \frac{\exp(z_X)}{1 + \exp(z_X)}$$

Ένα καλό μοντέλο;

Ποιο μοντέλο είναι καλύτερο;

$$Z_X = -10 + X_1 + X_2$$

$$\boxed{Z_X = -50 + 5 \cdot X_1 + 5 \cdot X_2}$$



Η συζήτηση «καλύτερου/χειρότερου» μοντέλου απαιτεί χρήση συνάρτησης απώλειας.

Λογιστική Παλινδρόμηση και Max Likelihood

Έχουμε δεδομένα: $\{(X_i, y_i)\}$ όπου $y = (1,0,1,1,0)$

Μοντέλο A έχει παραμέτρους (β_0^A, β^A): εκτιμεί $\hat{y} = (0.6,0.4,0.55,0.52,0.45)$

Μοντέλο B έχει παραμέτρους (β_0^B, β^B): εκτιμεί $\hat{y} = (0.9,0.55,0.45,0.85,0.1)$

Ποιο μοντέλο είναι «καλύτερο»;

Ακρίβεια:

Λογιστική Παλινδρόμηση και Max Likelihood

Έχουμε δεδομένα: $\{(X_i, y_i)\}$ όπου $y = (1,0,1,1,0)$

Μοντέλο A έχει παραμέτρους (β_0^A, β^A): εκτιμεί $\hat{y} = (0.6, 0.4, 0.55, 0.52, 0.45)$

Μοντέλο B έχει παραμέτρους (β_0^B, β^B): εκτιμεί $\hat{y} = (0.9, 0.55, 0.45, 0.85, 0.1)$

Ποιο μοντέλο είναι «καλύτερο»;

Ακρίβεια: Μοντέλο A – 5/5 vs 3/5

Πιθανοφάνεια (Likelihood):

Μοντέλο A. $0.6 \cdot (1 - 0.4) \cdot 0.55 \cdot 0.52 \cdot (1 - 0.45) = 0.05663$

Μοντέλο B. $0.9 \cdot (1 - 0.55) \cdot 0.45 \cdot 0.85 \cdot (1 - 0.1) = 0.13122$

Λογιστική Παλινδρόμηση και Max Likelihood

Έχουμε δεδομένα: $\{(X_i, y_i)\}$

Πως υπολογίζουμε την πιθανοφάνεια του μοντέλου με παραμέτρους $(\beta_0, \beta\ )$?

$$Z_X\ = \beta_0 + \beta^\top X \qquad P(Y = 1|X) = \mathrm{sigmoid}(z_X) = \frac{\exp(z_X)}{1 + \exp(z_X)}$$

Πιθανοφάνεια: $\prod_i \underbrace{\mathrm{sigmoid}\left(z_{X_i}\right)^{y_i}}_{\text{Η πιθανοφάνεια όταν έχουμε } y=1} \cdot \underbrace{\left(1 - \mathrm{sigmoid}\left(z_{X_i}\right)\right)^{1-y_i}}_{\text{Η πιθανοφάνεια όταν έχουμε } y=0}$

Λογιστική Παλινδρόμηση και Max Likelihood

Έχουμε δεδομένα: $\{(X_i, y_i)\}$

Πως υπολογίζουμε την πιθανοφάνεια του μοντέλου με παραμέτρους $(\beta_0, \beta\ )$?

$$Z_X = \beta_0 + \beta^\top X \qquad P(Y = 1|X) = \mathrm{sigmoid}(z_X) = \frac{\exp(z_X)}{1 + \exp(z_X)}$$

Πιθανοφάνεια: $\prod_i \mathrm{sigmoid}\left(z_{X_i}\right)^{y_i} \cdot \left(1 - \mathrm{sigmoid}\left(z_{X_i}\right)\right)^{1-y_i}$

Λογαριθμική Πιθανοφάνεια: $\sum_i y_i \cdot \log \mathrm{sigmoid}\left(z_{X_i}\right) + (1 - y_i) \log \left(1 - \mathrm{sigmoid}\left(z_{X_i}\right)\right)$

Η πιθανοφάνεια όταν έχουμε *y=1 ή y=0*

Λογιστική Παλινδρόμηση και Max Likelihood

Έχ

Πω ?

Αρνητική Λογαριθμική Πιθανοφάνεια

$$-\sum_i y_i \cdot \log \text{sigmoid}\left(z_{X_i}\right) + (1 - y_i) \log \left(1 - \text{sigmoid}\left(z_{X_i}\right)\right)$$

Σαν συνάρτηση του (β_0, β), η αρνητική λογαριθμική πιθανοφάνεια (negative log likelihood) ισοδυναμεί με την Cross Entropy σαν συνάρτηση του $\text{softmax}(\beta_0 + \beta^\top x)$ (logits)

$\text{oid}\left(z_{X_i}\right)\big)$

Λογιστική Παλινδρόμηση και Max Likelihood

Αρνητική Λογαριθμική Πιθανοφάνεια

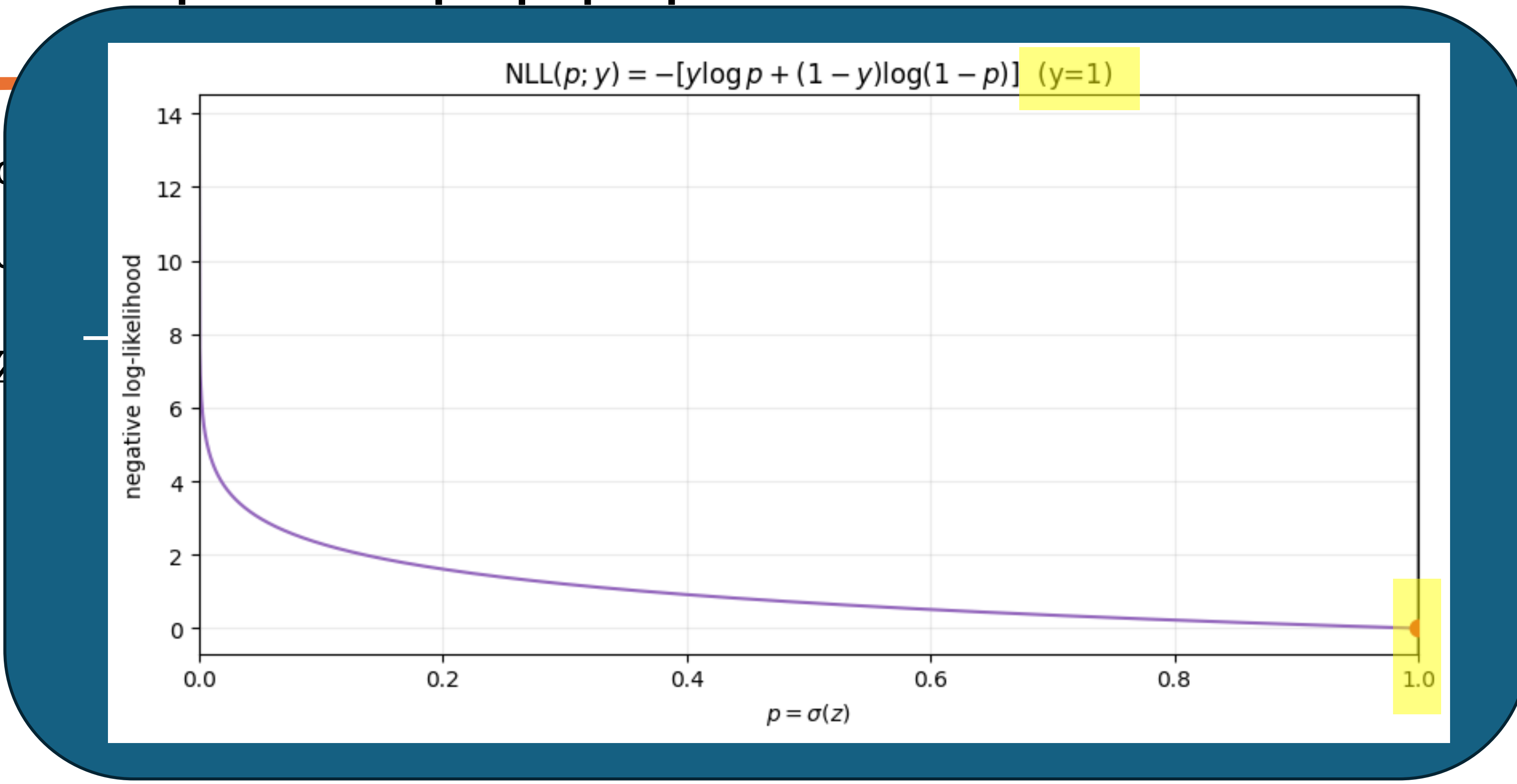
$$\sum_i -\left(y_i \log \frac{\exp(\beta_0 + \beta^\top x_i)}{1 + \exp(\beta_0 + \beta^\top x_i)} + (1 - y_i)\log \frac{1}{1 + \exp(\beta_0 + \beta^\top x_i)} \right)$$

Cross Entropy σαν συνάρτηση του $\mathrm{softmax}(\beta_0 + \beta^\top x)$ (logits) $\mathrm{oid}\left(z_{X_i}\right)\Big)$

Λογιστική Παλινδρόμηση και Max Likelihood

Έχ

Πω ?

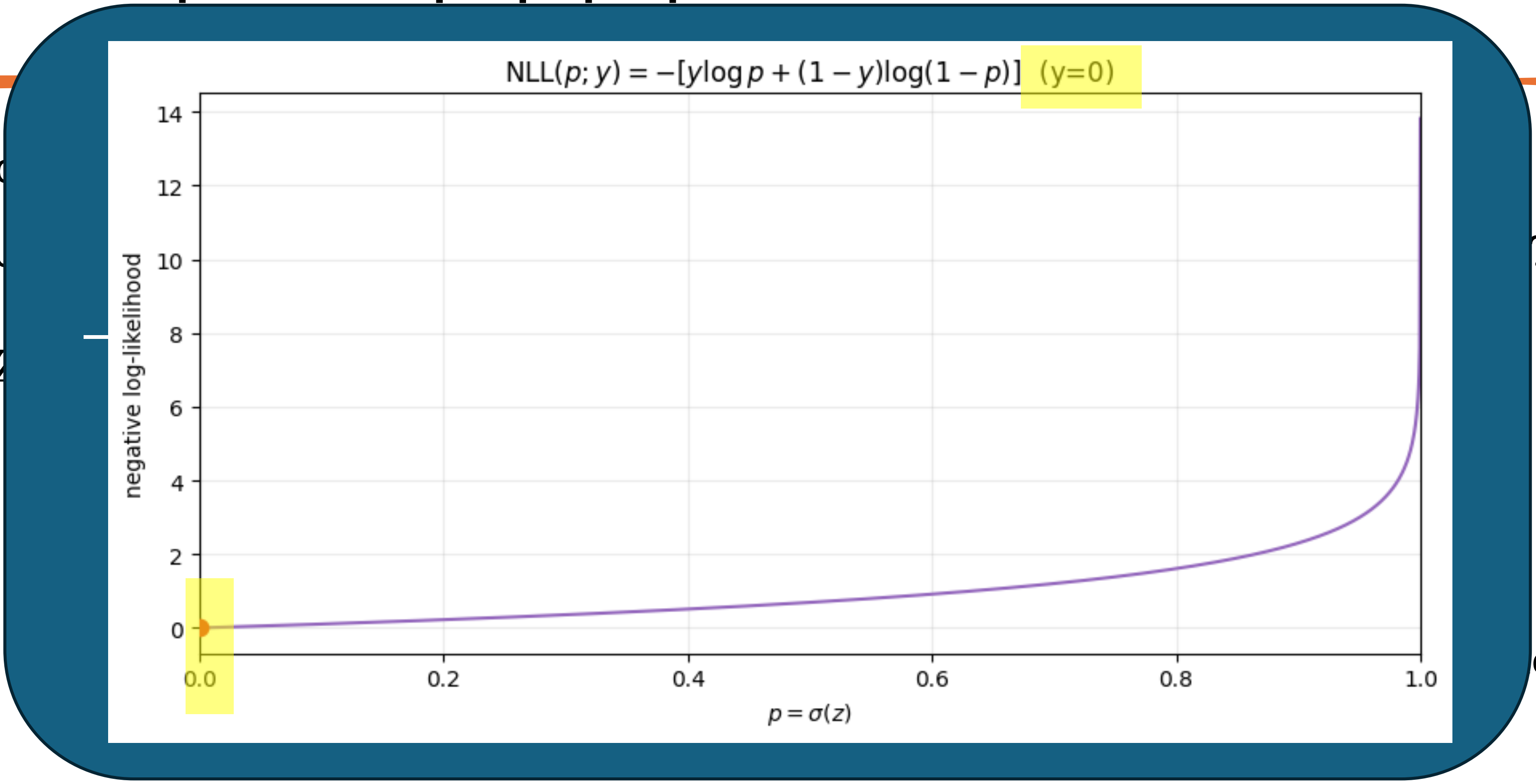


oid$(z_{X_i})\big)$

Λογιστική Παλινδρόμηση και Max Likelihood

Έχ

Πω ?



$\mathrm{oid}\left(z_{X_i}\right)\Big)$

Λογιστική Παλινδρόμηση και Max Likelihood

Αρνητική Λογαριθμική Πιθανοφάνεια

$$\sum_i -\left(y_i \log \frac{\exp(\beta_0 + \beta^\top x_i)}{1 + \exp(\beta_0 + \beta^\top x_i)} + (1 - y_i)\log \frac{1}{1 + \exp(\beta_0 + \beta^\top x_i)}\right)$$

Η συνάρτηση απώλειας έχει νόημα και όταν έχουμε $y \in [0,1]$ και όχι μόνο $y \in \{0,1\}$

Λο… $\mathrm{sigmoid}(z_{X_i}))$

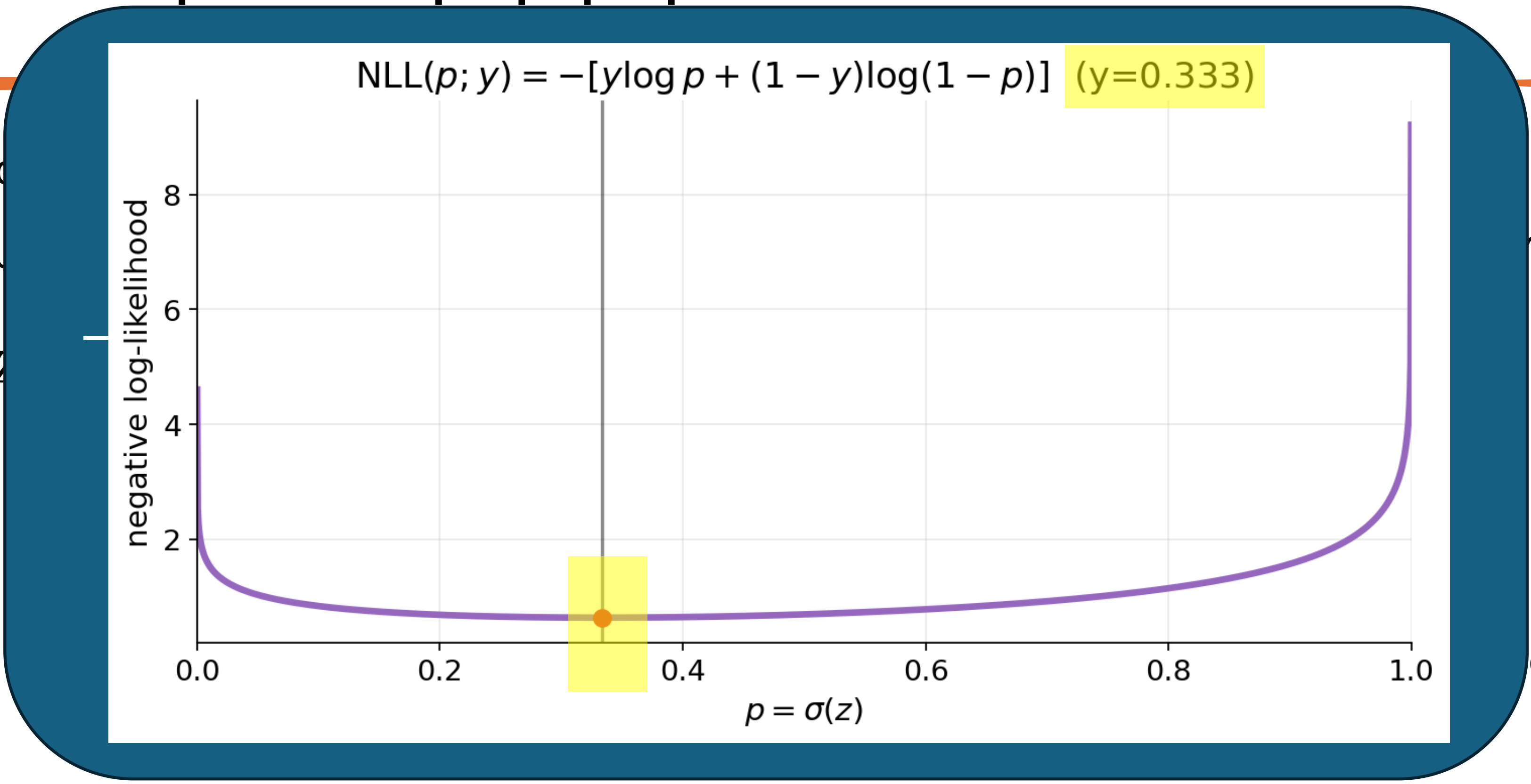
Λογιστική Παλινδρόμηση και Max Likelihood

Έχ

Πω ?

Ζ

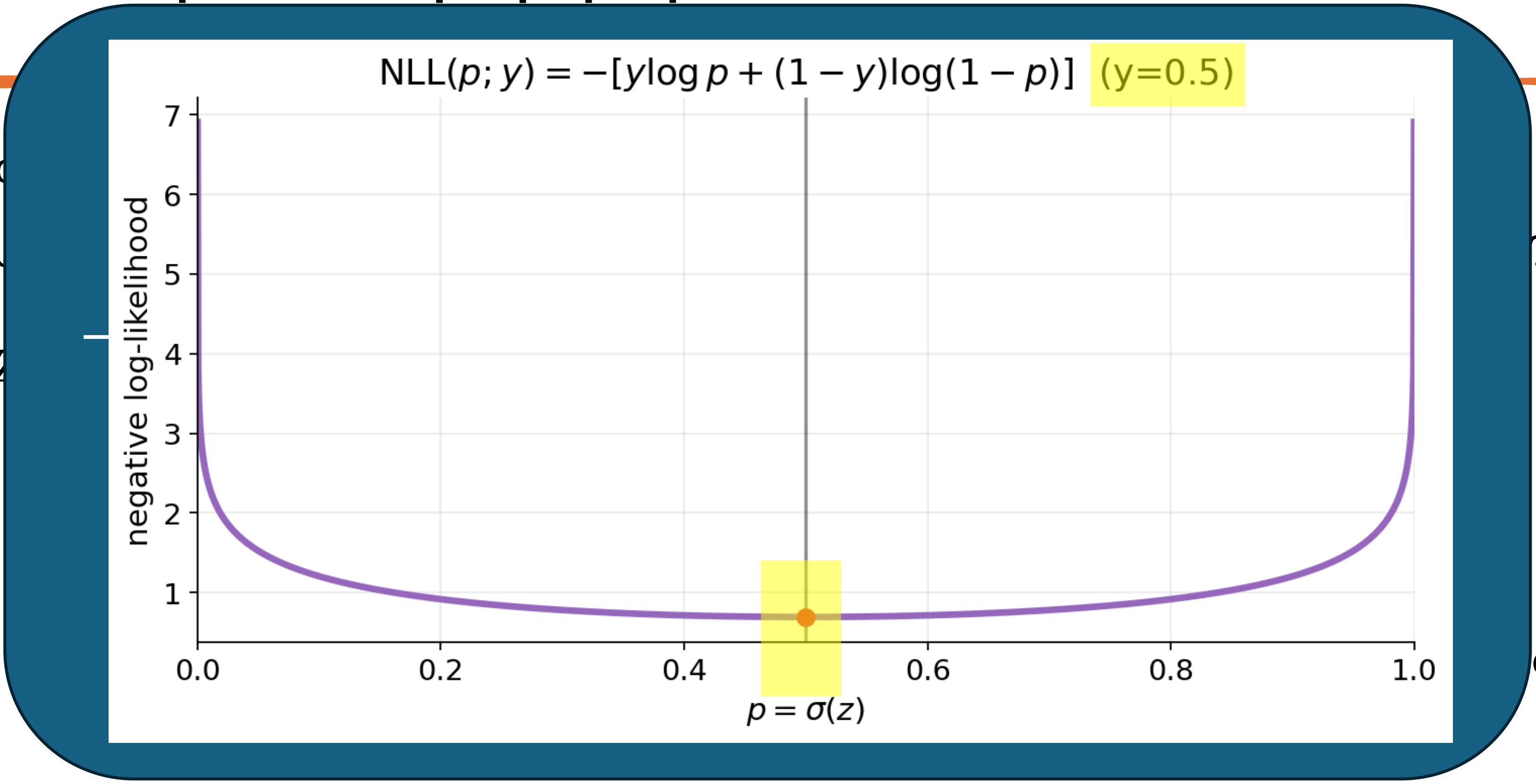
oid$\left(z_{X_i}\right)\big)$



Λογιστική Παλινδρόμηση και Max Likelihood

Έχ

Πω ?

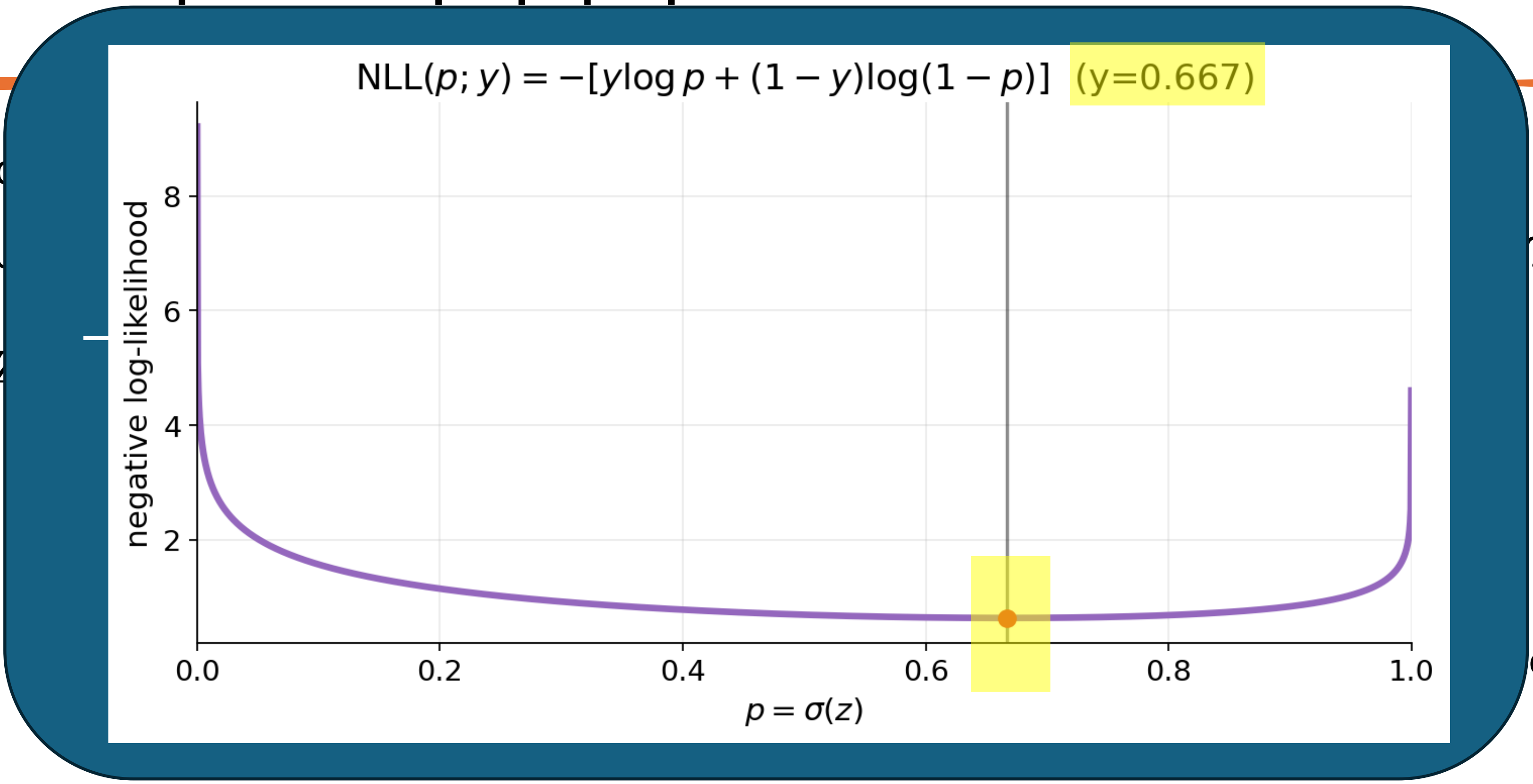


$\text{oid}\left(z_{X_i}\right)\Big)$

Λογιστική Παλινδρόμηση και Max Likelihood

Έχ

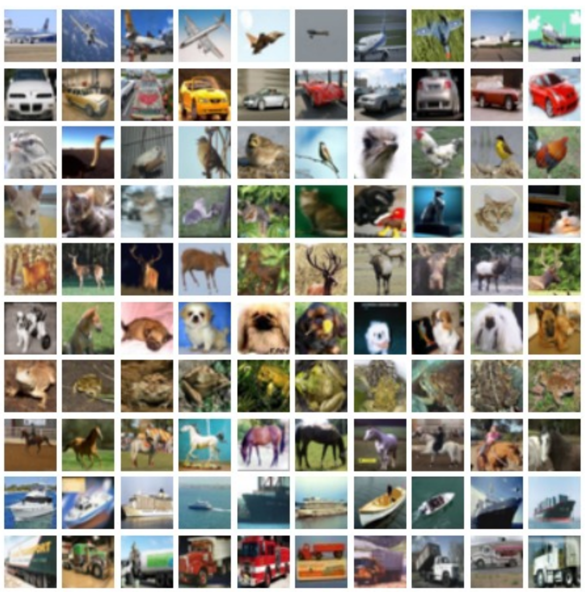
Πω ?



$\mathrm{oid}\left(z_{X_i}\right)\Big)$

Λογιστική Παλινδρόμηση και Max Likelihood

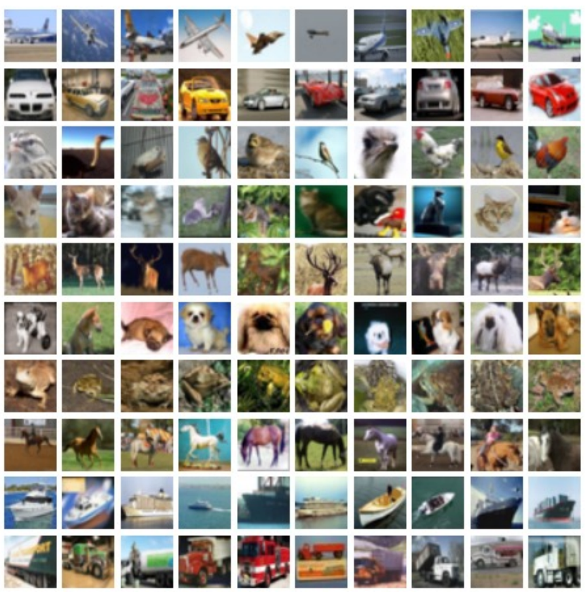
Μηχανική Όραση



$$
\left\{\begin{array}{l}
\text{σκύλος}\\
\text{γάτα}\\
\boxed{\text{ελάφι}}\\
\text{άλογο}\\
\text{αυτοκίνητο}\\
\text{φορτηγό}\\
\text{πλοίο}\\
\text{αεροπλάνο}\\
\text{πτηνό}\\
\text{βάτραχος}
\end{array}\right\}
\qquad
y = \left\{\begin{array}{c}
0\\ 0\\ \boxed{1}\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0
\end{array}\right\}
$$

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση

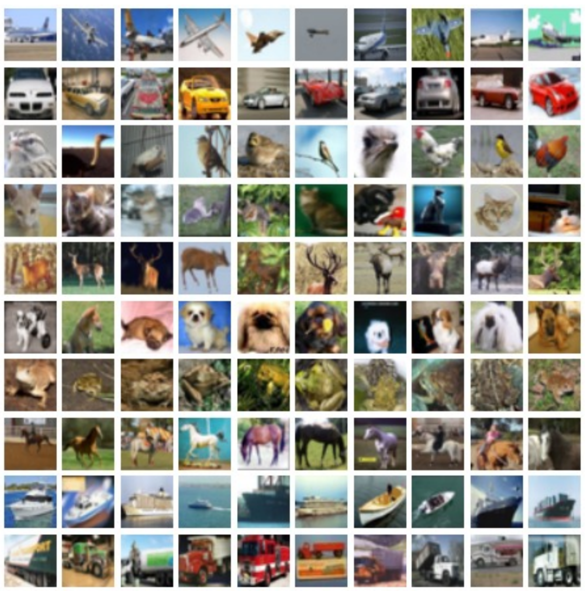


$$\left\{\begin{array}{l}\text{σκύλος}\\ \text{γάτα}\\ \text{ελάφι}\\ \boxed{\text{άλογο}}\\ \text{αυτοκίνητο}\\ \text{φορτηγό}\\ \text{πλοίο}\\ \text{αεροπλάνο}\\ \text{πτηνό}\\ \text{βάτραχος}\end{array}\right\}$$

$$y = \left\{\begin{array}{c}0\\0\\0\\\boxed{1}\\0\\0\\0\\0\\0\\0\end{array}\right\}$$

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση

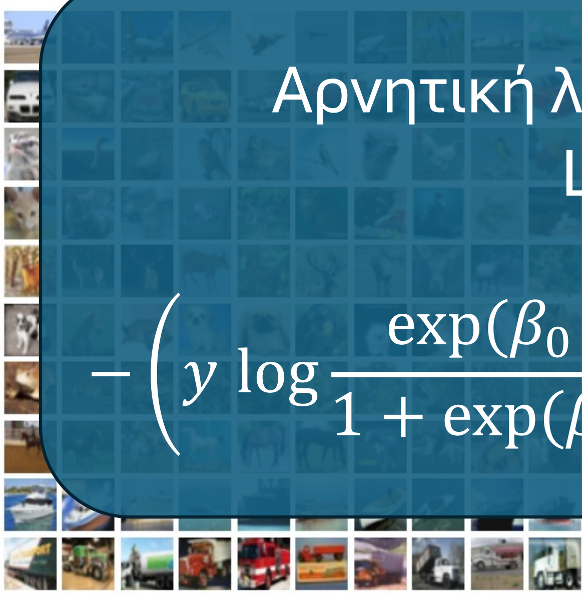


σκύλος
γάτα
ελάφι
άλογο
αυτοκίνητο
φορτηγό
πλοίο
αεροπλάνο
πτηνό
βάτραχος

$$y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση



σκύλος
γάτα
ελάφι
αυτοκίνητο
φορτηγό
πλοίο
αεροπλάνο
πτηνό
βάτραχος

0

Αρνητική λογαριθμική πιθανοφάνεια (Negative Log Likelihood) για δύο κατηγορίες

$$-\left(y \log \frac{\exp(\beta_0 + \beta^\top x)}{1 + \exp(\beta_0 + \beta^\top x)} + (1 - y)\log \frac{1}{1 + \exp(\beta_0 + \beta^\top x)}\right)$$

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση

Αρνητική λογαριθμική πιθανοφάνεια (Negative Log Likelihood) για δύο κατηγορίες

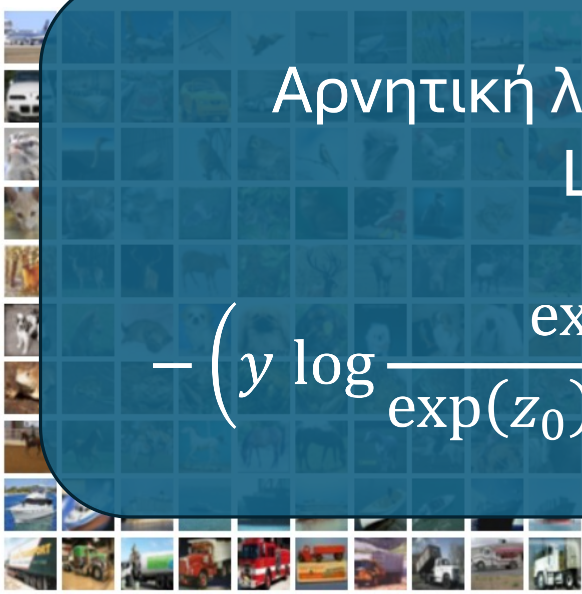
$$-\left(y \log \frac{\exp(z_1)}{1+\exp(z_1)} + (1-y)\log \frac{1}{1+\exp(z_1)}\right)$$

βάτραχος

0

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση



Αρνητική λογαριθμική πιθανοφάνεια (Negative Log Likelihood) για δύο κατηγορίες

$$-\left(y \log \frac{\exp(z_1)}{\exp(z_0)+\exp(z_1)} + (1-y)\log \frac{\exp(z_0)}{\exp(z_0)+\exp(z_1)}\right)$$

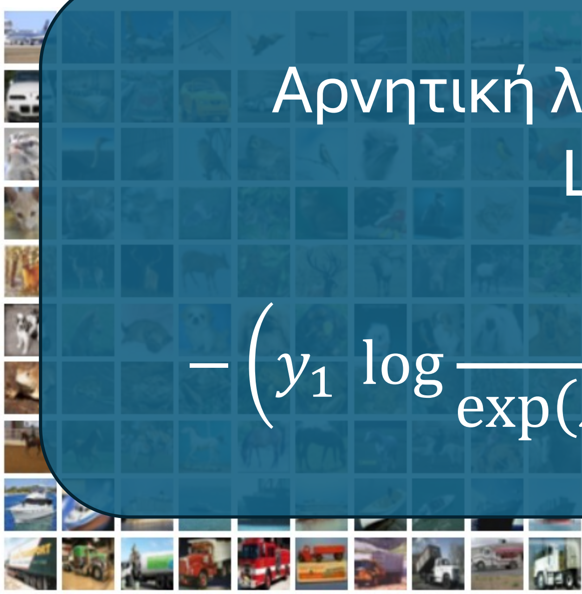
σκύλος
γάτα
ελάφι
άλογο
αυτοκίνητο
φορτηγό
πλοίο
αεροπλάνο
πτηνό
βάτραχος

y

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση



σκύλος
γάτα
ελάφι
αυτοκίνητο
φορτηγό
αεροπλάνο
πτηνό
βάτραχος

Αρνητική λογαριθμική πιθανοφάνεια (Negative Log Likelihood) για δύο κατηγορίες

$$-\left(y_1 \log \frac{\exp(z_1)}{\exp(z_0)+\exp(z_1)} + y_0 \log \frac{\exp(z_0)}{\exp(z_0)+\exp(z_1)}\right)$$

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση

Αρνητική λογαριθμική πιθανοφάνεια (Negative Log Likelihood) για δύο κατηγορίες

$$-\sum_{i=0}^{1}\left(y_i \log \frac{\exp(z_i)}{\exp(z_0)+\exp(z_1)}\right)$$

σκύλος
γάτα
ελάφι
αυτοκίνητο
πλοίο
αεροπλάνο
πτηνό
βάτραχος

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση

Αρνητική λογαριθμική πιθανοφάνεια (Negative Log Likelihood) για K κατηγορίες

$$-\sum_{i=1}^{K}\left(y_i \log \frac{\exp(z_i)}{\exp(z_1)+\cdots+\exp(z_K)}\right)$$

Λογιστική Παλινδρόμηση και Max Likelihood

Μηχανική Όραση

CIFAR-10: 10 Κατηγορίες
ImageNet: 1000 Κατηγορίες

Φυσική Γλώσσα: 50.000 Κατηγορίες

Λογιστική Παλινδρόμηση και Max Likelihood

Επεξεργασία Φυσικής Γλώσσας

Είναι σημαντικό πρόβλημα της επεξεργασίας φυσικής γλώσσας η επέκταση μιας [redacted] ⇒ $y = \left\{ \begin{array}{l} 0.2\text{: σκέψης} \\ 0.25\text{: πρότασης} \\ 0.3\text{: ιδέας} \\ 0.001\text{: γάτας} \\ \ldots \\ \ldots \\ \ldots \\ \ldots \end{array} \right\}$

Για την εκπαίδευση μεγάλων γλωσσικών μοντέλων πάλι χρησιμοποιούμε την ίδια απώλεια: Cross Entropy Loss / Negative Log Likelihood Loss

Maximum Likelihood & Παλινδρόμηση

Παλινδρόμιση

$$y = \beta^{\top} x + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

Μοντέλο: $x \to \hat{y} = \hat{\beta}^{\top} x$

Likelihood: $\left(y - \hat{\beta}^{\top} x\right) \sim N(0, \sigma^2) \propto \exp\left(-\left(\frac{y - \hat{\beta}^{\top} x}{\sigma^2}\right)^2\right)$

-log Likelihood: $\left(\frac{y - \hat{\beta}^{\top} x}{\sigma^2}\right)^2$

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

Maximum Likelihood & Παλινδρόμηση

Παλινδρόμιση

Αρχή Μέγιστης Πιθανοφάνειας (Maximum Likelihood)

(Λογιστική) Παλινδρόμηση: καλό μοντέλο → μεγάλη πιθανοφάνεια

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

Maximum Likelihood & Λογιστική Παλινδρόμηση

Λογιστική Παλινδρόμιση

$$y = 0 \text{ ή } y = 1 \rightarrow y = \binom{1}{0} \text{ ή } y = \binom{0}{1}$$

$$x \rightarrow \beta^\top x \rightarrow \text{softmax}(\beta^\top x) = \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix} = \begin{pmatrix} \hat{p} \\ 1-\hat{p} \end{pmatrix} = \begin{pmatrix} \dfrac{\exp(-\beta^\top x)}{1-\exp(-\beta^\top x)} \\ \dfrac{1}{1-\exp(-\beta^\top x)} \end{pmatrix}$$

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

?

Maximum Likelihood & Λογιστική Παλινδρόμηση

Λογιστική Παλινδρόμισ

Πιθανοφάνεια (Likel

$\boldsymbol{y = 0}$ ή $\boldsymbol{1}$: $\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{1-y} \end{pmatrix}$

Likelihood: $y =$

$\hat{p}^{y}(1-\hat{p})^{1-y}$

$\boldsymbol{Loss\big((y_1, y_2), (\widehat{p}_1, \widehat{p}_2)\big) = -\big(y_1 \log(\widehat{p}_1) + y_2 \log(\widehat{p}_2)\big)}$

\- log Likelihood: $-(y \log \hat{p} + (1-y)\log(1-\hat{p}))$

Η απώλεια της Λογιστικής Παλινδρόμησης είναι ακριβώς το **- Log Likelihood**

Maximum Likelihood & Λογιστική Παλινδρόμηση

Παράδειγμα με 3 δεδομένα: $(x_1, y_1 = 1), (x_2, y_2 = 0), (x_3, y_3 = 0)$

Μοντέλο: $x_1 \rightarrow \binom{\hat{p}_1}{1-\hat{p}_1}, x_2 \rightarrow \binom{\hat{p}_2}{1-\hat{p}_2}, x_3 \rightarrow \binom{\hat{p}_3}{1-\hat{p}_3}$

$$y_1 = \binom{1}{0}, y_2 = \binom{0}{1}, y_3 = \binom{0}{1}$$

Πιθανοφάνεια σύμφωνα με το μοντέλο μας: $\hat{p}_1 \cdot (1 - \hat{p}_2) \cdot (1 - \hat{p}_3)$

Καλό μοντέλο → μεγάλη πιθανοφάνεια

Maximum Likelihood & Λογιστική Παλινδρόμηση

Καλό μοντέλο → μεγάλη πιθανοφάνεια

$y_1 = 1, y_2 = 0, y_3 = 0$

Πιθανοφάνεια (Likelihood): $\hat{p}_1 \cdot (1 - \hat{p}_2) \cdot (1 - \hat{p}_3)$

-Log Likelihood: $-(\log\ \hat{p}_1 + \log(1 - \hat{p}_2) + \log(1 - \hat{p}_3))$

$$= -\boxed{(\log\ \hat{p}_1} + \boxed{\log(1 - \hat{p}_2)} + \boxed{\log(1 - \hat{p}_3))}$$

$$= -\boxed{(y_1 \log\ \hat{p}_1 + (1 - y_1)\log(1 - \hat{p}_1)}$$
$$\boxed{+ y_2 \log\ \hat{p}_2 + (1 - y_2)\log(1 - \hat{p}_2)}$$
$$\boxed{+ y_3 \log\ \hat{p}_3 + (1 - y_3)\log(1 - \hat{p}_3))}$$

Maximum Likelihood & Λογιστική Παλινδρόμηση

Καλό μοντέλο → μεγάλη πιθανοφάνεια

Παράδειγμα με 3 δεδομένα: Λογιστική Παλινδρόμηση $(x_1, y_1 = 1), (x_2, y_2 = 0), (x_3, y_3 = 0)$

$y_1 = 1, y_2 = 0, y_3 = 0$

$$\sum_i -\left(y_i \log \frac{\exp(\beta^\top x_i)}{1+\exp(\beta^\top x_i)} + (1-y_i) \frac{1}{1+\exp(\beta^\top x_i)} \right)$$

$$\boxed{+y_3 \log \hat{p}_3 + (1-y_3)\log(1-\hat{p}_3))}$$