ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Αναπλ. Καθηγ. Στελιος Ζήμερας
Τμημα Στατιστικης και Αναλογιστικων – Χρηματοοικονομικων Μαθηματικων
Πανεπιστημιο Αιγαιου
Σαμος

2021

Εισαγωγή

Το μοντέλο της Λογιστικής παλινδρόμησης (logistic regression) αποτελεί ειδική περίπτωση των γενικευμένων γραμμικών μοντέλων

Η Λογιστική παλινδρόμηση είναι χρήσιμη σε καταστάσεις στις οποίες επιθυμούμε την πρόβλεψη ύπαρξης ή της απουσίας ενός χαρακτηριστικού ή ενός συμβάντος.

Η πρόβλεψη αυτή βασίζεται στην κατασκευή ενός γραμμικού μοντέλου και συγκεκριμένα στον προσδιορισμό των τιμών που παίρνουν οι συντελεστές ενός συνόλου ανεξάρτητων μεταβλητών που χρησιμοποιούνται ως μεταβλητές πρόβλεψης

Εισαγωγή

Σε πολλές εφαρμογές η εξαρτημένη μεταβλητή παίρνει δυο μόνο τιμές , οι οποίες αντιστοιχούν σε δυο ενδεχόμενα. Για παράδειγμα , το αν ο ασθενής ζει ή απεβίωσε

Οι τιμές της μεταβλητής αποτελούν μια αυθαίρετη κωδικοποίηση των δυο ενδεχομένων , συνήθως 0 και 1

ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΙΣΗ

Τα μη γραμμικά μοντέλα έχουν την πιο κάτω μορφή:

$$Y_i = f\left(X_i, \gamma\right) + \varepsilon_i$$

η μορφή αυτή μοιάζει με τη μορφή που έχουμε για τα γραμμικά μοντέλα (δηλαδή η παρατήρηση Y_i είναι το άθροισμα της αναμενόμενης συνάρτησης $f(X_i,\gamma)$ με τυχαία σφάλματα ε_i.

η διαφορά είναι ότι η αναμενόμενη συνάρτηση εδώ είναι μη γραμμική

ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΙΣΗ

- Τα σφάλματα είναι τυχαίες μεταβλητές με τις πιο κάτω υποθέσεις:
- $E(\varepsilon_i)=0$
- Σταθερή διασπορά
- Ανά δύο τα σφάλματα είναι ασυσχέτιστα $E(\varepsilon_i,\varepsilon_j)=0$, $\forall i \neq j$
- Επίσης πολλές φορές υποθέτουμε ότι είναι κανονικές μεταβλητές

ΑΝΕΞΑΡΤΗΤΕΣ ΜΕΤΑΒΛΗΤΕΣ

ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΙΣΗ

Το λογιστικό μοντέλο είναι ένα

1. μη γραμμικό μοντέλο
2. τα σφάλματα δεν ακολουθούν κανονική κατανομή και
3. η μεταβλητή απόκρισης είναι διακριτή.

- Η λογιστική παλινδρόμηση χρησιμοποιείται σε περιπτώσεις στις οποίες επιθυμούμε να προβλέψουμε την απουσία ή την παρουσία ενός χαρακτηριστικού, ή ενός συμβάντος. Είναι μια γενίκευση της απλής γραμμικής παλινδρόμησης για την περίπτωση όπου η εξαρτημένη μεταβλητή (Y) είναι δίτιμη (δηλαδή παίρνει την τιμή 0 όταν απουσιάζει το χαρακτηριστικό ή την τιμή 1 όταν υπάρχει το χαρακτηριστικό).

Εισαγωγή

Εάν ορίσουμε την τιμή $y = 1$ σαν «επιτυχία» και την τιμή $y = 0$ σαν «αποτυχία» , τότε η y είναι τ.μ της κατανομής Bernoulli , δηλαδή $y \sim B(p)$, με μέση τιμή $E(y) = p$ και διασπορά $V(y) = p(1-p)$.

Γενικεύοντας σε μια σειρά από n επαναλήψεις (δηλαδή πραγματοποιήσεων των ενδεχομένων), ορίζουμε την τ.μ

$$y = \alpha\rho\iota\theta\mu\acute{o}\varsigma\ \varepsilon\pi\iota\tau\upsilon\chi\iota\acute{\omega}\nu\ \sigma\varepsilon\ n\ \delta o\kappa\iota\mu\acute{\varepsilon}\varsigma$$

Υπό την υπόθεση ότι η πιθανότητα επιτυχίας p είναι ίδια σε κάθε δοκιμή και οι δοκιμές είναι ανεξάρτητες μεταξύ τους , τότε ισχύει η Διωνυμική (binomial) κατανομή

$$y \sim b(n, p)$$

Εισαγωγή

σππ

$$f(y) = \binom{n}{p} p^y (1-p)^{n-y} \, , y = 0,1,2, \dots , n$$

p η πιθανότητα επιτυχίας η οποία είναι παράμετρος της κατανομής.

Η Διωνυμική κατανομή αποτελεί τη βασική κατανομή για την περιγραφή και ανάλυση μιας μεταβλητής αυτής της φύσης. Η μέση τιμή της y είναι ίση με $E(y) = np$ και η διασπορά με $V(y) = np(1-p)$. Στην ειδική περίπτωση που $n = 1$ μιλάμε για *δυαδικά δεδομένα* , αλλιώς για *διωνυμικά δεδομένα.*

ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΙΣΗ

Ερμηνεία

- Απλό γραμμικό μοντέλο:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Δίτιμη μεταβλητή (0, 1)

- Επειδή ισχύει $E(\varepsilon_i)=0$

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 X_i + \varepsilon_i) \\ &= E(\beta_0 + \beta_1 X_i) + E(\varepsilon_i) \\ &= \beta_0 + \beta_1 X_i \end{aligned}$$

ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΙΣΗ

Επίσης αφού είναι δίτιμη μεταβλητή η Y_i, θα είναι μια μεταβλητή Bernoulli με

- Όταν το $Y_i = 1$ έχουμε $\mathrm{P}(Y_i = 1) = \pi_i$
- Όταν το $Y_i = 0$ έχουμε $\mathrm{P}(Y_i = 0) = 1 - \pi_i$

Με βάση τον ορισμό της αναμενόμενης τιμής έχουμε

$$E(Y_i) = 1\pi_i + 0(1 - \pi_i) = \pi_i$$

Εξισώνοντας τους τύπους των δύο αναμενόμενων τιμών έχουμε

$$E(Y_i) = \beta_0 + \beta_1 X_i = \pi_i$$

Όταν το $Y_i = 1$ έχουμε $\mathrm{P}(Y_i = 1) = \pi_i$

Όταν ανεξάρτητη μεταβλητή η X_i

ΓΙΑΤΙ ΌΧΙ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ

1. Τα σφάλματα δεν είναι κανονικά

Έχουμε

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \Leftrightarrow \varepsilon_i = Y_i - (\beta_0 + \beta_1 X_i)$$

Όταν

$$Y_i = 0: \ \varepsilon_i = -\beta_0 - \beta_1 X_i$$

$$Y_i = 1: \ \varepsilon_i = 1 - \beta_0 - \beta_1 X_i$$

Όχι κανονική κατανομή σφαλμάτων

ΓΙΑΤΙ ΌΧΙ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ

2. Τα σφάλματα έχουν άνισες διασπορές

Όταν η αποκρινόμενη μεταβλητή παίρνει τις τιμές 0 ή 1 τα σφάλματα δεν έχουν ίσες διασπορές.

$$Var(\varepsilon_i) = Var(Y_i - \pi_i) = Var(Y_i) + Var(-\pi_i) = Var(Y_i) + 0 = Var(Y_i)$$

$$Var(Y_i) = E\left\{(Y_i - E(Y_i))^2\right\}$$

$$= (1-\pi_i)^2 \pi_i + (0-\pi_i)^2 (1-\pi_i)$$

$$= \pi_i (1-\pi_i)\left[(1-\pi_i) + \pi_i\right]$$

$$= \pi_i (1-\pi_i)$$

$$= (E(Y_i))(1 - E(Y_i))$$

$$= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i)$$

$$= Var(\varepsilon_i) \qquad (4)$$

Από την σχέση βλέπουμε πως η διασπορά των σφαλμάτων εξαρτάται από τα X_i, άρα η τιμή της διασποράς θα είναι διαφορετική για κάθε διαφορετικό X_i

ΓΙΑΤΙ ΌΧΙ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ

- **Περιορισμός στη συνάρτηση απόκρισης**

Η συνάρτηση απόκρισης επειδή παριστάνει πιθανότητες θα πρέπει να ισχύει ο περιορισμός

$$0 \le E(Y) = \pi \le 1 .$$

ΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

Το μοντέλο που χρησιμοποιούμε όταν η Y_i είναι δίτιμη είναι το λογιστικό, το οποίο ορίζεται ως εξής:

$$Y_i = E(Y_i) + \varepsilon_i$$

όπου Y_i ανεξάρτητη τ.μ. Bernoulli

$$E(Y_i) = \pi_i = \frac{e^{(\beta_0+\beta_1 X_i)}}{1+e^{(\beta_0+\beta_1 X_i)}} = \left[1+e^{(-\beta_0-\beta_1 X_i)}\right]^{-1}$$

ΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

Είδαμε πως η αναμενόμενη συνάρτηση πρέπει να παίρνει τιμές στο διάστημα [0,1]

$$E(Y_i) = \beta_0 + \beta_1 X_i = \pi_i .$$

Οι τιμές όμως της $E(Y_i)$ κυμαίνονται σε όλο το σύνολο των πραγματικών αριθμών.

Για να αντιμετωπίσουμε αυτό το πρόβλημα, μια σκέψη θα ήταν να αντικαταστήσουμε την πιθανότητα π_i της επιτυχίας του γεγονότος με τη σχετική πιθανότητα επιτυχίας, δηλαδή με το λόγο της πιθανότητας επιτυχίας του γεγονότος προς την πιθανότητα αποτυχίας του γεγονότος

$$\frac{\pi_i}{1-\pi_i} .$$

ΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

- Το μοντέλο

$$\frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 X_i$$

πάλι δεν είναι απόλυτα σωστό γιατί παίρνει τιμές από (0, +∞). Επομένως προτείνεται ο μετασχηματισμός

$$\pi_i' = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

$$\pi_i' = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_i$$

ΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

Έχουμε

$$\pi_i' = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_i$$

$$\Leftrightarrow \frac{\pi_i}{1-\pi_i} = e^{\beta_0+\beta_1 X_i} \Leftrightarrow \pi_i = e^{\beta_0+\beta_1 X_i}\left(1-\pi_i\right)$$

$$\Leftrightarrow \pi_i + \pi_i e^{\beta_0+\beta_1 X_i} = e^{\beta_0+\beta_1 X_i} \Leftrightarrow \pi_i(1+e^{\beta_0+\beta_1 X_i}) = e^{\beta_0+\beta_1 X_i}$$

$$\Leftrightarrow \pi_i = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}$$

$$\Leftrightarrow \boxed{E\left(Y_i\right) = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}}$$

ΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

Επίσης ισχύει

$$E(Y_i) = \left(1 + e^{-\beta_0 - \beta_1 X_i}\right)^{-1} :$$

Απόδειξη

$$E(Y_i) = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} = \left(\frac{1 + e^{\beta_0 + \beta_1 X_i}}{e^{\beta_0 + \beta_1 X_i}} \right)^{-1} = \left(\frac{1}{e^{\beta_0 + \beta_1 X_i}} + \frac{e^{\beta_0 + \beta_1 X_i}}{e^{\beta_0 + \beta_1 X_i}} \right)^{-1}$$

$$= \left(e^{-\beta_0 - \beta_1 X_i} + 1\right)^{-1}$$

ΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

Ορισμός

Ο λόγος $\frac{\pi_i}{1-\pi_i}$

ονομάζεται odds ενώ ο μετασχηματισμός $\pi_i' = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$

ονομάζεται logit μετασχηματισμός της πιθανότητας

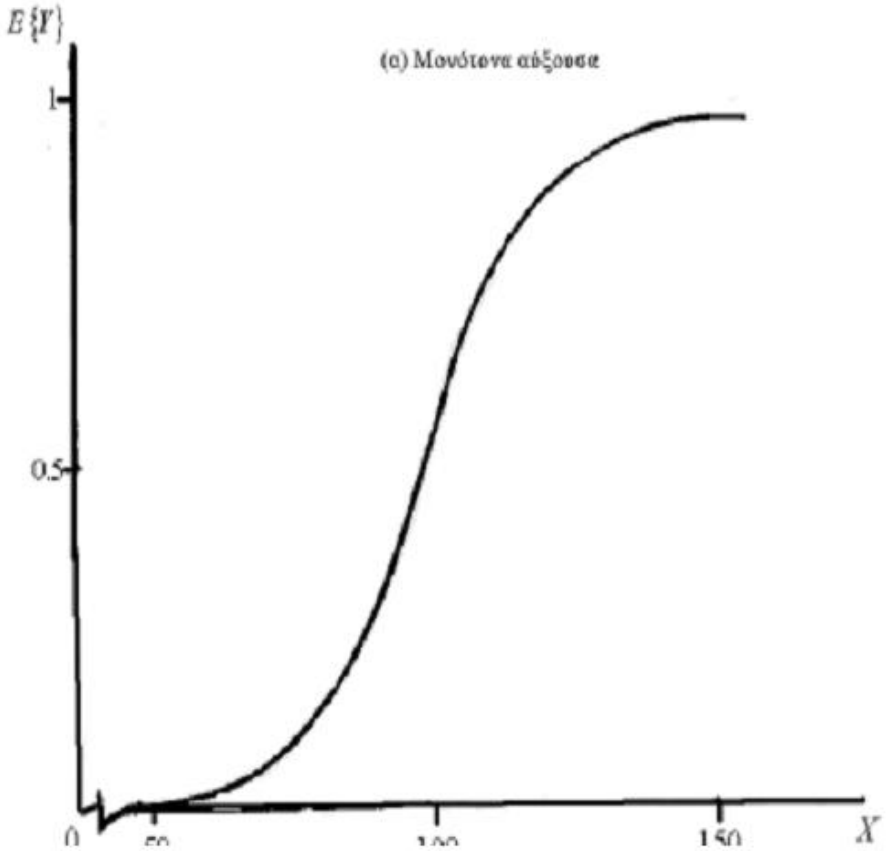
Η αναμενόμενη λογιστική συνάρτηση είναι:

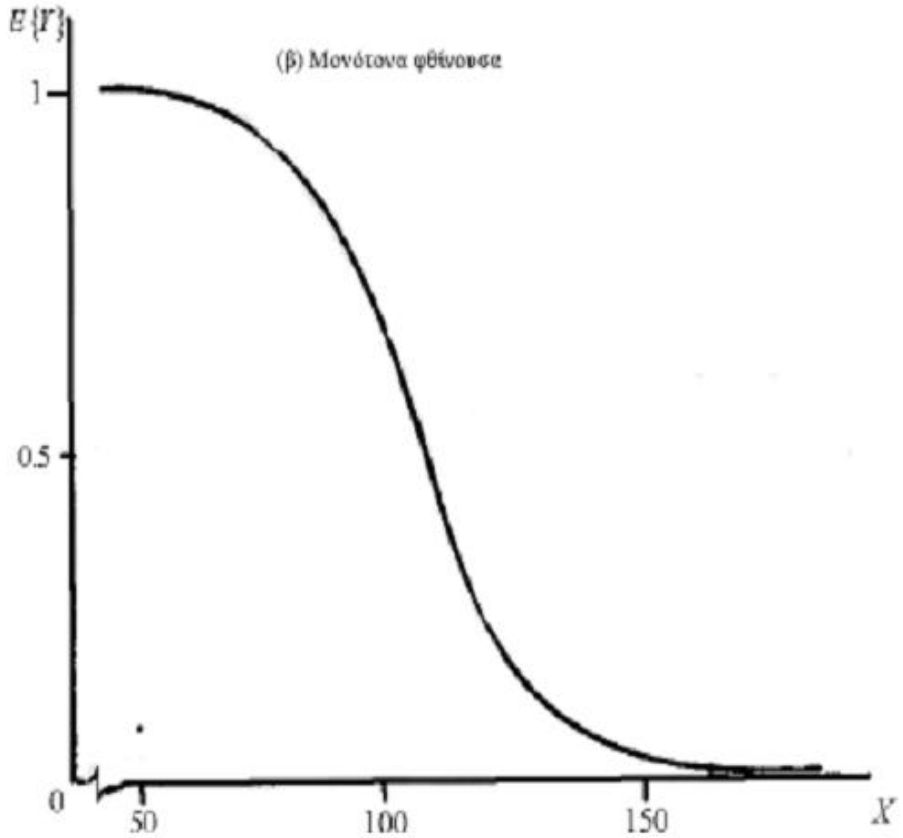
a) Είτε μονότονα αύξουσα συνάρτηση είτε μονότονα φθίνουσα,

b) Είναι σχεδόν γραμμική στην περιοχή [0.2, 0.8],

c) Πλησιάζει το 0 και 1 στις ακραίες τιμές της εμβέλειας του Χ

ΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ





ΣΥΝΑΡΤΗΣΗ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

Αφού τα Y_i είναι τυχαίες μεταβλητές Bernoulli όπου

$$\mathrm{P}\left(Y_i = 1\right) = \pi_i$$

$$\mathrm{P}\left(Y_i = 0\right) = 1 - \pi_i$$

η συνάρτηση πυκνότητας πιθανότητας είναι:

$$f_i\left(Y_i\right) = \pi_i^{Y_i}\left(1-\pi_i\right)^{1-Y_i} \qquad Y_i = 0,1 \quad \text{και} \quad i = 1,\ldots,n$$

ΣΥΝΑΡΤΗΣΗ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

Οι παρατηρήσεις Y_i είναι ανεξάρτητες οπότε η από κοινού συνάρτησης πιθανότητας θα είναι:

$$g(Y_1,\ldots\ldots,Y_n)=\prod_{i=1}^{n} f_i(Y_i)=\prod_{i=1}^{n}\pi_i^{Y_i}(1-\pi_i)^{1-Y_i}$$

$$\ln g(Y_1,\ldots\ldots,Y_n)=\ln\prod_{i=1}^{n}\pi_i^{Y_i}(1-\pi_i)^{1-Y_i}$$

$$=\sum_{i=1}^{n}\left[Y_i\ln\left(\frac{\pi_i}{1-\pi_i}\right)\right]+\sum_{i=1}^{n}\ln(1-\pi_i)$$

$$\pi_i'=\ln\left(\frac{\pi_i}{1-\pi_i}\right)=\beta_0+\beta_1X_i$$

$$E(Y_i)=\pi_i=\frac{e^{(\beta_0+\beta_1X_i)}}{1+e^{(\beta_0+\beta_1X_i)}}=\left[1+e^{(-\beta_0-\beta_1X_i)}\right]^{-1}$$

ΣΥΝΑΡΤΗΣΗ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

$$\ln L(\beta_0,\beta_1)=\sum_{i=1}^{n}Y_i(\beta_0+\beta_1X_i)+\sum_{i=1}^{n}\ln\left(1-\frac{e^{\beta_0+\beta_1X_i}}{1+e^{\beta_0+\beta_1X_i}}\right)$$

$$=\sum_{i=1}^{n}Y_i(\beta_0+\beta_1X_i)+\sum_{i=1}^{n}\ln\left(\frac{1+e^{\beta_0+\beta_1X_i}-e^{\beta_0+\beta_1X_i}}{1+e^{\beta_0+\beta_1X_i}}\right)$$

$$=\sum_{i=1}^{n}Y_i(\beta_0+\beta_1X_i)+\sum_{i=1}^{n}\ln\left(1+e^{\beta_0+\beta_1X_i}\right)^{-1}$$

$$=\sum_{i=1}^{n}Y_i(\beta_0+\beta_1X_i)-\sum_{i=1}^{n}\ln\left(1+e^{\beta_0+\beta_1X_i}\right)$$

Εκτιμήσεις β_0 και β_1 → $\hat{\pi}=\dfrac{e^{b_0+b_1X}}{1+e^{b_0+b_1X}}$ → $\hat{\pi}'=\ln\left(\dfrac{\hat{\pi}}{1-\hat{\pi}}\right)$

$$\hat{\pi}'=b_0+b_1X$$

ΕΡΜΗΝΕΙΑ

συντελεστής παλινδρόμησης b_1

Η ερμηνεία προέρχεται από την ιδιότητα που έχει ο εκτιμώμενος λόγος πιθανοτήτων (odds) $\frac{\pi_i}{1-\pi_i}$ ο οποίος πολλαπλασιάζεται με το e^{b_1} για κάθε μοναοα που αυξάνεται το Χ

$$OR = \frac{odds_2}{odds_1} = e^{b_1}$$

Αν το b_1 είναι θετικό, ο παράγοντας e^{b_1} είναι μεγαλύτερος από τη μονάδα, δηλαδή ο εκτιμώμενος λόγος πιθανοτήτων αυξάνεται. Αν το b_1 είναι αρνητικό, ο παράγοντας e^{b_1} είναι μικρότερος της μονάδας, και άρα ο εκτιμώμενος λόγος πιθανοτήτων μειώνεται.

Εισαγωγή

Σε πολλές περιπτώσεις η τ.μ y ενδέχεται να εξαρτάται από κάποιες επεξηγηματικές μεταβλητές. Η εξάρτηση της y από τις επεξηγηματικές μεταβλητές x (ανεξάρτητες μεταβλητές ή συμμεταβλητές) εισάγεται μέσω της εξάρτησης της πιθανότητας επιτυχίας p από τις x

Το μοντέλο λογιστικής παλινδρόμησης , το οποίο είναι ένα γενικευμένο γραμμικό μοντέλο εκφράζεται μέσω της σχέσης

$$n_x = g\big(E(y_x)\big) = g(\mu_x) = \boldsymbol{x'\beta}$$

Εισαγωγή

$$n_x = g\big(E(y_x)\big) = g(\mu_x) = \boldsymbol{x}'\boldsymbol{\beta}$$

με την ακόλουθη δομή:

1. $y_x \sim b(n_x, \mu_x)$ $(n_x > 1\,,\ διωνυμικά\ δεδομένα)$

ή $y_x \sim B(n_x, \mu_x)$ $(n_x = 1\,, δυαδικά\ δεδομένα)$

2. $n_x = g(\mu_x) = \ln\frac{\mu_x}{n_{x-\mu_x}} = \ln\frac{p_x}{1-p_x} = logit(p_x) = \boldsymbol{x}'\boldsymbol{\beta}$ $(συνάρτηση\ Logit)$
3. Ανεξαρτησία μεταξύ των παρατηρήσεων y_x **,**

n_x είναι ο αριθμός των επαναλήψεων της τιμής του διανύσματος $\boldsymbol{x}$ των επεξηγηματικών μεταβλητών.

Εισαγωγή

Αντιστρέφοντας τη συνάρτηση σύνδεσης προκύπτει:

$$p_x = e^{n_x}/(1 + e^{n_x})$$

για την οποία ισχύει ο περιορισμός $0 < p_x < 1$.

Για κάθε παρατήρηση i το μοντέλο γράφεται ως:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}\,, i = 1, \ldots, n$$

πιθανότητα «επιτυχίας»

$$p_i = p_{x_i} = \frac{exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})}{1 + exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})} = \frac{1}{1 + e^{-\boldsymbol{x_i'\beta}}}$$

linear predictor

$$x_i'\beta = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

$$E(y_i) = n_i p_i = n_i \frac{e^{\boldsymbol{x_i'\beta}}}{1 + e^{\boldsymbol{x_i'\beta}}}$$

Εκτίμηση παραμέτρων

Ας υποθέσουμε ότι τα δεδομένα μας είναι χωρισμένα σε κατηγορίες. Δηλαδή , έχουμε n_i στο πλήθος πειραματικές μονάδες στο i-οστό σημείο δεδομένων

Το μοντέλο

$$p_i = p_{x_i} = \frac{exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})}{1 + exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})} = \frac{1}{1 + e^{-x_i'\boldsymbol{\beta}}}$$

$$E(y_i) = n_i P(x_i) = n_i \frac{1}{1 + e^{-x_i'\boldsymbol{\beta}}} \, , i = 1,2, \ldots m$$

Εκτίμηση παραμέτρων

Με $y_1, y_2, \dots, y_m$ να είναι οι παρατηρούμενες τιμές των ανεξάρτητων διωνυμικών τυχαίων μεταβλητών. Σε αυτήν την περίπτωση ισχύει

$$var(y_i) = n_i P(x_i)[1 - P(x_i)]$$

και το άθροισμα

$$\sum_{i=1}^{m} n_i = n$$

είναι το συνολικό πλήθος του δείγματός μας.

Η συνάρτηση πιθανότητας μιας απλής διωνυμικής τυχαίας μεταβλητής y με παραμέτρους n , P δίνεται από τον τύπο:

$$\binom{n}{y} P^y (1 - P)^{n-y}$$

Εκτίμηση παραμέτρων

Ωστόσο , ο όρος $\binom{n}{y}$ δεν περιλαμβάνει το $\boldsymbol{\beta}$, οπότε δεν μπορεί να χρησιμοποιηθεί.

η log πιθανοφάνεια για το λογιστικό μοντέλο παλινδρόμησης

$$\ln[\mathcal{L}(\boldsymbol{P}; y)] = \sum_{i=1}^{m} \left\{ y_i \ln\left[\frac{P(x_i)}{1 - P(x_i)}\right] + n_i \ln[1 - P(x_i)] \right\}$$

Εκτίμηση παραμέτρων

Ο όρος $\ln\left[\frac{P(x_i)}{1-P(x_i)}\right]$ ονομάζεται $logit$ και γράφεται ως:

$$\ln\left[\frac{P(x_i)}{1-P(x_i)}\right] = x_i'\beta = \beta_0 + \sum_{j=1}^{k} x_{ij}\,\beta_j \quad , \qquad i = 1,2,\dots,m, \quad m \geq k+1$$

η log πιθανοφάνεια

$$\ln[\mathcal{L}(\beta; y)] = \sum_{i=1}^{m}\sum_{j=1}^{k} y_i x_{ij}\beta_j - \sum_{i=1}^{m} n_i \ln\left(1 + \exp\sum_{j=1}^{k} x_{ij}\beta_j\right)$$

Εκτίμηση παραμέτρων

μορφή πινάκων

$$\ln[\mathcal{L}(\beta; y)] = \beta' X y - \sum_{i=1}^{m} n_i \ln\left(1 + \exp(x_i'\beta)\right)$$

X είναι ο κλασσικός πίνακας του μοντέλου που συναντάμε και στην γραμμική παλινδρόμηση και

y το διάνυσμα της απόκρισης.

Εκτίμηση παραμέτρων

Παραγωγίζουμε ως προς β

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\beta_i}; \boldsymbol{y})}{\partial \boldsymbol{\beta}} = \boldsymbol{X'y} - \sum_{i=1}^{m} \left[\frac{n_i}{1 + e^{\boldsymbol{x_i'\beta}}}\right] e^{\boldsymbol{x_i'\beta}} \boldsymbol{x_i}$$

$$\frac{e^{\boldsymbol{x_i'\beta}}}{1 + e^{\boldsymbol{x_i'\beta}}} = \frac{1}{1 + e^{-\boldsymbol{x_i'\beta}}} = P(x_i)$$

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\beta_i}; \boldsymbol{y})}{\partial \boldsymbol{\beta}} = \boldsymbol{X'y} - \sum_{i=1}^{m} n_i P(x_i) x_i \quad \Rightarrow \quad ; \mathrm{X}'(\mathrm{y} - \mu)$$

$n_i P(\mathbf{x_i})$ αποτελεί τον μέσο της διωνυμικής τυχαίας μεταβλητής

εκτιμητής μέγιστης πιθανοφάνειας $\Rightarrow$ $\boldsymbol{X'(y - \mu) = 0}$

Εκτίμηση παραμέτρων

$$p_i = p_{x_i} = \frac{exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})}{1 + exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})} = \frac{1}{1 + e^{-x_i'\boldsymbol{\beta}}}$$

$$P(x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

λόγους πιθανοτήτων

$$Log\left[\frac{P}{(1-P)}\right]$$

logit

$$\ln\left[\frac{P(\boldsymbol{x_i})}{1 - P(\boldsymbol{x_i})}\right] = \boldsymbol{x_i'\beta}$$

$$\ln\left[\frac{P(\boldsymbol{x_i})}{1 - P(\boldsymbol{x_i})}\right] = \beta_0 + \beta_1 \boldsymbol{x_i}$$

Μεθοδο Wald

Η πρώτη εφαρμογή της μεθόδου Wald χρησιμοποιείται στον έλεγχο υποθέσεων για κάθε ξεχωριστό συντελεστή του μοντέλου της λογιστικής παλινδρόμησης. Πιο συγκεκριμένα , θέλουμε να ελέγξουμε:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

με το β_j να εμφανίζεται στον linear predictor $\boldsymbol{\chi_i}'\boldsymbol{\beta}$ του λογιστικού μοντέλου

Μέθοδος Wald

Για έναν εκτιμητή μέγιστης πιθανοφάνειας b_j ισχύει ότι:

$$z_j = \frac{b_j - \beta_j}{\sigma b_j}$$

Αυτός ακολουθεί την τυπική κανονική κατανομή $N(0,1)$ και έτσι ισχύει ότι το

$${z_j}^2 = \left(\frac{b_j}{\sigma b_j}\right)^2$$

Ακολουθεί ασυμπτωτικά την ${x_1}^2$ κατανομή , υπό την H_0 υπόθεση , όπου σb_jείναι το κατάλληλο διαγώνιο στοιχείο του ασυμπτωτικού πίνακα variance-covariance των b.

Μεθοδο καλης προσαρμογης

Με τη συμπερασματολογία πιθανοφάνειας , μπορούμε να ενισχύσουμε τον έλεγχο υποθέσεων , χρησιμοποιώντας την log $likelihood$. Η χρήση της μοιάζει αρκετά με τη χρήση της αρχής του επιπλέον αθροίσματος τετραγώνων (extra sum of squares principles) των γραμμικών μοντέλων. Για παράδειγμα , στα γραμμικά μοντέλα μπορούμε να χρησιμοποιήσουμε κάτω από τη μηδενική υπόθεση ένα μοντέλο ελαττωμένο (reduced model) , δηλαδή η μηδενική υπόθεση θέτει σε ένα υποσύνολο συντελεστών παλινδρόμησης την τιμή μηδέν. Ο έλεγχος χρησιμοποιεί τη διαφορά στο άθροισμα τετραγώνων του σφάλματος:

$$SS_E(reduced) - SS_E(full)$$

Ασυμπτωτικά

$$-2\ln\left[\frac{\mathcal{L}(reduced)}{\mathcal{L}(full)}\right] \sim {\chi_\Delta}^2$$

το $\mathcal{L}(.)$ είναι η πιθανοφάνεια και στην περίπτωσή μας , ζητούμε την πιθανοφάνεια για το πλήρες και για το ελαττωμένο μοντέλο.

η παράμετρος Δ είναι η διαφορά στον αριθμό των παραμέτρων ανάμεσα στο πλήρες και το ελαττωμένο μοντέλο.

Μεθοδο καλης προσαρμογης

Υποθέτουμε ότι ο linear predictor είναι $\beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \beta_3\chi_3$ και ενδιαφερόμαστε να εξετάσουμε την αρχική υπόθεση $H_0 : \beta_1 = \beta_2 = 0$

Το στατιστικό ελέγχου για το λόγο πιθανοφάνειας (likelihood ratio test statistic) δίνεται από τον τύπο:

$$2[\ln \mathcal{L}[b_0, b_1, b_2, b_3] - \ln \mathcal{L}[b_0^*, b_3^*]]$$

$\mathcal{L}(b_0^*, b_3^*)$ είναι η πιθανοφάνεια για το λογιστικό μοντέλο στο οποίο έχουμε επικαλεστεί την μηδενική υπόθεση (δηλαδή $\beta_1 = \beta_2 = 0$)

Παράδειγμα

Συγκέντρωση x (g/100cc)	Αριθμός εντόμων N	Αριθμός εντόμων που απεβίωσαν y	Ποσοστό
0.10	47	8	17.0
0.15	53	14	26.4
0.20	55	24	43.6
0.30	52	32	61.5
0.50	46	38	82.6
0.70	54	50	92.6
0.95	52	50	96.2

Θα εκτιμήσουμε την ED_{50}

αποτελεσματικές δόσεις

ED_p είναι η τιμή του x

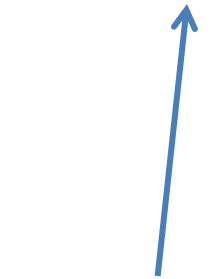
για την οποία η πιθανότητα του θανάτου μιας μύγας των φρούτων παίρνει την τιμή P.

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Wald chi-square	Pr>Chi-square	Standardized Estimate
INTERCPT	*1*	*-1.7361*	*0.2420*	*51.4482*	*0.0001*	
X	*1*	*6.2954*	*0.7422*	*71.9399*	*0.0001*	*1.024917*
INTERCPT	*1*	*3.1236*	*0.3349*	*86.9818*	*0.0001*	
LOGX	*1*	*2.1279*	*0.2214*	*92.3628*	*0.0001*	*0.898802*

Παράδειγμα

$$p_i = p_{x_i} = \frac{exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})}{1 + exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})} = \frac{1}{1 + e^{-x_i'\boldsymbol{\beta}}}$$

. ονтέλο 1



$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \ln x)}}$$

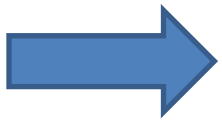
Μοντέλο 2

Παράδειγμα

$b_0 + b_1 x$	$b_0' + b_i' \ln x$
0.1844	0.2440
0.1607	0.1763
0.1428	0.1439
0.1336	0.1408
0.2139	0.2041
0.3432	0.2646
0.5194	0.3246

Για το μοντελο $b_0 + b_1 x$, ο $\widehat{ED}_{50}$ δίνεται από την εξίσωση:

$$\widehat{ED}_{50} = \frac{b_0}{b_1}$$



0.277g/100cc.

Για το μοντέλο $b_0' + b_1' \ln x$, το $\widehat{ED}_{50}$ δίνεται από την εξίσωση

$$\widehat{ED}_{50} = e^{-1.42} = 0.242g/100cc$$

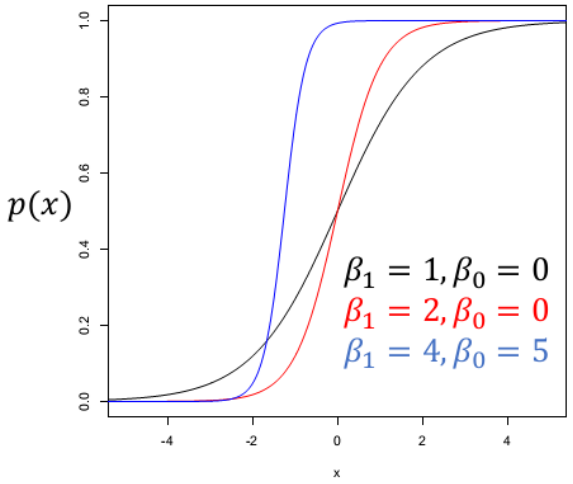
Παράδειγμα

- X: *ποσοτική επεξηγηματική* μεταβλητή
  - Εάν δίτιμη κατηγορική, μετατρέπεται σε ποσοτική με τιμές {0,1}
- Y: *δίτιμη* κατηγορική μεταβλητή απόκρισης, {0,1} από υπόθεση
  - Εάν λαμβάνει πάνω από 2 τιμές χρησιμοποιείται πολυωνυμική λογιστική παλινδρόμηση (multinomial logistic regression)
- *Παράμετρος πληθυσμού*: $p(x) =$ ποσοστό '1' στον υποπληθυσμό όπου $X = x$
- Λόγος πιθανοτήτων (*odds*) '1' προς '0': $\frac{p(x)}{1-p(x)}$
- Λογάριθμος odds: $\log\left(\frac{p(x)}{1-p(x)}\right)$
- Υπόθεση για πληθυσμό: $\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_1 x + \beta_0$ για κάθε x

  ή ισοδύναμα, $p(x)/(1-p(x)) = e^{\beta_1 x+\beta_0}$

  ή ισοδύναμα, $p(x) = \frac{1}{1+e^{-(\beta_1 x+\beta_0)}}$

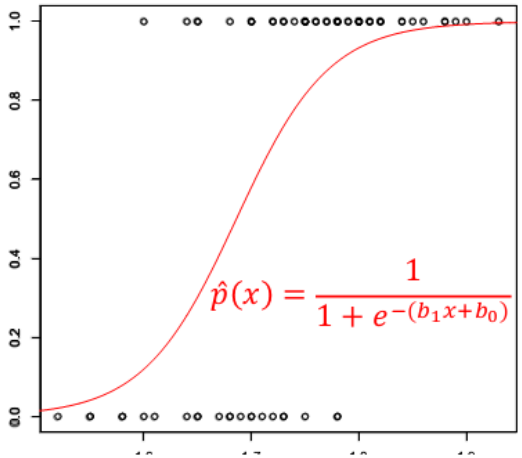
- *Odds ratio*: $e^{\beta_1} = \frac{\frac{p(x+1)}{1-p(x+1)}}{\frac{p(x)}{1-p(x)}}$



Παράδειγμα

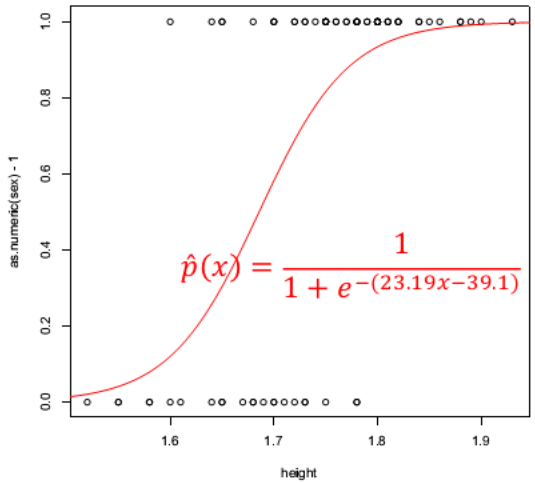
- Δεδομένα: για οποιεσδήποτε n γνωστές τιμές $x_1, \dots, x_n$ της X
  - Από κάθε υποπληθυσμό με $X = x_i$, λαμβάνουμε ένα τυχαίο δείγμα y_i μεγέθους 1

- Εκτιμητές
  - $\hat{p}(x)$: εκτιμητής ποσοστού υποπληθυσμού x
  - b_1: εκτιμητής της κλίσης β_1
  - b_0: εκτιμητής της σταθεράς β_0
  - Υπολογισμός από λογισμικό



Παράδειγμα

- Ποια σχέση έχει το ύψος (επεξηγηματική) με το φύλο (απόκριση) στους φοιτητές πληροφορικής;
- Απλό τυχαίο δείγμα (από ερωτηματολόγιο): $n = 80$
- $b_1 = 23.19, b_0 = -39.1$

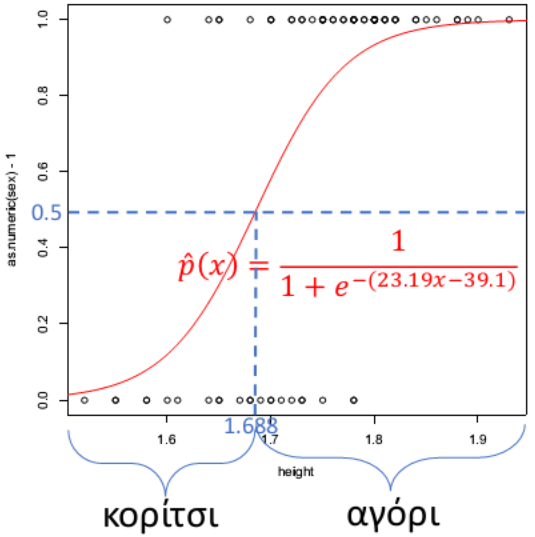


Παράδειγμα

- Ποια σχέση έχει το ύψος (επεξηγηματική) με το φύλο (απόκριση) στους φοιτητές πληροφορικής;
- Απλό τυχαίο δείγμα (από ερωτηματολόγιο): $n = 80$

- $b_1 = 23.19, b_0 = -39.1$

- Πρόβλεψη: $\hat{y} = \begin{cases} \text{αγορι}, & x > 1.688 \\ \text{κοριτσι}, & x < 1.688 \end{cases}$



Παράδειγμα

- Απλός υπολογισμός των εκτιμητών b_0, b_1 για δίτιμη *επεξηγηματική* μεταβλητή $X \in \{0,1\}$
  - Πχ, σχέση φύλου και επιτυχίας στις Πιθανότητες στους φοιτητές πληροφορικής που θα παίρνουν τ
    μάθημα της Στατιστικής στην Πληροφορική;

- Έστω $p(A), p(K)$ ποσοστό επιτυχίας σε αγόρια και κορίτσια αντίστοιχα

- Υπόθεση για πληθυσμό:
  - $\log\left(\frac{p(A)}{1-p(A)}\right) = \beta_1 \cdot 1 + \beta_0$
  - $\log\left(\frac{p(K)}{1-p(K)}\right) = \beta_1 \cdot 0 + \beta_0$

Παράδειγμα

- Απλός υπολογισμός των εκτιμητών b_0, b_1 για δίτιμη *επεξηγηματική* μεταβλητή $X \in \{0,1\}$
  - Πχ, σχέση φύλου και επιτυχίας στις Πιθανότητες στους φοιτητές πληροφορικής που θα παίρνουν το μάθημα της Στατιστικής στην Πληροφορική;
- Δεδομένα: ερωτηματολόγιο
  - Λόγος επιτυχίας/αποτυχίας στα αγόρια: $\frac{46}{11} = 4.18$
  - Λόγος επιτυχίας/αποτυχίας στα κορίτσια: $\frac{16}{8} = 2$

	Κορίτσι	Αγόρι
FALSE	8	11
TRUE	16	46

- Γραμμική παλινδρόμηση:
  - $\log(4.18) = b_1 \cdot 1 + b_0$
  - $\log(2) = b_1 \cdot 0 + b_0$
- Άρα, $b_1 = \log(2.09) = 0.737, b_0 = \log(2) = 0.693$
- Εκτιμητής odds ratio = δειγματικό odds ratio = $e^{b_1} = 2.09 = \frac{\frac{46}{11}}{\frac{16}{8}}$

Παράδειγμα

- $C\%$ διάστημα εμπιστοσύνης για *κλίση* β_1: $b_1 \pm z_* SE_{b_1}$
- $C\%$ διάστημα εμπιστοσύνης για *odds ratio* e^{β_1}: $e^{b_1 \pm z_* SE_{b_1}}$
  - z_* δίνεται από την τυπική κανονική κατανομή
  - SE_{b_1}: δειγματική τυπική απόκλιση εκτιμητή
    - Υπολογίζεται από λογισμικό

- Πχ, 95% διάστημα εμπιστοσύνης για odds ratio επιτυχίας/αποτυχίας σε αγόρια προς κορίτσια:
  1. Από δεδομένα: $b_1 = 0.737, SE_{b_1} = 0.548$
  2. $z_* = 1.96$
  3. $e^{z_* SE_{b_1}} = 2.927$
  4. Διάστημα = $\left[\frac{2.09}{2.927}, 2.09 \times 2.927\right] = [0.7, 6.13]$

Παράδειγμα

- Έλεγχος για την ύπαρξη σχέσης μεταξύ των μεταβλητών:

  $H_0: \beta_1 = 0$

  $H_a: \beta_1 \neq 0$

- Στατιστικό ελέγχου (*Wald statistic*) $z = \frac{b_1}{SE_{b_1}}$
- *p value* δίνεται από τυπική κανονική κατανομή
- Πχ, έχει σχέση το ύψος με το φύλο;
  1. Από δεδομένα: $b_1 = 23.18, SE_{b_1} = 5.41$
  2. $z = \frac{23.18}{5.41} = 4.289$
  3. p value $= 1.8 \times 10^{-5} \Rightarrow$ απορρίπτεται η $\beta_1 = 0$
- Πχ, έχει σχέση το φύλο με την επιτυχία στις Πιθανότητες;
  - p value=0.178. (Προσοχή! Δεν είναι ισοδύναμος έλεγχος με δίπλευρο z για σύγκριση ποσοστών μεταξύ δύο πληθυσμών ή χ^2, όπου δίνουν 0.173)

Πιθανολογική ταξινόμηση

Έστω ότι δίνονται N επισημασμένα (labeled) παραδείγματα εκπαίδευσης

$$\{\boldsymbol{x}_n, y_n\}_{n=1}^{N},\ \boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \{0,1\}$$

Στόχος είναι να εκπαιδευτεί ένας ταξινομητής ώστε να είναι σε θέση να προβλέπει το δυαδικό label y για ένα νέο input x

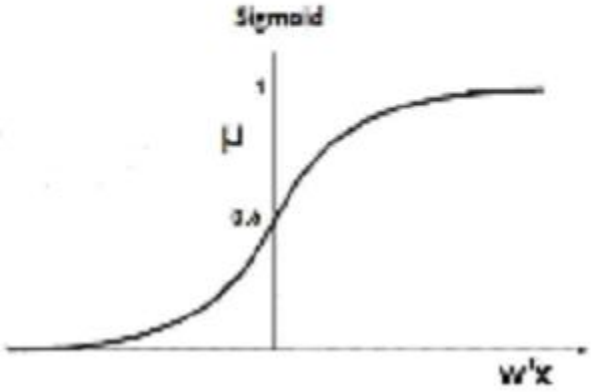
Θα θέλαμε επίσης ένα πιθανολογικό μοντέλο (probabilistic model) να μπορεί να προβλέπει τις πιθανότητες των labels

$$\begin{aligned} p(y_n = 1|\boldsymbol{x}_n, \boldsymbol{w}) &= \mu_n \\ p(y_n = 0|\boldsymbol{x}_n, \boldsymbol{w}) &= 1 - \mu_n \end{aligned}$$

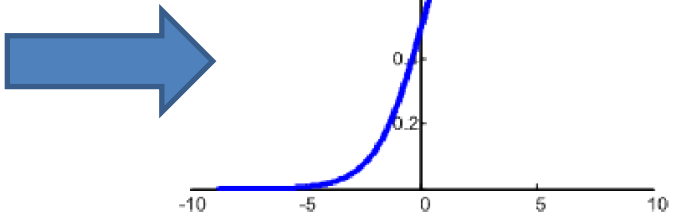
Πιθανολογική ταξινόμηση

Στη λογιστική παλινδρόμηση το μ ορίζεται μέσω της σιγμοειδούς συνάρτησης (sigmoid function)

$$\mu = \sigma(w^\top x) = \frac{1}{1+\exp(-w^\top x)} = \frac{\exp(w^\top x)}{1+\exp(w^\top x)}$$



Νευρωνικά δίκτυα $f(x) = \frac{1}{1+e^{-x}}$



Πιθανολογική ταξινόμηση

Έτσι, έχουμε:

$$p(y=1|x,w) = \mu = \sigma(w^\top x) = \frac{1}{1+\exp(-w^\top x)} = \frac{\exp(w^\top x)}{1+\exp(w^\top x)}$$

$$p(y=0|x,w) = 1-\mu = 1-\sigma(w^\top x) = \frac{1}{1+\exp(w^\top x)}$$