# Παλινδρόμηση (Regression)

Μηχανική Μάθηση

ΔΠΜΣ Επιστήμης Δεδομένων & Μηχανικής Μάθησης

Γιώργος Αλεξανδρίδης – gealexan@mail.ntua.gr

# Εισαγωγικές Έννοιες

# Ταξινόμηση και Παλινδρόμηση

- Δύο κύριες κατηγορίες της επιβλεπόμενης μάθησης (supervised learning)
  - Γνωστή και ως **μάθηση με παραδείγματα** (learning by examples)
  - Το σύστημα καλείται να μάθει την περιγραφή του μοντέλου από ένα *επιγεγραμμένο* (labelled) σύνολο δεδομένων, το οποίο αποτελείται από στιγμιότυπα για τα οποία **γνωρίζουμε** την επιθυμητή έξοδο
- **Ταξινόμηση** (Classification)
  - Η επιθυμητή έξοδος εντάσσεται σε *μια* ή *περισσότερες* διακριτές μεταξύ τους κατηγορίες
- **Παλινδρόμηση** (Regression)
  - Η επιθυμητή έξοδος έχει ένα *συνεχές* πεδίο τιμών
  - λχ δεδομένης της σημερινής ισοτιμίας δολαρίου και ευρώ, ποια θα είναι η αυριανή ισοτιμία;

# Παράδειγμα παλινδρόμησης

- Δεδομένα *μιας* διάστασης, τα οποία προέρχονται *ομοιόμορφα* από το πεδίο τιμών $x \in \mathbb{R}$

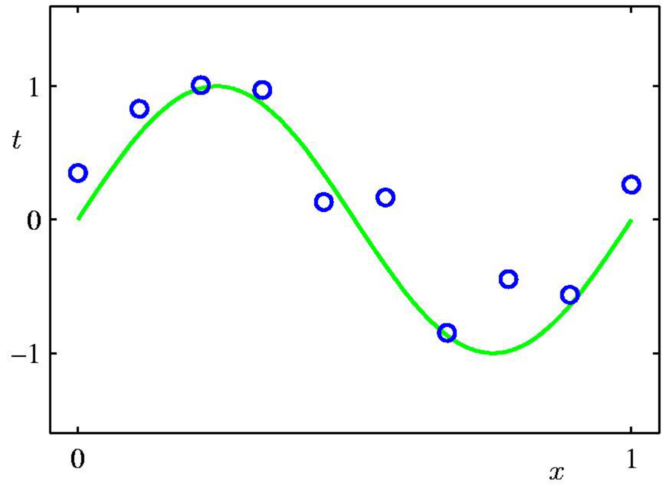
- Στην ετικέτα $y$ μπορεί να έχει εμφιλοχωρήσει *θόρυβος*, δηλαδή

$$t(x) = f(x) + \epsilon$$

- Στο διπλανό σχήμα, η πράσινη καμπύλη είναι αυτή που περιγράφει τη σχέση (συνάρτηση) εισόδου-εξόδου

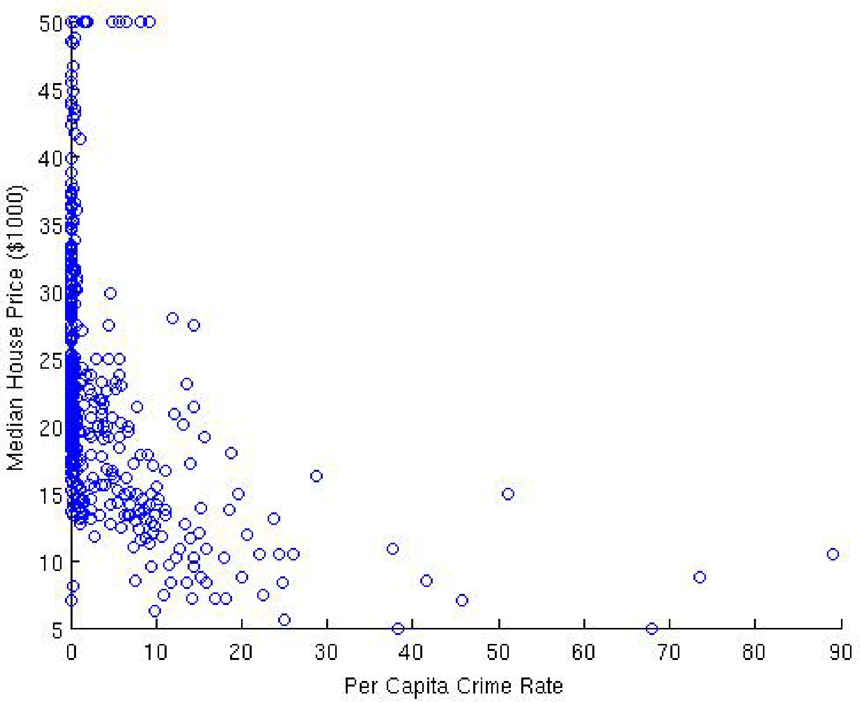
- Στόχος της παλινδρόμησης είναι να προσδιοριστεί αυτή η σχέση (συνάρτηση) από τις διαθέσιμες παρατηρήσεις (μπλε κουκίδες)

- **Ερωτήματα**
  1. *Πως* μπορούμε να παραμετροποιήσουμε το εν λόγω μοντέλο;
  2. *Ποια* συνάρτηση σφάλματος μπορούμε να χρησιμοποιήσουμε για να αξιολογήσουμε την προσαρμογή του μοντέλου στα δεδομένα
  3. *Πως* μπορούμε να γενικεύσουμε (generalization) σε νέα δεδομένα;



# Παράδειγμα: Τιμές ακινήτων στη Βοστώνη

- Θέλουμε να εκτιμήσουμε το μέσο κόστος απόκτησης κατοικίας στη Βοστώνη συναρτήσει διαφόρων χαρακτηριστικών
- Επιλέγουμε ως πρώτο χαρακτηριστικό την εγκληματικότητα
- Αποτελεί καλό χαρακτηριστικό για το σκοπό μας;



# Γραμμική Παλινδρόμηση

Linear Regression

# Τυπικός Ορισμός για μονοδιάστατα δεδομένα

- Έστω ότι έχουμε συλλογή δεδομένων $\mathcal{D}$ που αποτελείται από $N$ ζεύγη $\{(x^{(1)}, t^{(1)}), \cdots (x^{(N)}, t^{(N)})\}$, όπου
  1. $x \in \mathbb{R}$ το πεδίο ορισμού της εισόδου
     - Στο προηγούμενο παράδειγμα, ο δείκτης εγκληματικότητας
  2. $t \in \mathbb{R}$ το πεδίο τιμών των ετικετών (συνεχής τιμές)
     - Στο προηγούμενο παράδειγμα, η τιμή του ακινήτου

- Μοντελοποιούμε τη σχέση εισόδου-εξόδου υπό τη μορφή *πρωτοβάθμιας (γραμμικής) εξίσωσης*

$$y = w_0 + w_1 x$$

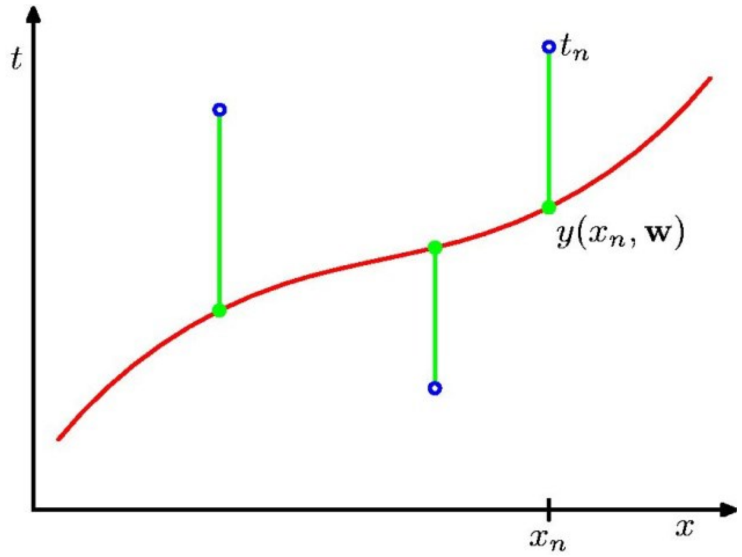
- Χωρίζουμε τη συλλογή δεδομένων μας σε δύο μη επικαλυπτόμενα σύνολα
  1. <u>Δεδομένα εκπαίδευσης</u>: Χρησιμοποιούνται για την κατασκευή της υπόθεσης
     - Δηλαδή της συνάρτησης που απεικονίζει τα δεδομένα εισόδου $x$ στην έξοδο $y$
  2. <u>Δεδομένα ελέγχου</u>: Επαληθεύουν την υπόθεση

# Θόρυβος

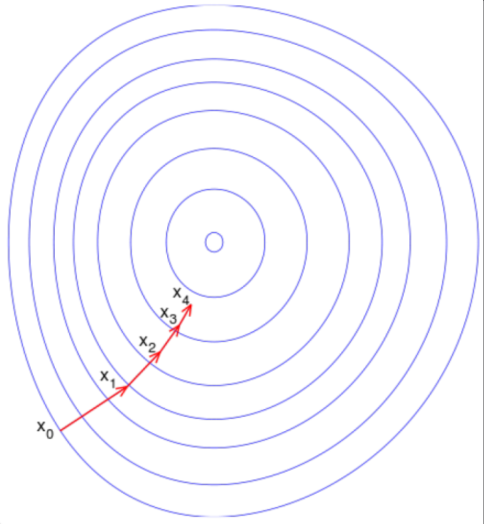
- Το γραμμικό μοντέλο είναι αρκετά *απλό* και μπορεί *να μην προσαρμόζεται* στα δεδομένα
  - Αυτή η έλλειψη προσαρμογής μπορεί να μοντελοποιηθει ως **θόρυβος** (*noise*)
- **Πηγές** θορύβου
  1. Θόρυβος **εισόδου**: Ανακρίβειες στις ιδιότητες/χαρακτηριστικά των δεδομένων
     - πχ θόρυβος στην καταγραφή των δεικτών εγκληματικότητας
  2. Θόρυβος **εξόδου**: Λανθασμένη επισημείωση των δεδομένων
     - πχ θόρυβος στην καταγραφή της ακριβούς τιμής των ακινήτων
  3. **Λανθάνουσες μεταβλητές** (latent variables): Επιπλέον χαρακτηριστικά που δεν έχουν ληφθεί υπόψη επηρεάζουν τη σχέση εισόδου-εξόδου
     - Ποιο άλλο μέγεθος θα μπορούσε να επηρεάσει τις τιμές των ακινήτων;
  4. **Χωρητικότητα** (capacity) μοντέλου: το μοντέλο μπορεί να είναι πολύ απλό για να περιγράψει τη σχέση εισόδου-εξόδου

# Εύρεση Παραμέτρων: 1ος Τρόπος

- Μέθοδος **συνήθων ελαχίστων τετραγώνων** (*ordinary least squares*)
- *Ελαχιστοποιεί* το άθροισμα των τετραγώνων της διαφοράς (υπολοίπου) μεταξύ σημείου και ευθείας
- Οδηγεί σε λύση κλειστής μορφής
$$\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{t}$$
- Προσέγγιση *αμερόληπτη* και *συνεπής* <u>αν</u>
    1. Τα σφάλματα έχουν *πεπερασμένη* διακύμανση
        - Αυτό συνήθως <u>ισχύει</u> στην πράξη
    2. *Ασυσχέτιστα* με το μοντέλο
        - Αυτό συνήθως <u>δεν ισχύει</u> στην πράξη, καθώς μπορεί να υπάρχουν λανθάνουσες συμμεταβλητές (latent covariates) που σχετίζονται με την παρατηρούμενη μεταβλητή και την έξοδο
            - πχ άλλοι παράγοντες που επηρεάζουν τις τιμές των ακινήτων πλην της εγκληματικότητας



# Εύρεση παραμέτρων: 2$^{ος}$ Τρόπος




- **Κατάβαση κλήσης** (*gradient descend*)
  - Αρχικοποίηση $\boldsymbol{w}$ σε τυχαίες τιμές
  - Ενημέρωση $\boldsymbol{w}$ σε κατεύθυνση αντίθετη της κλήσης $\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \frac{\partial L(y,t)}{\partial \boldsymbol{w}}$
  - $\eta$ ρυθμός μάθησης
- $L\left(y^{(n)}, t^{(n)}\right) = \left(t^{(n)} - y^{(n)}\right)^2$ συνάρτηση *σφάλματος* (τετραγωνικό σφάλμα)
- **Στοχαστική κατάβαση κλίσης** (*stochastic gradient descend* – SGD)
  - Κατά την $t$-οστή επανάληψη του αλγορίθμου εκπαίδευσης εξετάζουμε το $n$-οστό παράδειγμα εκπαίδευσης
  - $\boldsymbol{w}^{(t)} \leftarrow \boldsymbol{w}^{(t-1)} + 2\eta\left(t^{(n)} - y^{(n)}\right)x^{(n)}$
  - <u>Σφάλμα</u> $\epsilon = t^{(n)} - y^{(n)}$
  - Όσο το σφάλμα προσεγγίζει το 0, η ενημέρωση περιορίζεται (το $\boldsymbol{w}$ σταματά να μεταβάλλεται)
  - Αναλογία με μέθοδο ελαχίστων τετραγώνων
    - Ποια άλλη αναλογία βλέπετε;
- **Ενημέρωση κατά δέσμη** (*batch update*)
  - $N$ μέγεθος δέσμης (ομάδας παραδειγμάτων)
  - $\boldsymbol{w}^{(t)} \leftarrow \boldsymbol{w}^{(t-1)} + 2\eta \frac{1}{N} \sum_{n=1}^{N}\left(t^{(n)} - y^{(n)}\right)x^{(n)}$

# Πολυδιάστατα δεδομένα

- Ένας τρόπος επέκτασης του μοντέλου είναι να λάβουμε υπόψη μας τον γραμμικό συνδυασμό <u>και άλλων</u> διαστάσεων της εισόδου (χαρακτηριστικών)

$$y(\boldsymbol{x}) = w_0 + w_1 x_1 + w_2 x_2$$

  - Στο παράδειγμα των τιμών ακινήτων στη Βοστώνη, μπορούμε να λάβουμε υπόψη μας την έκταση του ακινήτου, τον αριθμό των δωματίων, …

- Στη γενική περίπτωση, τα δεδομένα εισόδου έχουν $d$ διαστάσεις.

$$\boldsymbol{x} = (x_1, \cdots, x_d)$$

- <u>Μοντέλο</u>: $y(\boldsymbol{x}) = w_0 + \sum_{j=1}^{d} w_j x_j = \boldsymbol{w}^T \boldsymbol{x}$

# Πολυωνυμική Παλινδρόμηση

Polynomial Regression

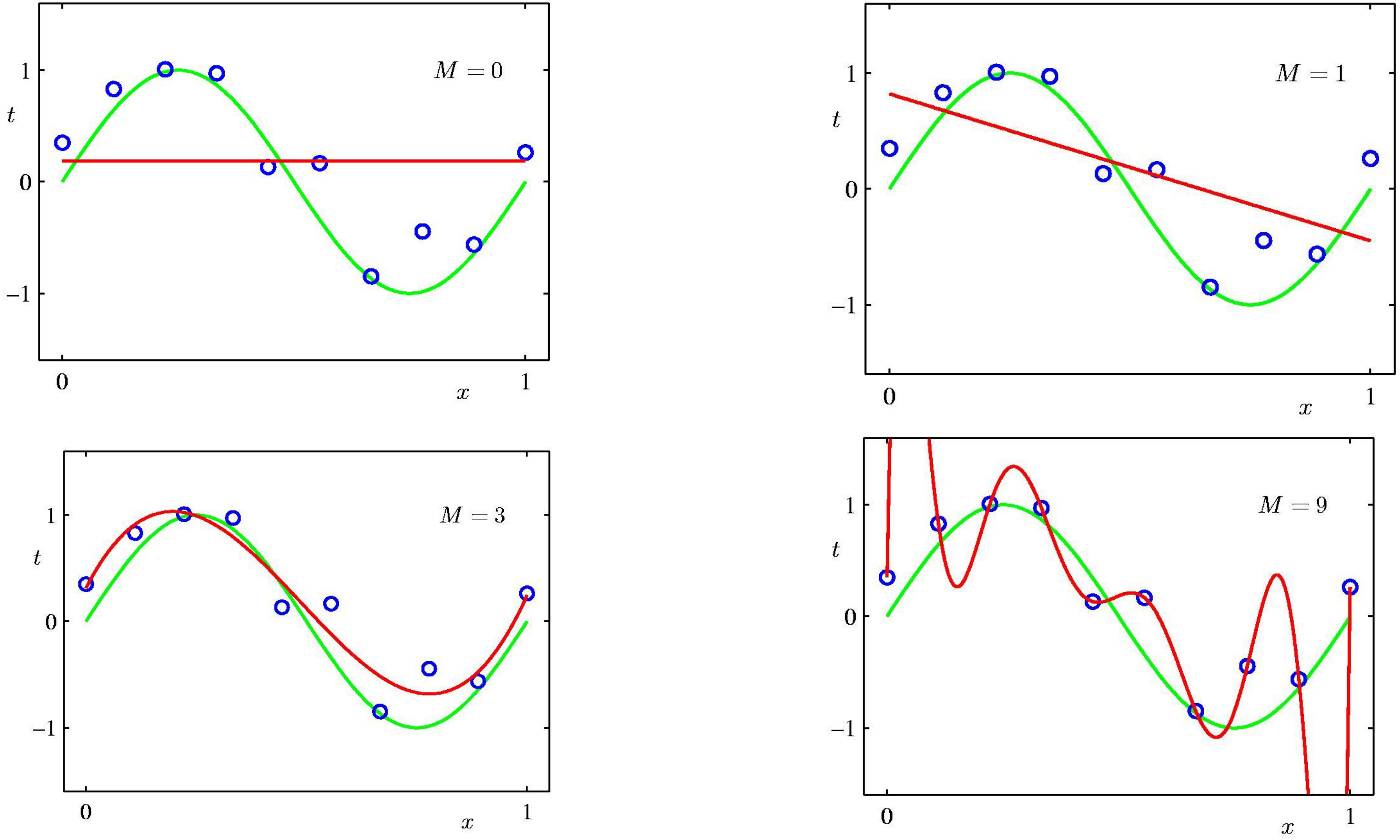
# Πολυωνυμικά μοντέλα

- *Πιο σύνθετα* μοντέλα, μοντελοποιούν και μη-γραμμικούς *συνδυασμούς* μεταξύ των χαρακτηριστικών της εισόδου
- Πολυωνυμικό μοντέλο $M$-οστής τάξης

$$y(\boldsymbol{x}, \boldsymbol{w}) = w_0 + \sum_{j=1}^{M} w_j \boldsymbol{x}^j$$

  - Γραμμική παλινδρόμηση ειδική περίπτωση πολυωνυμικής παλινδρόμησης για $M = 1$
- Γιατί μας ενδιαφέρουν;
  - <u>Θεώρημα προσέγγισης Weirstrass</u>: *Κάθε συνεχής συνάρτηση* $f$ *που λαμβάνει πραγματικές τιμές εντός διαστήματος* $[a, b]$ *μπορεί να προσεγγιστεί υπό τη μορφή πολυωνύμου για οποιοδήποτε επιθυμητό μέγεθος ακρίβειας* $\varepsilon > 0$
    - K. Weierstrass (1885). Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, 1885 (II).
- Ποιες άλλες θεωρήματα/τεχνικές προσέγγισης γνωρίζετε;

# Ποιο μοντέλο προσαρμόζεται καλύτερα στα δεδομένα;



# Τρόπος Εργασίας

1. **Αφελής εξαντλητική αναζήτηση** (Naïve Exhaustive Search)
   - Ξεκινώντας από τη γραμμική παλινδρόμηση, *αυξάνουμε συνεχώς* την τάξη του μοντέλου μέχρι το σφάλμα στα *δεδομένα επαλήθευσης* (validation data) να σταματήσει να μειώνεται.

2. **Περιορισμός** της **χωρητικότητας** του μοντέλου
   - Ξεκινάμε από ένα μοντέλο **πολύ μεγάλης** χωρητικότητας και μέσω των δεδομένων εκπαίδευσης και επαλήθευσης προσπαθούμε να **περιορίσουμε** τον χώρο υποθέσεών του
   - Στην προηγούμενη διαφάνεια, το πολυωνυμικό μοντέλο 9ης τάξης έκανε πιο περίπλοκες υποθέσεις για τη σχέση εισόδου-εξόδου από την πραγματικότητα, με αποτέλεσμα να **υπερπροσαρμόζεται** (*overfit*) στα δεδομένα εκπαίδευσης
   - Οι συντελεστές του πολυωνύμου *λάμβαναν πολύ μεγάλες*, κατ’ απόλυτη τιμή, τιμές
   - Ένας τρόπος αποφυγής της υπερπροσαρμογής είναι να κάνεις το μοντέλο να ψάχνει για λύσεις που έχουν μικρά, κατ’ απόλυτη τιμή, βάρη ⇒ **Ομαλοποίηση** (*Regularization*)

# Αμφικλινής παλινδρόμηση (Ridge Regression)

- <u>Εναλλακτικές ονομασίες</u>: **Φθορά βαρών** (*weight decay*) και **ομαλοποίηση Tikhonov** (*Tikhonov regularization*)
- Προσθήκη όρου ομαλοποίησης στη συνάρτηση σφάλματος $L$, ο οποίος είναι ανάλογος του τετραγώνου του διανύσματος των βαρών $\Omega(\boldsymbol{w}) = \frac{1}{2}\|w\|_2^2 = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$
  - Νόρμα Frobenius ή Ευκλείδεια νόρμα
- $\tilde{L}(\boldsymbol{w}; \mathbf{X}, \mathbf{y}) = L(\boldsymbol{w}; \mathbf{X}, \mathbf{y}) + \alpha\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$, όπου $\tilde{L}$ ομαλοποιημένη συνάρτηση σφάλματος
- Ενημέρωση βαρών μέσω στοχαστικής κατάβασης κλίσης (SGD)
  - Υπολογισμός κλίσης: $\nabla_w \tilde{L}(w; \mathbf{X}, \mathbf{y}) = \nabla_w L(w; \mathbf{X}, \mathbf{y}) + \alpha\boldsymbol{w}$
  - Ενημέρωση βαρών
    - $\boldsymbol{w}^{(t+1)} \leftarrow \boldsymbol{w}^{(t)} - \eta\left(\nabla_{\boldsymbol{w}^{(t)}} J\left(\boldsymbol{w}^{(t)}; \mathbf{X}, \mathbf{y}\right) + \alpha\boldsymbol{w}^{(t)}\right) \Rightarrow \boldsymbol{w}^{(t+1)} \leftarrow (1 - \eta\alpha)\boldsymbol{w}^{(t)} - \eta\nabla_{\boldsymbol{w}^{(t)}} J\left(\boldsymbol{w}^{(t)}; \mathbf{X}, \mathbf{y}\right)$
  - Το αποτέλεσμα είναι να μειώνεται («φθείρεται») το εύρος του διανύσματος των βαρών κατά παράγοντα $(1 - \eta\alpha)$ σε κάθε βήμα.
    - α βαθμός ομαλοποίησης (υπερ-παράμετρος του μοντέλου)

# Αμφικλινής παλινδρόμηση : Επίδραση στη διαδικασία μάθησης (1/3)

- Έστω $\boldsymbol{w}^*$ το **διάνυσμα βαρών** που **ελαχιστοποιεί** την $L$
- **Αντικατάσταση** της $L$ από την **τετραγωνική** της **προσέγγιση** $\hat{L}$ γύρω από το $\boldsymbol{w}^*$
  - Ειδικά για προβλήματα **γραμμικής παλινδρόμησης** όπου η αντικειμενική συνάρτηση υπολογίζει διαφορές τετραγώνων (π.χ. MSE), η προσέγγιση είναι τέλεια
  - $\hat{L}(\boldsymbol{w}) = L(\boldsymbol{w}^*) + \frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}^*)^T \boldsymbol{H}(\boldsymbol{w} - \boldsymbol{w}^*)$
  - Δεν υπάρχει πρωτοβάθμιος όρος μιας εξ' ορισμού είναι **ελάχιστο** και άρα η **κλίση** γύρω από το $\boldsymbol{w} - \boldsymbol{w}^*$ είναι (σχεδόν) μηδενική
  - $\boldsymbol{H}$: **Εσσιανός Πίνακας** (*Hessian Matrix*) όλων των $\frac{\vartheta^2 L}{\vartheta w_i \vartheta w_j}$
  - Επειδή βρισκόμαστε γύρω από ελάχιστο, ο $\boldsymbol{H}$ είναι θετικά ημι-καθορισμένος

# Αμφικλινής παλινδρόμηση : Επίδραση στη διαδικασία μάθησης (2/3)

- Η $\hat{L}$ ελαχιστοποιείται στα σημεία όπου η **κλίση** της $\nabla_w \hat{L}(w; \mathbf{X}, \mathbf{y}) = \boldsymbol{H}(\boldsymbol{w} - \boldsymbol{w}^*)$ γίνεται **ίση** με το 0
- Προσθέτοντας τον όρο ποινής $\alpha \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w}$ **το τοπικό ελάχιστο αλλάζει** και γίνεται πλέον $\widetilde{\boldsymbol{w}}$
  - $\nabla_w \left( \hat{L}(\widetilde{w}) + \frac{\alpha}{2} \widetilde{w}^T \widetilde{w} \right) = 0 \Longrightarrow H(\widetilde{w} - w^*) + \alpha \widetilde{w} = 0 \Longrightarrow \widetilde{w} = (H + \alpha I)^{-1} H w^*$
- Όταν $\alpha \longrightarrow 0$, το $\widetilde{\boldsymbol{w}}$ προσεγγίζει το $\boldsymbol{w}^*$
- Όταν $\alpha \neq 0$, χρησιμοποιούμε *αποσύνθεση ιδιοτιμών* (eigen decomposition) $\boldsymbol{H} = \boldsymbol{Q \Lambda Q}^T$
  - Εφικτό μιας και $\boldsymbol{H}$ συμμετρικός με πραγματικές τιμές
  - $\boldsymbol{\Lambda}$ **διαγώνιος** πίνακας **ιδιοτιμών**, $\boldsymbol{Q}$ **ορθο-κανονικός** πίνακας **ιδιοδιανυσμάτων**
    - $\widetilde{w} = Q(\Lambda + \alpha I)^{-1} \Lambda Q^T w^*$

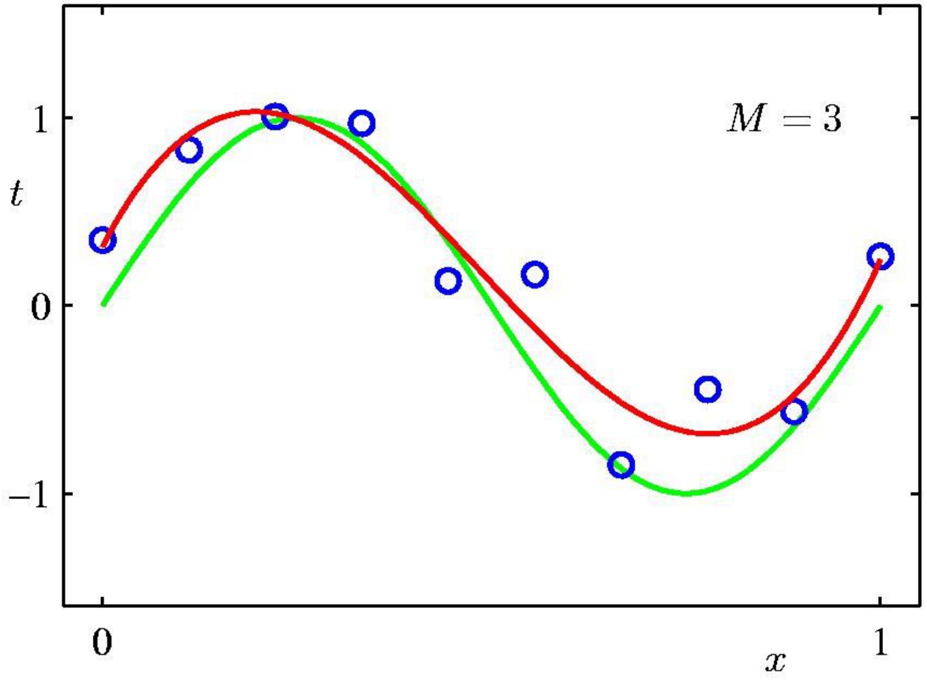
# Αμφικλινής παλινδρόμηση : Επίδραση στη διαδικασία μάθησης (3/3)

- Ουσιαστικά το $\boldsymbol{w}^*$ αναπροσαρμόζεται στην κατεύθυνση των αξόνων που ορίζονται από τα ιδιοδιανύσματα του $\boldsymbol{H}$ κατά ένα **παράγοντα** $\frac{\lambda_i}{\lambda_i+\alpha}$
  - Για μεγάλες ιδιοτιμές ($\lambda_i \gg \alpha$), η επίδραση της ομαλοποίησης είναι πολύ μικρή
  - Για μικρές ιδιοτιμές ($\lambda_i \ll \alpha$), η επίδραση της ομαλοποίησης είναι πολύ μεγάλη
    - Συρρικνώνει την επίδραση των αντίστοιχων ιδιοδιανυσμάτων στο 0

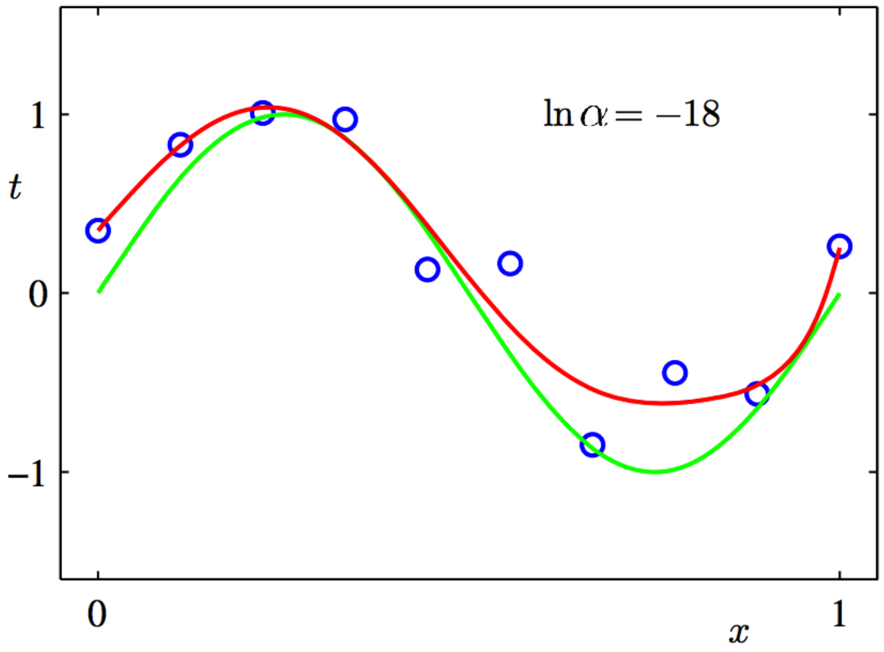
# Αμφικλινής Γραμμική Παλινδρόμηση

- **Συνάρτηση σφάλματος**: τετραγωνικό σφάλμα
  - $\mathrm{L}(\boldsymbol{w}) = (\boldsymbol{X}\boldsymbol{w} - y)^T(\boldsymbol{X}\boldsymbol{w} - y)$
  - Βέλτιστο διάνυσμα βαρών: $w^* = (X^T X)^{-1} X^T y$
- Προσθήκη **όρου ομαλοποίησης** αμφικλινούς πάλινδρόμησης
  - $L(\boldsymbol{w}) = (\boldsymbol{X}\boldsymbol{w} - y)^T(\boldsymbol{X}\boldsymbol{w} - y) + \frac{\alpha}{2} w^T w$
  - Βέλτιστο διάνυσμα βαρών: $\widetilde{w} = (X^T X + \alpha \mathrm{I})^{-1} X^T y$
- Ο όρος $X^T X$ είναι **ανάλογος** του **πίνακα συνδιασποράς** των χαρακτηριστικών της εισόδου
  - **Διαγώνιες τιμές**: Αντιστοιχούν στη διακύμανση των χαρακτηριστικών της εισόδου
  - Η αμφικλινής παλινδρόμηση αναγκάζει τον αλγόριθμο μάθησης να θεωρήσει ότι η είσοδος παρουσιάζει μεγαλύτερη διακύμανση και συνεπώς να **μικρύνει** τα **βάρη** εκείνα που εμφανίζουν μικρότερη συνδιασπορά.

# Παράδειγμα αμφικλινούς παλινδρόμησης




Πολυωνυμικό μοντέλο 3ης τάξης




Πολυωνυμικό μοντέλο 9ης τάξης με χρήση αμφικλινούς παλινδρόμησης

# Παλινδρόμηση Lasso (Lasso Regression)

- Προσθήκη όρου ομαλοποίησης στη συνάρτηση σφάλματος $L$, ο οποίος είναι ανάλογος της **απόλυτης τιμής** του διανύσματος των βαρών $\Omega(\boldsymbol{w}) = \|w\|_1$
- Ομαλοποιημένη αντικειμενική συνάρτηση
  - $\tilde{L}(w; \mathbf{X}, \mathbf{y}) = L(w; \mathbf{X}, \mathbf{y}) + \alpha\|w\|_1$
- Υπολογισμός κλίσης: $\nabla_w \tilde{L}(w; \mathbf{X}, \mathbf{y}) = \nabla_w L(w; \mathbf{X}, \mathbf{y}) + \alpha \mathrm{sgn}(\boldsymbol{w})$
- **Διαφορά** ως προς αμφικλινή παλινδρόμηση
  - Η συνεισφορά της ομαλοποίησης στην κλίση δεν είναι πλέον ανάλογη του εύρους $w_i$ της κάθε παραμέτρου, αλλά του προσήμου της
  - Για να προχωρήσουμε, κάνουμε την **επιπλέον παραδοχή** ότι ο εσσιανός πίνακας είναι διαγώνιος με $H_{i,i} > 0$
    - Έχουν αφαιρεθεί οι συσχετίσεις μεταξύ των χαρακτηριστικών της εισόδου λ.χ. με την προεπεξεργασία μέσω PCA

# Χαρακτηριστικά Παλινδρόμησης Lasso

- Τετραγωνική προσέγγιση $\hat{L}$ της $L$
  - $\hat{L}(\boldsymbol{w}) = L(\boldsymbol{w}^*) + \sum_i \left[\frac{1}{2} H_{i,i}(\boldsymbol{w} - \boldsymbol{w}^*)^2 + \alpha |w_i|\right]$
- Τοπικό ελάχιστο: $\widetilde{w}_i = \operatorname{sgn}(w_i^*) \max\left(|w_i^*| - \frac{\alpha}{H_{i,i}}, 0\right)$
  - Όταν $|w_i^*| \leq \frac{\alpha}{H_{i,i}}$, τότε το $\widetilde{w}_i$ γίνεται 0
  - Όταν $|w_i^*| > \frac{\alpha}{H_{i,i}}$, τότε το $\widetilde{w}_i$ «σύρεται» προς το 0 κατά έναν όρο $\frac{\alpha}{H_{i,i}}$
- Η παλινδρόμηση Lasso οδηγεί σε πιο **αραιές** (*sparse*) αναπαραστάσεις σε σύγκριση με την αμφικλινή παλινδρόμηση
  - Υπό την έννοια ότι **περισσότερες παράμετροι** έχουν μηδενικές τιμές
  - Χρησιμοποιείται ως μηχανισμός **επιλογής χαρακτηριστικών** (*feature selection*)

# Λογιστική Παλινδρόμηση

Logistic Regression

# Λογιστική Παλινδρόμηση

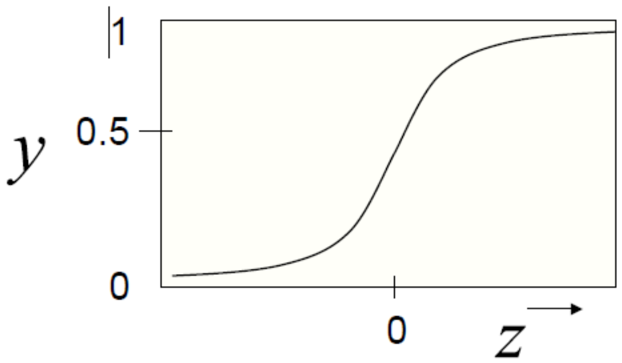
- Χρήση σε προβλήματα τα οποία έχουν **διακριτή** έξοδο, ωστόσο εμάς μας *ενδιαφέρει* να την *εκφράσουμε* υπό το πρίσμα **συνεχούς τιμής**
  - Πχ για συγκεκριμένες τιμές δεικτών στις εξετάσεις αίματος ασθενούς, ποια η πιθανότητα να εμφανίσει στεφανιαία νόσο;

- *Επέκταση* γραμμικής παλινδρόμησης μέσω της προσθήκης *σιγμοειδούς* συνάρτησης

$$y = \sigma(\boldsymbol{w}^T \boldsymbol{x} + w_0)$$

- Σιγμοειδής συνάρτηση

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$




- Έξοδος αποτελεί *ομαλή* συνάρτηση εισόδου και βαρών
  - Λαμβάνει τιμές στο [0,1]

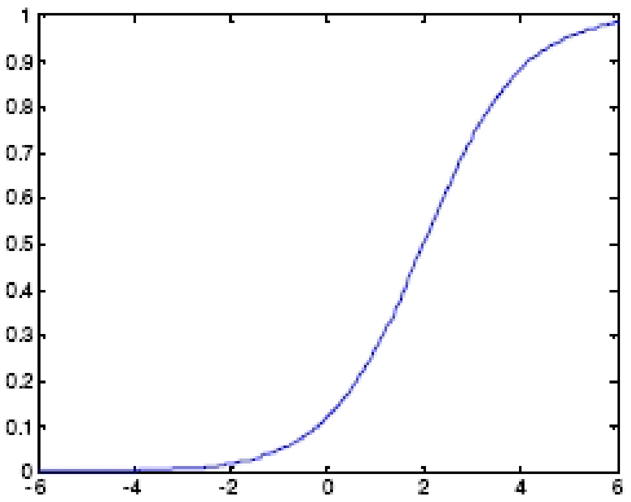
- Τι μας θυμίζει;

# Μορφή λογιστικής παλινδρόμησης

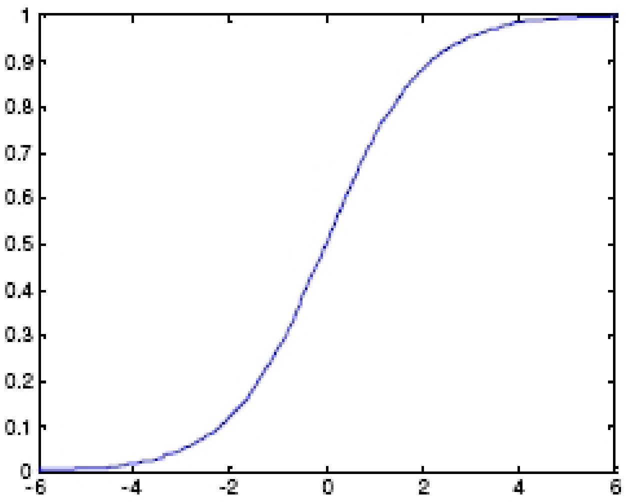
- Η μεταβολή των βαρών $\boldsymbol{w}$ αλλάζει τη μορφή της συνάρτησης

$$y = \sigma(w_1 x + w_0)$$

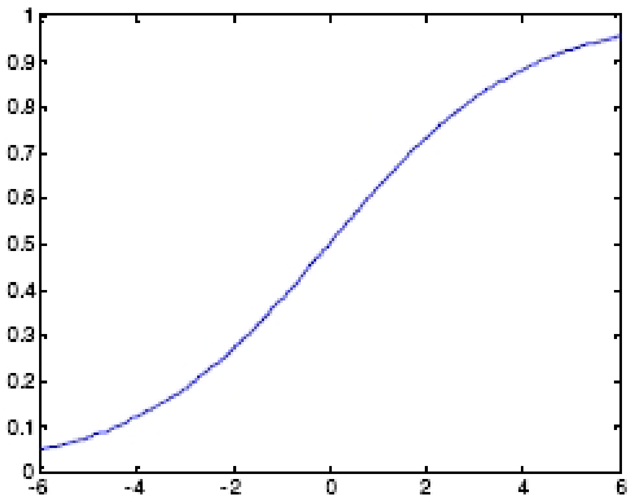
$w_0 = 0, w_1 = 1$



$w_0 = 0, w_1 = 0.5$



$w_0 = -2, w_1 = 1$

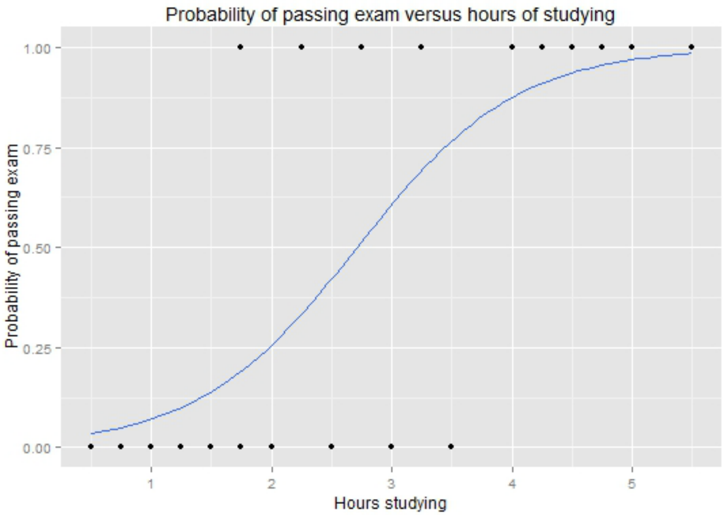


# Παράδειγμα

- Γνωρίζοντας τις ώρες που αφιέρωσε στη μελέτη ένας σπουδαστής/στρια, θα περάσει το διαγώνισμα;
- Δεδομένα εκπαίδευσης

| Hours | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 | 3.00 | 3.25 | 3.50 | 4.00 | 4.25 | 4.50 | 4.75 | 5.00 | 5.50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pass | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

- Πρόβλημα λογιστικής παλινδρόμησης
  - Εύρεση βαρών ***w*** (πχ μέσω κατάβασης κλίσης) που θα μας επιτρέπουν να κάνουμε προβλέψεις




| Ώρες Μελέτης | Πιθανότητα επιτυχίας στο διαγώνισμα |
|---|---|
| 1 | 0.07 |
| 2 | 0.26 |
| 3 | 0.61 |
| 4 | 0.87 |
| 5 | 0.97 |

# Βιβλιογραφία


- *Christopher M. Bishop* “Pattern Recognition and Machine Learning” – Springer (https://link.springer.com/book/9780387310732)
  - Γραμμικά Μοντέλα για Παλινδρόμηση (§3)
- *Ian Goodfellow, Yoshua Bengio, Aaron Courville* “Deep Learning” – MIT Press (https://www.deeplearningbook.org/)
  - Ομαλοποίηση μέσω της προσθήκης όρων ποινής (§7.1)