

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

**Εισαγωγικά περί Μηχανικής Μάθησης
Επισκόπηση Linear & Logistic Regression**

καθ. Βασίλης Μάγκλαρης
maglaris@netmode.ntua.gr
www.netmode.ntua.gr

(μέσω Webex)

Πέμπτη 25/2/2021

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Βιβλιογραφικές Αναφορές

1. Simon Haykin, “**Neural Networks & Learning Machines**”, 3rd Edition, Pearson Education, 2009
2. Simon Haykin, “**Νευρωνικά Δίκτυα & Μηχανική Μάθηση**”, 3η Έκδοση, Παπασωτηρίου, 2010 (Ελληνική μετάφραση)
3. Μιχάλης Λουλάκης, “**Στοχαστικές Διαδικασίες**”, ΣΕΑΒ 2015
http://repfiles.kallipos.gr/html_books/9759/TOC.html
4. Βασίλης Μάγκλαρης, “**Σημειώσεις Μαθήματος Συστήματα Αναμονής**”, Συλλογή διαφανειών για το προπτυχιακό μάθημα της ΣΗΜΜΥ – ΕΜΠ, 2018
http://www.netmode.ntua.gr/courses/undergraduate/queues/documents/Queuing_Systems_2018.pdf
5. Kevin P. Murphy, “**Machine Learning: A Probabilistic Perspective**”, MIT Press, 2012
6. Ian Goodfellow, Yoshua Bengio, Aaron Courville, “**Deep Learning**”, MIT Press, 2016
7. Daniel Jurafsky, James H. Martin, “**Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics & Speech Recognition**”, 3rd Edition draft, 2018
8. Andrew Ng, “**CS229 Lecture Notes**”, Stanford University, Fall 2018
<http://cs229.stanford.edu/notes/cs229-notes1.pdf>
9. James Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, “**An Introduction to Statistical Learning**”, Springer 2013 <https://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf>
10. Richard Sutton, Andrew Barto, “**Reinforcement Learning: An Introduction**”, MIT Press, 2018
11. Christopher Bishop, “**Pattern Recognition & Machine Learning**”, Springer 2006
12. Tom Mitchell, “**Machine Learning**”, McGraw Hill 1997 <http://www.cs.cmu.edu/~tom/mlbook.html>
13. Frank Kelly, “**Reversibility and Stochastic Networks**”, Wiley, 1979
<http://www.statslab.cam.ac.uk/~frank/BOOKS/book/whole.pdf>
14. Sheldon Ross, “**Applied Probability Models with Optimization Applications**”, Dover, 1992

Στα υλικά του μαθήματος αναπαράγονται σχήματα από τις αναφορές [1], [3] και [4] χωρίς περεταίρω ειδική μνεία

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

- Επισκόπηση Μηχανικής Μάθησης: Επιβλεπόμενη, μη επιβλεπόμενη, ανταγωνιστική, ενισχυτική μάθηση. Linear & Logistic Regression, προσδιορισμός παραμέτρων Νευρωνικών Δικτύων με επιβλεπόμενη μάθηση, Back-Propagation Algorithm (ύλη βασισμένη στο μάθημα της ΣΗΜΜΥ – Ε.Μ.Π. «Μηχανική Μάθηση» <https://dsml.ece.ntua.gr/studies/courses/mechanike-mathese>)
- Μη Επιβλεπόμενη Μάθηση: K-Means Clustering, Ανάλυση Κυρίων Συνιστωσών (Principal Components Analysis - PCA) (ύλη βασισμένη στα κεφ. 5 & 8 των αναφορών [1], [2])
- Μη Επιβλεπόμενη Μάθηση και Βασικές Έννοιες Στατιστικής Μηχανικής: Αλυσίδες Markov, ταξινόμηση καταστάσεων, πιθανότητες μετάβασης, εξισώσεις Chapman - Kolmogorov, επαναληπτικότητα - παροδικότητα, αναλλοίωτες κατανομές, ασυμπτωτική συμπεριφορά. Μέθοδοι Monte Carlo προσομοίωσης αλυσίδων Markov, αλγόριθμος Metropolis - Hastings. Προσομοιωμένη Ανόπτηση (Simulated Annealing), δειγματοληψία Gibbs. Παραγωγικά Μοντέλα Μάθησης (Generative Models), Μηχανή Boltzmann, Restricted Boltzmann Machine (RBM), Δίκτυα Πεποίθησης Μεγάλου Βάθους (Deep Belief Nets - DBN) (ύλη βασισμένη στο κεφ. 11 των αναφορών [1], [2] και στις αναφορές [3], [4])
- Ενισχυτική Μάθηση και Δυναμικός Προγραμματισμός: Διαδικασίες Απόφασης Markov (Markov Decision Processes), κριτήριο βελτιστοποίησης Bellman (Bellman's Optimality Criterion), αλγόριθμοι βελτιστοποίησης Δυναμικού Προγραμματισμού (Policy Iteration & Value Iteration Algorithms). Προσεγγιστικές μέθοδοι δυναμικού προγραμματισμού, Q-Learning (ύλη βασισμένη στο κεφ. 12 των αναφορών [1], [2])
- Ενισχυτική Μάθηση για Δρομολόγηση στο Internet: Αλγόριθμος Bellman-Ford, Border Gateway Protocols (BGP) (ύλη βασισμένη στο μάθημα της ΣΗΜΜΥ – Ε.Μ.Π. «Διαχείριση Δικτύων – Ευφυή Δίκτυα» http://mycourses.ntua.gr/course_description/index.php?cidReq=ECE1046)
- Πιθανοτική Ταξινόμηση: Κανόνας Bayes, Προσεγγιστικές Μέθοδοι - Naive Bayes Algorithm (ύλη βασισμένη στα κεφ. 1 & 2 των αναφορών [1], [2] και στο κεφ. 6 της αναφοράς [12])
- Μέθοδοι Ταξινόμησης Πυρήνα (Kernel Methods): Διαχωρισμότητα Προτύπων - Cover's Theorem, Radial-Basis Function (RBF) Networks, K-Means Clustering, Support Vector Machines (SVM) (ύλη βασισμένη στα κεφ. 5 & 6 των αναφορών [1], [2] και στο κεφ. 6 της αναφοράς [12])
- Δένδρα Αποφάσεων (Decision Trees), Random Forests (ύλη βασισμένη στο κεφ. 8 της αναφοράς [9])
- Αλγόριθμοι με βάση Ακολουθιακά Μοντέλα Συσχετισμένων Δεδομένων Μάθησης: Time-series & Speech Processing Datasets, Recurrent Neural Nets (RNN), Δίκτυα Long-Short Term Memory (LSTM) (ύλη βασισμένη στο κεφ. 15 των αναφορών [1], [2] και στο κεφ. 10 της αναφοράς [6])

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μερικοί Θεμελιωτές του Κλάδου (1/2)

Thomas Bayes (1701 -1761): Συνδυαστικές Πιθανότητες, Στατιστικές Εκτιμήσεις

https://en.wikipedia.org/wiki/Thomas_Bayes



Johann Carl Friedrich Gauss (1777 -1855): Άλγεβρα, Θεωρία Αριθμών, Κατανομή Σφάλματος Παρατηρήσεων

https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss



Josiah Willard Gibbs (1839 -1903): Στατιστική Μηχανική, Θερμοδυναμική

https://en.wikipedia.org/wiki/Josiah_Willard_Gibbs



Ludwig Boltzmann (1844 -1906): Στατιστική Μηχανική

https://en.wikipedia.org/wiki/Ludwig_Boltzmann



Andrey Markov (1856 -1922): Θεωρία Πιθανοτήτων, Στοχαστικές Διεργασίες

https://en.wikipedia.org/wiki/Andrey_Markov



Andrey Kolmogorov (1903 -1987): Θεωρία Πιθανοτήτων

https://en.wikipedia.org/wiki/Andrey_Kolmogorov



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μερικοί Θεμελιωτές του Κλάδου (2/2)

Donald Hebb (1904 - 1985): Νευροφυσιολογία, Κανόνες Μάθησης

https://en.wikipedia.org/wiki/Donald_O._Hebb



Richard Bellman (1920 - 1984): Εφαρμοσμένα Μαθηματικά, Δυναμικός Προγραμματισμός

https://en.wikipedia.org/wiki/Richard_E._Bellman



Nicholas Metropolis - Μητρόπουλος (1915 - 1999): Προσομοίωση Monte Carlo, Simulated Annealing

https://en.wikipedia.org/wiki/Nicholas_Metropolis



Frank Rosenblatt (1928 - 1972): Ψυχολογία, Τεχνητή Νοημοσύνη (AI), Νευρωνικά Δίκτυα, Perceptron

https://en.wikipedia.org/wiki/Frank_Rosenblatt



David Rumelhart (1942 - 2011): Ψυχολογία, AI, Back Propagation Algorithm

https://en.wikipedia.org/wiki/David_Rumelhart



Geoffrey Hinton (1947): AI, Back Propagation Algorithm, Μηχανή Boltzmann, Deep Belief Networks

https://en.wikipedia.org/wiki/Geoffrey_Hinton



Vladimir Vapnik (1936): Στατιστική Μάθηση, Support Vector Machines

https://en.wikipedia.org/wiki/Vladimir_Vapnik



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (1/3)

Λόγος ανάπτυξης Μηχανικής Μάθησης:

- Η κατακλυσμαία ζήτηση για αυτοματοποιημένες μεθόδους ανάλυσης δεδομένων με στόχο την ταξινόμηση βάσει χαρακτηριστικών και την εξαγωγή προβλέψεων
- Η ανάγκη εξαγωγής συμπερασμάτων που αποκρυσταλλώνουν την εμπειρία και αξιοποιούν διαθέσιμα στατιστικά ιστορικά δεδομένα
- Η αλματώδης συσσώρευση τεράστιου όγκου ετερογενών δεδομένων που απαιτούν την ανάπτυξη ευφυών αλγορίθμων εξόρυξης

Ορισμός:

Το σύνολο μεθόδων αυτοματοποιημένης ανίχνευσης (detection), επεξεργασίας και αποθήκευσης προτύπων - σχεδίων δεδομένων (data patterns) μέσω διαδικασίας μάθησης (*learning phase*) από γνωστά *training sets* για γενίκευση σε *test sets* (*generalization*):

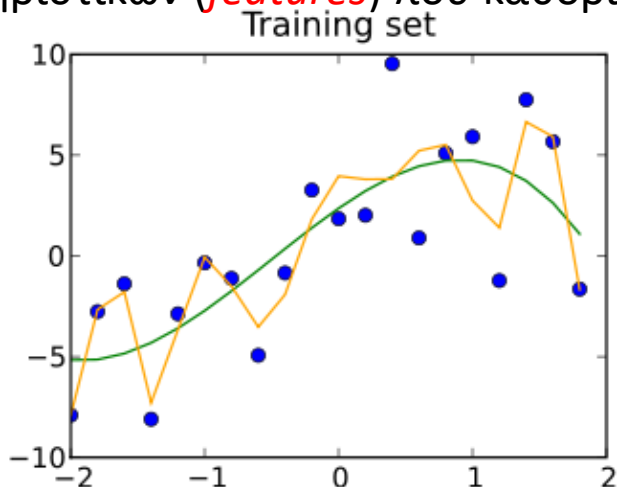
- *Συσχέτιση, ταξινόμηση και ομαδοποίηση* μεγάλου όγκου δεδομένων με επιλογή κυρίαρχων χαρακτηριστικών τους (features, patterns)
- *Αναγνώριση προτύπων* με βάση κύρια χαρακτηριστικά τους (pattern recognition)
- *Προσέγγιση συναρτήσεων* από ζεύγη τιμών δεδομένων
- *Προβλέψεις εμφάνισης* σύμφωνα με τα σχέδια που προέκυψαν
- *Στήριξη λήψης αποφάσεων* σε περιβάλλον αβεβαιότητας με χρήση εργαλείων θεωρίας πιθανοτήτων – στοχαστικών διαδικασιών και στατιστικής

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (2/3)

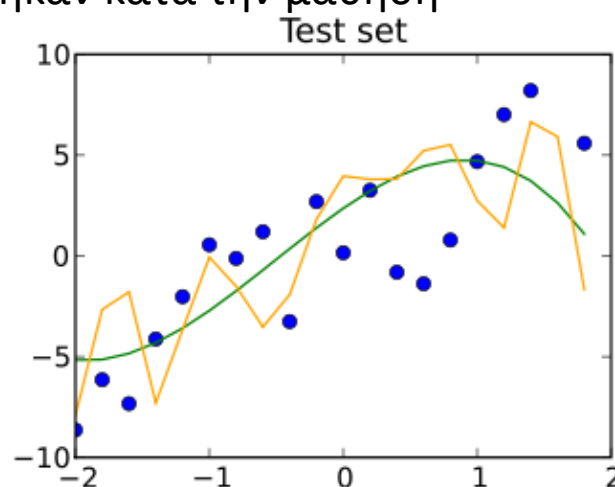
Ορισμοί Συνόλων Δεδομένων (https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets)

- **Training Datasets:** Δείγμα (*sample*) παραδειγμάτων που χρησιμοποιείται στη διαδικασία μάθησης για εκτίμηση παραμέτρων του συστήματος. Στοιχεία του συνόλου του δείγματος μάθησης αναφέρονται σαν δειγματικά στοιχεία μάθησης (*training sample points, examples*)
- **Validation Datasets:** Δείγμα παραδειγμάτων (δειγματικών στοιχείων) με χαρακτηριστικά όμοια με του δείγματος μάθησης για επικύρωση της σύγκλισης της διαδικασίας μάθησης
- **Test Datasets:** Δείγμα παραδειγμάτων που δεν χρησιμοποιήθηκαν στην διαδικασία μάθησης και εισάγονται σε ρυθμισμένο σύστημα. Αξιολογούν την ακρίβεια (*accuracy*) αντιστοιχήσεων (συσχετίσεων/ταξινομήσεων/προβλέψεων...) και την ικανότητα γενίκευσης (*generalization*) των χαρακτηριστικών (*features*) που καθορίστηκαν κατά την μάθηση



Training Dataset (μπλε σημεία μάθησης):

- Λεπτομερής **κίτρινη** καμπύλη εκτίμησης με απόκλιση $MSE=4$
- Απλή **πράσινη** καμπύλη με απόκλιση $MSE=9$



Test Dataset (μπλε σημεία γενίκευσης):

- Απόκλιση από **κίτρινη** καμπύλη $MSE=15$ (από 4) – **OVERFITTING**
- Απόκλιση από **πράσινη** καμπύλη $MSE=13$ (από 9)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (3/3)

Γενικές Κατηγορίες Συστημάτων Μηχανικής Μάθησης:

- **Επιβλεπόμενη Μάθηση με Εκπαιδευτή - Supervised Learning**
 - Χρήση δεδομένων μάθησης με συνημμένες επιθυμητές αποκρίσεις εξόδου (*labeled training sample points*) που εκπαιδεύουν σε πρώτη φάση το σύστημα Μηχανικής Μάθησης μέσω εξωτερικού εκπαιδευτή για αναζήτηση απόκρισης (ταξινόμηση, πρόβλεψη) σε επόμενη φάση γενίκευσης με νέα δεδομένα εισόδου
- **Μάθηση χωρίς Εκπαιδευτή**
 - Μη Επιβλεπόμενη Μάθηση - **Unsupervised Learning** όπου το σύστημα αυτορυθμίζεται ανακαλύπτοντας από μόνο του ενδιαφέρουσες στατιστικές δομές (*stochastic features, patterns*) σε μεγάλο όγκο μη χαρακτηρισμένων δεδομένων (*unlabeled datasets*) ώστε να προκύπτουν μοντέλα, μέθοδοι επεξεργασίας, αποθήκευσης και ταξινόμησής, π.χ. σε ομάδες (*clusters*)
 - Ενισχυτική Μάθηση - **Reinforcement Learning** όπου το σύστημα αντιδρά σε σήματα επιβράβευσης/αποθάρρυνσης μέσω *agents* από το περιβάλλον εισόδου, προς το οποίο κοινοποιεί ενέργειές του (*actions*) που επηρεάζουν την εξέλιξη της κατάστασης του περιβάλλοντος για την επίτευξη μακροπρόθεσμου στόχου

Η επιβλεπόμενη μάθηση προσφέρει απόδοση, αξιοπιστία και ταχύτητα για προβλήματα που αφορούν σε αποφάσεις χειρισμού δεδομένων μετά από διαδικασία μάθησης αλλά απαιτεί **labeled learning data sets** που δεν είναι εύκολα διαθέσιμα

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Γενικό Μοντέλο Επιβλεπόμενης Μάθησης - Supervised Learning

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

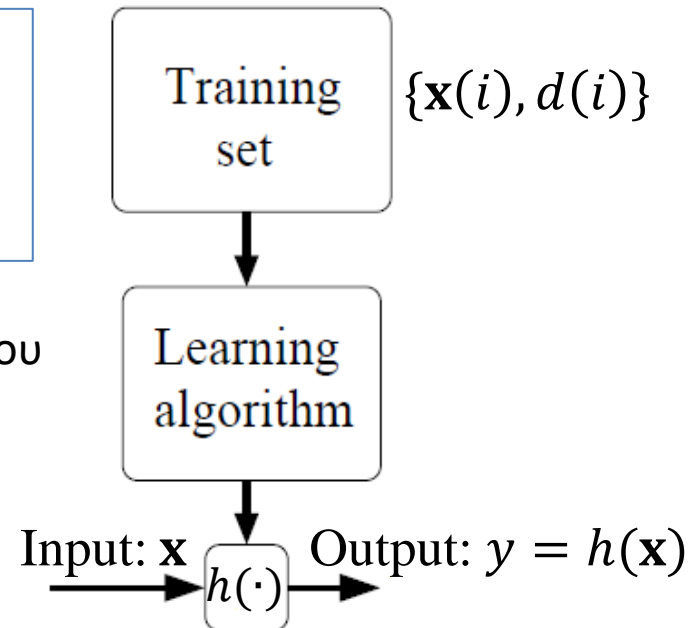
- Στόχος του συστήματος είναι η αντιστοίχιση ενός δειγματικού στοιχείου εισόδου (**input sample point, example**) $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ σε τιμές εξόδου y που εκτιμούν επιθυμητές (**desired**) τιμές d (π.χ. πρόβλεψη ή ταξινόμηση). Τα στοιχεία x_i είναι αριθμητικές τιμές που κωδικοποιούν m ειδοποιά χαρακτηριστικά (**features**) του δειγματικού στοιχείου \mathbf{x}

Ζητείται ο προσδιορισμός της συνάρτησης εισόδου - εξόδου $y = h(\mathbf{x}) \cong d$ που προκύπτει από δείγμα μάθησης (**Training Set**) N **labeled** ζευγών $\{\mathbf{x}(i), d(i)\}, i = 1, 2, \dots, N$ γνωστών σε εξωτερικό εκπαιδευτή (**supervisor**)

- Η σχεδίαση της $h(\cdot)$ βασίζεται σε αλγόριθμο μάθησης, με προσαρμογή της μορφής και των παραμέτρων ενός μοντέλου ώστε να προσεγγίζεται ο στόχος της υπόθεσης

$$d(i) \cong y(i) = h(\mathbf{x}(i))$$

- Αν ο στόχος ικανοποιείται με μικρό αριθμό διακριτών επιλογών της y πρόκειται για πρόβλημα Ταξινόμησης, **Classification** (για δύο επιλογές έχουμε δυαδική ταξινόμηση)
- Αν η έξοδος y λαμβάνει συνεχείς τιμές, το πρόβλημα αναφέρεται σαν Παλινδρόμηση, **Regression**



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

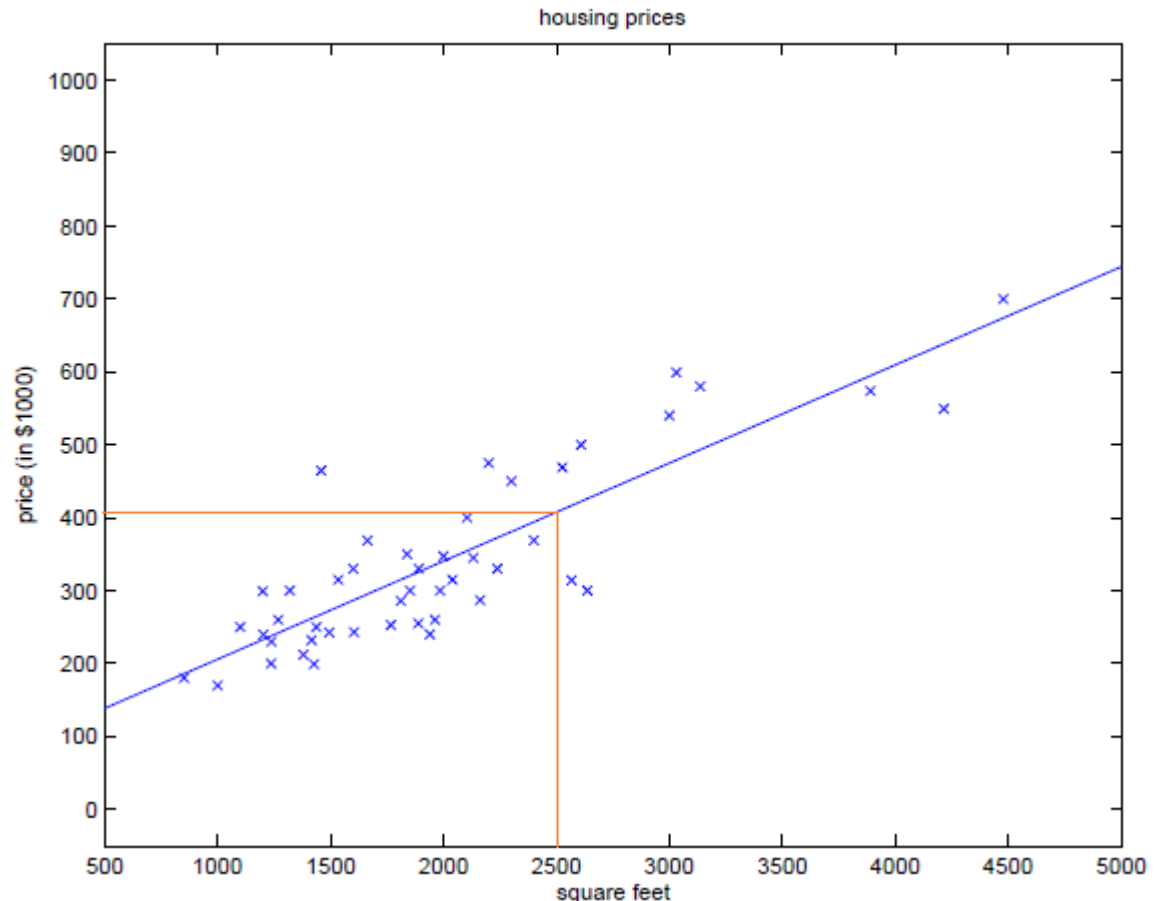
Παράδειγμα Επιβλεπόμενης Μάθησης – Linear Regression

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

- Ζητείται γραμμική συνάρτηση $h(x)$ που να προσεγγίζει την τιμή κατοικίας $y = h(x)$ με βάση την επιφάνεια x και εμπειρικά δεδομένα $\{x(i), d(i)\}, i = 1, 2, \dots, N$ (**Training Sample**) για καταγραμμένες περιπτώσεις κατοικιών στην ίδια περιοχή

ΕΜΒΑΔΟΝ (square feet)	ΤΙΜΗ (1000\$)
2104	400
1600	330
2400	369
1416	232
3000	540
...	...

Ενδεικτικά Ζεύγη από $N = 47$ Περιπτώσεις



Γραμμική Παλινδρόμηση – Linear Regression: $y = h(x) = 0.1392x + 89.6$
Πρόβλεψη τιμής κατοικίας 2500 τετραγωνικών: \$437,000

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός Παραμέτρων Linear Regression (1/2)

- Το διάνυσμα του δειγματικού στοιχείου εισόδου $\mathbf{x} = [x_0 \ x_1 \ \dots \ x_m]^T$ ορίζεται με τιμές που κωδικοποιούν m χαρακτηριστικά (**features**) του: x_1, x_2, \dots, x_m με $x_0 \triangleq 1$ (**intercept term**)
- Το σύστημα linear regression προσδιορίζει τις παραμέτρους $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_m]^T$ της συνάρτησης $y = h_{\mathbf{w}}(\mathbf{x}) = w_0 x_0 + w_1 x_1 + \dots + w_m x_m = \mathbf{w}^T \mathbf{x}$ ώστε η y να έχει μικρές **αποκλίσεις** εφαρμοζόμενη στο δείγμα μάθησης (**Training Set**) $\{\mathbf{x}(i), d(i)\}, i = 1, 2, \dots, N$
 - $\mathbf{x}(i)$: Διάνυσμα τιμών εισόδου (χαρακτηριστικών) στοιχείου μάθησης i (**regressors**)
 - $d(i)$: Τιμή εξόδου (απόκριση) στοιχείου μάθησης i (**regressand**)
 - $y(i) = h_{\mathbf{w}}(\mathbf{x}(i))$: Εκτίμηση εξόδου του συστήματος για διάνυσμα εισόδου $\mathbf{x}(i)$
 - $\varepsilon(i) = d(i) - y(i)$: Απόκλιση (**error**) εκτίμησης για το ζεύγος $\{\mathbf{x}(i), d(i)\}, i = 1, 2, \dots, N$
 - Οι $\mathbf{x}(i), d(i), \varepsilon(i)$ μπορούν να θεωρηθούν δειγματικές τιμές τυχαίων μεταβλητών
- Κοινό κριτήριο σύγκλησης αφορά στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (**Least Mean Square, LMS**) ως προς τις παραμέτρους \mathbf{w} της $h_{\mathbf{w}}(\mathbf{x})$ ή αντίστοιχα της συνάρτησης κόστους:

$$J(\mathbf{w}) \triangleq \frac{1}{2} \sum_{i=1}^N [\varepsilon(i)]^2 = \frac{1}{2} \sum_{i=1}^N [d(i) - h_{\mathbf{w}}(\mathbf{x}(i))]^2$$

- Παράδειγμα **Linear Regression** μίας μεταβλητής εισόδου x :

$$\mathbf{x} = [1 \ x]^T, \quad \mathbf{w} = [w_0 \ w_1]^T, \quad y = h_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} = w_0 + w_1 x$$

$$J(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N [d(i) - (w_0 + w_1 x(i))]^2$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός Παραμέτρων Linear Regression (2/2)

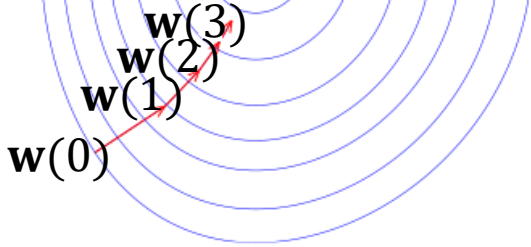
Απεικόνιση Σύγκλισης Gradient Descent:

Ελαχιστοποίηση της συνάρτησης $J(\mathbf{w})$, με παραμέτρους το διάνυσμα \mathbf{w} , μέσω διαδοχικής προσέγγισης στο βήμα $k \rightarrow k + 1$ προς την κλίση (Gradient) $\nabla J(\mathbf{w})$ σταθμισμένο κατά α :

$$\mathbf{w}(k + 1) = \mathbf{w}(k) - \alpha \nabla J(\mathbf{w}(k))$$

Αν υπάρχει σύγκλιση: $\mathbf{w} = \lim_{k \rightarrow \infty} \mathbf{w}(k)$

Σημείωση: Για linear regression υπάρχει πάντα σύγκλιση



Κανόνας Μάθησης LMS (Widrow-Hoff)

- **Batch Gradient Descent:** Προσδιορισμός του $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_m]^T$ που ελαχιστοποιεί το σφάλμα $J(\mathbf{w})$ σε κάθε βήμα για **όλο** το δείγμα μάθησης $\{\mathbf{x}(i), d(i)\}, i = 1, 2, \dots, N$

$$w_j := w_j - \alpha \frac{\partial J(\mathbf{w})}{\partial w_j} = w_j + \alpha \sum_{i=1}^N [d(i) - h_{\mathbf{w}}(\mathbf{x}(i))] x_j(i), \quad j = 0, 1, 2, \dots, m \quad \forall i$$

- **Stochastic (Incremental) Gradient Descent, Stochastic Approximations:** Προσδιορισμός του \mathbf{w} με τυχαία (στοχαστική) διαδοχική εισαγωγή **στοιχείων** του δείγματος μάθησης $\{\mathbf{x}(i), d(i)\}, i = 1, 2, \dots, N$ μέχρι να ικανοποιηθεί κριτήριο σύγκλισης

$$w_j := w_j + \alpha [d(i) - h_{\mathbf{w}}(\mathbf{x}(i))] x_j(i), \quad j = 0, 1, 2, \dots, m \quad i = 1, 2, \dots, N$$

- Η στοχαστική μέθοδος δίνει συνήθως ικανοποιητικά αποτελέσματα με μικρή επιβάρυνση υπολογιστικών πόρων και **προτιμάται για μηχανική μάθηση**
- Το βήμα α στις επαναλήψεις ορίζει τον ρυθμό της μάθησης (**learning rate**). Για σταθεροποίηση της σύγκλισης μπορεί να μεταβάλλεται στην πορεία των επαναλήψεων π.χ. μεγάλη τιμή στα πρώτα βήματα, μικρότερη όσο πλησιάζουμε στη σύγκλιση

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Polynomial Regression: Πολυωνυμική Προσέγγιση Χαρακτηριστικού

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

- Γραμμική προσέγγιση (**Linear Regression**) μίας μεταβλητής εισόδου $\mathbf{x} = [1 \ x]^T$:

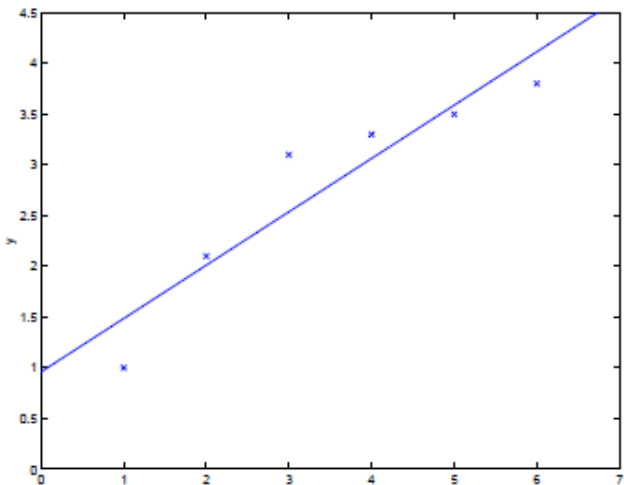
$$y = h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1x$$

- Προσέγγιση με γραμμικό πολυώνυμο **2^{ου} βαθμού**:

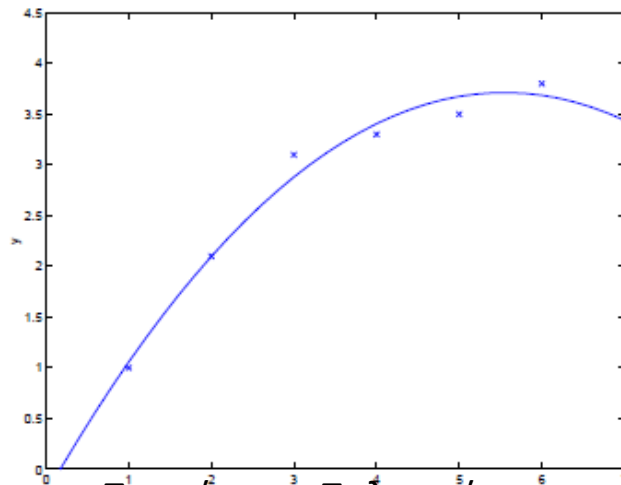
$$y = h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1x + w_2x^2$$

- Προσέγγιση με γραμμικό πολυώνυμο **K βαθμού**:

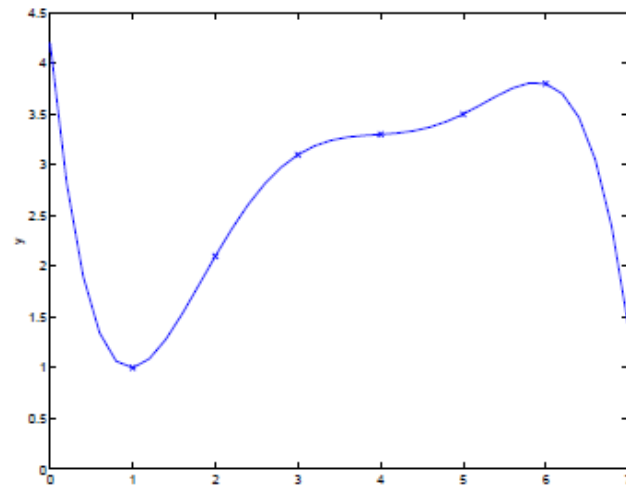
$$y = h_{\mathbf{w}}(\mathbf{x}) = \sum_{j=0}^K w_j x^j$$



Γραμμική Προσέγγιση
(**Underfitting**)



Προσέγγιση Πολυωνύμου
2^{ου} Βαθμού
(**OK**)

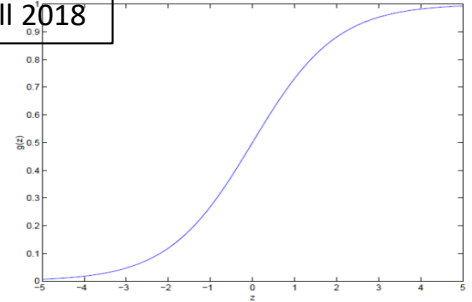


Προσέγγιση Πολυωνύμου
5^{ου} Βαθμού
(**Overfitting**)

Κίνδυνοι Υπεραπλούστευσης (**Underfitting**) & Υπερβολής (**Overfitting**)

Ταξινόμηση – Classification (1/2)

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018



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

Logistic/Sigmoid Function

Δειγματικά στοιχεία \mathbf{x} με m διαστάσεις (χαρακτηριστικά, **features**)

Διαδικές Κλάσεις Εξόδου (**Classes, Labels**) $y \in \{0,1\}$ ή $y \in \{-, +\}$

Training Set: $\{\mathbf{x}(i), d(i)\}, i = 1, 2, \dots, N$

• **Μοντέλο Logistic Regression:** $h_{\mathbf{w}}(\mathbf{x}) = g(\mathbf{w}^T \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}}$

$$\mathbf{w}^T \mathbf{x} = \sum_{j=0}^m w_j x_j = w_0 + \sum_{j=1}^m w_j x_j$$

Οι πιθανότητες τυχαίας μεταβλητής εξόδου $y \in \{0,1\}$ υπό συνθήκη μεταβλητών εισόδου \mathbf{x} και με συνάρτηση **Logistic Regression** $h_{\mathbf{w}}(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}}$ ακολουθούν κατανομή **Bernoulli** και οδηγούν σε εκτίμηση της εξόδου y μετά τον προσδιορισμό των παραμέτρων \mathbf{w} :

$$P(y = 1 | \mathbf{x}; \mathbf{w}) = h_{\mathbf{w}}(\mathbf{x}), \quad P(y = 0 | \mathbf{x}; \mathbf{w}) = 1 - h_{\mathbf{w}}(\mathbf{x})$$

$$\text{ή } p(y | \mathbf{x}; \mathbf{w}) = (h_{\mathbf{w}}(\mathbf{x}))^y (1 - h_{\mathbf{w}}(\mathbf{x}))^{1-y}$$

Κανόνας Εκτίμησης y :

$y = 1$ αν $h_{\mathbf{w}}(\mathbf{x}) > 1/2$

$y = 0$ αν $h_{\mathbf{w}}(\mathbf{x}) < 1/2$

Κριτήριο σύγκλησης λόγω μη γραμμικής $h_{\mathbf{w}}(\mathbf{x})$ προτιμάται της ελαχιστοποίησης του τετραγωνικού σφάλματος (**LMS**) η **μεγιστοποίηση** του λόγου πιθανοφάνειας (**Likelihood Ratio**) $L(\mathbf{w})$ των στοιχείων του συνόλου μάθησης $\{\mathbf{x}(i), d(i)\}, i = 1, 2, \dots, N$. Θεωρούμε πως οι τιμές εξόδου $d(i)$ είναι ανεξάρτητες δυαδικές τυχαίες μεταβλητές για το δείγμα μάθησης $\mathbf{X} = [\mathbf{x}(1) \ \mathbf{x}(2) \ \dots \ \mathbf{x}(N)]$ και με παραμέτρους \mathbf{w} :

$$L(\mathbf{w}) \triangleq p\{(d(1), d(2), \dots, d(N)) | (\mathbf{X}; \mathbf{w})\} = \prod_{i=1}^N p\{d(i) | (\mathbf{x}(i); \mathbf{w})\}$$

Ταξινόμηση – Classification (2/2)

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

- **Μοντέλο Logistic Regression (συνέχεια):**

$$L(\mathbf{w}) = \prod_{i=1}^N p\{d(i)|(\mathbf{x}(i); \mathbf{w})\} = \prod_{i=1}^N \left\{ (h_{\mathbf{w}}(\mathbf{x}(i)))^{d(i)} (1 - h_{\mathbf{w}}(\mathbf{x}(i)))^{1-d(i)} \right\}$$

Αντί της μεγιστοποίησης του $L(\mathbf{w})$ μεγιστοποιούμε τον λογάριθμο $l(\mathbf{w}) = \log L(\mathbf{w})$:

$$l(\mathbf{w}) = \sum_{i=1}^N \{d(i) \log h_{\mathbf{w}}(\mathbf{x}(i)) + (1 - d(i)) \log(1 - h_{\mathbf{w}}(\mathbf{x}(i)))\}$$

Εφαρμόζουμε μεγιστοποίηση **Gradient Ascent** στο βήμα $k \rightarrow k + 1$ με βήμα α θετικό:

$$\mathbf{w}(k + 1) = \mathbf{w}(k) + \alpha \nabla l(\mathbf{w}(k))$$

Για τον υπολογισμό της $\nabla l(\mathbf{w}(k))$ και την εφαρμογή του στοχαστικού προσεγγιστικού κανόνα (**Stochastic Gradient Ascent**) προσδιορισμού των παραμέτρων w_j με

διαδοχική εφαρμογή στα στοιχεία $i = 1, 2, \dots, N$ του **Training Set** υπολογίζουμε την

μερική παράγωγο $\frac{\partial}{\partial w_j} l(\mathbf{w}) = \dots = [d(i) - h_{\mathbf{w}}(\mathbf{x}(i))]x_j(i) \Leftrightarrow$

$$w_j := w_j + \alpha [d(i) - h_{\mathbf{w}}(\mathbf{x}(i))]x_j(i), \quad j = 1, 2, \dots, m$$

(ίδιας μορφής **Επαναληπτικός Κανόνας Μάθησης** με τον κανόνα **LMS!**)

- **Μοντέλο Perceptron:** $h_{\mathbf{w}}(\mathbf{x}) = g(\mathbf{w}^T \mathbf{x})$

$$g(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \text{ Threshold Function}$$

Προκύπτει παρόμοιος **Επαναληπτικός Κανόνας Μάθησης** παραμέτρων w_j

$$w_j := w_j + \alpha [d(i) - h_{\mathbf{w}}(\mathbf{x}(i))]x_j(i), \quad j = 1, 2, \dots, m$$