ε.δε.μ²
ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μηχανική Μάθηση & Τεχνητή Νοημοσύνη

Ορισμοί Συνόλων Δεδομένων (Datasets)

Διακριτικά (Discriminative) & Παραγωγικά (Generative) Μοντέλα, το ChatGPT

Επιβλεπόμενη Μάθηση, Linear & Logistic Regression

καθ. Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

Αίθουσα 002, Νέα Κτίρια ΣΗΜΜΥ

Τρίτη 28/2/2023

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
3. Bernhard Mehlig, "**Machine learning with neural networks**", Cambridge Univ. Press 2021 https://arxiv.org/pdf/1901.05639.pdf
4. Μιχάλης Λουλάκης, "**Στοχαστικές Διαδικασίες**", ΣΕΑΒ 2015 http://repfiles.kallipos.gr/html_books/9759/TOC.html
5. Βασίλης Μάγκλαρης, "**Σημειώσεις Μαθήματος Συστήματα Αναμονής**", Συλλογή διαφανειών ΣΗΜΜΥ – ΕΜΠ, 2018 http://www.netmode.ntua.gr/courses/undergraduate/queues/documents/Queuing_Systems_2018.pdf
6. Kevin P. Murphy, "**Machine Learning: A Probabilistic Perspective**", MIT Press, 2012
7. Ian Goodfellow, Yoshua Bengio, Aaron Courville, "**Deep Learning**", MIT Press, 2016 https://www.deeplearningbook.org/
8. Daniel Jurafsky, James H. Martin, "**Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics & Speech Recognition**", 3rd Edition draft, 2018
9. Andrew Ng, "**CS229 Lecture Notes**", Stanford University, Fall 2018 https://see.stanford.edu/materials/aimlcs229/cs229-notes1.pdf
10. James Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, "**An Introduction to Statistical Learning**", 2nd Edition, Springer 2021 https://hastie.su.domains/ISLR2/ISLRv2_website.pdf
11. Richard Sutton, Andrew Barto, "**Reinforcement Learning: An Introduction**", MIT Press, 2018
12. Christopher Bishop, "**Pattern Recognition & Machine Learning**", Springer 2006
13. Tom Mitchell, "**Machine Learning**", McGraw Hill 1997 http://www.cs.cmu.edu/~tom/mlbook.html
14. Charu C. Aggarwal, "**Outlier Analysis**", Springer 2013 https://link.springer.com/book/10.1007/978-3-030-68640-6
15. Leonida Gianfagna and Antonio Di Cecco, "**Explainable AI with Python**", Springer 2021 https://link.springer.com/book/10.1007/978-3-030-68640-6
16. Christoph Molnar, "**Interpretable Machine Learning**," Second Edition, Munich, 2022 https://christophm.github.io/interpretable-ml-book/
17. Frank Kelly, "**Reversibility and Stochastic Networks**", Wiley, 1979 http://www.statslab.cam.ac.uk/~frank/BOOKS/book/whole.pdf
18. Sheldon Ross, "**Applied Probability Models with Optimization Applications**", Dover, 1992
19. Dimitri P. Bertsekas and John Tsitsiklis, "**Neuro-Dynamic Programming**," Athena Scientific, Belmont MA 1996

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

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Περιγραφή Ύλης (1/2)

1. Επισκόπηση Αλγορίθμων Βελτιστοποίησης στη Μηχανική Μάθηση: Σχέση Μηχανικής Μάθησης (ML) και Τεχνητής Νοημοσύνης (AI). Σύνολα δεδομένων Training, Validation & Testing Datasets. Επιβλεπόμενη, μη επιβλεπόμενη, ενισχυτική μάθηση. Ορισμοί Διακριτικών (Discriminative) & Παραγωγικών (Generative) Μοντέλων, το ChatGPT (Chat Generative Pre-trained Transformer). Linear & Logistic Regression
2. Νευρωνικά Δίκτυα, κανόνας του Hebb. Προσδιορισμός παραμέτρων με επιβλεπόμενη μάθηση, Back-Propagation Algorithm (ύλη βασισμένη στο κεφ. 1 της [1], στο *κεφ. 1* της [3], στο *κεφ. 1* της [7] και στην [9])
3. Μη Επιβλεπόμενη Μάθηση: *K*-Means Clustering, Ανάλυση Κυρίων Συνιστωσών (Principal Components Analysis - PCA), Self-Organizing Maps (SOM), Autoencoders (ύλη βασισμένη στα *κεφ. 5* & *8* της [1] και στο *κεφ. 10* της [3])
4. Βασικές Έννοιες Στατιστικής Μηχανικής στη Μηχανική Μάθηση: Αλυσίδες Markov, ταξινόμηση καταστάσεων, πιθανότητες μετάβασης, εξισώσεις Chapman - Kolmogorov, επαναληπτικότητα - παροδικότητα, αναλλοίωτες κατανομές, ασυμπτωτική συμπεριφορά (ύλη βασισμένη στο *κεφ. 11* της [1], και στις [4], [5])
5. 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* της [1] και στο *κεφ. 4* της [3])
6. Ενισχυτική Μάθηση και Δυναμικός Προγραμματισμός: Διαδικασίες Απόφασης Markov (Markov Decision Processes), κριτήριο βελτιστοποίησης Bellman (Bellman's Optimality Criterion), αλγόριθμοι βελτιστοποίησης Δυναμικού Προγραμματισμού (Value & Policy Iteration algorithms). Προσεγγιστικές μέθοδοι δυναμικού προγραμματισμού, TD & Q-Learning (ύλη βασισμένη στο *κεφ. 12* της [1])

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Περιγραφή Ύλης (2/2)

7. Ενισχυτική Μάθηση για Δρομολόγηση στο Internet: Αλγόριθμος Bellman-Ford, Border Gateway Protocols (BGP) (ύλη βασισμένη στο μάθημα ΣΗΜΜΥ Ε.Μ.Π. «*Διαχείριση Δικτύων – Ευφυή Δίκτυα*» https://www.netmode.ntua.gr/wp-content/uploads/2023/01/NetMan_IP_Routing_2022_10_31.pdf)
8. Αλγόριθμοι Πυρήνα και Διαχωρισιμότητα Προτύπων: Θεώρημα του Cover, εφαρμογές σε Radial-Basis Function (RBF) Networks, Υβριδική Μάθηση, Support Vector Machines (SVM) (ύλη βασισμένη στα *κεφ. 5* & *6* της [1] και στο *κεφ. 6* της [13])
9. Μη-παραμετρικοί Ταξινομητές, ταξινόμηση σύμφωνα με γνωστές κλάσεις *K* γειτονικών στοιχείων μάθησης, *K*-Nearest Neighbors (KNN) (ύλη βασισμένη στο *κεφ. 2* της [10] και στο *κεφ. 8* της [13])
10. Στατιστική αξιολόγηση δυαδικής ταξινόμησης, Confusion Matrix, Receiver Operating Charactrisitcs (ROC) & Area Under the Curve (AUC), Παραμετρική Πιθανοτική Ταξινόμηση - κανόνας Bayes, προσεγγιστικές μέθοδοι, αλγόριθμος Naïve Bayes (ύλη βασισμένη στο *κεφ. 6* της [13] και στο *κεφ. 5* της [7])
11. Δένδρα Αποφάσεων (Decision Trees): Αλγόριθμοι διαμόρφωσης CART (Classification And Regression Trees), Gini Index, Random Forests, Αλγόριθμοι Bagging (Bootstrap & aggregating (ύλη βασισμένη στο *κεφ. 8* της [10])
12. Ακολουθιακά Μοντέλα και Αλγόριθμοι βασισμένοι σε Time-series & Speech Processing Datasets: Recurrent Neural Nets (RNN), δίκτυα Hopfield, Long-Short Term Memory (LSTM) Nets (ύλη βασισμένη στο *κεφ. 15* της [1], στο κεφ. 2 της [3] και στο *κεφ. 10* της [7])
13. Επεξηγησιμότητα Τεχνητής Νοημοσύνης – eXplainable AI (XAI), αλγοριθμικές προσεγγίσεις: Permutation Feature Importance, LIME (Local Interpretable Model Agnostic Explanation), SHAP (SHapley Additive exPlanations). Εφαρμογή σε ανίχνευση κυβερνοπεπιθέσεων στο Διαδίκτυο (ύλη βασισμένη στα *κεφ. 1, 2* & *4* της [15])

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

Alan Turing (1912 -1954): Υπολογιστική Μηχανική & Τεχνητή Νοημοσύνη (Computing Machinery & Intelligence)
https://en.wikipedia.org/wiki/Alan_Turing

Andrey Kolmogorov (1903 -1987): Θεωρία Πιθανοτήτων
https://en.wikipedia.org/wiki/Andrey_Kolmogorov

Richard Bellman (1920 - 1984): Εφαρμοσμένα Μαθηματικά, Δυναμικός Προγραμματισμός
https://en.wikipedia.org/wiki/Richard_E._Bellman

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μερικοί Θεμελιωτές του Κλάδου (2/2)

Nicholas Metropolis - Μητρόπουλος (1915 - 1999): Προσομοίωση Monte Carlo, Simulated Annealing
https://en.wikipedia.org/wiki/Nicholas_Metropolis

Donald Hebb (1904 - 1985): Νευροφυσιολογία, Κανόνες Μάθησης
https://en.wikipedia.org/wiki/Donald_O._Hebb

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

John Hopfield (1933): Φυσική, Βιολογία, AI, Recurrent Neural Networks (RNN)
https://en.wikipedia.org/wiki/John_Hopfield

Geoffrey Hinton (1947): AI, Back Propagation Algorithm, Μηχανή Boltzmann, Deep Belief Networks
ttps://en.wikipedia.org/wiki/Geoffrey_Hinton

Vladimir Vapnik (1936): Στατιστική Μάθηση, Support Vector Machines (SVM)
https://en.wikipedia.org/wiki/Vladimir_Vapnik

Teuvo Kohonen (1934 - 2021): Self-Organizing Maps (SOM)
https://en.wikipedia.org/wiki/Teuvo_Kohonen

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (1/6)

Λόγοι ανάπτυξης Μηχανικής Μάθησης:

- Η κατακλυσμιαία ανάπτυξη υπολογιστικών υποδομών αποθήκευσης και επεξεργασίας δεδομένων, επιτρέπει σήμερα την υλοποίηση αλγορίθμων στατιστικής ανάλυσης και στοχαστικής βελτιστοποίησης με βάση ιστορικά στοιχεία δείγματος μάθησης

- Η αλματώδης συσσώρευση τεράστιου όγκου πολυδιάστατων δεδομένων (*big data*) με πολλά χαρακτηριστικά, απαιτεί την ανάπτυξη ευφυών αλγορίθμων εξόρυξης εκτιμήσεων, προβλέψεων και ταξινόμησης νεοεμφανιζόμενων δειγματικών στοιχείων

- Η κατανόηση μεθόδων μάθησης σε βιολογικά συστήματα οδηγεί σε αλγορίθμους τεχνητής νοημοσύνης για συμπλήρωση και ελεγχόμενη πρόβλεψη (ή/και δημιουργία) δειγματικών στοιχείων, συμπεριλαμβανομένων ακολουθιών και χρονοσειρών (στοχαστικών διαδικασιών, *stochastic processes*) με βάση παρεμφερή στατιστικά χαρακτηριστικά αποθηκευμένου δείγματος μάθησης

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (2/6)

Ορισμός Τεχνητής Νοημοσύνης (Artificial Intelligence - AI):
Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind (IBM: https://www.ibm.com/topics/artificial-intelligence)

Ορισμός Μηχανικής Μάθησης (Machine Learning - ML):
Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy (IBM: https://www.ibm.com/topics/machine-learning)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (3/6)

Ορισμοί Συνόλων Δεδομένων (Datasets)

https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets

- ***Training Datasets***: Δείγμα (*sample*) παραδειγμάτων (σύνολο δειγματικών στοιχείων, *examples*, *sample points*) που χρησιμοποιείται στη διαδικασία ***μάθησης*** για ρύθμιση παραμέτρων ενός μοντέλου μηχανικής μάθησης με συγκεκριμένη διάρθρωση

- ***Validation Datasets***: Δείγμα παραδειγμάτων με χαρακτηριστικά όμοια με του δείγματος μάθησης για ***επικύρωση*** της σύγκλισης της διαδικασίας μάθησης. Οδηγεί στη επιλογή παραμέτρων διάρθρωσης ενός μοντέλου αποφάσεων (*hyperameter selection*) μέσω συγκρίσεων της ακρίβειας (*accuracy*) προτεινόμενων μοντέλων, ενώ ελέγχει περιορισμούς για ***γενίκευση*** (*generalization*) που μπορεί να οφείλονται σε υπερβολική ακρίβεια (*overfitting*) στα δεδομένα του *training dataset*. Το *validation dataset* μπορεί να παραλείπεται ή να είναι υποσύνολο του *training dataset* με επιλογή μικρότερου αριθμού στοιχείων (10 - 20%)

- ***Test Datasets***: Δείγμα παραδειγμάτων που δεν χρησιμοποιήθηκαν στη διαδικασία μάθησης και εισάγονται σε τελικά ρυθμισμένο σύστημα. Αξιολογούν την ικανότητα ***γενίκευσης*** (*generalization*) των χαρακτηριστικών (*features*) που καθορίστηκαν κατά την μάθηση

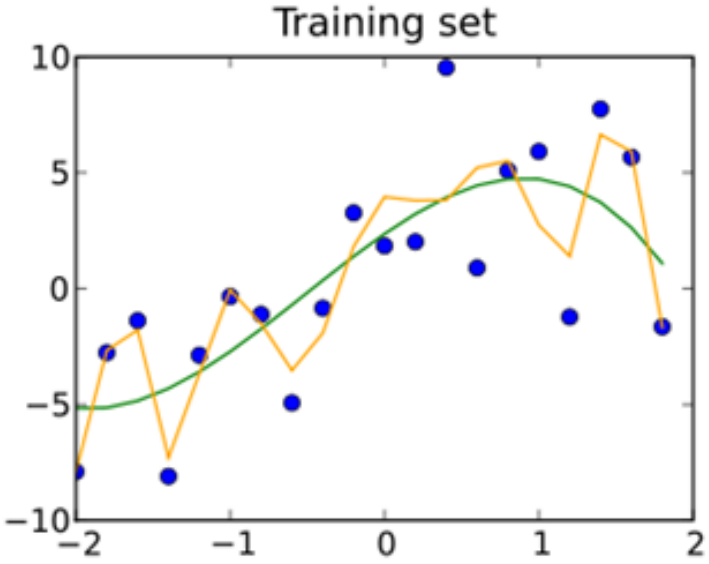
- Σε περίπτωση που δεν ορίζεται *validation dataset*, η ακρίβεια και η ικανότητα γενίκευσης ενός μοντέλου που προκύπτει από το *training dataset* κρίνεται απευθείας μέσω του *test dataset*

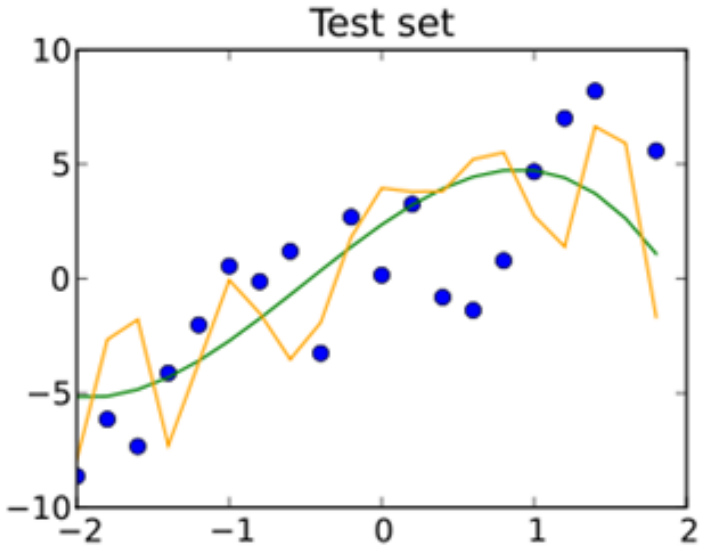
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (4/6)

Ορισμοί Συνόλων Δεδομένων (Datasets)

https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets





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 (5/6)

Ορισμοί Παραμέτρων και Υπερπαραμέτρων

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

- Οι ρυθμιζόμενοι παράμετροι (*parameters*) κατά τη φάση μάθησης αφορούν σε συγκεκριμένη δομή μοντέλου (π.χ. προσδιορισμός συναπτικών βαρών νευρωνικού δικτύου). Ορίζονται με επαναλήψεις για τη συγκεκριμενοποίηση του μοντέλου εισόδου/εξόδου ώστε να εξάγει ακριβή (κατά τεκμήριο) συμπεράσματα όταν τροφοδοτείται από τα δειγματικά στοιχεία μάθησης
- Παράμετροι που αφορούν στη δομή του μοντέλου (π.χ. αριθμός νευρώνων, στρώματα κρυφών νευρώνων) και σε κριτήρια σύγκλισης αναφέρονται σαν υπερπαράμετροι (*hyperparameters*)
- Οι *hyperparameters* επιλέγονται με βάση την εμπειρία του σχεδιαστή ή/και με δοκιμαστικές επαναλήψεις της διαδικασίας μάθησης (*training*) και επικύρωσης (*validation*) για να βελτιωθεί η ακρίβεια (*accuracy*) της τελικής διάρθρωσης συστήματος μηχανικής μάθησης πριν τη φάση *testing*
- Συνήθεις μέθοδοι για αναζήτηση *hypermarameters* μέσω επαναλαμβανόμενων δοκιμών: Exhaustive search, Grid search, random search… ανάλογα με τον επιτρεπτό αριθμό δοκιμών του μοντέλου με ρυθμισμένες παραμέτρους στη διαδικασία μάθησης
- Η επιλογή *hyperparameters* αν δεν υπάρχει διάκριση υποσυνόλου *validation dataset* στο training dataset μπορεί να γίνεται και με επαναληπτικές δοκιμές σε *test datasets*

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εισαγωγικά περί Μηχανικής Μάθησης (6/6)

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

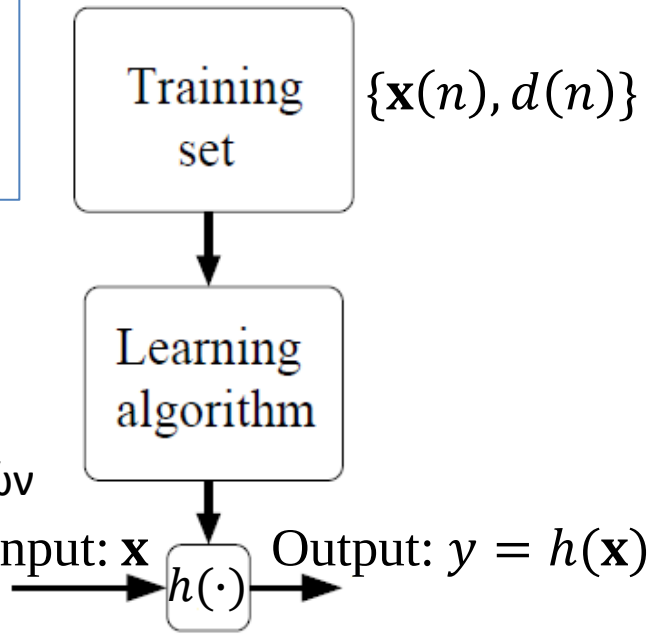
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Γενικό Μοντέλο Επιβλεπόμενης Μάθησης - Supervised Learning

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

- Στόχος του συστήματος είναι η αντιστοίχηση ενός δειγματικού στοιχείου εισόδου (***input sample point***, ***example***, ***instance***) $\mathbf{x} = [x_1\ x_2\ \dots x_m]^{\mathrm{T}}$ σε τιμές εξόδου y που εκτιμούν επιθυμητές τιμές d (***labels***, ***targets***) 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 $\{\mathbf{x}(n), d(n)\}$, $n = 1,2, \dots, N$ γνωστών σε εξωτερικό εκπαιδευτή (***supervisor***)



- Η μορφή και οι παράμετροι της $h(\cdot)$ προσδιορίζονται με αλγόριθμο μάθησης που συγκλίνει σε προσέγγιση του στόχου της υπόθεσης για τα N στοιχεία του δείγματος μάθησης

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

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

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μοντέλα Μηχανικής Μάθησης

Διακριτικά Μοντέλα (**Discriminative Models**):

Μέθοδοι ταξινόμησης (*classification*) ή εκτίμησης (παλινδρόμηση, *regression*) δειγματικών στοιχείων (*data elements*) μέσω υπό συνθήκη πιθανότητας (*conditional density*) εξόδου (*label*) βάσει χαρακτηριστικών (*features*) του, όπως αυτές προσεγγίστηκαν σε στοιχεία δείγματος μάθησης (*training sample*) για γενίκευση σε *test datasets* (*generalization*)

Ενδεικτικές Εφαρμογές:

- *Ταξινόμηση δειγματικών στοιχείων* με βάση συνάρτηση χαρακτηριστικών τους
- *Αναγνώριση προτύπων* με βάση κύρια χαρακτηριστικά τους (*pattern recognition*)
- *Εκτίμηση εξόδου* συμβατή με διαθέσιμα ζεύγη εισόδου - στόχου (*regression*)

Παραγωγικά Μοντέλα (**Generative Models**):

Μέθοδοι εκτίμησης τρόπων παραγωγής (*generation*) δειγματικών στοιχείων, στατιστικά συμβατών με ιδιότητες του δείγματος μάθησης (*training sample*) μέσω συνδυασμένων πιθανοτήτων (*joint probabilities*) εξόδου (*output*) και χαρακτηριστικών (*features*) εισόδου, όπως υπολογίστηκαν στα στοιχεία μάθησης

Ενδεικτικές Εφαρμογές:

- *Δημιουργία προσομοιωμένων στοιχείων*: κειμένων (συμβατών με αποδεκτά μοντέλα Natural Language processing - NLP), εικόνων, κινούμενων σχεδίων, ιδεατών τοπίων…
- *Εμπλουτισμός Μηχανών Αναζήτησης* (*Google*, *MS Bing* + *OpenAI Chat Generative Pre-trained Transformer - ChatGPT*)
- *Επικράτηση αληθοφανών εναλλακτικών εκτιμήσεων* σε συνέργεια με εργαλεία θεωρίας παιγνίων (*Generative Adversarial Networks – GAN*)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

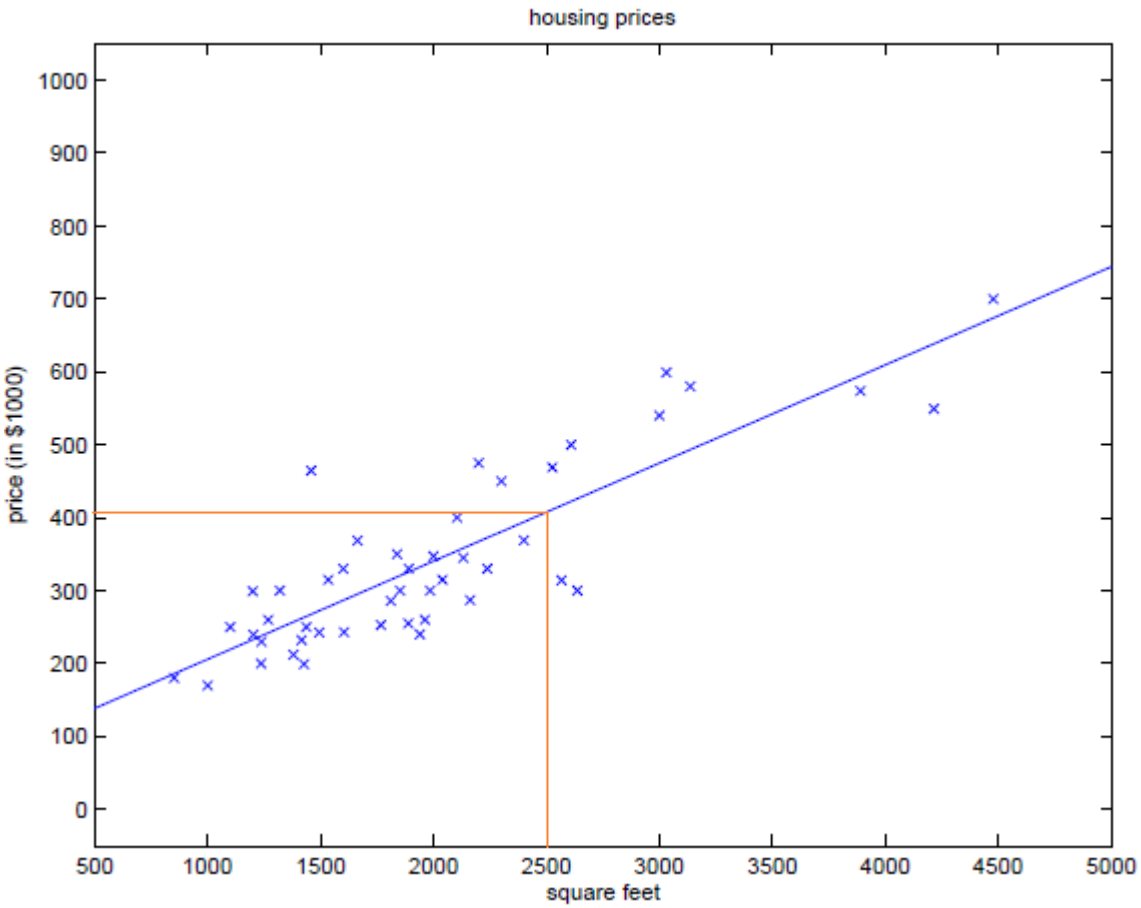
Παράδειγμα Επιβλεπόμενης Μάθησης – Linear Regression

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

Ζητείται γραμμική συνάρτηση $h(x)$ που προσεγγίζει την τιμή κατοικίας $y = h(x)$ με βάση την επιφάνεια x και *labeled* δεδομένα του Δείγματος Μάθησης (***Training Sample***) $\mathcal{D} = \{(x(1), d(1)), \dots, (x(N), d(N))\}$ καταγραμμένων περιπτώσεων κατοικιών στη 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: $y = h(x) = 0.1392\,x + 89.6$

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 $\mathbf{x} = [x_0\ x_1\ \dots x_m]^{\mathrm{T}}$ 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 $\mathbf{w} = [w_0\ w_1\ \dots w_m]^{\mathrm{T}}$ της συνάρτησης $y = h_{\boldsymbol{w}}(\mathbf{x}) = w_0 x_0 + w_1 x_1 + \dots + w_m\, x_m = \mathbf{w}^{\mathbf{T}}\mathbf{x}$ ώστε η y να έχει μικρές ***αποκλίσεις*** για το δείγμα μάθησης (***Training Set***) $\boldsymbol{\mathcal{D}} = \{(\mathbf{x}(1), d(1)), \dots, (\mathbf{x}(N), d(N))\}$
  - $\mathbf{x}(n)$: Διάνυσμα τιμών εισόδου (χαρακτηριστικών) στοιχείου μάθησης n (***regressors***)
  - $d(n)$: Τιμή εξόδου (*label*) στοιχείου μάθησης n (***regressand***)
  - $y(n) = h_{\boldsymbol{w}}(\mathbf{x}(n))$: Εκτίμηση εξόδου του συστήματος για διάνυσμα εισόδου $\mathbf{x}(n)$
  - $\varepsilon(n) = d(n) - y(n)$: Απόκλιση (***error***) εκτίμησης για το $\{\mathbf{x}(n), d(n)\}$, $n = 1,2, \dots, N$
  - Οι $\mathbf{x}(n), d(n), \varepsilon(n)$ 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:

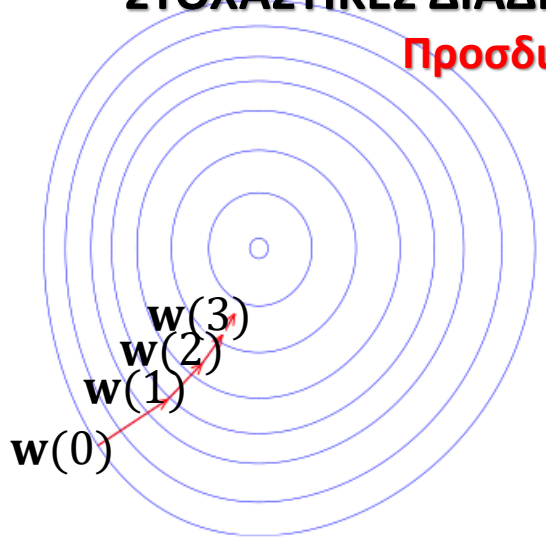
$$J(\mathbf{w}) \triangleq \frac{1}{2}\sum_{n=1}^{N}[\varepsilon(n)]^2 = \frac{1}{2}\sum_{n=1}^{N}[d(n) - h_{\boldsymbol{w}}(\mathbf{x}(n))]^2$$

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

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

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

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 $k \to k+1$ προς την κλίση (Gradient) $\nabla J(\mathbf{w})$ σταθμισμένο κατά την *hyperparameter* α:

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

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

Σημείωση: Για linear regression υπάρχει πάντα σύγκλιση

Κανόνας Μάθησης LMS (Widrow-Hoff)

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

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

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

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

  - 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 $\mathbf{x} = [1\ x]^{\mathrm{T}}$:

$$\mathrm{y} = h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1 x$$

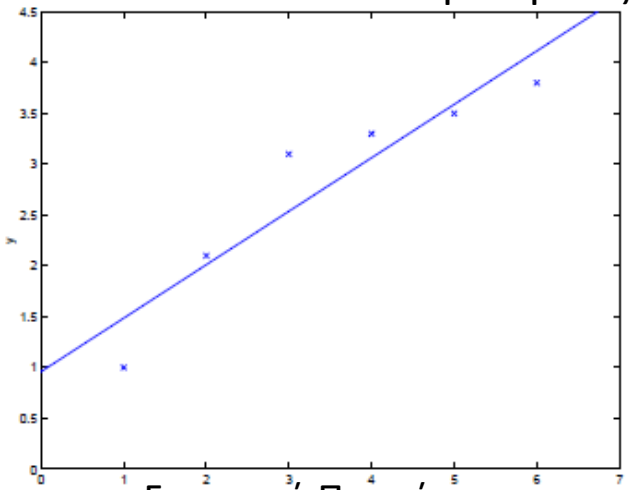
- Προσέγγιση με γραμμικό πολυώνυμο ***2ου βαθμού***:

$$\mathrm{y} = h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1 x + w_2 x^2$$

- Προσέγγιση με γραμμικό πολυώνυμο ***K* βαθμού** (*hyperparameter **K***):

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

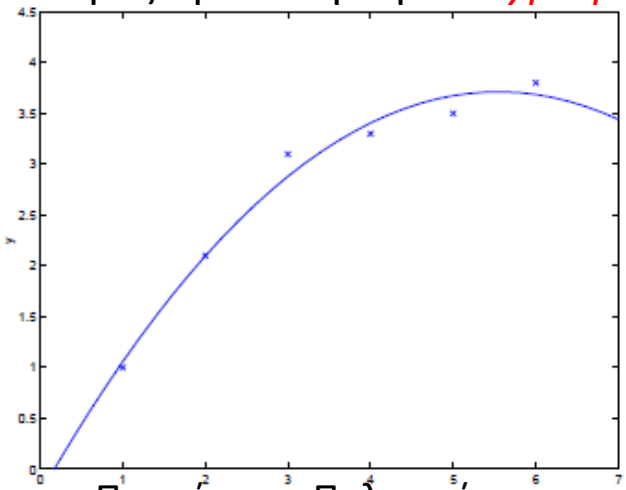
Εμπειρικές δοκιμές προσδιορισμού *hyperparameter **K***



Γραμμική Προσέγγιση

$K = 1$

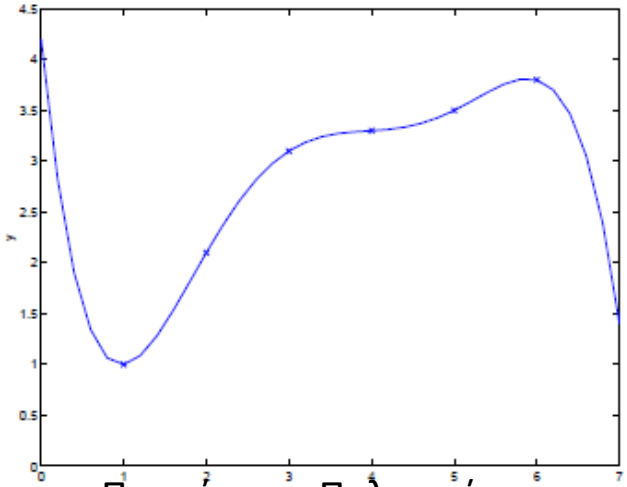
(***Underfitting***)



Προσέγγιση Πολυωνύμου

2ου Βαθμού, $K = 2$

(***OK***)



Προσέγγιση Πολυωνύμου

5ου Βαθμού, $K = 5$

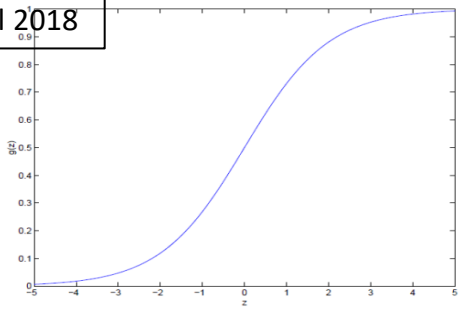
(***Overfitting***)

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

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ταξινόμηση – Classification (1/2)

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



Δειγματικά στοιχεία $\mathbf{x}$ με m διαστάσεις (χαρακτηριστικά, ***features***)
Δυαδικές Κλάσεις Εξόδου (***Classes***, ***Labels***) $y \in \{0,1\}$ ή $y \in \{-,+\}$
Training Set: $\{(\mathbf{x}(1), d(1)), \dots, (\mathbf{x}(N), d(N))\}$

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

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

$$g(z) = \frac{1}{1+e^{-z}}$$
Logistic/Sigmoid Function

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

$$P(y=1|\mathbf{x};\mathbf{w}) = h_{\mathbf{w}}(\mathbf{x}), \qquad P(y=0|\mathbf{x};\mathbf{w}) = 1 - h_{\mathbf{w}}(\mathbf{x})$$
$$\text{ή } p(y|\mathbf{x};\mathbf{w}) = \big(h_{\mathbf{w}}(\mathbf{x})\big)^{y}\big(1-h_{\mathbf{w}}(\mathbf{x})\big)^{1-y}$$

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 $\{\mathbf{x}(n), d(n)\}, n = 1,2,\dots,N$. Θεωρούμε πως οι τιμές εξόδου $d(n)$ είναι ανεξάρτητες δυαδικές τυχαίες μεταβλητές για το δείγμα μάθησης $\mathbf{X} = [\mathbf{x}(1)\ \mathbf{x}(2)\ \dots\ \mathbf{x}(N)]$ και με παραμέτρους $\mathbf{w}$:

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

Ταξινόμηση – Classification (2/2)

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

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

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

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

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

Εφαρμόζουμε ***Gradient Ascent*** στο βήμα $k \to k+1$ με *hyperparameter* α θετική:

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

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 $n = 1,2,\dots,N$ του ***Training Set*** υπολογίζουμε την μερική παράγωγο $\frac{\partial}{\partial w_j} l(\mathbf{w}) = \cdots = \left[d(n) - h_{\mathbf{w}}\left(\mathbf{x}(n)\right)\right]x_j(n)$ ⇨

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

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

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

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

Προκύπτει παρόμοιος ***Επαναληπτικός Κανόνας Μάθησης*** παραμέτρων w_j

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