To Aalto
Λεξικό της Μηχανικής Μάθησης

Alexander Jung[1], Konstantina Olioumtsevits[1], και Juliette Gronier[2]
[1]Aalto University [2]ENS Lyon

Μετάφραση από την Konstantina Olioumtsevits

June 27, 2025

αναφορά ως: A. Jung, K. Olioumtsevits, and J. Gronier, *To Aalto Λεξικό της Μηχανικής Μάθησης* [The Aalto Dictionary of Machine Learning]. Espoo, Finland: Aalto University, 2025.

Ευχαριστίες

Αυτό το λεξικό της μηχανικής μάθησης αναπτύχθηκε κατά τον σχεδιασμό και την υλοποίηση διαφορετικών μαθημάτων, συμπεριλαμβανομένων των CS-E3210 Machine Learning: Basic Principles, CS-C3240 Machine Learning, CS-E4800 Artificial Intelligence, CS-EJ3211 Machine Learning with Python, CS-EJ3311 Deep Learning with Python, CS-E4740 Federated Learning, και CS-E407507 Human-Centered Machine Learning. Αυτά τα μαθήματα προσφέρονται στο Aalto University `https://www.aalto.fi/en`, σε ενήλικους/ες σπουδαστές/σπουδάστριες μέσω του The Finnish Institute of Technology (FITech) `https://fitech.io/en/`, και σε διεθνείς φοιτητές/φοιτήτριες μέσω της European University Alliance Unite! `https://www.aalto.fi/en/unite`.

Είμαστε ευγνώμονες στους/στις σπουδαστές/σπουδάστριες που παρείχαν πολύτιμα σχόλια που ήταν καθοριστικά για το συγκεκριμένο λεξικό. Ιδιαίτερες ευχαριστίες στον Mikko Seesto για τη σχολαστική του διόρθωση προσχεδίων. Κάποια από τα σχήματα στο λεξικό έχουν προετοιμαστεί με τη βοήθεια του Salvatore Rastelli.

Η μετάφραση στα ελληνικά βασίζεται ιδιαίτερα σε σχετικά σχολικά βιβλία λυκείου `https://ebooks.edu.gr/ebooks`, σε αρχεία από την Εθνική Υπηρεσία Πληροφοριών της Ελλάδας `https://www.nis.gr/en`, και σε σχετικά λεξικά:

Γ. Γεωργίου, *Αγγλοελληνικό Λεξικό Μαθηματικής Ορολογίας*, 1999. [Διαδικτυακά]. Διαθέσιμο: `https://www.mas.ucy.ac.cy/georgios/bookfiles/dict1.pdf`. Πρόσβαση: 30 Μαΐου 2025.

Α. Καλογεροπούλου, Μ. Γκίκας, Δ. Καραγιαννάκης, και Μ. Λάμπρου, *Αγγλοελληνικό Λεξικό Μαθηματικών Όρων*. Αθήνα, Ελλάδα: Τροχαλία, 1992.

Σ. Καπιδάκης, Κ. Τοράκη, Σ. Χατζημαρή, Κ. Βαλεοντής, και Υ. Κύττα, *Λε-*

ξικό Επιστήμης της Πληροφόρησης. Αθήνα, Ελλάδα: Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις, 2024.

Κατάλογοι Συμβόλων

Σύνολα και Συναρτήσεις

$a \in \mathcal{A}$	Το αντικείμενο a είναι ένα στοιχείο του συνόλου $\mathcal{A}$.
$a := b$	Χρησιμοποιούμε το a ως συντομογραφία για το b.
$\lvert\mathcal{A}\rvert$	Η καρδινικότητα (δηλαδή ο αριθμός των στοιχείων) ενός πεπερασμένου συνόλου $\mathcal{A}$.
$\mathcal{A} \subseteq \mathcal{B}$	Το $\mathcal{A}$ είναι ένα υποσύνολο του $\mathcal{B}$.
$\mathcal{A} \subset \mathcal{B}$	Το $\mathcal{A}$ είναι ένα αυστηρό υποσύνολο του $\mathcal{B}$.
$\mathbb{N}$	Οι φυσικοί αριθμοί $1, 2, \ldots$.
$\mathbb{R}$	Οι πραγματικοί αριθμοί x [1].
$\mathbb{R}_+$	Οι μη αρνητικοί πραγματικοί αριθμοί $x \geq 0$.
$\mathbb{R}_{++}$	Οι θετικοί πραγματικοί αριθμοί $x > 0$.
$\{0, 1\}$	Το σύνολο που αποτελείται από τους δύο πραγματικούς αριθμούς 0 και 1.
$[0, 1]$	Το κλειστό διάστημα των πραγματικών αριθμών x με $0 \leq x \leq 1$.

$\underset{\mathbf{w}}{\arg\min} f(\mathbf{w})$	Το σύνολο των ελαχιστοποιητών για μια συνάρτηση πραγματικής τιμής $f(\mathbf{w})$. Βλέπε επίσης: συνάρτηση.
$\mathbb{S}^{(n)}$	Το σύνολο των διανυσμάτων μοναδιαίας νόρμας στο $\mathbb{R}^{n+1}$. Βλέπε επίσης: νόρμα.
$\exp(a)$	Η εκθετική του πραγματικού αριθμού $a \in \mathbb{R}$.
$\log a$	Ο λογάριθμος του θετικού αριθμού $a \in \mathbb{R}_{++}$.
$h(\cdot): \mathcal{A} \to \mathcal{B} : a \mapsto h(a)$	Μία συνάρτηση (δηλαδή map) που δέχεται οποιοδήποτε στοιχείο $a \in \mathcal{A}$ από ένα σύνολο $\mathcal{A}$ ως είσοδο και δίνει ένα καλά ορισμένο στοιχείο $h(a) \in \mathcal{B}$ ενός συνόλου $\mathcal{B}$. Το σύνολο $\mathcal{A}$ είναι το πεδίο της συνάρτησης h και το σύνολο $\mathcal{B}$ είναι το πεδίο τιμών της h. Ο στόχος της μηχανικής μάθησης είναι να βρει (ή να μάθει) μία συνάρτηση h (δηλαδή μία υπόθεση) που διαβάζει τα χαρακτηριστικά $\mathbf{x}$ ενός σημείου δεδομένων και δίνει μία πρόβλεψη $h(\mathbf{x})$ για την ετικέτα y του. Βλέπε επίσης: συνάρτηση, map, ml, υπόθεση, χαρακτηριστικό, σημείο δεδομένων, πρόβλεψη, ετικέτα.
$\mathrm{epi}(f)$	Το επίγραμμα μίας συνάρτησης πραγματικής τιμής $f : \mathbb{R}^d \to \mathbb{R}$. Βλέπε επίσης: epigraph, συνάρτηση.

$\dfrac{\partial f(w_1, \ldots, w_d)}{\partial w_j}$	Η μερική παράγωγος (αν υπάρχει) μίας συνάρτησης πραγματικής τιμής $f : \mathbb{R}^d \to \mathbb{R}$ αναφορικά με το w_j [2, Κεφ. 9]. Βλέπε επίσης: συνάρτηση.
$\nabla f(\mathbf{w})$	Η κλίση μίας παραγωγίσιμης συνάρτησης πραγματικής τιμής $f : \mathbb{R}^d \to \mathbb{R}$ είναι το διάνυσμα $\nabla f(\mathbf{w}) = \big(\partial f/\partial w_1, \ldots, \partial f/\partial w_d\big)^T \in \mathbb{R}^d$ [2, Κεφ. 9]. Βλέπε επίσης: κλίση, παραγωγίσιμη, συνάρτηση.

Πίνακες και Διανύσματα

$\mathbf{x} = (x_1, \ldots, x_d)^T$	Ένα διάνυσμα μήκους d, με την j-στή του καταχώριση να είναι x_j.
$\mathbb{R}^d$	Το σύνολο των διανυσμάτων $\mathbf{x} = \big(x_1, \ldots, x_d\big)^T$ που αποτελούνται από d καταχωρίσεις πραγματικών τιμών $x_1, \ldots, x_d \in \mathbb{R}$.
$\mathbf{I}_{l \times d}$	Ένας γενικευμένος πίνακας ταυτότητας με l σειρές και d στήλες. Οι καταχωρίσεις του $\mathbf{I}_{l \times d} \in \mathbb{R}^{l \times d}$ είναι ίσες με 1 κατά μήκος της κύριας διαγωνίου και ίσες με 0 διαφορετικά.
$\mathbf{I}_d, \mathbf{I}$	Ένας τετραγωνικός πίνακας ταυτότητας μεγέθους $d \times d$. Αν το μέγεθος είναι προφανές από τα συμφραζόμενα, παραλείπουμε τον δείκτη.
$\lVert\mathbf{x}\rVert_2$	Η Ευκλείδειος (ή ℓ_2) νόρμα του διανύσματος $\mathbf{x} = \big(x_1, \ldots, x_d\big)^T \in \mathbb{R}^d$ ορίζεται ως $\lVert\mathbf{x}\rVert_2 := \sqrt{\sum_{j=1}^{d} x_j^2}$. Βλέπε επίσης: νόρμα.
$\lVert\mathbf{x}\rVert$	Κάποια νόρμα του διανύσματος $\mathbf{x} \in \mathbb{R}^d$ [3]. Εκτός αν προσδιορίζεται διαφορετικά, εννοούμε την Ευκλείδεια νόρμα $\lVert\mathbf{x}\rVert_2$. Βλέπε επίσης: νόρμα.
$\mathbf{x}^T$	Ο ανάστροφος πίνακας που έχει το διάνυσμα $\mathbf{x} \in \mathbb{R}^d$ ως μοναδική του στήλη.

$\mathbf{X}^T$	Ο ανάστροφος πίνακας $\mathbf{X} \in \mathbb{R}^{m \times d}$. Ένας τετραγωνικός πίνακας παραγματικών τιμών $\mathbf{X} \in \mathbb{R}^{m \times m}$ λέγεται συμμετρικός αν $\mathbf{X} = \mathbf{X}^T$.
$\mathbf{X}^{-1}$	Ο αντίστροφος πίνακας ενός πίνακα $\mathbf{X} \in \mathbb{R}^{d \times d}$. Βλέπε επίσης: αντίστροφος πίνακας.
$\mathbf{0} = \big(0, \ldots, 0\big)^T$	Το διάνυσμα στο $\mathbb{R}^d$ με κάθε καταχώριση να είναι ίση με μηδέν.
$\mathbf{1} = \big(1, \ldots, 1\big)^T$	Το διάνυσμα στο $\mathbb{R}^d$ με κάθε καταχώριση να είναι ίση με ένα.
$\big(\mathbf{v}^T, \mathbf{w}^T\big)^T$	Το διάνυσμα μήκους $d + d'$ που προκύπτει από την αλληλουχία των καταχωρίσεων του διανύσματος $\mathbf{v} \in \mathbb{R}^d$ με τις καταχωρίσεις του $\mathbf{w} \in \mathbb{R}^{d'}$.
$\mathrm{span}\{\mathbf{B}\}$	Το εύρος ενός πίνακα $\mathbf{B} \in \mathbb{R}^{a \times b}$, που είναι ο υποχώρος όλων των γραμμικών συνδυασμών των στηλών του $\mathbf{B}$, έτσι ώστε $\mathrm{span}\{\mathbf{B}\} = \big\{\mathbf{Ba} : \mathbf{a} \in \mathbb{R}^b\big\} \subseteq \mathbb{R}^a$.
$\det(\mathbf{C})$	Η ορίζουσα του πίνακα $\mathbf{C}$. Βλέπε επίσης: ορίζουσα.
$\mathbf{A} \otimes \mathbf{B}$	Το γινόμενο Kronecker των $\mathbf{A}$ και $\mathbf{B}$ [4].

Θεωρία Πιθανοτήτων

$\mathbf{x} \sim p(\mathbf{z})$	Η τυχαία μεταβλητή $\mathbf{x}$ κατανέμεται σύμφωνα με την κατανομή πιθανότητας $p(\mathbf{z})$ [5], [6]. Βλέπε επίσης: τυχαία μεταβλητή, κατανομή πιθανότητας.
$\mathbb{E}_p\{f(\mathbf{z})\}$	Η προσδοκία μίας τυχαίας μεταβλητής $f(\mathbf{z})$ που προκύπτει από την εφαρμογή μίας ντετερμινιστικής συνάρτησης f σε μία τυχαία μεταβλητή $\mathbf{z}$ της οποίας η κατανομή πιθανότητας είναι $p(\mathbf{z})$. Αν η κατανομή πιθανότητας είναι προφανής από τα συμφραζόμενα, γράφουμε απλώς $\mathbb{E}\{f(\mathbf{z})\}$. Βλέπε επίσης: προσδοκία, τυχαία μεταβλητή, συνάρτηση, κατανομή πιθανότητας.
$p(\mathbf{x}, y)$	Μία (από κοινού) κατανομή πιθανότητας μίας τυχαίας μεταβλητής της οποίας οι πραγματώσεις είναι σημεία δεδομένων με χαρακτηριστικά $\mathbf{x}$ και ετικέτα y. Βλέπε επίσης: κατανομή πιθανότητας, τυχαία μεταβλητή, πραγμάτωση, data point, feature, ετικέτα.
$p(\mathbf{x}\|y)$	Μία κατανομή πιθανότητας υπό συνθήκη μίας τυχαίας μεταβλητής $\mathbf{x}$ δεδομένης της τιμής μίας άλλης τυχαίας μεταβλητής y [7, Sec. 3.5]. Βλέπε επίσης: κατανομή πιθανότητας, τυχαία μεταβλητή.

$p(\mathbf{x};\mathbf{w})$	Μία παραμετροποιημένη κατανομή πιθανότητας μίας τυχαίας μεταβλητής $\mathbf{x}$. Η κατανομή πιθανότητας εξαρτάται από ένα παραμετρικό διάνυσμα $\mathbf{w}$. Για παράδειγμα, $p(\mathbf{x};\mathbf{w})$ θα μπορούσε να είναι μία πολυμεταβλητή κανονική κατανομή με το παραμετρικό διάνυσμα $\mathbf{w}$ που δίνεται από τις καταχωρίσεις του διανύσματος μέσης τιμής $\mathbb{E}\{\mathbf{x}\}$ και τον πίνακα συνδιακύμανσης $\mathbb{E}\Big\{\big(\mathbf{x}-\mathbb{E}\{\mathbf{x}\}\big)\big(\mathbf{x}-\mathbb{E}\{\mathbf{x}\}\big)^T\Big\}$. Βλέπε επίσης: κατανομή πιθανότητας, τυχαία μεταβλητή, παράμετρος, πολυμεταβλητή κανονική κατανομή, μέση τιμή, πίνακας συνδιακύμανσης.
$\mathcal{N}(\mu,\sigma^2)$	Η κατανομή πιθανότητας μίας Gaussian τυχαίας μεταβλητής $x \in \mathbb{R}$ με μέση τιμή (ή προσδοκία) $\mu = \mathbb{E}\{x\}$ και διακύμανση $\sigma^2 = \mathbb{E}\big\{(x-\mu)^2\big\}$. Βλέπε επίσης: κατανομή πιθανότητας, Gaussian random variable (Gaussian RV), μέση τιμή, expectation, διακύμανση.
$\mathcal{N}(\boldsymbol{\mu},\mathbf{C})$	Η πολυμεταβλητή κανονική κατανομή μίας Gaussian τυχαίας μεταβλητής τιμής διανύσματος $\mathbf{x} \in \mathbb{R}^d$ με μέση τιμή (ή προσδοκία) $\boldsymbol{\mu} = \mathbb{E}\{\mathbf{x}\}$ και πίνακα συνδιακύμανσης $\mathbf{C} = \mathbb{E}\big\{\big(\mathbf{x}-\boldsymbol{\mu}\big)\big(\mathbf{x}-\boldsymbol{\mu}\big)^T\big\}$. Βλέπε επίσης: πολυμεταβλητή κανονική κατανομή, Gaussian RV, μέση τιμή, expectation, πίνακας συνδιακύμανσης.

Μηχανική Μάθηση

r	Ένας δείκτης $r = 1, 2, \ldots$ που απαριθμεί τα σημεία δεδομένων. Βλέπε επίσης: data point.
m	Ο αριθμός των σημείων δεδομένων σε ένα σύνολο δεδομένων (δηλαδή το μέγεθός του). Βλέπε επίσης: data point, σύνολο δεδομένων.
$\mathcal{D}$	Ένα σύνολο δεδομένων $\mathcal{D} = \{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}\}$ είναι μία λίστα μεμονωμένων σημείων δεδομένων $\mathbf{z}^{(r)}$, for $r = 1, \ldots, m$. Βλέπε επίσης: σύνολο δεδομένων, data point.
d	Ο αριθμός των χαρακτηριστικών που χαρακτηρίζουν ένα σημείο δεδομένων. Βλέπε επίσης: feature, data point.
x_j	Το j-στό χαρακτηριστικό ενός σημείου δεδομένων. Το πρώτο χαρακτηριστικό δηλώνεται x_1, το δεύτερο χαρακτηριστικό x_2, και ούτω καθεξής. Βλέπε επίσης: data point, feature.
$\mathbf{x}$	Το διάνυσμα χαρακτηριστικών $\mathbf{x} = \big(x_1, \ldots, x_d\big)^T$ ενός σημείου δεδομένων του οποίου οι καταχωρίσεις είναι τα μεμονωμένα χαρακτηριστικά ενός σημείου δεδομένων. Βλέπε επίσης: διάνυσμα χαρακτηριστικών, data point, feature.

$\mathcal{X}$	Ο χώρος χαρακτηριστικών $\mathcal{X}$ είναι το σύνολο όλων των πιθανών τιμών που μπορούν να πάρουν τα χαρακτηριστικά $\mathbf{x}$ ενός σημείου δεδομένων. Βλέπε επίσης: χώρος χαρακτηριστικών, feature, data point.
$\mathbf{z}$	Αντί του συμβόλου $\mathbf{x}$, χρησιμοποιούμε μερικές φορές $\mathbf{z}$ ως ένα άλλο σύμβολο για να δηλώσουμε ένα διάνυσμα του οποίου οι καταχωρίσεις είναι τα μεμονωμένα χαρακτηριστικά ενός σημείου δεδομένων. Χρειαζόμαστε δύο διαφορετικά σύμβολα για να διακρίνουμε τα ακατέργαστα χαρακτηριστικά από αυτά που έχουν μαθευτεί [8, Κεφ. 9]. Βλέπε επίσης: feature, data point.
$\mathbf{x}^{(r)}$	Το διάνυσμα χαρακτηριστικών του r-στού σημείου δεδομένων εντός ενός συνόλου δεδομένων. Βλέπε επίσης: feature, data point, σύνολο δεδομένων.
$x_j^{(r)}$	Το j-στό χαρακτηριστικό του r-στού σημείου δεδομένων εντός ενός συνόλου δεδομένων. Βλέπε επίσης: feature, data point, σύνολο δεδομένων.
$\mathcal{B}$	Μία μικρο-δέσμη (ή υποσύνολο) τυχαία επιλεγμένων σημείων δεδομένων. Βλέπε επίσης: δέσμη, data point.
B	Το μέγεθος μίας μικρο-δέσμης (δηλαδή ο αριθμός των σημείων δεδομένων σε αυτή). Βλέπε επίσης: data point, δέσμη.

y	Η ετικέτα (ή η ποσότητα ενδιαφέροντος) ενός σημείου δεδομένων. Βλέπε επίσης: ετικέτα, data point.
$y^{(r)}$	Η ετικέτα του r-στού σημείου δεδομένων. Βλέπε επίσης: ετικέτα, data point.
$\left(\mathbf{x}^{(r)}, y^{(r)}\right)$	Τα χαρακτηριστικά και η ετικέτα του r-στού σημείου δεδομένων. Βλέπε επίσης: feature, ετικέτα, data point.
$\mathcal{Y}$	Ο χώρος ετικετών $\mathcal{Y}$ μίας μεθόδου μηχανικής μάθησης αποτελείται από όλες τις πιθανές τιμές ετικετών που ένα σημείο δεδομένων μπορεί να φέρει. Ο ονομαστικός χώρος ετικετών μπορεί να είναι μεγαλύτερος από το σύνολο των διαφορετικών τιμών ετικετών που προκύπτουν σε ένα συγκεκριμένο σύνολο δεδομένων (π.χ. ένα σύνολο εκπαίδευσης). Προβλήματα (ή μέθοδοι) μηχανικής μάθησης που χρησιμοποιούν έναν αριθμητικό χώρο ετικετών, όπως $\mathcal{Y} = \mathbb{R}$ ή $\mathcal{Y} = \mathbb{R}^3$, αναφέρονται ως προβλήματα (ή μέθοδοι) παλινδρόμησης. Προβλήματα (ή μέθοδοι) μηχανικής μάθησης που χρησιμοποιούν έναν διακριτό χώρο ετικετών, όπως $\mathcal{Y} = \{0, 1\}$ ή $\mathcal{Y} = \{\textit{γάτα}, \textit{σκύλος}, \textit{ποντίκι}\}$, αναφέρονται ως προβλήματα (ή μέθοδοι) ταξινόμησης. Βλέπε επίσης: χώρος ετικετών, ml, ετικέτα, data point, σύνολο δεδομένων, σύνολο εκπαίδευσης, regression, ταξινόμηση.

η	Ο ρυθμός μάθησης (ή το μέγεθος βήματος) που χρησιμοποιείται από τις μεθόδους με βάση την κλίση. Βλέπε επίσης: ρυθμός μάθησης, μέγεθος βήματος, μέθοδοι με βάση την κλίση.
$h(\cdot)$	Μία map υπόθεσης που διαβάζει τα χαρακτηριστικά $\mathbf{x}$ ενός σημείου δεδομένων και παραδίδει μία πρόβλεψη $\hat{y} = h(\mathbf{x})$ για την ετικέτα y του σημείου δεδομένων. Βλέπε επίσης: υπόθεση, map, feature, data point, πρόβλεψη, ετικέτα.
$\mathcal{Y}^{\mathcal{X}}$	Δεδομένων δύο συνόλων $\mathcal{X}$ και $\mathcal{Y}$, δηλώνουμε ως $\mathcal{Y}^{\mathcal{X}}$ το σύνολο όλων των πιθανών map υπόθεσης $h : \mathcal{X} \to \mathcal{Y}$. Βλέπε επίσης: υπόθεση, map.
$\mathcal{H}$	Ένας χώρος υποθέσεων ή μοντέλο που χρησιμοποιείται από μία μέθοδο μηχανικής μάθησης. Ο χώρος υποθέσεων αποτελείται από διαφορετικές map υπόθεσης $h : \mathcal{X} \to \mathcal{Y}$, μεταξύ των οποίων η μέθοδος μηχανικής μάθησης πρέπει να επιλέξει. Βλέπε επίσης: χώρος υποθέσεων, μοντέλο, ml, υπόθεση, map.
$d_{\mathrm{eff}}(\mathcal{H})$	Η αποτελεσματική διάσταση ενός χώρου υποθέσεων $\mathcal{H}$. Βλέπε επίσης: αποτελεσματική διάσταση, χώρος υποθέσεων.

B^2	Η τετραγωνική μεροληψία μίας υπόθεσης $\hat{h}$ που έχει μαθευτεί, ή των παραμέτρων της. Σημείωση ότι η $\hat{h}$ γίνεται μία τυχαία μεταβλητή αν μαθαίνεται από σημεία δεδομένων που είναι τυχαίες μεταβλητές. Βλέπε επίσης: μεροληψία, υπόθεση, παράμετρος, τυχαία μεταβλητή, data point.
V	Η διακύμανση μίας υπόθεσης $\hat{h}$ που έχει μαθευτεί, ή των παραμέτρων της. Σημείωση ότι η $\hat{h}$ γίνεται μία τυχαία μεταβλητή αν μαθαίνεται από σημεία δεδομένων που είναι τυχαίες μεταβλητές. Βλέπε επίσης: διακύμανση, υπόθεση, παράμετρος, τυχαία μεταβλητή, data point.
$L\left((\mathbf{x}, y), h\right)$	Η απώλεια που προκαλείται από την πρόβλεψη της ετικέτας y ενός σημείου δεδομένων χρησιμοποιώντας την πρόβλεψη $\hat{y} = h(\mathbf{x})$. Η πρόβλεψη $\hat{y}$ προκύπτει από την αξιολόγηση της υπόθεσης $h \in \mathcal{H}$ για το διάνυσμα χαρακτηριστικών $\mathbf{x}$ του σημείου δεδομένων. Βλέπε επίσης: απώλεια, ετικέτα, data point, πρόβλεψη, υπόθεση, διάνυσμα χαρακτηριστικών.
E_v	Το σφάλμα επικύρωσης μίας υπόθεσης h, που είναι η μέση της απώλεια που προκαλείται σε ένα σύνολο επικύρωσης. Βλέπε επίσης: σφάλμα επικύρωσης, υπόθεση, loss, σύνολο επικύρωσης.

$\widehat{L}(h\|\mathcal{D})$	Η εμπειρική διακινδύνευση ή η μέση απώλεια που προκαλείται από την υπόθεση h σε ένα σύνολο δεδομένων $\mathcal{D}$. Βλέπε επίσης: εμπειρική διακινδύνευση, loss, υπόθεση, σύνολο δεδομένων.
E_t	Το σφάλμα εκπαίδευσης μίας υπόθεσης h, που είναι η μέση της απώλεια που προκαλείται σε ένα σύνολο εκπαίδευσης. Βλέπε επίσης: training error, υπόθεση, loss, σύνολο εκπαίδευσης.
t	Ένας δείκτης διακριτού χρόνου $t = 0, 1, \ldots$ που χρησιμοποιείται για την απαρίθμηση ακολουθιακών γεγονότων (ή χρονικών στιγμών).
t	Ένας δείκτης που απαριθμεί τις εργασίες μάθησης εντός ενός προβλήματος μάθησης πολυδιεργασίας. Βλέπε επίσης: εργασία μάθησης, μάθηση πολυδιεργασίας.
α	Μία παράμετρος ομαλοποίησης που ελέγχει το ποσό της ομαλοποίησης. Βλέπε επίσης: παράμετρος, ομαλοποίηση.
$\lambda_j(\mathbf{Q})$	Η j-στή ιδιοτιμή (ταξινομημένη σε αύξουσα ή φθίνουσα σειρά) ενός θετικά ημιορισμένου πίνακα $\mathbf{Q}$. Χρησιμοποιούμε επίσης τη συντομογραφία λ_j αν ο αντίστοιχος πίνακας είναι προφανής από τα συμφραζόμενα. Βλέπε επίσης: ιδιοτιμή, θετικά ημιορισμένος.

$\sigma(\cdot)$	Η συνάρτηση ενεργοποίησης που χρησιμοποιείται από έναν τεχνητό νευρώνα εντός ενός τεχνητού νευρωνικού δικτύου. Βλέπε επίσης: συνάρτηση ενεργοποίησης, τεχνητό νευρωνικό δίκτυο.
$\mathcal{R}_{\hat{y}}$	Μία περιοχή αποφάσεων εντός ενός χώρου χαρακτηριστικών. Βλέπε επίσης: περιοχή αποφάσεων, χώρος χαρακτηριστικών.
$\mathbf{w}$	Ένα παραμετρικό διάνυσμα $\mathbf{w} = \big(w_1, \ldots, w_d\big)^T$ ενός μοντέλου, π.χ. τα βάρη ενός γραμμικού μοντέλου ή σε ένα τεχνητό νευρωνικό δίκτυο. Βλέπε επίσης: παράμετρος, model, βάρη, γραμμικό μοντέλο, ΤΝΔ.
$h^{(\mathbf{w})}(\cdot)$	Μία map υπόθεσης που περιλαμβάνει παραμέτρους μοντέλου $w_1, \ldots, w_d$ που μπορούν να ρυθμιστούν στοιβαγμένες στο διάνυσμα $\mathbf{w} = \big(w_1, \ldots, w_d\big)^T$. Βλέπε επίσης: υπόθεση, map, παράμετροι μοντέλου.
$\phi(\cdot)$	Μία χάρτης χαρακτηριστικών $\phi : \mathcal{X} \rightarrow \mathcal{X}' : \mathbf{x} \mapsto \mathbf{x}' := \phi\big(\mathbf{x}\big) \in \mathcal{X}'$. Βλέπε επίσης: χάρτης χαρακτηριστικών.
$K\big(\cdot, \cdot\big)$	Δεδομένου κάποιου χώρου χαρακτηριστικών $\mathcal{X}$, ένας πυρήνας είναι μία map $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{C}$ που είναι θετικά ημιορισμένη. Βλέπε επίσης: χώρος χαρακτηριστικών, πυρήνας, map, θετικά ημιορισμένος.

Ομοσπονδιακή Μάθηση

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	Ένας μη κατευθυνόμενος γράφος του οποίου οι κόμβοι $i \in \mathcal{V}$ αντιπροσωπεύουν συσκευές εντός ενός δικτύου ομοσπονδιακής μάθησης. Οι μη κατευθυνόμενες σταθμισμένες ακμές $\mathcal{E}$ αντιπροσωπεύουν τη συνεκτικότητα μεταξύ συσκευών και τις στατιστικές ομοιότητες μεταξύ των συνόλων δεδομένων τους και των εργασιών μάθησης. Βλέπε επίσης: γράφος, συσκευή, δίκτυο ομοσπονδιακής μάθησης, σύνολο δεδομένων, εργασία μάθησης.
$i \in \mathcal{V}$	Ένας κόμβος που αντιπροσωπεύει κάποια συσκευή εντός ενός δικτύου ομοσπονδιακής μάθησης. Η συσκευή μπορεί να έχει πρόσβαση σε ένα τοπικό σύνολο δεδομένων και να εκπαιδεύσει ένα τοπικό μοντέλο. Βλέπε επίσης: συσκευή, δίκτυο ομοσπονδιακής μάθησης, τοπικό σύνολο δεδομένων, local model.
$\mathcal{G}^{(\mathcal{C})}$	Ο επαγόμενος υπογράφος του $\mathcal{G}$ χρησιμοποιώντας τους κόμβους στο $\mathcal{C} \subseteq \mathcal{V}$.
$\mathbf{L}^{(\mathcal{G})}$	Ο πίνακας Laplace ενός γράφου $\mathcal{G}$. Βλέπε επίσης: πίνακας Laplace, graph.
$\mathbf{L}^{(\mathcal{C})}$	Ο πίνακας Laplace του επαγόμενου γράφου $\mathcal{G}^{(\mathcal{C})}$. Βλέπε επίσης: πίνακας Laplace, graph.

$\mathcal{N}^{(i)}$	Η γειτονιά ενός κόμβου i σε έναν γράφο $\mathcal{G}$. Βλέπε επίσης: neighborhood, graph.
$d^{(i)}$	Ο σταθμισμένος βαθμός $d^{(i)} := \sum_{i' \in \mathcal{N}^{(i)}} A_{i,i'}$ ενός κόμβου i σε έναν γράφο $\mathcal{G}$. Βλέπε επίσης: graph.
$d_{\max}^{(\mathcal{G})}$	Ο μέγιστος σταθμισμένος βαθμός κόμβου ενός γράφου $\mathcal{G}$. Βλέπε επίσης: μέγιστο, βαθμός κόμβου, graph.
$\mathcal{D}^{(i)}$	Το τοπικό σύνολο δεδομένων $\mathcal{D}^{(i)}$ που φέρει ο κόμβος $i \in \mathcal{V}$ ενός δικτύου ομοσπονδιακής μάθησης. Βλέπε επίσης: τοπικό σύνολο δεδομένων, δίκτυο ομοσπονδιακής μάθησης.
m_i	Ο αριθμός των σημείων δεδομένων (δηλαδή το μέγεθος δείγματος) που περιέχονται στο τοπικό σύνολο δεδομένων $\mathcal{D}^{(i)}$ στον κόμβο $i \in \mathcal{V}$. Βλέπε επίσης: data point, μέγεθος δείγματος, τοπικό σύνολο δεδομένων.
$\mathbf{x}^{(i,r)}$	Τα χαρακτηριστικά του r-στού σημείου δεδομένων στο τοπικό σύνολο δεδομένων $\mathcal{D}^{(i)}$. Βλέπε επίσης: feature, data point, τοπικό σύνολο δεδομένων.

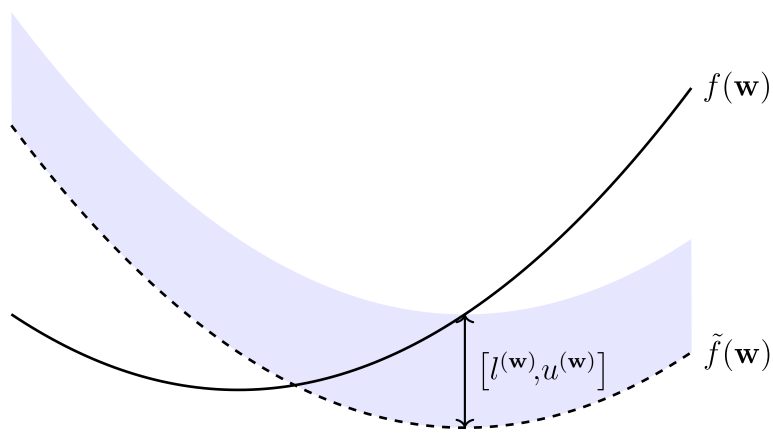
$y^{(i,r)}$	Η ετικέτα του r-στού σημείου δεδομένων στο τοπικό σύνολο δεδομένων $\mathcal{D}^{(i)}$. Βλέπε επίσης: ετικέτα, data point, τοπικό σύνολο δεδομένων.
$\mathbf{w}^{(i)}$	Οι τοπικοί παράμετροι μοντέλου της συσκευής i εντός ενός δικτύου ομοσπονδιακής μάθησης. Βλέπε επίσης: παράμετροι μοντέλου, συσκευή, δίκτυο ομοσπονδιακής μάθησης.
$L_i\left(\mathbf{w}\right)$	Η τοπική συνάρτηση απώλειας που χρησιμοποιείται από την συσκευή i για να μετρήσει τη χρησιμότητα κάποιας επιλογής $\mathbf{w}$ για τις παραμέτρους μοντέλου. Βλέπε επίσης: συνάρτηση απώλειας, συσκευή, παράμετροι μοντέλου.
$L^{(\mathrm{d})}\left(\mathbf{x}, h\big(\mathbf{x}\big), h'\big(\mathbf{x}\big)\right)$	Η απώλεια που προκαλείται από μία υπόθεση h' σε ένα σημείο δεδομένων με χαρακτηριστικά $\mathbf{x}$ και ετικέτα $h\big(\mathbf{x}\big)$ που προκύπτει από μία άλλη υπόθεση. Βλέπε επίσης: loss, υπόθεση, data point, feature, ετικέτα.
$\mathrm{stack}\big\{\mathbf{w}^{(i)}\big\}_{i=1}^{n}$	Το διάνυσμα $\left(\big(\mathbf{w}^{(1)}\big)^T, \ldots, \big(\mathbf{w}^{(n)}\big)^T\right)^T \in \mathbb{R}^{dn}$ που προκύπτει από την κάθετη στοίβαξη των τοπικών παραμέτρων μοντέλου $\mathbf{w}^{(i)} \in \mathbb{R}^d$. Βλέπε επίσης: παράμετροι μοντέλου.

Έννοιες Μηχανικής Μάθησης

αβεβαιότητα Η αβεβαιότητα αναφέρεται στον βαθμό εμπιστοσύνης—ή στην έλλειψη αυτής—που σχετίζεται με μία ποσότητα όπως μία πρόβλεψη μοντέλου, μία εκτίμηση παραμέτρου, ή ένα παρατηρούμενο σημείο δεδομένων. Στη μηχανική μάθηση, αβεβαιότητα ανακύπτει από διάφορες πηγές, συμπεριλαμβανομένων θορυβωδών δεδομένων, περιορισμένων δειγμάτων εκπαίδευσης, ή αμφισημίας στις υποθέσεις μοντέλου. Η θεωρία πιθανοτήτων προσφέρει ένα ηθικό πλαίσιο για την αναπαράσταση και την ποσοτικοποίηση τέτοιας αβεβαιότητας.

Βλέπε επίσης: model, πρόβλεψη, παράμετρος, data point, ml, δεδομένα, δείγμα, probability.

αισιοδοξία παρά την αβεβαιότητα Οι μέθοδοι μηχανικής μάθησης μαθαίνουν παραμέτρους μοντέλου $\mathbf{w}$ σύμφωνα με κάποιο κριτήριο επίδοσης $\bar{f}(\mathbf{w})$. Ωστόσο, δεν μπορούν να έχουν άμεση στο $\bar{f}(\mathbf{w})$ αλλά βασίζονται σε μία εκτίμηση (ή προσέγγιση) $f(\mathbf{w})$ του $\bar{f}(\mathbf{w})$. Ως ένα χαρακτηριστικό παράδειγμα, οι μέθοδοι βασιμένες στην ελαχιστοποίηση εμπειρικής διακινδύνευσης χρησιμοποιούν τη μέση απώλεια σε ένα συγκεκριμένο σύνολο δεδομένων (δηλαδή το σύνολο εκπαίδευσης) ως μία εκτίμηση για τη διακινδύνευση μίας υπόθεσης. Χρησιμοποιώντας ένα πιθανοτικό μοντέλο, μπορεί κανείς να κατασκευάσει ένα διάστημα εμπιστοσύνης $\left[l^{(\mathbf{w})}, u^{(\mathbf{w})}\right]$ για κάθε επιλογή $\mathbf{w}$ για τις παραμέτρους μοντέλου. Μία απλή κατασκευή είναι $l^{(\mathbf{w})} := f(\mathbf{w}) - \sigma/2$, $u^{(\mathbf{w})} := f(\mathbf{w}) + \sigma/2$, με το σ να είναι ένα μέτρο της (αναμενόμενης) απόκλισης του $f(\mathbf{w})$ από το $\bar{f}(\mathbf{w})$. Μπορούμε επίσης να χρησιμοποιήσουμε άλλες κατασκευές για αυτό το διάστημα εφόσον εξασφαλίζουν ότι $\bar{f}(\mathbf{w}) \in \left[l^{(\mathbf{w})}, u^{(\mathbf{w})}\right]$ με αρκετά υψηλή πιθανότητα. Ένας αισιόδοξος επιλέγει τις παραμέτρους μοντέλου σύμφωνα με την πιο ευνοϊκή—αλλά εύλογη—τιμή $\tilde{f}(\mathbf{w}) := l^{(\mathbf{w})}$ του κριτηρίου επίδοσης. Δύο παραδείγματα μεθόδων μηχανικής μάθησης που χρησιμοποιούν μία τέτοια αισιόδοξη κατασκευή μίας αντικειμενικής συνάρτησης είναι οι μέθοδοι δομημένης ελαχιστοποίησης διακινδύνευσης [9, Κεφ. 11] και άνω φράγματος εμπιστοσύνης για διαδοχική λήψη αποφάσεων [10, Sec. 2.2].



Σχ. 1. Οι μέθοδοι μηχανικής μάθησης μαθαίνουν παραμέτρους μοντέλου $\mathbf{w}$ χρησιμοποιώντας κάποια εκτίμηση του $f(\mathbf{w})$ για το τελικό κριτήριο επίδοσης $\bar{f}(\mathbf{w})$. Χρησιμοποιώντας ένα πιθανοτικό μοντέλο, κανείς μπορεί να χρησιμοποιήσει το $f(\mathbf{w})$ για να κατασκευάσει διαστήματα εμπιστοσύνης $\left[l^{(\mathbf{w})}, u^{(\mathbf{w})}\right]$ που περιέχουν το $\bar{f}(\mathbf{w})$ με υψηλή πιθανότητα. Το καλύτερο εύλογο μέτρο επίδοσης για μία συγκεκριμένη επιλογή $\mathbf{w}$ των παραμέτρων μοντέλου είναι $\tilde{f}(\mathbf{w}) := l^{(\mathbf{w})}$.

Βλέπε επίσης: ml, παράμετροι μοντέλου, ελαχιστοποίηση εμπειρικής διακινδύνευσης, loss, σύνολο δεδομένων, σύνολο εκπαίδευσης, διακινδύνευση, υπόθεση, πιθανοτικό μοντέλο, probability, αντικειμενική συνάρτηση, δομημένη ελαχιστοποίηση διακινδύνευσης, άνω φράγμα εμπιστοσύνης.

ακρίβεια Θεωρούμε σημεία δεδομένων που χαρακτηρίζονται από χαρακτηριστικά $\mathbf{x} \in \mathcal{X}$ και μία κατηγορική ετικέτα y που παίρνει τιμές από ένα πεπερασμένο χώρο ετικετών $\mathcal{Y}$. Η ακρίβεις μίας υπόθεσης $h : \mathcal{X} \rightarrow \mathcal{Y}$, όταν εφαρμόζεται στα σημεία δεδομένων ενός συνόλου δεδομένων $\mathcal{D} = \{(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(m)}, y^{(m)})\}$, ορίζεται τότε ως

$$1 - (1/m) \sum_{r=1}^{m} L^{(0/1)} \left((\mathbf{x}^{(r)}, y^{(r)}), h \right)$$

χρησιμοποιώντας την 0/1 απώλεια $L^{(0/1)}(\cdot, \cdot)$.

Βλέπε επίσης: data point, feature, ετικέτα, χώρος ετικετών, υπόθεση, σύνολο δεδομένων, 0/1 απώλεια.

ακραία τιμή Many ml methods are motivated by the παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων, which interprets data points as πραγμάτωσηs of ανεξάρτητες και ταυτόσημα κατανεμημένες τυχαία μεταβλητήs with a common κατανομή πιθανότητας. The παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων is useful for applications where the statistical properties of the data generation process are stationary (or time-invariant) [114]. However, in some applications the data consists of a majority of regular data points that conform with an παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων as well as a small number of data points that have fundamentally different statistical properties compared to the regular data points. We refer to a data point that substantially deviates from the statistical properties of most data points

as an outlier. Different methods for outlier detection use different measures for this deviation. Statistical learning theory studies fundamental limits on the ability to mitigate outliers reliably [115], [116].
Βλέπε επίσης: ml, παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων, data point, πραγμάτωση, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, κατανομή πιθανότητας, data.

αλγόριθμος Ένας αλγόριθμος (algorithm) είναι μία ακριβής, βήμα προς βήμα προδιαγραφή για το πώς να παραχθεί μία έξοδος (output) από μία συγκεκριμένη είσοδο (input) εντός ενός πεπερασμένου αριθμού υπολογιστικών βημάτων [11]. Για παράδειγμα, ένας αλγόριθμος για την εκπαίδευση ενός γραμμικού μοντέλου περιγράφει ρητά πώς να μετασχηματιστεί ένα δεδομένο σύνολο εκπαίδευσης σε παραμέτρους μοντέλου μέσω μίας ακολουθίας βημάτων κλίσης. Για να μελετήσουμε αλγόριθμους ενδελεχώς, μπορούμε να τους αναπαραστήσουμε (ή να τους προσεγγίσουμε) με διαφορετικές μαθηματικές δομές [12]. Μία προσέγγιση είναι να αναπαραστήσουμε έναν αλγόριθμο ως μία συλλογή πιθανών εκτελέσεων. Κάθε μεμονωμένη εκτέλεση είναι μία ακολουθία της ακόλουθης μορφής:

$$\text{input}, s_1, s_2, \ldots, s_T, \text{output}.$$

Αυτή η ακολουθία ξεκινάει από μία είσοδο και προοδεύει μέσω ενδιάμεσων βημάτων μέχρι να παραδοθεί μία έξοδος. Είναι κρίσιμο ότι ένας αλγόριθμος συμπεριλαμβάνει περισσότερα από απλώς μία αντιστοίχιση από είσοδο σε έξοδο· περιλαμβάνει επίσης ενδιάμεσα υπολογιστικά βήματα $s_1, \ldots, s_T$.
Βλέπε επίσης: γραμμικό μοντέλο, σύνολο εκπαίδευσης, παράμετροι μο-

ντέλου, βήμα κλίσης, model, stochastic.

αλγόριθμος k-μέσων Ο αλγόριθμος k-μέσων (k-means) είναι μία μέθοδος σκληρής συσταδοποίησης που αποδίδει κάθε σημείο δεδομένων ενός συνόλου δεδομένων σε ακριβώς μία από τις k διαφορετικές συστάδες. Η μέθοδος εναλλάσσεται μεταξύ της ενημέρωσης των αποδόσεων συστάδων (με τη συστάδα με την πλησιέστερη μέση τιμή) και, δεδομένων των ενημερωμένων αποδόσεων συστάδων, του επανυπολογισμού των μέσων τιμών των συστάδων [8, Κεφ. 8].

Βλέπε επίσης: μέση τιμή, αλγόριθμος, hard clustering, data point, σύνολο δεδομένων, συστάδα.

αμοιβαίες πληροφορίες Οι αμοιβαίες πληροφορίες (mutual information - MI) $I\left(\mathbf{x};y\right)$ μεταξύ δύο τυχαίων μεταβλητών $\mathbf{x}$, y που ορίζονται στον ίδιο χώρο πιθανοτήτων δίνονται από [13]

$$I\left(\mathbf{x};y\right) := \mathbb{E}\left\{\log\frac{p(\mathbf{x},y)}{p(\mathbf{x})p(y)}\right\}.$$

Αποτελεί μέτρο του πόσο καλά μπορούμε να εκτιμήσουμε την y βάσει μόνο του $\mathbf{x}$. Μία μεγάλη τιμή του $I\left(\mathbf{x};y\right)$ υποδεικνύει ότι η y μπορεί να προβλεφθεί καλά μόνο από το $\mathbf{x}$. Αυτή η πρόβλεψη θα μπορούσε να προκύψει από μία υπόθεση που μαθαίνεται από μία μέθοδο μηχανικής μάθησης βασισμένη στην ελαχιστοποίηση εμπειρικής διακινδύνευσης.

Βλέπε επίσης: τυχαία μεταβλητή, χώρος πιθανοτήτων, πρόβλεψη, υπόθεση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, ml.

αμφικλινής παλινδρόμηση Η αμφικλινής παλινδρόμηση μαθαίνει τα βάρη

$\mathbf{w}$ μίας γραμμικής map υπόθεσης $h^{(\mathbf{w})}(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$. Η ποιότητα μίας συγκεκριμένης επιλογής για τις παραμέτρους μοντέλου $\mathbf{w}$ μετριέται από το άθροισμα των δύο συνιστωστών. Η πρώτη συνιστώσα είναι η μέση απώλεια τετραγωνικού σφάλματος που προκαλείται από την $h^{(\mathbf{w})}$ σε ένα σύνολο σημείων δεδομένων με ετικέτες (δηλαδή το σύνολο εκπαίδευσης). Η δεύτερη συνιστώσα είναι η ανηγμένη τετραγωνική Ευκλείδεια νόρμα $\alpha\|\mathbf{w}\|_2^2$ με μία παράμετρο ομαλοποίησης $\alpha > 0$. Η προσθήκη $\alpha\|\mathbf{w}\|_2^2$ στη μέση απώλεια τετραγωνικού σφάλματος είναι ισοδύναμη με την αντικατάσταση αρχικών σημείων δεδομένων από τις πραγματώσεις (άπειρα πολλών) ανεξάρτητων και ταυτόσημα κατανεμημένων τυχαίων μεταβλητών που είναι κεντρικές γύρω από αυτά τα σημεία δεδομένων (βλέπε ομαλοποίηση).

Βλέπε επίσης: regression, βάρη, υπόθεση, map, παράμετροι μοντέλου, απώλεια τετραγωνικού σφάλματος, σημείο δεδομένων με ετικέτα, σύνολο εκπαίδευσης, νόρμα, ομαλοποίηση, παράμετρος, data point, πραγμάτωση, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή.

ανάλυση ιδιαζουσών τιμών Η ανάλυση ιδιαζουσών τιμών (singular value decomposition - SVD) για έναν πίνακα $\mathbf{A} \in \mathbb{R}^{m\times d}$ είναι μία παραγοντοποίηση της μορφής

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{U}^T$$

με ορθοκανονικούς πίνακες $\mathbf{V} \in \mathbb{R}^{m\times m}$ και $\mathbf{U} \in \mathbb{R}^{d\times d}$ [3]. Ο πίνακας $\mathbf{\Lambda} \in \mathbb{R}^{m\times d}$ είναι μη μηδενικός μόνο κατά την κύρια διαγώνιο, της οποίας οι καταχωρίσεις $\Lambda_{j,j}$ είναι μη αρνητικές και αναφέρονται ως ιδιάζουσες τιμές.

ανάλυση ιδιοτιμών Η ανάλυση ιδιοτιμών (eigenvalue decomposition - EVD) για έναν τετραγωνικό πίνακα $\mathbf{A} \in \mathbb{R}^{d \times d}$ είναι μία παραγοντοποίηση της μορφής

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}.$$

Οι στήλες του πίνακα $\mathbf{V} = \big(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(d)}\big)$ είναι τα ιδιοδιανύσματα του πίνακα $\mathbf{V}$. Ο διαγώνιος πίνακας $\mathbf{\Lambda} = \mathrm{diag}\big\{\lambda_1, \ldots, \lambda_d\big\}$ περιέχει τις ιδιοτιμές λ_j που αντιστοιχούν στα ιδιοδιανύσματα $\mathbf{v}^{(j)}$. Σημείωση ότι η παραπάνω ανάλυση υπάρχει μόνο αν ο πίνακας $\mathbf{A}$ είναι διαγωνοποιήσιμος.
Βλέπε επίσης: ιδιοδιάνυσμα, ιδιοτιμή.

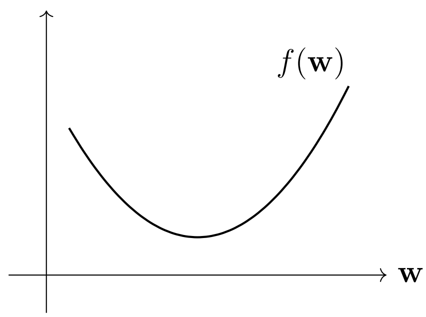
ανάλυση κυρίων συνιστωσών Η ανάλυση κυρίων συνιστωσών (principal component analysis - PCA) καθορίζει έναν γραμμικό χάρτη χαρακτηριστικών, έτσι ώστε τα νέα χαρακτηριστικά να μας επιτρέπουν να ξανακατασκευάσουμε τα αρχικά χαρακτηριστικά με το ελάχιστο σφάλμα ανακατασκευής [8].
Βλέπε επίσης: χάρτης χαρακτηριστικών, feature, ελάχιστο.

ανεξάρτητες και ταυτόσημα κατανεμημένες Μπορεί να είναι χρήσιμο να ερμηνεύσουμε σημεία δεδομένων $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$ ως πραγματώσεις ανεξάρτητων και ταυτόσημα κατανεμημένων (independent and identically distributed - i.i.d.) τυχαίων μεταβλητών με κοινή κατανομή πιθανότητας. Αν αυτές οι τυχαίες μεταβλητές είναι συνεχούς τιμής, η από κοινού τους συνάρτηση πυκνότητας πιθανότητας είναι $p\big(\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}\big) = \prod_{r=1}^{m} p\big(\mathbf{z}^{(r)}\big)$, με την $p(\mathbf{z})$ να είναι η κοινή περιθωριακή συνάρτηση πυκνότητας πιθανότητας των υποκείμενων τυχαίων μεταβλητών.
Βλέπε επίσης: data point, πραγμάτωση, τυχαία μεταβλητή, κατανομή

πιθανότητας, συνάρτηση πυκνότητας πιθανότητας.

ανταμοιβή Μία ανταμοιβή αναφέρεται σε κάποια παρατηρούμενη (ή μετρημένη) ποσότητα που μας επιτρέπει να εκτιμήσουμε την απώλεια που προκύπτει από την πρόβλεψη (ή απόφαση) μίας υπόθεσης $h(\mathbf{x})$. Για παράδειγμα, σε μία εφαρμογή μηχανικής μάθησης σε αυτοοδηγούμενα οχήματα, η $h(\mathbf{x})$ θα μπορούσε να αναπαριστά την τρέχουσα κατεύθυνση οδήγησης ενός οχήματος. Θα μπορούσαμε να κατασκευάσουμε μία ανταμοιβή από τις μετρήσεις ενός αισθητήρα σύγκρουσης που υποδεικνύει αν το όχημα κινείται προς ένα εμπόδιο. Ορίζουμε μία χαμηλή ανταμοιβή για την κατεύθυνση οδήγησης $h(\mathbf{x})$ αν το όχημα κινείται επικίνδυνα προς ένα εμπόδιο. Βλέπε επίσης: loss, πρόβλεψη, υπόθεση, ml.

αντικειμενική συνάρτηση Μια αντικειμενική συνάρτηση είναι μία map που αποδίδει μία αριθμητική αντικειμενική τιμή $f(\mathbf{w})$ σε κάθε επιλογή $\mathbf{w}$ κάποιας μεταβλητής που θέλουμε να βελτιστοποιήσουμε (βλέπε Σχ. 2). Στο πλαίσιο της μηχανικής μάθησης, η μεταβλητή βελτιστοποίησης θα μπορούσε να είναι οι παράμετροι μοντέλου μίας υπόθεσης $h^{(\mathbf{w})}$. Κοινές αντικειμενικές συναρτήσεις περιλαμβάνουν τη διακινδύνευση (δηλαδή την προσδοκώμενη απώλεια) ή την εμπειρική διακινδύνευση (δηλαδή τη μέση απώλεια πάνω σε ένα σύνολο εκπαίδευσης). Οι μέθοδοι μηχανικής μάθησης εφαρμόζουν τεχνικές βελτιστοποίησης, όπως τις μεθόδους με βάση την κλίση, για να βρούν την επιλογή $\mathbf{w}$ με τη βέλτιστη τιμή (π.χ., το ελάχιστο ή το μέγιστο) της αντικειμενικής συνάρτησης.



Σχ. 2. Μία αντικειμενική συνάρτηση αντιστοιχίζει κάθε πιθανή τιμή $\mathbf{w}$ μίας μεταβλητής βελτιστοποίησης, όπως οι παράμετροι μοντέλου ενός μοντέλου μηχανικής μάθησης, σε μία τιμή που μετράει τη χρησιμότητα της $\mathbf{w}$.

Βλέπε επίσης: συνάρτηση, map, ml, παράμετροι μοντέλου, υπόθεση, διακινδύνευση, loss, empirical risk, σύνολο εκπαίδευσης, μέθοδοι με βάση την κλίση, ελάχιστο, maximum, model, συνάρτηση απώλειας.

αντίστροφος πίνακας An inverse matrix $\mathbf{A}^{-1}$ is defined for a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ that is of full rank, meaning its columns are linearly independent. In this case, $\mathbf{A}$ is said to be invertible, and its inverse satisfies

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

A square matrix is invertible if and only if its ορίζουσα is non-zero. Inverse matrices are fundamental in solving systems of linear equations and in the closed-form solution of γραμμική παλινδρόμηση [120], [118]. The concept of an inverse matrix can be extended to matrices that are not square or not full rank. One may define a "left inverse" $\mathbf{B}$ satisfying $\mathbf{B}\mathbf{A} = \mathbf{I}$, or a "right inverse" $\mathbf{C}$ satisfying $\mathbf{A}\mathbf{C} = \mathbf{I}$. For general rectangular or singular matrices, the Moore–Penrose pseudoinverse $\mathbf{A}^{+}$

provides a unified concept of generalized inverse matrix [3].

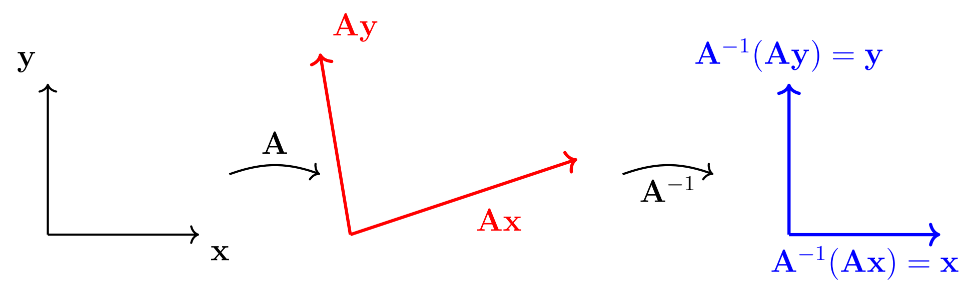


Fig. 3. A matrix $\mathbf{A}$ represents a linear transformation of $\mathbb{R}^2$. The inverse matrix $\mathbf{A}^{-1}$ represents the inverse transformation.

Βλέπε επίσης: ορίζουσα, γραμμική παλινδρόμηση, pseudoinverse.

άνω φράγμα εμπιστοσύνης (ΑΦΕ) Consider an ml application that requires selecting, at each time step k, an action a_k from a finite set of alternatives $\mathcal{A}$. The utility of selecting action a_k is quantified by a numeric ανταμοιβή signal $r^{(a_k)}$. A widely used πιθανοτικό μοντέλο for this type of sequential decision-making problem is the stochastic MAB setting [10]. In this model, the ανταμοιβή $r^{(a)}$ is viewed as the πραγμάτωση of an τυχαία μεταβλητή with unknown μέση τιμή $\mu^{(a)}$. Ideally, we would always choose the action with the largest expected ανταμοιβή $\mu^{(a)}$, but these μέση τιμήs are unknown and must be estimated from observed data. Simply choosing the action with the largest estimate $\widehat{\mu}^{(a)}$ can lead to suboptimal outcomes due to estimation αβεβαιότητα. The UCB (upper confidence bound; UCB) strategy addresses this by selecting actions not only based on their estimated μέση τιμήs but also by incorporating a term that reflects the αβεβαιότητα in these estimates—favoring actions with a high potential ανταμοιβή and high αβεβαιότητα. Theoretical guarantees for the performance of UCB strategies, including logarithmic regret bounds, are established in [10].

Βλέπε επίσης: ml, ανταμοιβή, πιθανοτικό μοντέλο, stochastic, MAB, model, πραγμάτωση, τυχαία μεταβλητή, μέση τιμή, data, αβεβαιότητα, regret.

αξιόπιστη τεχνητή νοημοσύνη (αξιόπιστη ΤΝ) Εκτός από τις υπολογιστικές διαστάσεις και τις στατιστικές διαστάσεις, μία τρίτη κύρια διάσταση σχεδιασμού μεθόδων μηχανικής μάθησης είναι η αξιοπιστία τους [14]. Η Ευρωπαϊκή Ένωση (ΕΕ) έχει διατυπώσει επτά βασικές απαιτήσεις για αξιόπιστη τεχνητή νοημοσύνη (trustworthy artificial intelligence - trustworthy AI) (που συνήθως χτίζουν πάνω σε μεθόδους μηχανικής μάθησης) [15]:

1) Ανθρώπινη παρέμβαση και εποπτεία·
2) Τεχνική ευρωστία και ασφάλεια·
3) Ιδιωτικότητα και διακυβέρνηση των δεδομένων·
4) Διαφάνεια·
5) Διαφορετικότητα, απαγόρευση των διακρίσεων και δικαιοσύνη·
6) Κοινωνιακή και περιβαλλοντική ευημερία·
7) Λογοδοσία.

Βλέπε επίσης: υπολογιστικές διαστάσεις, στατιστικές διαστάσεις, ml, τεχνητή νοημοσύνη (ΤΝ), ευρωστία, data, transparency.

απόκλιση Θεωρούμε μία εφαρμογή ομοσπονδιακής μάθησης με networked data που αναπαριστώνται από ένα δίκτυο ομοσπονδιακής μάθησης. Οι μέθοδοι ομοσπονδιακής μάθησης χρησιμοποιούν ένα μέτρο απόκλισης για να συκρίνουν maps υπόθεσης από τοπικά μοντέλα σε κόμβους i, i' που συνδέονται με μία ακμή στο δίκτυο ομοσπονδιακής μάθησης.

Βλέπε επίσης: federated learning (FL), networked data, δίκτυο ομοσπονδιακής μάθησης, υπόθεση, map, local model.

απόκλιση Kullback-Leibler (απόκλιση KL) Η απόκλιση KL (Kullback-Leibler divergence - KL divergence) είναι ένα ποσοτικό μέτρο του πόσο διαφορετική είναι μία κατανομή πιθανότητας από μία άλλη κατανομή πιθανότητας [13].

Βλέπε επίσης: κατανομή πιθανότητας.

απόκλιση Rényi Η απόκλιση Rényi μετράει την (αν)ομοιότητα μεταξύ δύο κατανομών πιθανοτήτων [16].

Βλέπε επίσης: κατανομή πιθανότητας.

αποτελεσματική διάσταση Η αποτελεσματική διάσταση $d_{\text{eff}}(\mathcal{H})$ ενός άπειρου χώρου υποθέσεων $\mathcal{H}$ είναι ένα μέτρος του μεγέθους του. Σε γενικές γραμμές, η αποτελεσματική διάσταση είναι ίση με τον αποτελεσματικό αριθμό ανεξάρτητων παραμέτρων μοντέλου που μπορούν να ρυθμιστούν. Αυτές οι παράμετροι μπορεί να είναι συντελεστές που χρησιμοποιούνται σε μία linear map ή τα βάρη και οι όροι μεροληψίας ενός τεχνητού νευρωνικού δικτύου.

Βλέπε επίσης: χώρος υποθέσεων, παράμετροι μοντέλου, παράμετρος, linear map, βάρη, μεροληψία, ΤΝΔ.

απώλεια Οι μέθοδοι μηχανικής μάθησης χρησιμοποιούν μία συνάρτηση απώλειας $L(\mathbf{z}, h)$ για να μετρήσουν το σφάλμα που προκαλείται από την εφαρμογή μίας συγκεκριμένης υπόθεσης σε ένα συγκεκριμένο σημείο δεδομένων. Με μία μικρή κατάχρηση του συμβολισμού, χρησιμοποιούμε τον όρο απώλεια και για την ίδια τη συνάρτηση απώλειας L και για τη συγκεκριμένη τιμή $L(\mathbf{z}, h)$, για ένα σημείο δεδομένων $\mathbf{z}$ και μία υπόθεση h.

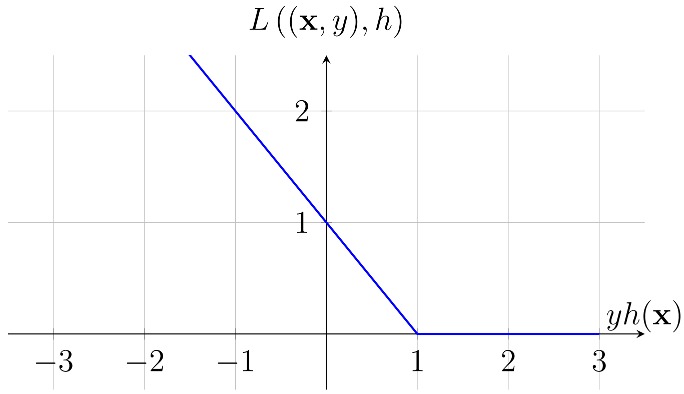
Βλέπε επίσης: ml, συνάρτηση απώλειας, υπόθεση, data point.

απώλεια απόλυτου σφάλματος Θεωρούμε ένα σημείο δεδομένων με χαρακτηριστικά $\mathbf{x} \in \mathcal{X}$ και αριθμητική ετικέτα $y \in \mathbb{R}$. Η απώλεια απόλυτου σφάλματος που προκαλείται από μία υπόθεση $h : \mathcal{X} \to \mathbb{R}$ ορίζεται ως $|y - h(\mathbf{x})|$, δηλαδή η απόλυτη διαφορά μεταξύ της πρόβλεψης $h(\mathbf{x})$ και της αληθούς ετικέτας y.

Βλέπε επίσης: data point, feature, ετικέτα, loss, υπόθεση, πρόβλεψη.

απώλεια άρθρωσης Θεωρούμε ένα σημείο δεδομένων που χαρακτηρίζεται από ένα διάνυσμα χαρακτηριστικών $\mathbf{x} \in \mathbb{R}^d$ και μία δυαδική ετικέτα $y \in \{-1, 1\}$. Η απώλεια άρθρωσης που προκαλείται από μία map υπόθεσης $h(\mathbf{x})$ πραγματικής τιμής ορίζεται ως

$$L((\mathbf{x}, y), h) := \max\{0, 1 - yh(\mathbf{x})\}. \tag{1}$$



Σχ. 4. Η απώλεια άρθρωσης που προκαλείται από την πρόβλεψη $h(\mathbf{x}) \in \mathbb{R}$ για ένα σημείο δεδομένων με ετικέτα $y \in \{-1, 1\}$. Μία ομαλοποιημένη παραλλαγή της απώλειας άρθρωσης χρησιμοποιείται από τη μηχανή διανυσμάτων υποστήριξης [17].

Βλέπε επίσης: data point, διάνυσμα χαρακτηριστικών, ετικέτα, loss, υπόθεση, map, πρόβλεψη, μηχανή διανυσμάτων υποστήριξης.

απώλεια τετραγωνικού σφάλματος Η απώλεια τετραγωνικού σφάλματος (squared error loss) μετράει το σφάλμα πρόβλεψης μίας υπόθεσης h όταν προβλέπει μία αριθμητική ετικέτα $y \in \mathbb{R}$ από τα χαρακτηριστικά $\mathbf{x}$ ενός σημείου δεδομένων. Ορίζεται ως

$$L((\mathbf{x}, y), h) := \big(y - \underbrace{h(\mathbf{x})}_{=\hat{y}}\big)^2.$$

Βλέπε επίσης: loss, πρόβλεψη, υπόθεση, ετικέτα, feature, data point.

απώλεια Huber Η απώλεια Huber ενώνει την απώλεια τετραγωνικού σφάλματος και την απώλεια απόλυτου σφάλματος.

Βλέπε επίσης: loss, απώλεια τετραγωνικού σφάλματος, απώλεια απόλυτου σφάλματος.

αριθμός συνθήκης Ο αριθμός συνθήκης $\kappa(\mathbf{Q}) \geq 1$ ενός θετικά ορισμένου πίνακα $\mathbf{Q} \in \mathbb{R}^{d \times d}$ είναι ο λόγος α/β μεταξύ της μεγαλύτερης α και της μικρότερης β ιδιοτιμής του $\mathbf{Q}$. Ο αριθμός συνθήκης είναι χρήσιμος για την ανάλυση μεθόδων μηχανικής μάθησης. Η υπολογιστική πολυπλοκότητα των μεθόδων με βάση την κλίση για γραμμική παλινδρόμηση εξαρτάται κρίσιμα από τον αριθμό συνθήκης του πίνακα $\mathbf{Q} = \mathbf{X}\mathbf{X}^T$, με τον πίνακα χαρακτηριστικών $\mathbf{X}$ του συνόλου εκπαίδευσης. Συνεπώς, από υπολογιστικής άποψης, προτιμούμε χαρακτηριστικά σημεία δεδομένων, έτσι ώστε ο $\mathbf{Q}$ να έχει έναν αριθμό συνθήκης κοντά στο 1.

Βλέπε επίσης: ιδιοτιμή, ml, μέθοδοι με βάση την κλίση, γραμμική παλινδρόμηση, πίνακας χαρακτηριστικών, σύνολο εκπαίδευσης, feature, data point.

αρχή της ελαχιστοποίησης των δεδομένων Ο Ευρωπαϊκός κανονισμός για την προστασία δεδομένων περιλαμβάνει μία αρχή ελαχιστοποίησης δεδομένων. Αυτή η αρχή απαιτεί έναν υπεύθυνο επεξεργασίας δεδομένων για να περιορίσει τη συλλογή προσωπικών πληροφοριών σε ό,τι είναι άμεσα σχετικό και απαραίτητο για την εκπλήρωση ενός προσδιορισμένου σκοπού. Τα δεδομένα πρέπει να φυλάσσονται μόνο για το χρονικό διάστημα που είναι απαραίτητα προκειμένου να εκπληρωθεί αυτός ο σκοπός [18, Άρθρο 5(1)(c)], [19].

Βλέπε επίσης: data.

αυτοκωδικοποιητής Ένας αυτοκωδικοποιητής (autoencoder) είναι μία μέθοδος μηχανικής μάθησης που μαθαίνει ταυτόχρονα έναν κωδικοποιητή map $h(\cdot) \in \mathcal{H}$ και έναν αποκωδικοποιητή map $h^*(\cdot) \in \mathcal{H}^*$. Είναι μία περίπτωση της ελαχιστοποίησης εμπειρικής διακινδύνευσης που χρησιμοποιεί μία απώλεια υπολογιζόμενη από το σφάλμα ανακατασκευής $\mathbf{x} - h^*\big(h(\mathbf{x})\big)$.

Βλέπε επίσης: ml, map, ελαχιστοποίηση εμπειρικής διακινδύνευσης, loss.

βαθμός κόμβου Ο βαθμός ενός κόμβου $d^{(i)}$ $i \in \mathcal{V}$ σε έναν μη κατευθυνόμενο γράφο είναι ο αριθμός των γειτόνων του, δηλαδή $d^{(i)} := \big|\mathcal{N}^{(i)}\big|$.

Βλέπε επίσης: graph, γείτονες.

βαθμός συσχέτισης Ο βαθμός συσχέτισης είναι ένας αριθμός που υποδεικνύει το κατά πόσο ένα σημείο δεδομένων ανήκει σε μία συστάδα [8, Κεφ. 8]. Ο βαθμός της συσχέτισης μπορεί να ερμηνευτεί ως μία μαλακή απόδοση συστάδας. Οι μέθοδοι μαλακής συσταδοποίησης μπορούν να κωδικοποιήσουν τον βαθμό συσχέτισης με έναν πραγματικό αριθμό στο διάστημα $[0, 1]$. Η σκληρή συσταδοποίηση προκύπτει ως η ακραία περίπτωση όταν ο βαθμός συσχέτισης παίρνει μόνο τιμές 0 or 1.

Βλέπε επίσης: data point, συστάδα, soft clustering, hard clustering.

βαθύ δίκτυο Ένα βαθύ δίκτυο είναι ένα τεχνητό νευρωνικό δίκτυο με έναν (σχετικά) μεγάλο αριθμό κρυφών στρωμάτων. Βαθιά μάθηση είναι ένας όρος-ομπρέλα για μεθόδους μηχανικής μάθησης που χρησιμοποιούν ένα βαθύ δίκτυο ως το μοντέλο τους [20].

Βλέπε επίσης: ΤΝΔ, ml, model.

βάρη Θεωρούμε έναν παραμετροποιημένο χώρο υποθέσεων $\mathcal{H}$. Χρησιμοποιούμε τον όρο βάρη για αριθμητικές παραμέτρους μοντέλου που χρησιμοποιούνται για να κλιμακώσουν χαρακτηριστικά ή τους μετασχηματισμούς τους προκειμένου να υπολογίσουμε $h^{(\mathbf{w})} \in \mathcal{H}$. Ένα γραμμικό μοντέλο χρησιμοποιεί βάρη $\mathbf{w} = \left(w_1, \ldots, w_d\right)^T$ για να υπολογίσει τον γραμμικό συνδυασμό $h^{(\mathbf{w})}(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$. Βάρη χρησιμοποιούνται επίσης σε τεχνητά νευρωνικά δίκτυα για να σχηματιστούν γραμμικοί συνδυασμοί χαρακτηριστικών ή των εξόδων νευρώνων σε κρυφά στρώματα.

Βλέπε επίσης: χώρος υποθέσεων, παράμετροι μοντέλου, feature, γραμμικό μοντέλο, ΤΝΔ.

βάρος ακμής Σε κάθε ακμή $\{i, i'\}$ ενός δικτύου ομοσπονδιακής μάθησης αποδίδεται ένα μη αρνητικό βάρος ακμής $A_{i,i'} \geq 0$. Ένα μηδενικό βάρος ακμής $A_{i,i'} = 0$ υποδεικνύει την απουσία μίας ακμής μεταξύ κόμβων $i, i' \in \mathcal{V}$.

Βλέπε επίσης: δίκτυο ομοσπονδιακής μάθησης.

βάση αναφοράς Consider some ml method that produces a learned υπόθεση (or trained model) $\hat{h} \in \mathcal{H}$. We evaluate the quality of a trained model by computing the average loss on a σύνολο ελέγχου. But how can we assess whether the resulting σύνολο ελέγχου performance is sufficiently good? How can we determine if the trained model performs close to optimal and there is little point in investing more resources (for data collection or computation) to improve it? To this end, it is useful to have a reference (or baseline) level against which we can compare the performance of the trained model. Such a reference value might be obtained from human performance, e.g., the misclassification rate of dermatologists who diagnose cancer from visual inspection of skin [21]. Another source for a baseline is an existing, but for some reason unsuitable, ml method. For example, the existing ml method might be computationally too expensive for the intended ml application. Nevertheless, its σύνολο ελέγχου error can still serve as a baseline. Another, somewhat more principled, approach to constructing a baseline is via a πιθανοτικό μοντέλο. In many cases, given a πιθανοτικό μοντέλο $p(\mathbf{x}, y)$, we can precisely determine the ελάχιστο achievable διακινδύνευση among any hypotheses (not even required to belong to the χώρος υποθέσεων $\mathcal{H}$) [22]. This ελάχιστο achievable διακινδύνευση (referred to as the διακινδύνευση Bayes) is the διακινδύνευση of the εκτιμήτρια Bayes for the ετικέτα y of a data point, given its features $\mathbf{x}$. Note that, for a given choice of συνάρτηση απώλειας, the εκτιμήτρια Bayes (if it exists) is completely determined by the κατανομή πιθανότητας $p(\mathbf{x}, y)$ [22, Ch. 4]. However, computing the εκτιμήτρια Bayes and διακινδύνευση Bayes presents two main challenges:

1) The κατανομή πιθανότητας $p(\mathbf{x}, y)$ is unknown and needs to be estimated.
2) Even if $p(\mathbf{x}, y)$ is known, it can be computationally too expensive to compute the διακινδύνευση Bayes exactly [23].

A widely used πιθανοτικό μοντέλο is the πολυμεταβλητή κανονική κατανομή $(\mathbf{x}, y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for data points characterized by numeric features and ετικέτας. Here, for the απώλεια τετραγωνικού σφάλματος, the εκτιμήτρια Bayes is given by the posterior μέση τιμή $\mu_{y|\mathbf{x}}$ of the ετικέτα y, given the features $\mathbf{x}$ [22], [24]. The corresponding διακινδύνευση Bayes is given by the posterior διακύμανση $\sigma^2_{y|\mathbf{x}}$ (see Fig. 5).

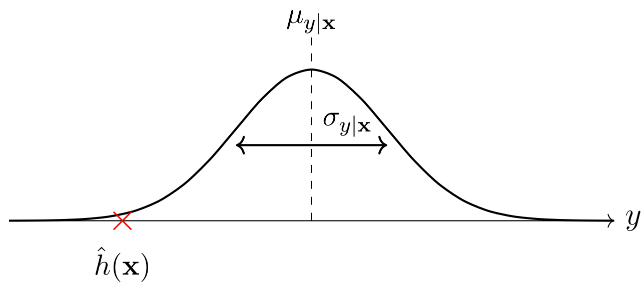


Fig. 5. If the features and the ετικέτα of a data point are drawn from a πολυμεταβλητή κανονική κατανομή, we can achieve the ελάχιστο διακινδύνευση (under απώλεια τετραγωνικού σφάλματος) by using the εκτιμήτρια Bayes $\mu_{y|\mathbf{x}}$ to predict the ετικέτα y of a data point with features $\mathbf{x}$. The corresponding ελάχιστο διακινδύνευση is given by the posterior διακύμανση $\sigma^2_{y|\mathbf{x}}$. We can use this quantity as a baseline for the average loss of a trained model $\hat{h}$.

Βλέπε επίσης: ml, υπόθεση, model, loss, σύνολο ελέγχου, data, πιθανοτικό μοντέλο, ελάχιστο, διακινδύνευση, χώρος υποθέσεων, διακινδύνευση Bayes, εκτιμήτρια Bayes, ετικέτα, data point, feature, συνάρτηση απώλειας, κατανομή πιθανότητας, πολυμεταβλητή κανονική κατανομή, απώλεια τετραγωνικού σφάλματος, μέση τιμή, διακύμανση.

βήμα κλίσης Given a παραγωγίσιμη real-valued συνάρτηση $f(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ and a vector $\mathbf{w} \in \mathbb{R}^d$, the gradient step updates $\mathbf{w}$ by adding the scaled negative gradient $\nabla f(\mathbf{w})$ to obtain the new vector (see Fig. 6)

$$\widehat{\mathbf{w}} := \mathbf{w} - \eta \nabla f(\mathbf{w}). \tag{2}$$

Mathematically, the gradient step is an operator $\mathcal{T}^{(f,\eta)}$ that is parametrized by the συνάρτηση f and the μέγεθος βήματος η.

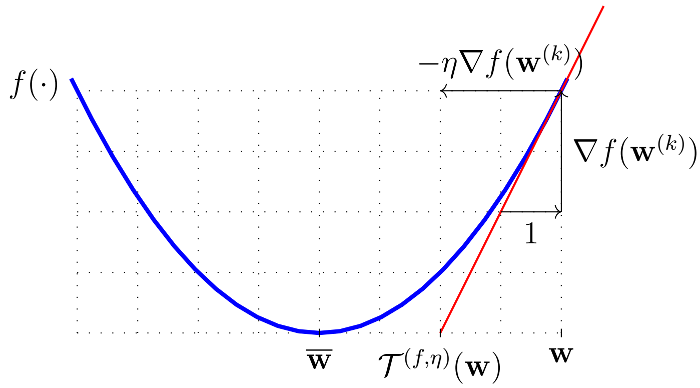


Fig. 6. The basic gradient step (2) maps a given vector $\mathbf{w}$ to the updated vector $\mathbf{w}'$. It defines an operator $\mathcal{T}^{(f,\eta)}(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d : \mathbf{w} \mapsto \widehat{\mathbf{w}}$.

Note that the gradient step (2) optimizes locally - in a neighborhood whose size is determined by the μέγεθος βήματος η - a linear approximation to the συνάρτηση $f(\cdot)$. A natural γενίκευση of (2) is to locally optimize the συνάρτηση itself - instead of its linear approximation - such that

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}' \in \mathbb{R}^d}{\arg\min} \, f(\mathbf{w}') + \frac{1}{\eta} \left\| \mathbf{w} - \mathbf{w}' \right\|_2^2 . \tag{3}$$

We intentionally use the same symbol η for the παράμετρος in (3) as we used for the μέγεθος βήματος in (2). The larger the η we choose in (3), the more progress the update will make towards reducing the συνάρτηση value $f(\widehat{\mathbf{w}})$. Note that, much like the gradient step (2), also the update (3) defines an operator that is parametrized by the συνάρτηση $f(\cdot)$ and the ρυθμός μάθησης η. For a κυρτός συνάρτηση $f(\cdot)$, this operator is known as the εγγύς τελεστής of $f(\cdot)$ [25].

Βλέπε επίσης: παραγωγίσιμη, συνάρτηση, gradient, μέγεθος βήματος, neighborhood, γενίκευση, παράμετρος, ρυθμός μάθησης, convex, εγγύς τελεστής.

γείτονες Οι γείτονες ενός κόμβου $i \in \mathcal{V}$ εντός ενός δικτύου ομοσπονδιακής μάθησης είναι εκείνοι οι κόμβοι $i' \in \mathcal{V} \setminus \{i\}$ που συνδέονται (μέσω μίας ακμής) με τον κόμβο i.

Βλέπε επίσης: δίκτυο ομοσπονδιακής μάθησης.

γειτονιά Η γειτονιά ενός κόμβου $i \in \mathcal{V}$ είναι το υποσύνολο κόμβων που αποτελούνται από τους γείτονες του i.

Βλέπε επίσης: γείτονες.

γενικευμένη ολική μεταβολή GTV (generalized total variation - GTV) is a measure of the variation of trained local models $h^{(i)}$ (or their παράμετροι μοντέλου $\mathbf{w}^{(i)}$) assigned to the nodes $i = 1, \ldots, n$ of an undirected weighted graph $\mathcal{G}$ with edges $\mathcal{E}$. Given a measure $d^{(h,h')}$ for the απόκλιση between υπόθεση maps h, h', the GTV is

$$\sum_{\{i,i'\} \in \mathcal{E}} A_{i,i'} d^{(h^{(i)},h^{(i')})}.$$

Here, $A_{i,i'} > 0$ denotes the weight of the undirected edge $\{i, i'\} \in \mathcal{E}$.

Βλέπε επίσης: local model, παράμετροι μοντέλου, graph, απόκλιση, υπόθεση, map.

γενίκευση Generalization refers to the ability of a model trained on a σύνολο εκπαίδευσης to make accurate πρόβλεψης on new, unseen data points. This is a central goal of ml and TN, i.e., to learn patterns that extend beyond the σύνολο εκπαίδευσης. Most ml systems use ελαχιστοποίηση εμπειρικής διακινδύνευσης to learn a υπόθεση $\hat{h} \in \mathcal{H}$ by minimizing the average loss over a σύνολο εκπαίδευσης of data points $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$, denoted $\mathcal{D}^{(\text{train})}$. However, success on the σύνολο εκπαίδευσης does not guarantee success on unseen data—this discrepancy is the challenge of generalization.

To study generalization mathematically, we need to formalize the notion of "unseen" data. A widely used approach is to assume a πιθανοτικό μοντέλο for data generation, such as the παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων. Here, we interpret data points as independent τυχαία μεταβλητήs with an identical κατανομή πιθανότητας $p(\mathbf{z})$. This κατανομή πιθανότητας, which is assumed fixed but unknown, allows us to define the διακινδύνευση of a trained model $\hat{h}$ as the expected loss

$$\bar{L}(\hat{h}) = \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})}\{L(\hat{h}, \mathbf{z})\}.$$

The difference between διακινδύνευση $\bar{L}(\hat{h})$ and empirical risk $\widehat{L}(\hat{h}|\mathcal{D}^{(\text{train})})$ is known as the generalization gap. Tools from probability theory, such as concentration inequalitys and uniform convergence, allow us to bound this gap under certain conditions [9].

Generalization without probability: Probability theory is one way to study how well a model generalizes beyond the σύνολο εκπαίδευσης, but it is not the only way. Another option is to use simple, deterministic changes to the data points in the σύνολο εκπαίδευσης. The basic idea is that a good model $\hat{h}$ should be robust, i.e., its πρόβλεψη $\hat{h}(\mathbf{x})$ should not change much if we slightly change the features $\mathbf{x}$ of a data point $\mathbf{z}$.

For example, an object detector trained on smartphone photos should still detect the object if a few random pixels are masked [111]. Similarly, it should deliver the same result if we rotate the object in the image [112].

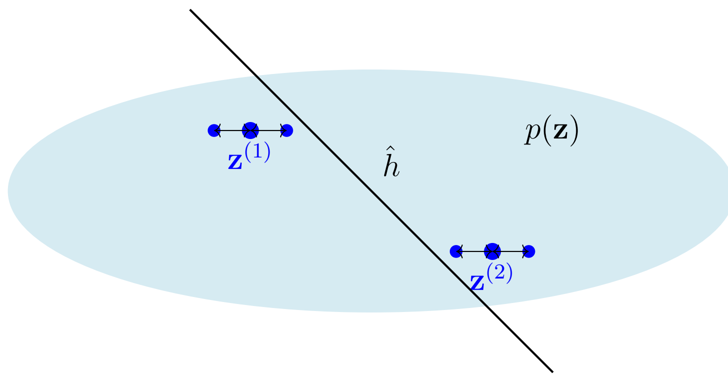


Fig. 7. Two data points $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}$ that are used as a σύνολο εκπαίδευσης to learn a υπόθεση $\hat{h}$ via ελαχιστοποίηση εμπειρικής διακινδύνευσης. We can evaluate $\hat{h}$ outside $\mathcal{D}^{(\text{train})}$ either by an παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων with some underlying κατανομή πιθανότητας $p(\mathbf{z})$ or by perturbing the data points.

Βλέπε επίσης: model, σύνολο εκπαίδευσης, πρόβλεψη, data point, ml, TN, ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπόθεση, loss, data, πιθανοτικό μοντέλο, παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων, τυχαία μεταβλητή, κατανομή πιθανότητας, διακινδύνευση, empirical risk, generalization gap, probability, concentration inequality, feature.

γενικός κανονισμός για την προστασία δεδομένων (ΓΚΠΔ) Ο ΓΚΠΔ (general data protection regulation - GDPR) θεσπίστηκε από την ΕΕ και τέθηκε σε ισχύ από τις 25 Μαΐου 2018 [18]. Διαφυλάσσει την ιδιωτικότητα και τα δικαιώματα δεδομένων των ατόμων στην ΕΕ.

Ο ΓΚΠΔ έχει σημαντικές επιπτώσεις για το πώς συλλέγονται δεδομένα, πώς αποθηκεύονται, και πώς χρησιμοποιούνται στις εφαρμογές μηχανικής μάθησης. Βασικές διατάξεις περιλαμβάνουν τα εξής:

- Αρχή της ελαχιστοποίησης των δεδομένων: Τα συστήματα μηχανικής μάθησης θα πρέπει να χρησιμοποιούν μόνο την απαραίτητη ποσότητα προσωπικών δεδομένων για τον σκοπό τους.
- Διαφάνεια και εξηγησιμότητα: Τα συστήματα μηχανικής μάθησης θα πρέπει να επιτρέπουν στους χρήστες τους να κατανοούν πώς τα συστήματα παίρνουν αποφάσεις που επηρεάζουν τους χρήστες.
- Δικαιώματα των υποκειμένων των δεδομένων: Οι χρήστες θα πρέπει να έχουν την ευκαιρία να έχουν πρόσβαση, να διορθώνουν, και να διαγράφουν τα προσωπικά τους δεδομένα, καθώς και να αντιτίθενται στην αυτοματοποιημένη λήψη αποφάσεων και στην κατάρτιση προφίλ.
- Λογοδοσία: Οι οργανισμοί πρέπει να εξασφαλίζουν την εύρωστη ασφάλεια δεδομένων και να αποδεικνύουν συμμόρφωση μέσω τεκμηρίων και τακτικών ελέγχων.

Βλέπε επίσης: data, ml, data minimization principle, transparency, εξηγησιμότητα.

γραμμικό μοντέλο Consider an ml application involving data points, each represented by a numeric διάνυσμα χαρακτηριστικών $\mathbf{x} \in \mathbb{R}^d$. A linear model defines a χώρος υποθέσεων consisting of all real-valued linear maps from $\mathbb{R}^d$ to $\mathbb{R}$ such that

$$\mathcal{H}^{(d)} := \left\{h : \mathbb{R}^d \to \mathbb{R} \mid h(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} \text{ for some } \mathbf{w} \in \mathbb{R}^d\right\}. \tag{4}$$

Each value of d defines a different χώρος υποθέσεων, corresponding to the number of features used to compute the πρόβλεψη $h(\mathbf{x})$. The choice of d is often guided by υπολογιστικές διαστάσεις (e.g., fewer features reduce computation), στατιστικές διαστάσεις (e.g., more features typically reduce μεροληψία and thus lower διακινδύνευση), and ερμηνευσιμότητα. A linear model using a small number of well-chosen features is generally considered more interpretable [26], [27]. The linear model is attractive because it can typically be trained using scalable convex μέθοδος βελτιστοποίησηςs [28], [29]. Moreover, linear models often permit rigorous statistical analysis, including fundamental limits on the ελάχιστο achievable διακινδύνευση [30]. They are also useful for analyzing more complex, non-linear models such as TNΔs. For instance, a βαθύ δίκτυο can be viewed as the composition of a χάρτης χαρακτηριστικών—implemented by the input and hidden layers—and a linear model in the output layer. Similarly, a decision tree can be interpreted as applying a one-hot encoded χάρτης χαρακτηριστικών based on περιοχή αποφάσεωνs, followed by a linear model that assigns a πρόβλεψη to each region. More generally, any trained model $\hat{h} \in \mathcal{H}$ that is παραγωγίσιμη at some $\mathbf{x}'$ can be locally approximated by a linear map $g(\mathbf{x})$. Figure 8 illustrates such a local linear approximation, defined by the gradient $\nabla \hat{h}(\mathbf{x}')$. Note that the gradient is only defined where $\hat{h}$ is παραγωγίσιμη. To ensure ευρωστία in the context of αξιόπιστη τεχνητή νοημοσύνη (αξιόπιστη TN), one may prefer models whose associated map $\hat{h}$ is Lipschitz continuous. A classic result in mathematical analysis—Rademacher's Theorem—states that if $\hat{h}$ is Lipschitz continuous with some constant L over an open set $\Omega \subseteq \mathbb{R}^d$, then $\hat{h}$ is παραγωγίσιμη almost everywhere in Ω [31, Th. 3.1].

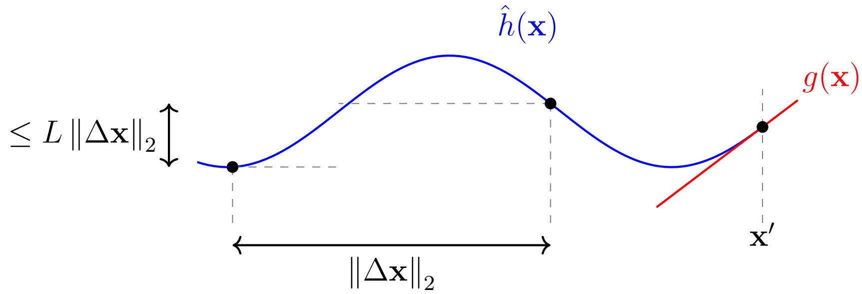


Fig. 8. A trained model $\hat{h}(\mathbf{x})$ that is παραγωγίσιμη at a point $\mathbf{x}'$ can be locally approximated by a linear map $g \in \mathcal{H}^{(d)}$. This local approximation is determined by the gradient $\nabla\hat{h}(\mathbf{x}')$.

Βλέπε επίσης: ml, data point, διάνυσμα χαρακτηριστικών, model, χώρος υποθέσεων, linear map, feature, πρόβλεψη, υπολογιστικές διαστάσεις, στατιστικές διαστάσεις, μεροληψία, διακινδύνευση, ερμηνευσιμότητα, convex, μέθοδος βελτιστοποίησης, ελάχιστο, ΤΝΔ, βαθύ δίκτυο, χάρτης χαρακτηριστικών, decision tree, περιοχή αποφάσεων, παραγωγίσιμη, gradient, ευρωστία, αξιόπιστη ΤΝ, map, LIME.

γραμμική παλινδρόμηση Η γραμμική παλινδρόμηση στοχεύει να μάθει μία γραμμική map υπόθεσης για να προβλέψει μία αριθμητική ετικέτα με βάση τα αριθμητικά χαρακτηριστικά ενός σημείου δεδομένων. Η ποιότητα μίας γραμμικής map υπόθεσης μετράται χρησιμοποιώντας τη μέση απώλεια τετραγωνικού σφάλματος που προκαλείται σε ένα σύνολο σημείων δεδομένων με ετικέτες, στο οποίο αναφερόμαστε ως το σύνολο εκπαίδευσης.

Βλέπε επίσης: regression, υπόθεση, map, ετικέτα, feature, data point, απώλεια τετραγωνικού σφάλματος, σημείο δεδομένων με ετικέτα, σύνολο εκπαίδευσης.

γραμμικός ταξινομητής Θεωρούμε σημεία δεδομένων που χαρακτηρίζονται από αριθμητικά χαρακτηριστικά $\mathbf{x} \in \mathbb{R}^d$ και μία ετικέτα $y \in \mathcal{Y}$ από κάποιον πεπερασμένο χώρο ετικετών $\mathcal{Y}$. Ένας γραμμικός ταξινομητής χαρακτηρίζεται από το γεγονός ότι έχει περιοχές αποφάσεων που διαχωρίζονται από υπερεπίπεδα στο $\mathbb{R}^d$ [8, Κεφ. 2].

Βλέπε επίσης: data point, feature, ετικέτα, χώρος ετικετών, ταξινομητής, περιοχή αποφάσεων.

γράφος Ένας γράφος $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ είναι ένα ζεύγος που αποτελείται από ένα σύνολο κόμβων $\mathcal{V}$ και ένα σύνολο ακμών $\mathcal{E}$. Στην πιο γενική του μορφή, ένας γράφος προσδιορίζεται από μία map που αποδίδει σε κάθε ακμή $e \in \mathcal{E}$ ένα ζεύγος κόμβων [32]. Μία σημαντική οικογένεια γράφων είναι οι απλοί μη κατευθυνόμενοι γράφοι. Ένας απλός μη κατευθυνόμενος γράφος προκύπτει από την ταυτοποίηση κάθε ακμής $e \in \mathcal{E}$ με δύο διαφορετικούς κόμβους $\{i, i'\}$. Οι σταθμισμένοι γράφοι προσδιορίζουν επίσης αριθμητικά βάρη A_e για κάθε ακμή $e \in \mathcal{E}$.

Βλέπε επίσης: map, βάρη.

γράφος ομοιότητας Some ml applications generate data points that are related by a domain-specific notion of similarity. These similarities can be represented conveniently using a similarity graph $\mathcal{G} = \big(\mathcal{V} := \{1, \ldots, m\}, \mathcal{E}\big)$. The node $r \in \mathcal{V}$ represents the r-th data point. Two nodes are connected by an undirected edge if the corresponding data points are similar.

Βλέπε επίσης: ml, data point, graph.

δεδομένα Τα δεδομένα αναφέρονται σε αντικείμενα που φέρουν πληροφορίες. Αυτά τα αντικείμενα μπορεί να είναι συγκεκριμένα φυσικά αντικείμενα (όπως άνθρωποι ή ζώα) ή αφηρημένες έννοιες (όπως αριθμοί). Συχνά χρησιμοποιούμε αναπαραστάσεις (ή προσεγγίσεις) των αρχικών δεδομένων που είναι πιο βολικές για την επεξεργασία των δεδομένων. Αυτές οι προσεγγίσεις βασίζονται σε διαφορετικά μοντέλα δεδομένων, με το σχεσιακό μοντέλο δεδομένων να είναι ένα από τα πιο ευρέως χρησιμοποιούμενα [33].

Βλέπε επίσης: model.

δείγμα Μία πεπερασμένη ακολουθία (ή λίστα) σημείων δεδομένων $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$ που προκύπτει ή ερμηνεύεται ως η πραγμάτωση m ανεξάρτητων και ταυτόσημα κατανεμημένων τυχαίων μεταβλητών με κοινή κατανομή πιθανότητας $p(\mathbf{z})$. Το μήκος m της ακολουθίας αναφέρεται ως το μέγεθος δείγματος.

Βλέπε επίσης: data point, πραγμάτωση, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, κατανομή πιθανότητας, μέγεθος δείγματος.

δέντρο αποφάσεων A decision tree is a flow-chart-like representation of a υπόθεση map h. More formally, a decision tree is a directed graph containing a root node that reads in the διάνυσμα χαρακτηριστικών $\mathbf{x}$ of a data point. The root node then forwards the data point to one of its children nodes based on some elementary test on the features $\mathbf{x}$. If the receiving child node is not a leaf node, i.e., it has itself children nodes, it represents another test. Based on the test result, the data point is forwarded to one of its descendants. This testing and forwarding of the data point is continued until the data point ends up in a leaf node without any children.

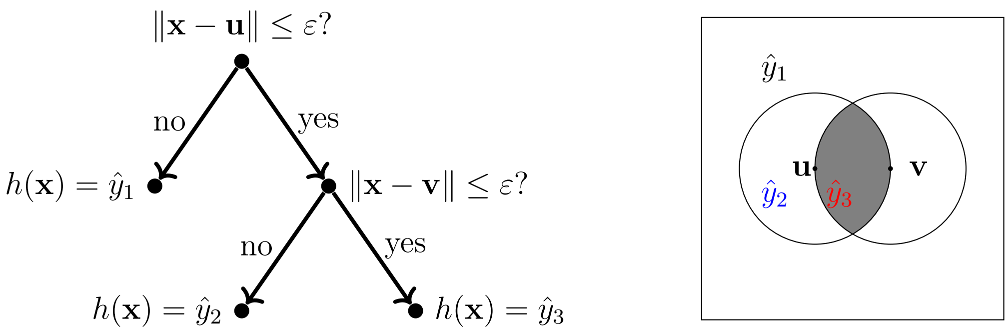


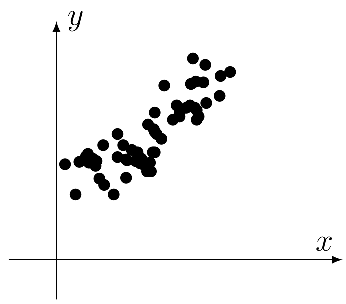
Fig. 9. Left: A decision tree is a flow-chart-like representation of a piecewise constant υπόθεση $h : \mathcal{X} \to \mathbb{R}$. Each piece is a περιοχή αποφάσεων $\mathcal{R}_{\hat{y}} := \{\mathbf{x} \in \mathcal{X} : h(\mathbf{x}) = \hat{y}\}$. The depicted decision tree can be applied to numeric διάνυσμα χαρακτηριστικώνs, i.e., $\mathcal{X} \subseteq \mathbb{R}^d$. It is parametrized by the threshold $\varepsilon > 0$ and the vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. Right: A decision tree partitions the χώρος χαρακτηριστικών $\mathcal{X}$ into περιοχή αποφάσεωνs. Each περιοχή αποφάσεων $\mathcal{R}_{\hat{y}} \subseteq \mathcal{X}$ corresponds to a specific leaf node in the decision tree.

Βλέπε επίσης: υπόθεση, map, graph, διάνυσμα χαρακτηριστικών, data point, feature, περιοχή αποφάσεων, χώρος χαρακτηριστικών.

δέσμη Στο πλαίσιο της στοχαστικής καθόδου κλίσης, μία δέσμη αναφέρεται σε ένα τυχαία επιλεγμένο υποσύνολο του γενικού συνόλου εκπαίδευσης. Χρησιμοποιούμε τα σημεία δεδομένων σε αυτό το υποσύνολο για να εκτιμήσουμε την κλίση του σφάλματος εκπαίδευσης και, με τη σειρά του, να ενημερώσουμε τις παραμέτρους μοντέλου.

Βλέπε επίσης: στοχαστική κάθοδος κλίσης, σύνολο εκπαίδευσης, data point, gradient, training error, παράμετροι μοντέλου.

διάγραμμα διασποράς Μία τεχνική οπτικοποίησης που απεικονίζει σημεία δεδομένων με σημεία σε ένα δισδιάστατο επίπεδο. Το Σχ. 10 απεικονίζει ένα παράδειγμα ενός διαγράμματος διασποράς.



Σχ. 10. Ένα διάγραμμα διασποράς κάποιων σημείων δεδομένων που αντιπροσωπεύουν καθημερινές καιρικές συνθήκες στη Φινλανδία. Κάθε σημείο δεδομένων χαρακτηρίζεται από την ελάχιστη θερμοκρασία της ημέρας x ως το χαρακτηριστικό του και τη μέγιστη θερμοκρασία της ημέρας y ως την ετικέτα του. Οι θερμοκρασίες έχουν μετρηθεί στον σταθμό καιρού του Φινλανδικού Μετεωρολογικού Ινστιτούτου στο Ελσίνκι Καισανιεμι κατά την περίοδο 1.9.2024 - 28.10.2024.

Ένα διάγραμμα διασποράς μπορεί να επιτρέψει τον οπτικό έλεγχο σημείων δεδομένων που αναπαριστώνται φυσικά από διανύσματα χαρακτηριστικών σε χώρους υψηλής διάστασης.

Βλέπε επίσης: data point, ελάχιστο, feature, maximum, ετικέτα, Φινλανδικό Μετεωρολογικό Ινστιτούτο, διάνυσμα χαρακτηριστικών, μείωση της διαστασιμότητας.

διακινδύνευση Θεωρούμε μία υπόθεση h που χρησιμοποιείται για να προβλεφθεί η ετικέτα y ενός σημείου δεδομένων βάσει των χαρακτηριστικών $\mathbf{x}$. Μετράμε την ποιότητα μίας συγκεκριμένης πρόβλεψης χρησιμοποιώντας μία συνάρτηση απώλειας $L((\mathbf{x}, y), h)$. Αν ερμηνεύσουμε τα σημεία δεδομένων ως τις πραγματώσεις ανεξάρτητων και ταυτόσημα κατανεμημένων τυχαίων μεταβλητών, τότε και η $L((\mathbf{x}, y), h)$ γίνεται η πραγμάτωση μίας τυχαίας μεταβλητής. Η παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων μας επιτρέπει να ορίσουμε τη διακινδύνευση μίας υπόθεσης ως την αναμενόμενη απώλεια $\mathbb{E}\big\{L((\mathbf{x}, y), h)\big\}$. Σημείωση ότι η διακινδύνευση της h εξαρτάται τόσο από την συγκεκριμένη επιλογή για την συνάρτηση απώλειας όσο και από την κατανομή πιθανότητας των σημείων δεδομένων. Βλέπε επίσης: υπόθεση, ετικέτα, data point, feature, πρόβλεψη, συνάρτηση απώλειας, πραγμάτωση, ανεξάρτητες και ταυτόσημα κατανεμημένες τυχαία μεταβλητή, παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων, loss, κατανομή πιθανότητας.

διακινδύνευση Bayes Θεωρούμε ένα πιθανοτικό μοντέλο με μία κοινή κατανομή πιθανότητας $p(\mathbf{x}, y)$ για τα χαρακτηριστικά $\mathbf{x}$ και την ετικέτα y ενός σημείου δεδομένων. Η διακινδύνευση Bayes (Bayes risk) είναι η ελάχιστη πιθανή διακινδύνευση που μπορεί να επιτευχθεί από οποιαδήποτε υπόθεση $h : \mathcal{X} \rightarrow \mathcal{Y}$. Οποιαδήποτε υπόθεση που επιτυγχάνει τη διακινδύνευση Bayes αναφέρεται ως μία εκτιμήτρια Bayes [22].
Βλέπε επίσης: πιθανοτικό μοντέλο, κατανομή πιθανότητας, feature, ετικέτα, data point, διακινδύνευση, ελάχιστο, υπόθεση, εκτιμήτρια Bayes.

διακύμανση Η διακύμανση μίας τυχαίας μεταβλητής πραγματικής τιμής x ορίζεται ως η προσδοκία $\mathbb{E}\big\{\big(x - \mathbb{E}\{x\}\big)^2\big\}$ της τετραγωνικής διαφοράς μεταξύ της x και της προσδοκίας της $\mathbb{E}\{x\}$. Επεκτείνουμε αυτόν τον ορισμό σε τυχαίες μεταβλητές διανυσματικής τιμής $\mathbf{x}$ ως $\mathbb{E}\big\{\big\|\mathbf{x} - \mathbb{E}\{\mathbf{x}\}\big\|_2^2\big\}$.
Βλέπε επίσης: τυχαία μεταβλητή, expectation.

διάνυσμα χαρακτηριστικών Το διάνυσμα χαρακτηριστικών αναφέρεται σε ένα διάνυσμα $\mathbf{x} = \left(x_1, \ldots, x_d\right)^T$ του οποίου οι καταχωρίσεις είναι ξεχωριστά χαρακτηριστικά $x_1, \ldots, x_d$. Πολλές μέθοδοι μηχανικής μάθησης χρησιμοποιούν διανύσματα χαρακτηριστικών που ανήκουν σε κάποιον Ευκλείδειο χώρο $\mathbb{R}^d$ πεπερασμένης διάστασης. Για κάποιες μεθόδους μηχανικής μάθησης, ωστόσο, μπορεί να είναι πιο βολικό να δουλεύουμε με διανύσματα χαρακτηριστικών που ανήκουν σε ένα διανυσματικό χώρο άπειρης διάστασης (π.χ. βλέπε τη μέθοδο πυρήνα).

Βλέπε επίσης: feature, ml, Ευκλείδειος χώρος, μέθοδος πυρήνα.

διαρροή ιδιωτικότητας Θεωρούμε μία εφαρμογή μηχανικής μάθησης που επεξεργάζεται ένα σύνολο δεδομένων $\mathcal{D}$ και δίνει κάποια έξοδο, όπως οι προβλέψεις που προκύπτουν για νέα σημεία δεδομένων. Διαρροή ιδιωτικότητας ανακύπτει αν η έξοδος φέρει πληροφορίες σχετικά με ένα ιδιωτικό (ή ευαίσθητο) χαρακτηριστικό ενός σημείου δεδομένων (που μπορεί να είναι άνθρωπος) ενός $\mathcal{D}$. Με βάση ένα πιθανοτικό μοντέλο για την παραγωγή δεδομένων, μπορούμε να μετρήσουμε τη διαρροή ιδιωτικότητας μέσω των αμοιβαίων πληροφοριών μεταξύ της εξόδου και του ευαίσθητου χαρακτηριστικού. Ένα άλλο ποιοτικό μέτρο διαρροής ιδιωτικότητας είναι η διαφορική ιδιωτικότητα. Οι σχέσεις μεταξύ διαφορετικών μέτρων διαρροής ιδιωτικότητας έχουν μελετηθεί στη βιβλιογραφία (βλέπε [34]).

Βλέπε επίσης: ml, σύνολο δεδομένων, πρόβλεψη, data point, feature, πιθανοτικό μοντέλο, data, αμοιβαίες πληροφορίες, διαφορική ιδιωτικότητα.

διασταυρούμενη επικύρωση k-συνόλων Η διασταυρούμενη επικύρωση k-συνόλων (k-fold cross-validation - k-fold CV) είναι μία μέθοδος για τη μάθηση και επικύρωση μίας υπόθεσης χρησιμοποιώντας ένα συγκεκριμένο σύνολο δεδομένων. Αυτή η μέθοδος διαιρεί το σύνολο δεδομένων ισότιμα σε k υποσύνολα και στη συνέχεια εκτελεί k επαναλήψεις εκπαίδευσης μοντέλου (π.χ. μέσω της ελαχιστοποίησης εμπειρικής διακινδύνευσης) και επικύρωσης. Κάθε επανάληψη χρησιμοποιεί ένα διαφορετικό υποσύνολο ως το σύνολο επικύρωσης και τα υπόλοιπα $k-1$ υποσύνολα ως σύνολο εκπαίδευσης. Η τελική έξοδος είναι ο μέσος όρος των σφαλμάτων επικύρωσης που προκύπτουν από τις k επαναλήψεις.

Βλέπε επίσης: υπόθεση, σύνολο δεδομένων, model, ελαχιστοποίηση εμπειρικής διακινδύνευσης, επικύρωση, σύνολο επικύρωσης, σύνολο εκπαίδευσης, σφάλμα επικύρωσης.

δίαυλος ιδιωτικότητας Ο δίαυλος ιδιωτικότητας είναι μία μέθοδος για τη μάθηση φιλικών προς την ιδιωτικότητα χαρακτηριστικών σημείων δεδομένων [35].

Βλέπε επίσης: feature, data point.

διαφάνεια Transparency is a fundamental requirement for αξιόπιστη ΤΝ [36]. In the context of ml methods, transparency is often used interchangeably with εξηγησιμότητα [37], [38]. However, in the broader scope of ΤΝ systems, transparency extends beyond εξηγησιμότητα and includes providing information about the system's limitations, reliability, and intended use. In medical diagnosis systems, transparency requires disclosing the confidence level for the πρόβλεψηs delivered by a trained model. In credit scoring, ΤΝ-based lending decisions should be accompanied by explanations of contributing factors, such as income level or credit history. These explanations allow humans (e.g., a loan applicant) to understand and contest automated decisions. Some ml methods inherently offer transparency. For example, λογιστική παλινδρόμηση provides a quantitative measure of ταξινόμηση reliability through the value $|h(\mathbf{x})|$. Decision trees are another example, as they allow human-readable decision rules [26]. Transparency also requires a clear indication when a user is engaging with an TN system. For example, TN-powered chatbots should notify users that they are interacting with an automated system rather than a human. Furthermore, transparency encompasses comprehensive documentation detailing the purpose and design choices underlying the TN system. For instance, model datasheets [39] and TN system cards [40] help practitioners understand the intended use cases and limitations of an TN system [41].

Βλέπε επίσης: αξιόπιστη TN, ml, εξηγησιμότητα, TN, πρόβλεψη, model, λογιστική παλινδρόμηση, ταξινόμηση, decision tree.

διαφορική ιδιωτικότητα Consider some ml method $\mathcal{A}$ that reads in a σύνολο δεδομένων (e.g., the σύνολο εκπαίδευσης used for ελαχιστοποίηση εμπειρικής διακινδύνευσης) and delivers some output $\mathcal{A}(\mathcal{D})$. The output could be either the learned παράμετροι μοντέλου or the πρόβλεψηs for specific data points. DP (differential privacy; DP) is a precise measure of διαρροή ιδιωτικότητας incurred by revealing the output. Roughly speaking, an ml method is differentially private if the κατανομή πιθανότητας of the output $\mathcal{A}(\mathcal{D})$ does not change too much if the ευαίσθητο ιδιοχαρακτηριστικό of one data point in the σύνολο εκπαίδευσης is changed. Note that DP builds on a πιθανοτικό μοντέλο for an ml method, i.e., we interpret its output $\mathcal{A}(\mathcal{D})$ as the πραγμάτωση of an τυχαία μεταβλητή. The randomness in the output can be ensured by intentionally adding the πραγμάτωση of an auxiliary τυχαία μεταβλητή (noise) to the output of the ml method.

Βλέπε επίσης: ml, σύνολο δεδομένων, σύνολο εκπαίδευσης, ελαχιστοποίηση εμπειρικής διακινδύνευσης, παράμετροι μοντέλου, πρόβλεψη, data point, διαρροή ιδιωτικότητας, κατανομή πιθανότητας, ευαίσθητο ιδιοχαρακτηριστικό, πιθανοτικό μοντέλο, πραγμάτωση, τυχαία μεταβλητή.

διεπαφή προγραμματισμού εφαρμογών An API (application programming interface; API) is a formal mechanism that allows software components to interact in a structured and modular way [42]. In the context of ml, APIs are commonly used to provide access to a trained ml model. Users—whether humans or machines—can submit the διάνυσμα χαρακτηριστικών of a data point and receive a corresponding πρόβλεψη. Suppose a trained ml model is defined as $\widehat{h}(x) := 2x + 1$. Through an API, a user can input $x = 3$ and receive the output $\widehat{h}(3) = 7$ without knowledge of the detailed structure of the ml model or its training. In practice, the model is typically deployed on a server connected to the internet. Clients send requests containing feature values to the server, which responds with the computed πρόβλεψη $\widehat{h}(\mathbf{x})$. APIs promote modularity in ml system design, i.e., one team can develop and train the model, while another team handles integration and user interaction. Publishing a trained model via an API also offers practical advantages:

- The server can centralize computational resources which are required to compute πρόβλεψηs.

- The internal structure of the model remains hidden—which is useful for protecting intellectual property or trade secrets.

However, APIs are not without διακινδύνευση. Techniques such as model inversion can potentially reconstruct a model from its πρόβλεψηs on carefully selected διάνυσμα χαρακτηριστικώνs.
Βλέπε επίσης: ml, model, διάνυσμα χαρακτηριστικών, data point, πρόβλεψη, feature, model inversion.

δίκτυο ομοσπονδιακής μάθησης An FL network (federated learning network - FL network) is an undirected weighted graph whose nodes represent data generators that aim to train a local (or personalized) model. Each node in an FL network represents some συσκευή capable of collecting a τοπικό σύνολο δεδομένων and, in turn, train a local model. FL methods learn a local υπόθεση $h^{(i)}$, for each node $i \in \mathcal{V}$, such that it incurs small loss on the τοπικό σύνολο δεδομένωνs.
Βλέπε επίσης: FL, graph, data, model, συσκευή, τοπικό σύνολο δεδομένων, local model, υπόθεση, loss.

δομημένη ελαχιστοποίηση διακινδύνευσης Η δομημένη ελαχιστοποίηση διακινδύνευσης (structural risk minimization - SRM) είναι μία περίπτωση ομαλοποιημένης ελαχιστοποίησης εμπειρικής διακινδύνευσης, με την οποία το μοντέλο $\mathcal{H}$ μπορεί να εκφραστεί ως μία μετρήσιμη ένωση υπομοντέλων: $\mathcal{H} = \bigcup_{n=1}^{\infty} \mathcal{H}^{(n)}$. Κάθε υπομοντέλο $\mathcal{H}^{(n)}$ επιτρέπει την παραγώγιση ενός προσεγγιστικού άνω φράγματος στο σφάλμα γενίκευσης που προκαλείται κατά την εφαρμογή ελαχιστοποίησης εμπειρικής διακινδύνευσης για την εκπαίδευση του $\mathcal{H}^{(n)}$. Αυτά τα μεμονωμένα φράγματα—ένα για κάθε υπομοντέλο—συνδυάζονται έπειτα για να σχηματίσουν έναν ομαλοποιητή που χρησιμοποιείται στον στόχο ομαλοποιημένης ελαχιστοποίησης εμπειρικής διακινδύνευσης. Αυτά τα προσεγγιστικά άνω φράγματα (ένα για κάθε $\mathcal{H}^{(n)}$) συνδυάζονται στη συνέχεια για να κατασκευάσουν έναν ομαλοποιητή για την ομαλοποιημένη ελαχιστοποίηση εμπειρικής διακινδύνευσης [9, Sec. 7.2].

Βλέπε επίσης: ομαλοποιημένη ελαχιστοποίηση εμπειρικής διακινδύνευσης, model, γενίκευση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, ομαλοποιητής, διακινδύνευση.

εγγύς τελεστής Given a convex συνάρτηση $f(\mathbf{w}')$, we define its proximal operator as [25], [43]

$$\mathbf{prox}_{f(\cdot),\rho}(\mathbf{w}) := \operatorname*{arg\,min}_{\mathbf{w}' \in \mathbb{R}^d} \left[f(\mathbf{w}') + \frac{\rho}{2} \left\| \mathbf{w} - \mathbf{w}' \right\|_2^2 \right] \text{ with } \rho > 0.$$

As illustrated in Fig. 11, evaluating the proximal operator amounts to minimizing a penalized variant of $f(\mathbf{w}')$. The penalty term is the scaled squared Euclidean distance to a given vector $\mathbf{w}$ (which is the input to the proximal operator). The proximal operator can be interpreted as a γενίκευση of the βήμα κλίσης, which is defined for a λεία convex συνάρτηση $f(\mathbf{w}')$. Indeed, taking a βήμα κλίσης with μέγεθος βήματος η at the current vector $\mathbf{w}$ is the same as applying the proximal operator of the συνάρτηση $\tilde{f}(\mathbf{w}') = \big(\nabla f(\mathbf{w})\big)^T (\mathbf{w}' - \mathbf{w})$ and using $\rho = 1/\eta$.

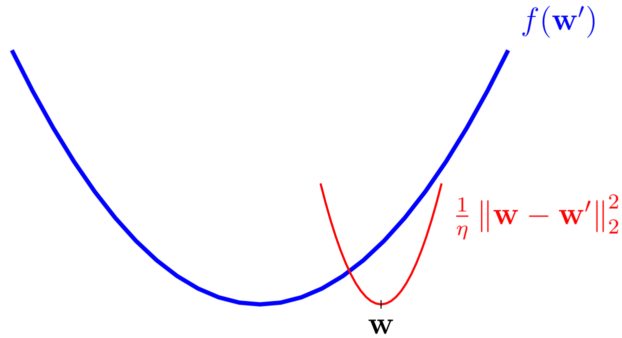


Fig. 11. A generalized βήμα κλίσης updates a vector $\mathbf{w}$ by minimizing a penalized version of the συνάρτηση $f(\cdot)$. The penalty term is the scaled squared Euclidean distance between the optimization variable $\mathbf{w}'$ and the given vector $\mathbf{w}$.

Βλέπε επίσης: convex, συνάρτηση, γενίκευση, βήμα κλίσης, λεία, μέγεθος βήματος.

εκκίνηση Για την ανάλυση μεθόδων μηχανικής μάθησης, είναι συχνά χρήσιμο να ερμηνεύουμε ένα συγκεκριμένο σύνολο σημείων δεδομένων $\mathcal{D} = \{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}\}$ ως πραγματώσεις ανεξάρτητων και ταυτόσημα κατανεμημένων τυχαίων μεταβλητών με κοινή κατανομή πιθανότητας $p(\mathbf{z})$. Γενικά, δεν γνωρίζουμε την $p(\mathbf{z})$ ακριβώς, αλλά χρειάζεται να την εκτιμήσουμε. Η εκκίνηση χρησιμοποιεί το ιστόγραμμα του $\mathcal{D}$ ως μία εκτιμήτρια για την υποκείμενη κατανομή πιθανότητας $p(\mathbf{z})$.

Βλέπε επίσης: ml, data point, πραγμάτωση, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, κατανομή πιθανότητας, ιστόγραμμα.

εκτιμήτρια Bayes Θεωρούμε ένα πιθανοτικό μοντέλο με μία από κοινού κατανομή πιθανότητας $p(\mathbf{x}, y)$ για τα χαρακτηριστικά $\mathbf{x}$ και την ετικέτα y ενός σημείου δεδομένων. Για μία δεδομένη συνάρτηση απώλειας $L(\cdot, \cdot)$, αναφερόμαστε σε μία υπόθεση h ως μία εκτιμήτρια Bayes αν η διακινδύνευσή της $\mathbb{E}\{L((\mathbf{x}, y), h)\}$ είναι το ελάχιστο [22]. Σημείωση ότι η ιδιότητα μίας υπόθεσης να είναι μία εκτιμήτρια Bayes εξαρτάται από την υποκείμενη κατανομή πιθανότητας και την επιλογή για την συνάρτηση απώλειας $L(\cdot, \cdot)$.

Βλέπε επίσης: πιθανοτικό μοντέλο, κατανομή πιθανότητας, feature, ετικέτα, data point, συνάρτηση απώλειας, υπόθεση, διακινδύνευση, ελάχιστο.

ελάχιστο Δεδομένου ενός συνόλου πραγματικών αριθμών, το ελάχιστο είναι ο μικρότερος από αυτούς τους αριθμούς. Σημείωση ότι για κάποια σύνολα, όπως το σύνολο αρνητικών πραγματικών αριθμών, το ελάχιστο δεν υφίσταται.

ελάχιστο άνω φράγμα (ή supremum) The supremum (supremum or least upper bound) of a set of real numbers is the smallest number that is greater than or equal to every element in the set. More formally, a real number a is the supremum of a set $\mathcal{A} \subseteq \mathbb{R}$ if: 1) a is an upper bound of $\mathcal{A}$; and 2) no number smaller than a is an upper bound of $\mathcal{A}$. Every non-empty set of real numbers that is bounded above has a supremum, even if it does not contain its supremum as an element [2, Sec. 1.4].

ελαχιστοποίηση εμπειρικής διακινδύνευσης Η ελαχιστοποίηση εμπειρικής διακινδύνευσης (empirical risk minimization - ERM) είναι το πρόβλημα βελτιστοποίησης της εύρεσης μίας υπόθεσης (από ένα μοντέλο) με την ελάχιστη μέση απώλεια (ή εμπειρική διακινδύνευση) σε ένα δεδομένο σύνολο δεδομένων $\mathcal{D}$ (δηλαδή το σύνολο εκπαίδευσης). Πολλές μέθοδοι μηχανικής μάθησης προκύπτουν από εμπειρική διακινδύνευση μέσω συγκεκριμένων επιλογών σχεδιασμού για το σύνολο δεδομένων, το μοντέλο, και την απώλεια [8, Κεφ. 3].

Βλέπε επίσης: πρόβλημα βελτιστοποίησης, υπόθεση, model, ελάχιστο, loss, empirical risk, σύνολο δεδομένων, σύνολο εκπαίδευσης, ml.

εμπειρική διακινδύνευση Η εμπειρική διακινδύνευση $\widehat{L}(h|\mathcal{D})$ μίας υπόθεσης σε ένα σύνολο δεδομένων $\mathcal{D}$ είναι η μέση απώλεια που προκαλείται από την h όταν εφαρμόζεται στα σημεία δεδομένων του $\mathcal{D}$.

Βλέπε επίσης: διακινδύνευση, υπόθεση, σύνολο δεδομένων, loss, data point.

εντροπία Στο πλαίσιο της μηχανικής μάθησης και της εξηγησιμότητας, η εντροπία ποσοτικοποιεί την αβεβαιότητα ή την τυχαιότητα στις προβλέψεις ενός μοντέλου. Συγκεκριμένα, η υπό συνθήκη (ή διαφορική) εντροπία μετράει τη μη προβλεψιμότητα των εξόδων του εκπαιδευμένου μοντέλου δεδομένης μίας άλλης πηγής προβλέψεων (π.χ. από έναν χρήστη που είναι άνθρωπος). Μία χαμηλότερη συνθήκη εντροπίας συνεπάγεται υψηλότερη εξηγησιμότητα, καθώς οι προβλέψεις του μοντέλου ευθυγραμμίζονται περισσότερο με την αντίληψη του χρήστη. Ωστόσο, ο όρος εντροπία χρησιμοποιείται επίσης σε άλλα πεδία με διαφορετικές σημασίες, όπως τη μέτρηση της αταξίας στη θερμοδυναμική ή την πολυπλοκότητας σε δυναμικά συστήματα.

Βλέπε επίσης: ml, εξηγησιμότητα, αβεβαιότητα, model, πρόβλεψη, πιθανοτικό μοντέλο.

εξήγηση Μία προσέγγιση για να καταστούν μέθοδοι της μηχανικής μάθησης

διαφανείς είναι να παρέχεται μία επεξήγηση μαζί με την πρόβλεψη που παραδίδεται από μία μέθοδο μηχανικής μάθησης. Οι επεξηγήσεις μπορούν να πάρουν πολλές διαφορετικές μορφές. Μία επεξήγηση θα μπορούσε να είναι κάποιο φυσικό κείμενο ή κάποιο ποσοτικό μέτρο για τη σημασία ξεχωριστών χαρακτηριστικών ενός σημείου δεδομένων [44]. Μπορούμε επίσης να χρησιμοποιήσουμε οπτικές μορφές επεξηγήσεων, όπως διαγράμματα έντασης για την ταξινόμηση εικόνων [45].
Βλέπε επίσης: ml, πρόβλεψη, feature, data point, ταξινόμηση.

εξηγήσιμη ελαχιστοποίηση εμπειρικής διακινδύνευσης EERM (explainable empirical risk minimization - EERM) is an instance of δομημένη ελαχιστοποίηση διακινδύνευσης that adds a ομαλοποίηση term to the average loss in the αντικειμενική συνάρτηση of ελαχιστοποίηση εμπειρικής διακινδύνευσης. The ομαλοποίηση term is chosen to favor υπόθεση maps that are intrinsically explainable for a specific user. This user is characterized by their πρόβλεψηs provided for the data points in a σύνολο εκπαίδευσης [47].
Βλέπε επίσης: δομημένη ελαχιστοποίηση διακινδύνευσης, ομαλοποίηση, loss, αντικειμενική συνάρτηση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπόθεση, map, πρόβλεψη, data point, σύνολο εκπαίδευσης.

εξηγήσιμη μηχανική μάθηση XML (explainable machine learning - XML) methods aim at complementing each πρόβλεψη with an εξήγηση of how the πρόβλεψη has been obtained. The construction of an explicit εξήγηση might not be necessary if the ml method uses a sufficiently simple (or interpretable) model [26].

Βλέπε επίσης: πρόβλεψη, εξήγηση, ml, model.

εξηγησιμότητα Ορίζουμε την (υποκειμενική) επεξηγησιμότητα μίας μεθόδου μηχανικής μάθησης ως το επίπεδο προσομοιωσιμότητας [46] των προβλέψεων που παραδίδονται από ένα σύστημα μηχανικής μάθησης σε έναν χρήστη που είναι άνθρωπος. Ποσοτικά μέτρα για την (υποκειμενική) επεξηγησιμότητα ενός εκπαιδευμένου μοντέλου μπορούν να κατασκευαστούν συγκρίνοντας τις προβλέψεις του με τις προβλέψεις που παρέχονται από έναν χρήστη σε ένα σύνολο ελέγχου [46], [47]. Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε πιθανοτικά μοντέλα για δεδομένα και να μετρήσουμε την επεξηγησιμότητα ενός εκπαιδευμένου μοντέλου μηχανικής μάθησης μέσω της υπό συνθήκης (διαφορικής) εντροπίας των προβλέψεών του, δεδομένων των προβλέψεων του χρήστη [37], [48].

Βλέπε επίσης: ml, πρόβλεψη, model, σύνολο ελέγχου, πιθανοτικό μοντέλο, data, εντροπία.

επαύξηση δεδομένων Data augmentation methods add synthetic data points to an existing set of data points. These synthetic data points are obtained by perturbations (e.g., adding noise to physical measurements) or transformations (e.g., rotations of images) of the original data points. These perturbations and transformations are such that the resulting synthetic data points should still have the same ετικέτα. As a case in point, a rotated cat image is still a cat image even if their διάνυσμα χαρακτηριστικώνs (obtained by stacking pixel color intensities) are very different (see Fig. 12). Data augmentation can be an efficient form of ομαλοποίηση.

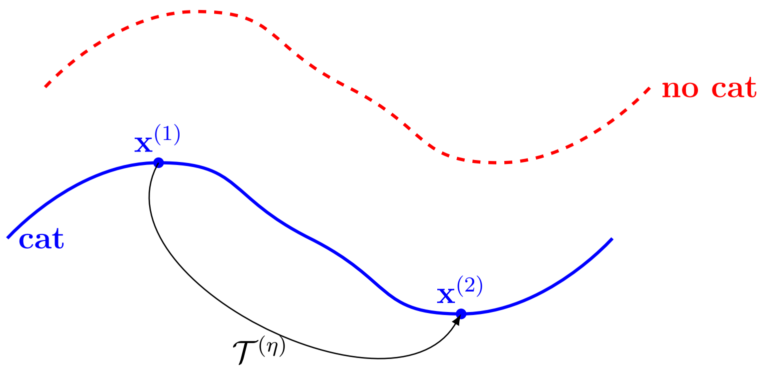


Fig. 12. Data augmentation exploits intrinsic symmetries of data points in some χώρος χαρακτηριστικών $\mathcal{X}$. We can represent a symmetry by an operator $\mathcal{T}^{(\eta)} : \mathcal{X} \to \mathcal{X}$, parametrized by some number $\eta \in \mathbb{R}$. For example, $\mathcal{T}^{(\eta)}$ might represent the effect of rotating a cat image by η degrees. A data point with διάνυσμα χαρακτηριστικών $\mathbf{x}^{(2)} = \mathcal{T}^{(\eta)}\big(\mathbf{x}^{(1)}\big)$ must have the same ετικέτα $y^{(2)} = y^{(1)}$ as a data point with διάνυσμα χαρακτηριστικών $\mathbf{x}^{(1)}$.

Βλέπε επίσης: data, data point, ετικέτα, διάνυσμα χαρακτηριστικών, ομαλοποίηση, χώρος χαρακτηριστικών.

επίθεση An attack on an ml system refers to an intentional action—either active or passive—that compromises the system's integrity, availability, or confidentiality. Active attacks involve perturbing components such as σύνολο δεδομένωνs (via data poisoning) or communication links between συσκευήs in a FL setting. Passive attacks, such as επίθεση της ιδιωτικότητας, aim to infer ευαίσθητο ιδιοχαρακτηριστικόs without modifying the system. Depending on their goal, we distinguish between επίθεση άρνησης υπηρεσιώνs, κερκόπορτα attacks, and επίθεση της ιδιωτικότηταςs.

Βλέπε επίσης: ml, σύνολο δεδομένων, data poisoning, συσκευή, FL, επίθεση της ιδιωτικότητας, ευαίσθητο ιδιοχαρακτηριστικό, επίθεση άρνησης υπηρεσιών, κερκόπορτα.

επίθεση άρνησης υπηρεσιών Μία επίθεση άρνησης υπηρεσιών στοχεύει (π.χ. μέσω data poisoning) να κατευθύνει την εκπαίδευση ενός μοντέλου, έτσι ώστε να έχει χαμηλή επίδοση για τυπικά σημεία δεδομένων.

Βλέπε επίσης: data poisoning, model, data point.

επίθεση της ιδιωτικότητας A privacy επίθεση on an ml system aims to infer ευαίσθητο ιδιοχαρακτηριστικόs of individuals by exploiting partial access to a trained ml model. One form of a privacy επίθεση user1 is model inversion.

Βλέπε επίσης: επίθεση user1, ml, ευαίσθητο ιδιοχαρακτηριστικό, model, model inversion, αξιόπιστη ΤΝ, γενικός κανονισμός για την προστασία δεδομένων (ΓΚΠΔ).

επικύρωση Θεωρούμε μία υπόθεση $\hat{h}$ που έχει μαθευτεί μέσω κάποιας μεθόδου μηχανικής μάθησης, π.χ. λύνοντας την ελαχιστοποίηση εμπειρικής διακινδύνευσης σε ένα σύνολο εκπαίδευσης $\mathcal{D}$. Η επικύρωση αναφέρεται στην πρακτική της αξιολόγησης της απώλειας που προκαλείται από την υπόθεση $\hat{h}$ σε ένα σύνολο σημείων δεδομένων που δεν περιέχονται στο σύνολο εκπαίδευσης $\mathcal{D}$.

Βλέπε επίσης: υπόθεση, ml, ελαχιστοποίηση εμπειρικής διακινδύνευσης, σύνολο εκπαίδευσης, loss, data point.

επιλογή μοντέλου In ml, model selection refers to the process of choosing between different candidate models. In its most basic form, model selection amounts to: 1) training each candidate model; 2) computing the σφάλμα επικύρωσης for each trained model; and 3) choosing the model with the smallest σφάλμα επικύρωσης [8, Ch. 6].

Βλέπε επίσης: ml, model, σφάλμα επικύρωσης.

εργασία μάθησης Consider a σύνολο δεδομένων $\mathcal{D}$ constituted by several data points, each of them characterized by features $\mathbf{x}$. For example, the σύνολο δεδομένων $\mathcal{D}$ might be constituted by the images of a particular database. Sometimes it might be useful to represent a σύνολο δεδομένων $\mathcal{D}$, along with the choice of features, by a κατανομή πιθανότητας $p(\mathbf{x})$. A learning task associated with $\mathcal{D}$ consists of a specific choice for the ετικέτα of a data point and the corresponding χώρος ετικετών. Given a choice for the συνάρτηση απώλειας and model, a learning task gives rise to an instance of ελαχιστοποίηση εμπειρικής διακινδύνευσης. Thus, we could define a learning task also via an instance of ελαχιστοποίηση εμπειρικής διακινδύνευσης, i.e., via an αντικειμενική συνάρτηση. Note that, for the same σύνολο δεδομένων, we obtain different learning tasks by using different choices for the features and ετικέτα of a data point. These learning tasks are related, as they are based on the same σύνολο δεδομένων, and solving them jointly (via μάθηση πολυδιεργασίας methods) is typically preferable over solving them separately [49], [50], [51].

Βλέπε επίσης: σύνολο δεδομένων, data point, feature, κατανομή πιθανότητας, ετικέτα, χώρος ετικετών, συνάρτηση απώλειας, model, ελαχιστοποίηση εμπειρικής διακινδύνευσης, αντικειμενική συνάρτηση, μάθηση πολυδιεργασίας.

ερμηνευσιμότητα Μία μέθοδος μηχανικής μάθησης είναι ερμηνεύσιμη για έναν συγκεκριμένο χρήστη αν μπορεί να αναμένει τις προβλέψεις που παραδίδονται από τη μέθοδο. Η έννοια της ερμηνευσιμότητας μπορεί να γίνει ακριβής με τη χρήση ποσοτικών μέτρων της αβεβαιότητας σχετικά με τις προβλέψεις [37].

Βλέπε επίσης: ml, πρόβλεψη, αβεβαιότητα.

ετικέτα Ένα υψηλότερου επιπέδου γεγονός ή ποσότητα ενδιαφέροντος που σχετίζεται με ένα σημείο δεδομένων. Για παράδειγμα, αν ένα σημείο δεδομένων είναι μία εικόνα, η ετικέτα θα μπορούσε να υποδεικνύει αν η εικόνα περιέχει μία γάτα ή όχι. Συνώνυμα του όρου ετικέτα, που χρησιμοποιούνται συχνά σε συγκεκριμένους τομείς, περιλαμβάνουν «μεταβλητή απόκρισης,» «μεταβλητή εξόδου,» και «στόχος» [52], [53], [54].

Βλέπε επίσης: data point.

ευαίσθητο ιδιοχαρακτηριστικό Η μηχανική μάθηση περιστρέφεται γύρω από τη μάθηση μίας map υπόθεσης που μας επιτρέπει να προβλέψουμε την ετικέτα ενός σημείου δεδομένων από τα χαρακτηριστικά του. Σε κάποιες εφαρμογές, πρέπει να εξασφαλίσουμε ότι η έξοδος που παραδίδεται από ένα σύστημα μηχανικής μάθησης δεν μας επιτρέπει να συμπεράνουμε ευαίσθητα ιδιοχαρακτηριστικά ενός σημείου δεδομένων. Ποιο μέρος ενός σημείου δεδομένων θεωρείται ευαίσθητο ιδιοχαρακτηριστικό είναι μία επιλογή σχεδιασμού που ποικίλλει μεταξύ διαφορετικών τομέων εφαρμογής.

Βλέπε επίσης: ml, υπόθεση, map, ετικέτα, data point, feature.

Ευκλείδειος χώρος Ο Ευκλείδειος χώρος $\mathbb{R}^d$ διάστασης $d \in \mathbb{N}$ αποτελείται από διανύσματα $\mathbf{x} = \big(x_1, \ldots, x_d\big)$, με d καταχωρίσεις πραγματικής τιμής $x_1, \ldots, x_d \in \mathbb{R}$. Ένας τέτοιος Ευκλείδειος χώρος είναι εξοπλισμένος με μία γεωμετρική δομή που ορίζεται από το εσωτερικό γινόμενο $\mathbf{x}^T\mathbf{x}' =$

$\sum_{j=1}^{d} x_j x'_j$ μεταξύ οποιωνδήποτε δύο διανυσμάτων $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ [2].

ευρωστία Robustness is a key requirement for αξιόπιστη TN. It refers to the property of an ml system to maintain acceptable performance even when subjected to different forms of perturbations. These perturbations can be to the features of a data point in order to manipulate the πρόβλεψη delivered by a trained ml model. Robustness also includes the stability of ελαχιστοποίηση εμπειρικής διακινδύνευσης-based methods against perturbations of the σύνολο εκπαίδευσης. Such perturbations can occur within data poisoning επίθεση user1s.

Βλέπε επίσης: αξιόπιστη TN, ml, feature, data point, πρόβλεψη, model, stability, ελαχιστοποίηση εμπειρικής διακινδύνευσης, σύνολο εκπαίδευσης, data poisoning, επίθεση user1.

θετικά ημιορισμένος Ένας συμμετρικός (πραγματικών τιμών) πίνακας $\mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{d \times d}$ αναφέρεται ως θετικά ημιορισμένος (positive semi-definite - psd) αν $\mathbf{x}^T \mathbf{Q} \mathbf{x} \geq 0$ για κάθε διάνυσμα $\mathbf{x} \in \mathbb{R}^d$. Η ιδιότητα του να είναι θετικά ημιορισμένος μπορεί να επεκταθεί από πίνακες σε συμμετρικές (πραγματικών τιμών) maps πυρήνα $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (με $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}', \mathbf{x})$) ως εξής: Για οποιοδήποτε πεπερασμένο σύνολο διανυσμάτων χαρακτηριστικών $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$, ο επακόλουθος πίνακας $\mathbf{Q} \in \mathbb{R}^{m \times m}$ με καταχωρίσεις $Q_{r,r'} = K\big(\mathbf{x}^{(r)}, \mathbf{x}^{(r')}\big)$ είναι θετικά ημιορισμένος [55].

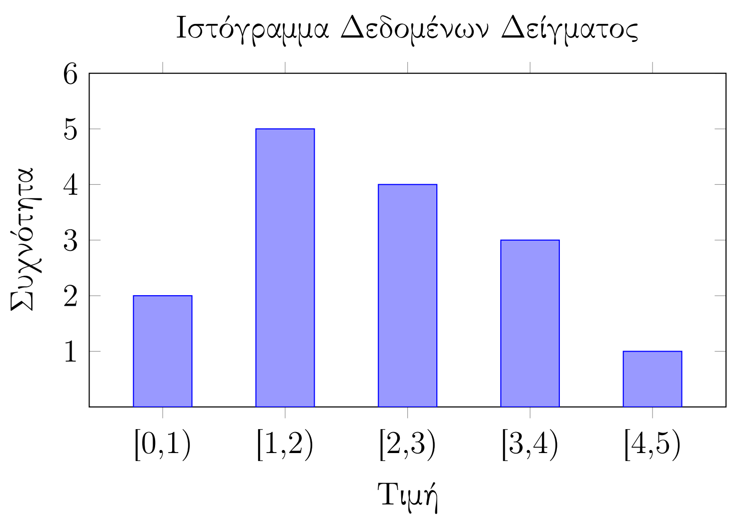
Βλέπε επίσης: πυρήνας, map, διάνυσμα χαρακτηριστικών.

ιδιοδιάνυσμα Ένα ιδιοδιάνυσμα ενός πίνακα $\mathbf{A} \in \mathbb{R}^{d \times d}$ είναι ένα μη μηδενικό διάνυσμα $\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ τέτοιο ώστε $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$ με κάποια ιδιοτιμή λ.

Βλέπε επίσης: ιδιοτιμή.

ιδιοτιμή Αναφερόμαστε σε έναν αριθμό $\lambda \in \mathbb{R}$ ως μία ιδιοτιμή ενός τετραγωνικού πίνακα $\mathbf{A} \in \mathbb{R}^{d \times d}$ αν υπάρχει ένα μη μηδενικό διάνυσμα $\mathbf{x} \in \mathbb{R}^{d} \setminus \{\mathbf{0}\}$ τέτοιο ώστε $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$.

ιστόγραμμα Θεωρούμε ένα σύνολο δεδομένων $\mathcal{D}$ που αποτελείται από m σημεία δεδομένων $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$, καθένα από τα οποία ανήκει σε κάποιο κελί $[-U, U] \times \ldots \times [-U, U] \subseteq \mathbb{R}^{d}$ με πλάγιο μήκος U. Χωρίζουμε αυτό το κελί ισότιμα σε μικρότερα στοιχειώδη κελιά με πλάγιο μήκος Δ. Το ιστόγραμμα του $\mathcal{D}$ αποδίδει κάθε στοιχειώδες κελί στο αντίστοιχο κλάσμα των σημεία δεδομένων του $\mathcal{D}$ που εμπίπτουν σε αυτό το στοιχειώδες κελί. Ένα οπτικό παράδειγμα ενός τέτοιο ιστογράμματος παρέχεται στο Σχ. 13.



Σχ. 13. Ένα ιστόγραμμα που αναπαριστά τη συχνότητα των σημείων δεδομένων που εμπίπτουν εντός πεδίων διακριτών τιμών (δηλαδή κάδων). Το ύψος κάθε ράβδου δείχνει τον αριθμό των δειγμάτων στο αντίστοιχο διάστημα.

Βλέπε επίσης: σύνολο δεδομένων, data point, δείγμα.

κάθοδος κλίσης GD (gradient descent; GD) is an iterative method for finding the ελάχιστο of a παραγωγίσιμη συνάρτηση $f(\mathbf{w})$ of a vector-valued argument $\mathbf{w} \in \mathbb{R}^d$. Consider a current guess or approximation $\mathbf{w}^{(k)}$ for the ελάχιστο of the συνάρτηση $f(\mathbf{w})$. We would like to find a new (better) vector $\mathbf{w}^{(k+1)}$ that has a smaller objective value $f(\mathbf{w}^{(k+1)}) < f\big(\mathbf{w}^{(k)}\big)$ than the current guess $\mathbf{w}^{(k)}$. We can achieve this typically by using a βήμα κλίσης

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \eta \nabla f(\mathbf{w}^{(k)}) \tag{5}$$

with a sufficiently small μέγεθος βήματος $\eta > 0$. Fig. 14 illustrates the effect of a single GD step (5).

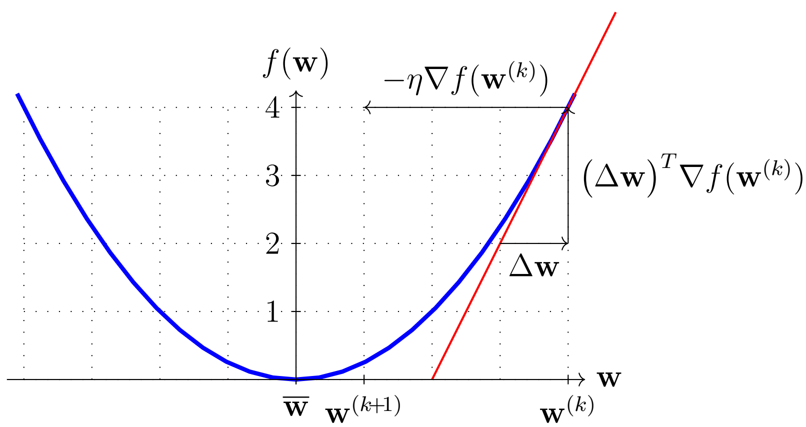


Fig. 14. A single βήμα κλίσης (5) towards the minimizer $\overline{\mathbf{w}}$ of $f(\mathbf{w})$.

Βλέπε επίσης: ελάχιστο, παραγωγίσιμη, συνάρτηση, βήμα κλίσης, μέγεθος βήματος, gradient.

κάθοδος υποκλίσης Subgradient descent is a γενίκευση of κάθοδος κλίσης that does not require differentiability of the συνάρτηση to be minimized. This γενίκευση is obtained by replacing the concept of a gradient with that of a subgradient. Similar to gradients, also subgradients allow us to construct local approximations of an αντικειμενική συνάρτηση. The αντικειμενική συνάρτηση might be the empirical risk $\widehat{L}\big(h^{(\mathbf{w})}\big|\mathcal{D}\big)$ viewed as a συνάρτηση of the παράμετροι μοντέλου $\mathbf{w}$ that select a υπόθεση $h^{(\mathbf{w})} \in \mathcal{H}$.

Βλέπε επίσης: subgradient, γενίκευση, κάθοδος κλίσης, συνάρτηση, gradient, αντικειμενική συνάρτηση, empirical risk, παράμετροι μοντέλου, υπόθεση.

κανονικοποίηση δεδομένων Η κανονικοποίηση δεδομένων αναφέρεται σε μετασχηματισμούς που εφαρμόζονται στα διανύσματα χαρακτηριστικών σημείων δεδομένων για να βελτιωθούν οι στατιστικές διαστάσεις ή οι υπολογιστικές διαστάσεις της μεθόδου μηχανικής μάθησης. Για παράδειγμα, στη γραμμική παλινδρόμηση με μεθόδους με βάση την κλίση που χρησιμοποιούν έναν σταθερό ρυθμό μάθησης, η σύγκλιση εξαρτάται από τον έλεγχο της νόρμας διανυσμάτων χαρακτηριστικών στο σύνολο εκπαίδευσης. Μία κοινή προσέγγιση είναι να κανονικοποιούμε τα διανύσματα χαρακτηριστικών, έτσι ώστε η νόρμα τους να μην υπερβαίνει το ένα [8, Κεφ. 5].

Βλέπε επίσης: data, διάνυσμα χαρακτηριστικών, data point, ml, στατιστικές διαστάσεις, υπολογιστικές διαστάσεις, γραμμική παλινδρόμηση, μέθοδοι με βάση την κλίση, ρυθμός μάθησης, νόρμα, σύνολο εκπαίδευσης.

κατανομή πιθανότητας Για να αναλύσουμε μεθόδους μηχανικής μάθησης, μπορεί να είναι χρήσιμο να ερμηνεύσουμε σημεία δεδομένων ως ανεξάρτητες και ταυτόσημα κατανεμημένες πραγματώσεις μίας τυχαίας μεταβλητής. Οι τυπικές ιδιότητες τέτοιων σημείων δεδομένων διέπονται τότε από την κατανομή πιθανότητας αυτής της τυχαίας μεταβλητής. Η κατανομή πιθανότητας μίας δυαδικής τυχαίας μεταβλητής $y \in \{0, 1\}$ προσδιορίζεται πλήρως από τις πιθανότητες $p(y = 0)$ και $p(y = 1) = 1 - p(y = 0)$. Η κατανομή πιθανότητας μίας τυχαίας μεταβλητής πραγματικής τιμής $x \in \mathbb{R}$ μπορεί να προσδιορίζεται από μία συνάρτηση πυκνότητας πιθανότητας $p(x)$, έτσι ώστε $p(x \in [a, b]) \approx p(a)|b - a|$. Στην πιο γενική περίπτωση, η κατανομή πιθανότητας ορίζεται από ένα μέτρο πιθανότητας [6], [24].

Βλέπε επίσης: ml, data point, ανεξάρτητες και ταυτόσημα κατανεμημένες, πραγμάτωση, τυχαία μεταβλητή, probability, συνάρτηση πυκνότητας πιθανότητας.

κερκόπορτα Μία επίθεση κερκόπορτας (backdoor) αναφέρεται στον σκόπιμο χειρισμό της διαδικασίας εκπαίδευσης που αποτελεί τη βάσης μίας μεθόδου μηχανικής μάθησης. Αυτός ο χειρισμός μπορεί να υλοποιηθεί με τη διαταραχή του συνόλου εκπαίδευσης (δηλαδή μέσω τ.. data poisoning) ή μέσω του αλγόριθμου βελτιστοποίησης που χρησιμοποιείται από μία μέθοδο βασισμένη στην ελαχιστοποίηση εμπειρικής διακινδύνευσης. Ο στόχος μίας επίθεσης κερκόπορτας είναι να ωθήσει την υπόθεση $\hat{h}$ που έχει μαθευτεί προς συκεκριμένες προβλέψεις για ένα ορισμένο πεδίο τιμών χαρακτηριστικών. Το συγκεκριμένο πεδίο τιμών χαρακτηριστικών χρησιμεύει ως το κλειδί (ή έναυσμα) για να ξεκλειδώσει μία κερκόπορτα με την έννοια της παροχής ανώμαλων προβλέψεων. Το κλειδί $\mathbf{x}$ και η σχετική ανώμαλη πρόβλεψη $\hat{h}(\mathbf{x})$ είναι γνωστά μόνο στον επιτιθέμενο.

Βλέπε επίσης: ml, σύνολο εκπαίδευσης, data poisoning, αλγόριθμος, ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπόθεση, πρόβλεψη, feature.

κλίση Για μία συνάρτηση πραγματικής τιμής $f : \mathbb{R}^d \to \mathbb{R} : \mathbf{w} \mapsto f(\mathbf{w})$, αν υφίσταται ένα διάνυσμα $\mathbf{g}$ τέτοιο ώστε

$$\lim_{\mathbf{w} \to \mathbf{w}'} f(\mathbf{w}) - \big(f(\mathbf{w}') + \mathbf{g}^T(\mathbf{w} - \mathbf{w}')\big)/\|\mathbf{w} - \mathbf{w}'\| = 0$$

αναφέρεται ως η κλίση της f στο $\mathbf{w}'$. Αν υφίσταται, η κλίση είναι μοναδική και δηλώνεται ως $\nabla f(\mathbf{w}')$ ή $\nabla f(\mathbf{w})\big|_{\mathbf{w}'}$ [2].
Βλέπε επίσης: συνάρτηση.

κριτήριο τερματισμού Πολλές μέθοδοι μηχανικής μάθησης χρησιμοποιούν επαναληπτικούς αλγόριθμους που κατασκευάζουν μία ακολουθία παραμέτρων μοντέλου (όπως τα βάρη μίας linear map ή τα βάρη ενός τεχνητού νευρωνικού δικτύου). Αυτές οι παράμετροι (ελπίζουμε να) συγκλίνουν σε μία βέλτιστη επιλογή για τις παραμέτρους μοντέλου. Στην πράξη, δεδομένων των περιορισμένων υπολογιστικών πόρων, χρειάζεται να σταματήσουμε την επανάληψη μετά από έναν πεπερασμένο αριθμό επαναλήψεων. Ένα κριτήριο τερματισμού είναι οποιαδήποτε καλά ορισμένη συνθήκη για να αποφασίσουμε πότε να σταματήσουμε την επανάληψη.
Βλέπε επίσης: ml, αλγόριθμος, παράμετροι μοντέλου, βάρη, linear map, ΤΝΔ, παράμετρος.

κυρτή συσταδοποίηση Θεωρούμε ένα σύνολο δεδομένων $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathbb{R}^d$. Η κυρτή συσταδοποίηση μαθαίνει διανύσματα $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(m)}$ ελαχιστοποιώντας το

$$\sum_{r=1}^{m} \left\| \mathbf{x}^{(r)} - \mathbf{w}^{(r)} \right\|_2^2 + \alpha \sum_{i,i' \in \mathcal{V}} \left\| \mathbf{w}^{(i)} - \mathbf{w}^{(i')} \right\|_p .$$

Εδώ, $\|\mathbf{u}\|_p := \left(\sum_{j=1}^{d} |u_j|^p \right)^{1/p}$ δηλώνει την p-νόρμα (για $p \geq 1$). Προκύπτει ότι πολλά από τα βέλτιστα διανύσματα $\widehat{\mathbf{w}}^{(1)}, \ldots, \widehat{\mathbf{w}}^{(m)}$ συμπίπτουν. Μία συστάδα τότε αποτελείται από αυτά τα σημεία δεδομένων $r \in \{1, \ldots, m\}$ με ταυτόσημα $\widehat{\mathbf{w}}^{(r)}$ [56], [57].

Βλέπε επίσης: σύνολο δεδομένων, convex, συσταδοποίηση, νόρμα, συστάδα, data point.

κυρτός Ένα υποσύνολο $\mathcal{C} \subseteq \mathbb{R}^d$ του Ευκλείδειου χώρου $\mathbb{R}^d$ αναφέρεται ως κυρτό αν περιέχει το ευθύγραμμο τμήμα μεταξύ οποιωνδήποτε δύο σημείων $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ σε ένα σύνολο. Μία συνάρτηση $f : \mathbb{R}^d \to \mathbb{R}$ είναι κυρτή αν το επίγραμμά της $\left\{ \left(\mathbf{w}^T, t \right)^T \in \mathbb{R}^{d+1} : t \geq f(\mathbf{w}) \right\}$ είναι ένα κυρτό σύνολο [58]. Παρουσιάζουμε ένα παράδειγμα ενός κυρτού συνόλου και μίας κυρτής συνάρτησης στο Σχ. 15.

Σχ. 15. Αριστερά: Ένα κυρτό σύνολο $\mathcal{C} \subseteq \mathbb{R}^d$. Δεξιά: Μία κυρτή συνάρτηση $f : \mathbb{R}^d \to \mathbb{R}$.

Βλέπε επίσης: Ευκλείδειος χώρος, συνάρτηση, epigraph.

λεία A real-valued συνάρτηση $f : \mathbb{R}^d \to \mathbb{R}$ is smooth if it is παραγωγίσιμη and its gradient $\nabla f(\mathbf{w})$ is continuous at all $\mathbf{w} \in \mathbb{R}^d$ [59], [60]. A smooth συνάρτηση f is referred to as β-smooth if the gradient $\nabla f(\mathbf{w})$ is Lipschitz continuous with Lipschitz constant β, i.e.,

$$\|\nabla f(\mathbf{w}) - \nabla f(\mathbf{w}')\| \leq \beta \|\mathbf{w} - \mathbf{w}'\|, \text{ for any } \mathbf{w}, \mathbf{w}' \in \mathbb{R}^d.$$

The constant β quantifies the amount of smoothness of the συνάρτηση f: the smaller the β, the smoother f is. Optimization problems with a smooth αντικειμενική συνάρτηση can be solved effectively by μέθοδοι με βάση την κλίση. Indeed, μέθοδοι με βάση την κλίση approximate the αντικειμενική συνάρτηση locally around a current choice $\mathbf{w}$ using its gradient. This approximation works well if the gradient does not change too rapidly. We can make this informal claim precise by studying the effect of a single βήμα κλίσης with μέγεθος βήματος $\eta = 1/\beta$ (see Fig. 16).

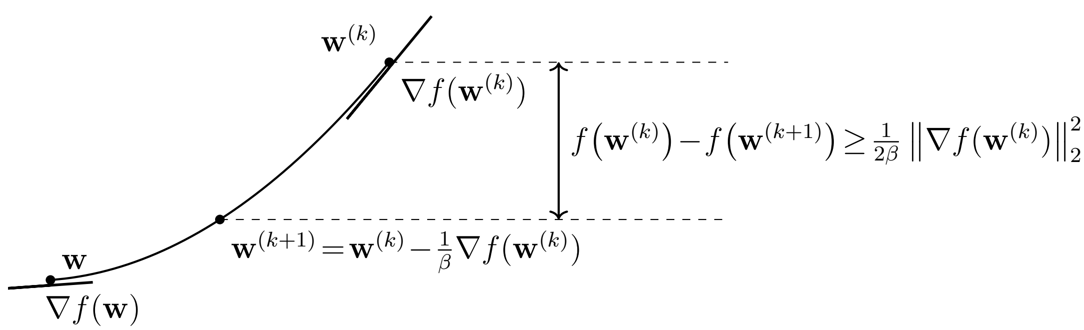


Fig. 16. Consider an αντικειμενική συνάρτηση $f(\mathbf{w})$ that is β-smooth. Taking a βήμα κλίσης, with μέγεθος βήματος $\eta = 1/\beta$, decreases the objective by at least $1/2\beta \left\|\nabla f(\mathbf{w}^{(k)})\right\|_2^2$ [59], [60], [61]. Note that the μέγεθος βήματος $\eta = 1/\beta$ becomes larger for smaller β. Thus, for smoother αντικειμενική συνάρτησηs (i.e., those with smaller β), we can take larger steps.

Βλέπε επίσης: συνάρτηση, παραγωγίσιμη, gradient, optimization problem, αντικειμενική συνάρτηση, μέθοδοι με βάση την κλίση, βήμα κλίσης, μέγεθος βήματος.

λογιστική απώλεια Consider a data point characterized by the features $\mathbf{x}$ and a binary ετικέτα $y \in \{-1, 1\}$. We use a real-valued υπόθεση h to predict the ετικέτα y from the features $\mathbf{x}$. The logistic loss incurred by this πρόβλεψη is defined as

$$L\left((\mathbf{x}, y), h\right) := \log\left(1 + e^{-yh(\mathbf{x})}\right). \tag{6}$$

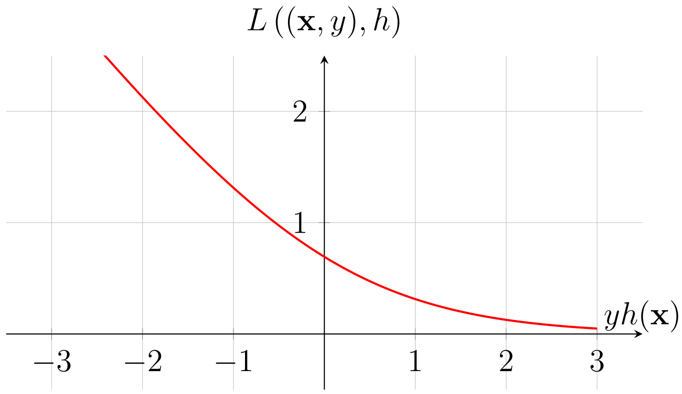


Fig. 17. The logistic loss incurred by the πρόβλεψη $h(\mathbf{x}) \in \mathbb{R}$ for a data point with ετικέτα $y \in \{-1, 1\}$.

Carefully note that the expression (6) for the logistic loss applies only for the χώρος ετικετών $\mathcal{Y} = \{-1, 1\}$ and when using the thresholding rule (10).

Βλέπε επίσης: data point, feature, ετικέτα, υπόθεση, loss, πρόβλεψη, χώρος ετικετών.

λογιστική παλινδρόμηση Η λογιστική παλινδρόμηση μαθαίνει μία γραμμική map υπόθεσης (ή έναν ταξινομητή) $h(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$ για να προβλέψει μία δυαδική ετικέτα y με βάση το αριθμητικό διάνυσμα χαρακτηριστικών $\mathbf{x}$ ενός σημείου δεδομένων. Η ποιότητα μίας γραμμικής map υπόθεσης μετράται από τη μέση λογιστική απώλεια σε κάποια σημεία δεδομένων με ετικέτες (δηλαδή το σύνολο εκπαίδευσης).

Βλέπε επίσης: regression, υπόθεση, map, ταξινομητής, ετικέτα, διάνυσμα χαρακτηριστικών, data point, λογιστική απώλεια, σημείο δεδομένων με ετικέτα, σύνολο εκπαίδευσης.

μάθηση πολυδιεργασίας Η μάθηση πολυδιεργασίας στοχεύει στην αξιοποίηση σχέσεων μεταξύ διαφορετικών εργασιών μάθησης. Θεωρούμε δύο εργασίες μάθησης που προκύπτουν από το ίδιο σύνολο δεδομένων λήψεων από κάμερα υπολογιστή. Η πρώτη εργασία είναι να προβλεφθεί η παρουσία ενός ανθρώπου, ενώ η δεύτερη εργασία είναι να προβλεφθεί η παρουσία ενός αυτοκινήτου. Μπορεί να είναι χρήσιμο να χρησιμοποιηθεί η ίδια δομή βαθιού δικτύου και για τις δύο εργασίες και να επιτραπεί μόνο τα βάρη του τελικού επιπέδου εξόδου να είναι διαφορετικά.

Βλέπε επίσης: εργασία μάθησης, σύνολο δεδομένων, βαθύ δίκτυο, βάρη.

μάθηση χαρακτηριστικών Θεωρούμε μία εφαρμογή μηχανικής μάθησης με σημεία δεδομένων που χαρακτηρίζονται από ακατέργαστα χαρακτηριστικά $\mathbf{x} \in \mathcal{X}$. Η μάθηση χαρακτηριστικών αναφέρεται στην εργασία της μάθησης μίας map

$$\mathbf{\Phi} : \mathcal{X} \to \mathcal{X}' : \mathbf{x} \mapsto \mathbf{x}'$$

που διαβάζει τα ακατέργαστα χαρακτηριστικά $\mathbf{x} \in \mathcal{X}$ ενός σημείου δεδομένων και παραδίδει νέα χαρακτηριστικά $\mathbf{x}' \in \mathcal{X}'$ από έναν νέο χώρο χαρακτηριστικών $\mathcal{X}'$. Διαφορετικές μέθοδοι μάθησης χαρακτηριστικών προκύπτουν για διαφορετικές επιλογές σχεδιασμού των $\mathcal{X}, \mathcal{X}'$, για έναν χώρο υποθέσεων $\mathcal{H}$ πιθανών maps $\mathbf{\Phi}$, και για ένα ποσοτικό μέτρο της χρησιμότητας μίας συγκεκριμένης $\mathbf{\Phi} \in \mathcal{H}$. Για παράδειγμα, η ανάλυση κυρίων συνιστωσών χρησιμοποιεί $\mathcal{X} := \mathbb{R}^d$, $\mathcal{X}' := \mathbb{R}^{d'}$ με $d' < d$, και έναν χώρο υποθέσεων

$$\mathcal{H} := \left\{ \mathbf{\Phi} : \mathbb{R}^d \to \mathbb{R}^{d'} : \mathbf{x}' := \mathbf{F}\mathbf{x} \text{ με κάποια } \mathbf{F} \in \mathbb{R}^{d' \times d} \right\}.$$

Η ανάλυση κυρίων συνιστωσών μετράει τη χρησιμότητα μίας συγκεκριμένης map $\mathbf{\Phi}(\mathbf{x}) = \mathbf{F}\mathbf{x}$ από το ελάχιστο γραμμικό σφάλμα ανακατασκευής που προκαλείται σε ένα σύνολο δεδομένων, έτσι ώστε

$$\min_{\mathbf{G} \in \mathbb{R}^{d \times d'}} \sum_{r=1}^{m} \left\| \mathbf{G}\mathbf{F}\mathbf{x}^{(r)} - \mathbf{x}^{(r)} \right\|_2^2 .$$

Βλέπε επίσης: ml, data point, feature, map, χώρος χαρακτηριστικών, χώρος υποθέσεων, ανάλυση κυρίων συνιστωσών, ελάχιστο, σύνολο δεδομένων.

μαλακή συσταδοποίηση Η μαλακή συσταδοποίηση αναφέρεται στην εργασία χωρισμού ενός συγκεκριμένου συνόλου σημείων δεδομένων σε (μερικές) αλληλεπικαλυπτόμενες συστάδες. Κάθε σημείο δεδομένων αποδίδεται σε αρκετές διαφορετικές συστάδες με μεταβαλλόμενους βαθμούς συσχέτισης. Οι μέθοδοι μαλακής συσταδοποίησης καθορίζουν τον βαθμό συσχέτισης (ή την απόδοση μαλακής συστάδας) για κάθε σημείο δεδομένων και κάθε συστάδα. Μία ηθική προσέγγιση στη μαλακή συσταδοποίηση είναι με την ερμηνεία σημείων δεδομένων ως ανεξάρτητες και ταυτόσημα κατανεμημένες πραγματώσεις ενός Gaussian mixture model (GMM). Προκύπτει τότε μία φυσική επιλογή για τον βαθμό συσχέτισης ως την υπό συνθήκη πιθανότητα ενός σημείου δεδομένων που ανήκει σε μία συγκεκριμένη συνιστώσα μίγματος.

Βλέπε επίσης: συσταδοποίηση, data point, συστάδα, βαθμός συσχέτισης, ανεξάρτητες και ταυτόσημα κατανεμημένες, πραγμάτωση, GMM, probability.

μεγάλο γλωσσικό μοντέλο Τα μεγάλα γλωσσικά μοντέλα (large language model - LLM) είναι ένα όρος-ομπρέλα για μεθόδους μηχανικής μάθησης που επεξεργάζονται και παράγουν κείμενο παρόμοιο με ανθρώπινο. Αυτές οι μέθοδοι συνήθως χρησιμοποιούν βαθιά δίκτυα με δισεκατομμύρια (ή ακόμα και τρισεκατομμύρια) παραμέτρους. Μία ευρέως χρησιμοποιούμενη επιλογή για την αρχιτεκτονική του δικτύου αναφέρεται ως Transformers [62]. Η εκπαίδευση μεγάλων γλωσσικών μοντέλων βασίζεται συχνά στην εργασία της πρόβλεψης μερικών λέξεων που σκόπιμα αφαιρούνται από ένα μεγάλο σώμα κειμένων. Έτσι, μπορούμε να κατασκευάσουμε σημεία δεδομένων με ετικέτες απλώς επιλέγοντας κάποιες λέξεις ενός κειμένου ως ετικέτες και τις υπόλοιπες λέξεις ως χαρακτηριστικά σημείων δεδομένων. Αυτή η κατασκευή απαιτεί πολύ λίγη ανθρώπινη εποπτεία και επιτρέπει την παραγωγή επαρκώς μεγάλων συνόλων εκπαίδευσης για μεγάλα γλωσσικά μοντέλα.

Βλέπε επίσης: ml, βαθύ δίκτυο, παράμετρος, σημείο δεδομένων με ετικέτα,

ετικέτα, feature, data point, σύνολο εκπαίδευσης, model.

μέγεθος βήματος Βλέπε ρυθμός μάθησης.

μέγεθος δείγματος Ο αριθμός των ξεχωριστών σημείων δεδομένων που περιέχονται σε ένα σύνολο δεδομένων.

Βλέπε επίσης: data point, σύνολο δεδομένων.

μέγιστο Το μέγιστο ενός συνόλου $\mathcal{A} \subseteq \mathbb{R}$ πραγματικών αριθμών είναι το μέγιστο στοιχείο σε αυτό το σύνολο, αν ένα τέτοιο στοιχείο υφίσταται. Ένα σύνολο $\mathcal{A}$ έχει ένα μέγιστο αν είναι άνω φραγμένο και επιτυγχάνει το ελάχιστο άνω φράγμα του [2, Sec. 1.4].

Βλέπε επίσης: ελάχιστο άνω φράγμα (ή supremum).

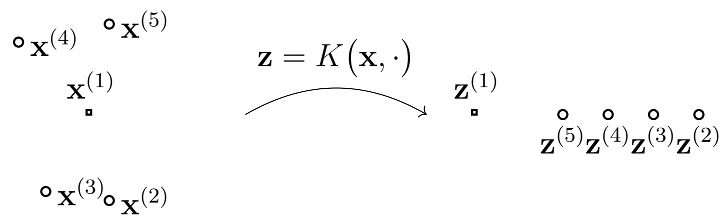
μέθοδοι με βάση την κλίση Οι μέθοδοι με βάση την κλίση είναι επαναληπτικές τεχνικές για την εύρεση του ελάχιστου (ή του μέγιστου) μίας παραγωγίσιμης αντικειμενικής συνάρτησης των παραμέτρων μοντέλου. Αυτές οι μέθοδοι κατασκευάζουν μία ακολουθία προσεγγίσεων σε μία βέλτιστη επιλογή παραμέτρων μοντέλου που οδηγεί σε μία ελάχιστη (ή μέγιστη) τιμή της αντικειμενικής συνάρτησης. Όπως το όνομά τους υποδεικνύει, οι μέθοδοι με βάση την κλίση χρησιμοποιούν τις κλίσεις της αντικειμενικής συνάρτησης που αξιολογούνται κατά τις προηγούμενες επαναλήψεις για να κατασκευάσουν νέες, (ελπίζοντας) βελτιωμένες παραμέτρους μοντέλου. Ένα σημαντικό παράδειγμα μίας μεθόδου με βάση την κλίση είναι η κάθοδος κλίσης.

Βλέπε επίσης: gradient, ελάχιστο, maximum, παραγωγίσιμη, αντικειμενική συνάρτηση, παράμετροι μοντέλου, κάθοδος κλίσης.

μέθοδος βελτιστοποίησης Μία μέθοδος βελτιστοποίησης είναι ένας αλγόριθμος που διαβάζει μία αναπαράσταση ενός προβλήματος βελτιστοποίησης και παραδίδει μία (προσεγγιστική) λύση ως την έξοδό του [29], [58], [59].

Βλέπε επίσης: αλγόριθμος, optimization problem.

μέθοδος πυρήνα Μία μέθοδος πυρήνα είναι μία μέθοδος μηχανικής μάθησης που χρησιμοποιεί έναν πυρήνα K για να αντιστοιχίσει το αρχικό (δηλαδή ακατέργαστο) διάνυσμα χαρακτηριστικών $\mathbf{x}$ ενός σημείου δεδομένων σε ένα νέο (μετασχηματισμένο) διάνυσμα χαρακτηριστικών $\mathbf{z} = K(\mathbf{x}, \cdot)$ [17], [55]. Το κίνητρο για τον μετασχηματισμό των διανυσμάτων χαρακτηριστικών είναι ότι, χρησιμοποιώντας έναν κατάλληλο πυρήνα, τα σημεία δεδομένων έχουν μία «πιο ευχάριστη» γεωμετρία στον μετασχηματισμένο χώρο χαρακτηριστικών. Για παράδειγμα, σε ένα πρόβλημα δυαδικής ταξινόμησης, η χρήση μετασχηματισμένων διανυσμάτων χαρακτηριστικών $\mathbf{z}$ μπορεί να μας επιτρέψει να χρησιμοποιήσουμε γραμμικά μοντέλα, ακόμα και αν τα σημεία δεδομένων δεν είναι γραμμικώς διαχωρίσιμα στον αρχικό χώρο χαρακτηριστικών (βλέπε Σχ. 18).



Σχ. 18. Πέντε σημεία δεδομένων που χαρακτηρίζονται από διανύσματα χαρακτηριστικών $\mathbf{x}^{(r)}$ και ετικέτες $y^{(r)} \in \{\circ, \Box\}$, για $r = 1, \ldots, 5$. Με αυτά τα διανύσματα χαρακτηριστικών, δεν υπάρχει τρόπος να διαχωρίσουμε τις δύο τάξεις με μία ευθεία γραμμή (που αναπαριστά το ορίο απόφασης ενός γραμμικού ταξινομητή). Αντίθετα, τα μετασχηματισμένα διανύσματα χαρακτηριστικών $\mathbf{z}^{(r)} = K(\mathbf{x}^{(r)}, \cdot)$ μας επιτρέπουν να διαχωρίσουμε τα σημεία δεδομένων χρησιμοποιώντας έναν γραμμικό ταξινομητή.

Βλέπε επίσης: πυρήνας, ml, διάνυσμα χαρακτηριστικών, data point, χώρος χαρακτηριστικών, ταξινόμηση, γραμμικό μοντέλο, ετικέτα, όριο απόφασης, γραμμικός ταξινομητής.

μείωση της διαστασιμότητας Dimensionality reduction refers to methods that learn a transformation $h : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ a (typically large) set of raw features $x_1, \ldots, x_d$ into a smaller set of informative features $z_1, \ldots, z_{d'}$. Using a smaller set of features is beneficial in several ways:

- Statistical benefit: It typically reduces the risk of υπερπροσαρμογή, as reducing the number of features often reduces the αποτελεσματική διάσταση of a model.
- Computational benefit: Using fewer features means less computation for the training of ml models. As a case in point, γραμμική παλινδρόμηση methods need to invert a matrix whose size is determined by the number of features.

- Visualization: Dimensionality reduction is also instrumental for data visualization. For example, we can learn a transformation that delivers two features z_1, z_2 which we can use, in turn, as the coordinates of a διάγραμμα διασποράς. Fig. 19 depicts the διάγραμμα διασποράς of hand-written digits that are placed according transformed features. Here, the data points are naturally represented by a large number of grayscale values (one value for each pixel).

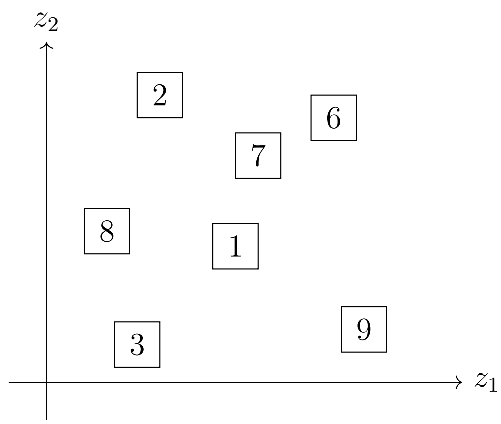


Fig. 19. Example of dimensionality reduction: High-dimensional image data (e.g., high-resolution images of hand-written digits) embedded into 2D using learned features (z_1, z_2) and visualized in a διάγραμμα διασποράς.

Βλέπε επίσης: feature, υπερπροσαρμογή, αποτελεσματική διάσταση, model, ml, γραμμική παλινδρόμηση, data, διάγραμμα διασποράς, data point.

μεροληψία Θεωρούμε μία μέθοδο μηχανικής μάθησης που χρησιμοποιεί έναν παραμετροποιημένο χώρο υποθέσεων $\mathcal{H}$. Μαθαίνει τις παραμέτρους μοντέλου $\mathbf{w} \in \mathbb{R}^d$ χρησιμοποιώντας το σύνολο δεδομένων

$$\mathcal{D} = \left\{ \left(\mathbf{x}^{(r)}, y^{(r)}\right) \right\}_{r=1}^{m}.$$

Για να αναλύσουμε τις ιδιότητες της μεθόδου μηχανικής μάθησης, συνήθως ερμηνεύουμε τα σημεία δεδομένων ως πραγματώσεις ανεξάρτητων και ταυτόσημα κατανεμημένων τυχαίων μεταβλητών,

$$y^{(r)} = h^{(\overline{\mathbf{w}})}\big(\mathbf{x}^{(r)}\big) + \boldsymbol{\varepsilon}^{(r)}, r = 1, \ldots, m.$$

Μπορούμε τότε να ερμηνεύσουμε τη μέθοδο μηχανικής μάθησης ως μία εκτιμήτρια $\widehat{\mathbf{w}}$ που υπολογίζεται από το $\mathcal{D}$ (π.χ. λύνοντας την ελαχιστοποίηση εμπειρικής διακινδύνευσης). Η (τετραγωνική) μεροληψία που προκαλείται από την εκτίμηση $\widehat{\mathbf{w}}$ ορίζεται τότε ως $B^2 := \big\|\mathbb{E}\{\widehat{\mathbf{w}}\} - \overline{\mathbf{w}}\big\|_2^2$. Βλέπε επίσης: ml, χώρος υποθέσεων, παράμετροι μοντέλου, σύνολο δεδομένων, data point, πραγμάτωση, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, ελαχιστοποίηση εμπειρικής διακινδύνευσης.

μέση τιμή Η μέση τιμή μίας τυχαίας μεταβλητής $\mathbf{x}$, που παίρνει τιμές σε έναν Ευκλείδειο χώρο $\mathbb{R}^d$, είναι η προσδοκία της $\mathbb{E}\{\mathbf{x}\}$. Ορίζεται ως το ολοκλήρωμα Lebesgue του $\mathbf{x}$ αναφορικά με την υποκείμενη κατανομή πιθανότητας P (π.χ. βλέπε [2] ή [6]), δηλαδή

$$\mathbb{E}\{\mathbf{x}\} = \int_{\mathbb{R}^d} \mathbf{x}\, \mathrm{d}P(\mathbf{x}).$$

Χρησιμοποιούμε επίσης τον όρο για να αναφερθούμε στον μέσο όρο μίας πεπερασμένης ακολουθίας $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathbb{R}^d$. Ωστόσο, αυτοί οι δύο ορισμοί είναι ουσιαστικά ίδιοι. Πράγματι, μπορούμε να χρησιμοποιήσουμε την ακολουθία $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathbb{R}^d$ για να κατασκευάσουμε μία διακριτή τυχαία μεταβλητή $\widetilde{\mathbf{x}} = \mathbf{x}^{(I)}$, με τον δείκτη I να επιλέγεται ομοιόμορφα στην τύχη από το σύνολο $\{1, \ldots, m\}$. Η μέση τιμή της $\widetilde{\mathbf{x}}$ είναι ακριβώς ο μέσος όρος $(1/m) \sum_{r=1}^{m} \mathbf{x}^{(r)}$.

Βλέπε επίσης: τυχαία μεταβλητή, Ευκλείδειος χώρος, expectation, κατανομή πιθανότητας.

μέση τιμή δείγματος Η μέση τιμή δείγματος $\mathbf{m} \in \mathbb{R}^d$ για ένα συγκεκριμένο σύνολο δεδομένων, με διανύσματα χαρακτηριστικών $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathbb{R}^d$, ορίζεται ως

$$\mathbf{m} = \frac{1}{m} \sum_{r=1}^{m} \mathbf{x}^{(r)}.$$

Βλέπε επίσης: δείγμα, μέση τιμή, σύνολο δεδομένων, διάνυσμα χαρακτηριστικών.

μέσο τετραγωνικό σφάλμα εκτίμησης Θεωρούμε μία μέθοδο μηχανικής μάθησης που μαθαίνει παραμέτρους μοντέλου $\widehat{\mathbf{w}}$ με βάση κάποιο σύνολο δεδομένων $\mathcal{D}$. Αν ερμηνεύσουμε τα σημεία δεδομένων στο $\mathcal{D}$ ως ανεξάρτητες και ταυτόσημα κατανεμημένες πραγματώσεις μίας τυχαίας μεταβλητής $\mathbf{z}$, ορίζουμε το σφάλμα εκτίμησης $\Delta \mathbf{w} := \widehat{w} - \overline{\mathbf{w}}$. Εδώ, $\overline{\mathbf{w}}$ δηλώνει τις αληθείς παραμέτρους μοντέλου της κατανομής πιθανότητας του $\mathbf{z}$. Το μέσο τετραγωνικό σφάλμα εκτίμησης (mean squared estimation error - MSEE) ορίζεται ως η προσδοκία $\mathbb{E}\{\|\Delta \mathbf{w}\|^2\}$ της τετραγωνικής Ευκλείδειας νόρμας του σφάλματος εκτίμησης [22], [63].

Βλέπε επίσης: ml, παράμετροι μοντέλου, σύνολο δεδομένων, data point, ανεξάρτητες και ταυτόσημα κατανεμημένες, πραγμάτωση, τυχαία μεταβλητή, σφάλμα εκτίμησης, κατανομή πιθανότητας, expectation, νόρμα, μέση τιμή,.

μετρική Στην πιο γενική της μορφή, μία μετρική είναι ένα ποσοτικό μέτρο που χρησιμοποιείται για τη σύγκριση ή αξιολόγηση αντικειμένων. Στα μαθηματικά, μία μετρική μετράει την απόσταση μεταξύ δύο σημείων και πρέπει να ακολουθεί συγκεκριμένους κανόνες, δηλαδή η απόσταση να είναι πάντα μη αρνητική, να είναι μηδενική μόνο αν τα σημεία είναι ίδια, να είναι συμμετρική, και να ικανοποιεί την τριγωνική ανισότητα [2]. Στη μηχανική μάθηση, μία μετρική είναι ένα ποσοτικό μέτρο του πόσο καλά επιδίδει ένα μοντέλο. Παραδείγματα περιλαμβάνουν την ακρίβεια, την precision, και τη μέση 0/1 απώλεια σε ένα σύνολο ελέγχου [20], [64]. Μία συνάρτηση απώλειας χρησιμοποιείται για να εκπαιδεύσει μοντέλα, ενώ μία μετρική χρησιμοποιείται για να συγκίνει εκαπιδευμένα μοντέλα.

See also: ml, model, ακρίβεια, 0/1 απώλεια, σύνολο ελέγχου, συνάρτηση απώλειας, loss, επιλογή μοντέλου.

μη λεία Αναφερόμαστε σε μία συνάρτηση ως μη λεία αν δεν είναι λεία [59].

Βλέπε επίσης: συνάρτηση, λεία.

μηχανή διανυσμάτων υποστήριξης (ΜΔΥ) The SVM (support vector machine; SVM) is a binary ταξινόμηση method that learns a linear υπόθεση map. Thus, like γραμμική παλινδρόμηση and λογιστική παλινδρόμηση, it is also an instance of ελαχιστοποίηση εμπειρικής διακινδύνευσης for the γραμμικό μοντέλο. However, the SVM uses a different συνάρτηση απώλειας from the one used in those methods. As illustrated in Fig. 20, it aims to maximally separate data points from the two different classes in the χώρος χαρακτηριστικών (i.e., maximum margin principle). Maximizing this separation is equivalent to minimizing a regularized variant of the απώλεια άρθρωσης (1) [64], [17], [65].

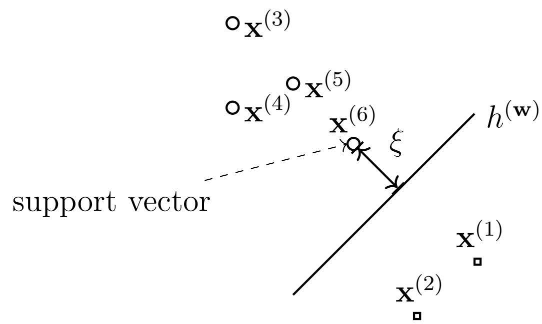


Fig. 20. The SVM learns a υπόθεση (or ταξινομητής) $h^{(\mathbf{w})}$ with minimal average soft-margin απώλεια άρθρωσης. Minimizing this loss is equivalent to maximizing the margin ξ between the όριο απόφασης of $h^{(\mathbf{w})}$ and each class of the σύνολο εκπαίδευσης.

The above basic variant of SVM is only useful if the data points from different categories can be (approximately) linearly separated. For an ml application where the categories are not derived from a πυρήνας.

Βλέπε επίσης: ταξινόμηση, υπόθεση, map, γραμμική παλινδρόμηση, λογιστική παλινδρόμηση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, γραμμικό μοντέλο, συνάρτηση απώλειας, data point, χώρος χαρακτηριστικών, maximum, απώλεια άρθρωσης, μηχανή διανυσμάτων υποστήριξης, ταξινομητής, loss, όριο απόφασης, σύνολο εκπαίδευσης, ml, πυρήνας.

μηχανική μάθηση Η μηχανική μάθηση (machine learning - ML) στοχεύει να προβλέψει μία ετικέτα από τα χαρακτηριστικά ενός σημείου δεδομένων. Οι μέθοδοι μηχανικής μάθησης το επιτυγχάνουν αυτό μαθαίνοντας μία υπόθεση από έναν χώρο υποθέσεων (ή μοντέλο) μέσω της ελαχιστοποίησης μίας συνάρτησης απώλειας [8], [66]. Μία ακριβής διατύπωση αυτής της αρχής είναι η ελαχιστοποίηση εμπειρικής διακινδύνευσης. Διαφορετικές μέθοδοι μηχανικής μάθησης προκύπτουν από διαφορετικές επιλογές σχεδιασμού για σημεία δεδομένων (δηλαδή τα χαρακτηριστικά και την ετικέτα τους), το μοντέλο, και τη συνάρτηση απώλειας [8, Κεφ. 3].

Βλέπε επίσης: ετικέτα, feature, data point, υπόθεση, χώρος υποθέσεων, model, συνάρτηση απώλειας, ελαχιστοποίηση εμπειρικής διακινδύνευσης.

μοντέλο Στο πλαίσιο της μηχανικής μάθησης, ο όρος μοντέλο συνήθως αναφέρεται στον υποκείμενο χώρο υποθέσεων μίας μεθόδου μηχανικής μάθησης [8], [9]. Ωστόσο, ο όρος χρησιμοποιείται επίσης σε άλλα πεδία αλλά με διαφορετική σημασία. Για παράδειγμα, ένα πιθανοτικό μοντέλο αναφέρεται σε ένα παραμετροποιημένο σύνολο κατανομών πιθανοτήτων.

Βλέπε επίσης: ml, χώρος υποθέσεων, πιθανοτικό μοντέλο, κατανομή πιθανότητας.

μοντέλο στοχαστικής ομάδας Το μοντέλο στοχαστικής ομάδας (stochastic block model - SBM) είναι ένα πιθανοτικό παραγωγικό μοντέλο για έναν μη κατευθυνόμενο γράφο $\mathcal{G} = \big(\mathcal{V}, \mathcal{E}\big)$ με ένα δεδομένο σύνολο κόμβων $\mathcal{V}$ [67]. Στην πιο βασική του παραλλαγή, το μοντέλο στοχαστικής ομάδας παράγει έναν γράφο πρώτα αποδίδοντας τυχαία κάθε κόμβο $i \in \mathcal{V}$ σε έναν δείκτη συστάδας $c_i \in \{1, \ldots, k\}$. Ένα ζεύγος διαφορετικών κόμβων στον γράφο συνδέεται με μία ακμή με πιθανότητα $p_{i,i'}$ που εξαρτάται μόνο από τις ετικέτες $c_i, c_{i'}$. Η παρουσία ακμών μεταξύ διαφορετικών ζευγών κομβων είναι στατιστικά ανεξάρτητη.

Βλέπε επίσης: model, graph, συστάδα, probability, ετικέτα.

νόμος των μεγάλων αριθμών Ο νόμος των μεγάλων αριθμών αναφέρε-

ται στη σύγκλιση του μέσου όρου ενός αυξανόμενου (μεγάλου) αριθμού ανεξάρτητων και ταυτόσημα κατανεμημένων τυχαίων μεταβλητών στη μέση τιμή της κοινής τους κατανομής πιθανότητας. Διαφορετικές περιπτώσεις του νόμου των μεγάλων αριθμών προκύπτουν από τη χρήση διαφορετικών εννοιών σύγκλισης [68].
Βλέπε επίσης: ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, μέση τιμή, κατανομή πιθανότητας.

νόρμα Μία νόρμα είναι μία συνάρτηση που αντιστοιχίζει κάθε (διανυσματικό) στοιχείο ενός διανυσματικού χώρου σε έναν μη αρνητικό αριθμό. Αυτή η συνάρτηση πρέπει να είναι ομογενής και ορισμένη, και πρέπει να ικανοποιεί την τριγωνική ανισότητα [69].
Βλέπε επίσης: συνάρτηση.

ολική μεταβολή Βλέπε γενικευμένη ολική μεταβολή.

ομαλοποιημένη ελαχιστοποίηση εμπειρικής διακινδύνευσης Basic ελαχιστοποίηση εμπειρικής διακινδύνευσης learns a υπόθεση (or trains a model) $h \in \mathcal{H}$ based solely on the empirical risk $\widehat{L}(h|\mathcal{D})$ incurred on a σύνολο εκπαίδευσης $\mathcal{D}$. To make ελαχιστοποίηση εμπειρικής διακινδύνευσης less prone to υπερπροσαρμογή, we can implement ομαλοποίηση by including a (scaled) ομαλοποιητής $\mathcal{R}\{h\}$ in the learning objective. This leads to RERM (regularized empirical risk minimization - RERM),

$$\hat{h} \in \operatorname*{arg\,min}_{h \in \mathcal{H}} \widehat{L}(h|\mathcal{D}) + \alpha \mathcal{R}\{h\}. \tag{7}$$

The παράμετρος $\alpha \geq 0$ controls the ομαλοποίηση strength. For $\alpha = 0$, we recover standard ελαχιστοποίηση εμπειρικής διακινδύνευσης without ομαλοποίηση. As α increases, the learned υπόθεση is increasingly biased toward small values of $\mathcal{R}\{h\}$. The component $\alpha\mathcal{R}\{h\}$ in the αντικειμενική συνάρτηση of (7) can be intuitively understood as a surrogate for the increased average loss that may occur when predicting ετικέτας for data points outside the σύνολο εκπαίδευσης. This intuition can be made precise in various ways. For example, consider a γραμμικό μοντέλο trained using απώλεια τετραγωνικού σφάλματος and the ομαλοποιητής $\mathcal{R}\{h\} = \|\mathbf{w}\|_2^2$. In this setting, $\alpha\mathcal{R}\{h\}$ corresponds to the expected increase in loss caused by adding Gaussian RVs to the διάνυσμα χαρακτηριστικώνs in the σύνολο εκπαίδευσης [8, Ch. 3]. A principled construction for the ομαλοποιητής $\mathcal{R}\{h\}$ arises from approximate upper bounds on the γενίκευση error. The resulting RERM instance is known as δομημένη ελαχιστοποίηση διακινδύνευσης [110, Sec. 7.2].

Βλέπε επίσης: ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπόθεση, model, empirical risk, σύνολο εκπαίδευσης, υπερπροσαρμογή, ομαλοποίηση, ομαλοποιητής, παράμετρος, αντικειμενική συνάρτηση, loss, ετικέτα, data point, γραμμικό μοντέλο, απώλεια τετραγωνικού σφάλματος, Gaussian RV, διάνυσμα χαρακτηριστικών, γενίκευση, δομημένη ελαχιστοποίηση διακινδύνευσης.

ομαλοποίηση A key challenge of modern ml applications is that they often use large models, which have an αποτελεσματική διάσταση in the order of billions. Training a high-dimensional model using basic ελαχιστοποίηση εμπειρικής διακινδύνευσης-based methods is prone to υπερπροσαρμογή: the learned υπόθεση performs well on the σύνολο εκπαίδευσης but poorly outside the σύνολο εκπαίδευσης. Regularization refers to modifications of a given instance of ελαχιστοποίηση εμπειρικής διακινδύνευσης in order to avoid υπερπροσαρμογή, i.e., to ensure that the learned υπόθεση performs not much worse outside the σύνολο εκπαίδευσης. There are three routes for implementing regularization:

1) Model pruning: We prune the original model $\mathcal{H}$ to obtain a smaller model $\mathcal{H}'$. For a parametric model, the pruning can be implemented via constraints on the παράμετροι μοντέλου (such as $w_1 \in [0.4, 0.6]$ for the weight of feature x_1 in γραμμική παλινδρόμηση).

2) Loss penalization: We modify the αντικειμενική συνάρτηση of ελαχιστοποίηση εμπειρικής διακινδύνευσης by adding a penalty term to the training error. The penalty term estimates how much larger the expected loss (or διακινδύνευση) is compared to the average loss on the σύνολο εκπαίδευσης.

3) Data augmentation: We can enlarge the σύνολο εκπαίδευσης $\mathcal{D}$ by adding perturbed copies of the original data points in $\mathcal{D}$. One example for such a perturbation is to add the πραγμάτωση of an τυχαία μεταβλητή to the διάνυσμα χαρακτηριστικών of a data point.

Fig. 21 illustrates the above three routes to regularization. These routes are closely related and sometimes fully equivalent: data augmentation using Gaussian RVs to perturb the διάνυσμα χαρακτηριστικώνs in the σύνολο εκπαίδευσης of γραμμική παλινδρόμηση has the same effect as adding the penalty $\lambda \|\mathbf{w}\|_2^2$ to the training error (which is nothing but αμφικλινής παλινδρόμηση). The decision on which route to use for regularization can be based on the available computational infrastructure. For example, it might be much easier to implement data augmentation than model pruning.

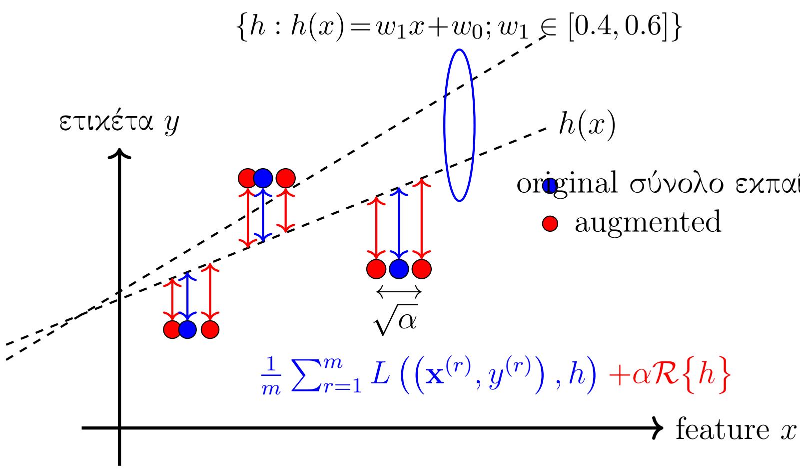


Fig. 21. Three approaches to regularization: 1) data augmentation; 2) loss penalization; and 3) model pruning (via constraints on παράμετροι μοντέλου).

Βλέπε επίσης: ml, model, αποτελεσματική διάσταση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπερπροσαρμογή, υπόθεση, σύνολο εκπαίδευσης, παράμετροι μοντέλου, feature, γραμμική παλινδρόμηση, loss, αντικειμενική συνάρτηση, training error, διακινδύνευση, data augmentation, data point, πραγμάτωση, τυχαία μεταβλητή, διάνυσμα χαρακτηριστικών, Gaussian RV, ridge regression, ετικέτα.

ομαλοποιητής A regularizer assigns each υπόθεση h from a χώρος υποθέσεων $\mathcal{H}$ a quantitative measure $\mathcal{R}\{h\}$ for how much its πρόβλεψη error on a σύνολο εκπαίδευσης might differ from its πρόβλεψη errors on data points outside the σύνολο εκπαίδευσης. Ridge regression uses the regularizer $\mathcal{R}\{h\} := \|\mathbf{w}\|_2^2$ for linear υπόθεση maps $h^{(\mathbf{w})}(\mathbf{x}) := \mathbf{w}^T\mathbf{x}$ [8, Ch. 3]. Lasso uses the regularizer $\mathcal{R}\{h\} := \|\mathbf{w}\|_1$ for linear υπόθεση maps $h^{(\mathbf{w})}(\mathbf{x}) := \mathbf{w}^T\mathbf{x}$ [8, Ch. 3].

Βλέπε επίσης: υπόθεση, χώρος υποθέσεων, πρόβλεψη, σύνολο εκπαίδευσης, data point, ridge regression, map, Lasso.

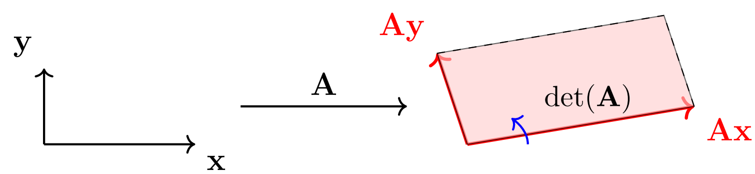
ομοσπονδιακή μάθηση Η ομοσπονδιακή μάθηση (federated learning - FL) είναι ένας όρος-ομπρέλα για μεθόδους μηχανικής μάθησης που εκπαιδεύουν μοντέλα με έναν συνεργατικό τρόπο χρησιμοποιώντας αποκεντρωμένα δεδομένα και υπολογισμό.

Βλέπε επίσης: ml, model, data.

οριζόντια ομοσπονδιακή μάθηση HFL (horizontal federated learning; HFL) uses τοπικό σύνολο δεδομένωνs constituted by different data points but uses the same features to characterize them [70]. For example, weather forecasting uses a network of spatially distributed weather (observation) stations. Each weather station measures the same quantities, such as daily temperature, air pressure, and precipitation. However, different weather stations measure the characteristics or features of different spatiotemporal regions. Each spatiotemporal region represents an individual data point, each characterized by the same features (e.g., daily temperature or air pressure).

Βλέπε επίσης: τοπικό σύνολο δεδομένων, data point, feature, FL, vertical federated learning (VFL), clustered federated learning (CFL).

ορίζουσα The determinant $\det(\mathbf{A})$ of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a scalar that characterizes how (the orientation of) volumes in $\mathbb{R}^n$ are altered by applying $\mathbf{A}$ [3], [120]. Note that a matrix $\mathbf{A}$ represents a linear transformation on $\mathbb{R}^n$. In particular, $\det(\mathbf{A}) > 0$ preserves orientation, $\det(\mathbf{A}) < 0$ reverses orientation, and $\det(\mathbf{A}) = 0$ collapses volume entirely, indicating that $\mathbf{A}$ is non-invertible. The determinant also satisfies $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$, and if $\mathbf{A}$ is diagonalizable with ιδιοτιμήs $\lambda_1, \ldots, \lambda_n$, then $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$ [69]. For the special cases $n = 2$ (2D) and $n = 3$ (3D), the determinant can be interpreted as an oriented area or volume spanned by the column vectors of $\mathbf{A}$.



Βλέπε επίσης: ιδιοτιμή, αντίστροφος πίνακας.

όριο απόφασης Consider a υπόθεση map h that reads in a feature vector $\mathbf{x} \in \mathbb{R}^d$ and delivers a value from a finite set $\mathcal{Y}$. The decision boundary of h is the set of vectors $\mathbf{x} \in \mathbb{R}^d$ that lie between different περιοχή αποφάσεωνs. More precisely, a vector $\mathbf{x}$ belongs to the decision boundary if and only if each neighborhood $\{\mathbf{x}' : \|\mathbf{x} - \mathbf{x}'\| \leq \varepsilon\}$, for any $\varepsilon > 0$, contains at least two vectors with different συνάρτηση values.

Βλέπε επίσης: υπόθεση, map, feature, περιοχή αποφάσεων, neighborhood, συνάρτηση.

παλινδρόμηση Τα προβλήματα παλινδρόμησης περιστρέφονται γύρω από την πρόβλεψη μίας αριθμητικής ετικέτας μόνο από τα χαρακτηριστικά ενός σημείου δεδομένων [8, Κεφ. 2].

Βλέπε επίσης: πρόβλεψη, ετικέτα, feature, data point.

παλινδρόμηση ελάχιστης απόλυτης απόκλισης Η παλινδρόμηση ελάχιστης απόλυτης απόκλισης είναι μία περίπτωση της ελαχιστοποίησης εμπειρικής διακινδύνευσης που χρησιμοποιεί την απώλεια απόλυτου σφάλματος. Είναι μία ειδική περίπτωση της παλινδρόμησης Huber.

Βλέπε επίσης: ελαχιστοποίηση εμπειρικής διακινδύνευσης, απώλεια απόλυτου σφάλματος, παλινδρόμηση Huber.

παλινδρόμηση Huber Η παλινδρόμηση Huber αναφέρεται σε μεθόδους βασισμένες στην ελαχιστοποίηση εμπειρικής διακινδύνευσης που χρησιμοποιούν την απώλεια Huber ως μέτρο του σφάλματος πρόβλεψης. Δύο σημαντικές ειδικές περιπτώσεις της παλινδρόμησης Huber είναι η παλινδρόμηση ελάχιστης απόλυτης απόκλισης και η γραμμική παλινδρόμηση. Η ρύθμιση της παραμέτρου-κατωφλιού της απώλειας Huber επιτρέπει στον χρήστη να ανταλλάξει την ευρωστία της απώλειας απόλυτου σφάλματος με τα υπολογιστικά οφέλη της λείας απώλειας τετραγωνικού σφάλματος.

Βλέπε επίσης: regression, ελαχιστοποίηση εμπειρικής διακινδύνευσης, απώλεια Huber, πρόβλεψη, regression, παλινδρόμηση ελάχιστης απόλυτης απόκλισης, γραμμική παλινδρόμηση, παράμετρος, ευρωστία, απώλεια απόλυτου σφάλματος, λεία, απώλεια τετραγωνικού σφάλματος.

παραγωγίσιμη Μία συνάρτηση πραγματικής τιμής $f : \mathbb{R}^d \to \mathbb{R}$ είναι παραγωγίσιμη αν μπορεί, σε οποιοδήποτε σημείο, να προσεγγιστεί τοπικά από μία γραμμική συνάρτηση. Η τοπική γραμμική προσέγγιση στο σημείο $\mathbf{x}$ καθορίζεται από την κλίση $\nabla f(\mathbf{x})$ [2].

Βλέπε επίσης: συνάρτηση, gradient.

παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων Η παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων (independent and identically distributed assumption - i.i.d. assumption) ερμηνεύει σημεία δεδομένων ενός συνόλου δεδομένων ως τις πραγματώσεις ανεξάρτητων και ταυτόσημα κατανεμημένων τυχαίων μεταβλητών.

Βλέπε επίσης: ανεξάρτητες και ταυτόσημα κατανεμημένες, data point, σύνολο δεδομένων, πραγμάτωση, τυχαία μεταβλητή.

παραδοχή συσταδοποίησης The συσταδοποίηση assumption postulates that data points in a σύνολο δεδομένων form a (small) number of groups or συστάδας. Data points in the same συστάδα are more similar to each other than those outside the συστάδα [71]. We obtain different συσταδοποίηση methods by using different notions of similarity between data points.

Βλέπε επίσης: συσταδοποίηση, data point, σύνολο δεδομένων, συστάδα.

παράμετρος The parameter of an ml model is a tunable (i.e., learnable or adjustable) quantity that allowa us to choose between different υπόθεση maps. For example, the γραμμικό μοντέλο $\mathcal{H} := \{h^{(\mathbf{w})} : h^{(\mathbf{w})}(x) = w_1 x + w_2\}$ consists of all υπόθεση maps $h^{(\mathbf{w})}(x) = w_1 x + w_2$ with a particular choice for the parameters $\mathbf{w} = \big(w_1, w_2\big)^T \in \mathbb{R}^2$. Another example of a model parameter is the βάρη assigned to a connection between two neurons of an ΤΝΔ.

Βλέπε επίσης: ml, model, υπόθεση, map, γραμμικό μοντέλο, βάρη, ΤΝΔ.

παράμετροι μοντέλου Οι παράμετροι μοντέλου είναι ποσότητες που χρησιμοποιούνται για να επιλεχθεί μία συγκεκριμένη map υπόθεσης από ένα μοντέλο. Μπορούμε να σκεφτούμε μία λίστα παραμέτρων μοντέλου ως ένα μοναδικό αναγνωριστικό για μία map υπόθεσης όμοια με το πώς ένας αριθμός κοινωνικής ασφάλισης ταυτοποιεί ένα άτομο στην Ελλάδα.

Βλέπε επίσης: model, παράμετρος, υπόθεση, map.

περιοχή αποφάσεων Θεωρούμε μία map υπόθεσης h που δίνει τιμές από ένα πεπερασμένο σύνολο $\mathcal{Y}$. Για κάθε τιμή ετικέτας (δηλαδή κατηγορία) $a \in \mathcal{Y}$, η υπόθεση h καθορίζει ένα υποσύνολο τιμών χαρακτηριστικών $\mathbf{x} \in \mathcal{X}$ που οδηγούν στις ίδιες εξόδους $h(\mathbf{x}) = a$. Αναφερόμαστε σε αυτό το υποσύνολο ως μία περιοχή αποφάσεων της υπόθεσης h.

Βλέπε επίσης: υπόθεση, map, ετικέτα, feature.

πιθανότητα Αποδίδουμε μία τιμή πιθανότητας, συνήθως επιλεγμένη στο διάστημα $[0, 1]$, σε κάθε γεγονός που μπορεί να συμβεί σε ένα τυχαίο πείραμα [6], [7], [72], [73].

πιθανοτικό μοντέλο Ένα πιθανοτικό μοντέλο ερμηνεύει σημεία δεδομένων ως πραγματώσεις τυχαίων μεταβλητών με κοινή κατανομή πιθανότητας. Αυτή η κοινή κατανομή πιθανότητας συνήθως περιλαμβάνει παραμέτρους που πρέπει να επιλεχθούν χειρωνακτικά ή να μαθευτούν μέσω μεθόδων στατιστικής συμπερασματολογίας όπως η εκτίμηση μέγιστης πιθανοφάνειας [22].

Βλέπε επίσης: model, data point, πραγμάτωση, τυχαία μεταβλητή, κατανομή πιθανότητας, παράμετρος, μέγιστη πιθανοφάνεια.

πίνακας σύγχυσης Consider data points, which are characterized by features $\mathbf{x}$ and ετικέτα y, having values from the finite χώρος ετικετών

$\mathcal{Y} = \{1, \ldots, k\}$. For a given υπόθεση h, the confusion matrix is a $k \times k$ matrix with rows representing the elements of $\mathcal{Y}$. The columns of a confusion matrix correspond to the πρόβλεψη $h(\mathbf{x})$. The (c, c')-th entry of the confusion matrix is the fraction of data points with ετικέτα $y = c$ and resulting in a πρόβλεψη $h(\mathbf{x}) = c'$.

Βλέπε επίσης: data point, feature, ετικέτα, χώρος ετικετών, υπόθεση, πρόβλεψη.

πίνακας συνδιακύμανσης Ο πίνακας συνδιακύμανσης μίας τυχαίας μεταβλητής $\mathbf{x} \in \mathbb{R}^d$ ορίζεται ως $\mathbb{E}\Big\{ \big(\mathbf{x} - \mathbb{E}\{\mathbf{x}\}\big)\big(\mathbf{x} - \mathbb{E}\{\mathbf{x}\}\big)^T \Big\}$.

Βλέπε επίσης: τυχαία μεταβλητή.

πίνακας συνδιακύμανσης δείγματος Ο πίνακας συνδιακύμανσης δείγματος $\widehat{\mathbf{\Sigma}} \in \mathbb{R}^{d \times d}$ για ένα δεδομένο σύνολο διανυσμάτων χαρακτηριστικών $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathbb{R}^d$ ορίζεται ως

$$\widehat{\mathbf{\Sigma}} = \frac{1}{m} \sum_{r=1}^{m} (\mathbf{x}^{(r)} - \widehat{\mathbf{m}})(\mathbf{x}^{(r)} - \widehat{\mathbf{m}})^T.$$

Εδώ χρησιμοποιούμε τη μέση τιμή δείγματος $\widehat{\mathbf{m}}$.

Βλέπε επίσης: δείγμα, πίνακας συνδιακύμανσης, διάνυσμα χαρακτηριστικών, μέση τιμή δείγματος.

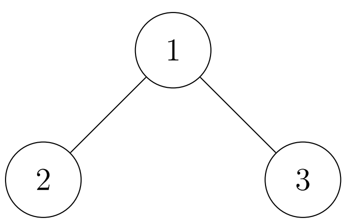
πίνακας χαρακτηριστικών Θεωρούμε ένα σύνολο δεδομένων $\mathcal{D}$ με m σημεία δεδομένων με διανύσματα χαρακτηριστικών $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathbb{R}^d$. Είναι βολικό να συγκεντρώσουμε τα μεμονωμένα διανύσματα χαρακτηριστικών σε έναν πίνακα χαρακτηριστικών $\mathbf{X} := \big(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\big)^T$ μεγέθους $m \times d$.

Βλέπε επίσης: σύνολο δεδομένων, data point, διάνυσμα χαρακτηριστικών, feature.

πίνακας Laplace Η δομή ενός γράφου $\mathcal{G}$, με κόμβους $i = 1, \ldots, n$, μπορεί να αναλυθεί χρησιμοποιώντας τις ιδιότητες ειδικών πινάκων που σχετίζονται με τον $\mathcal{G}$. Ένας τέτοιος πίνακας είναι ο πίνακας Laplace γράφου $\mathbf{L}^{(\mathcal{G})} \in \mathbb{R}^{n \times n}$, ο οποίος ορίζεται για έναν μη κατευθυνόμενο και σταθμισμένο γράφο [74], [75]. Από άποψη στοιχείων ορίζεται ως (βλέπε Σχ. 22)

$$L^{(\mathcal{G})}_{i,i'} := \begin{cases} -A_{i,i'}, & \text{for } i \neq i', \{i,i'\} \in \mathcal{E}; \\ \sum_{i'' \neq i} A_{i,i''}, & \text{for } i = i'; \\ 0, & \text{else.} \end{cases} \tag{8}$$

Εδώ, $A_{i,i'}$ δηλώνει το βάρος ακμής μίας ακμής $\{i, i'\} \in \mathcal{E}$.



$$\mathbf{L}^{(\mathcal{G})} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Σχ. 22. Αριστερά: Κάποιος μη κατευθυνόμενος γράφος $\mathcal{G}$ με τρεις κόμβους $i = 1, 2, 3$. Δεξιά: Ο πίνακας Laplace $\mathbf{L}^{(\mathcal{G})} \in \mathbb{R}^{3 \times 3}$ του $\mathcal{G}$.

Βλέπε επίσης: graph, βάρος ακμής.

πλησιέστερος γείτονας NN (nearest neighbor; NN) methods learn a υπόθεση $h : \mathcal{X} \rightarrow \mathcal{Y}$ whose συνάρτηση value $h(\mathbf{x})$ is solely determined by the NNs within a given σύνολο δεδομένων. Different methods use different metrics for determining the NNs. If data points are characterized by numeric διάνυσμα χαρακτηριστικώνs, we can use their Euclidean distances as the metric.

Βλέπε επίσης: υπόθεση, συνάρτηση, σύνολο δεδομένων, data point, διάνυσμα χαρακτηριστικών, γείτονες.

πολυμεταβλητή κανονική κατανομή The multivariate normal distribution, which is denoted $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, is a fundamental πιθανοτικό μοντέλο for numerical διάνυσμα χαρακτηριστικώνs of fixed dimension d. It defines a family of κατανομή πιθανότηταςs over vector-valued τυχαία μεταβλητήs $\mathbf{x} \in \mathbb{R}^d$ [7], [24], [76]. Each distribution in this family is fully specified by its μέση τιμή vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and πίνακας συνδιακύμανσης $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$. When the πίνακας συνδιακύμανσης $\boldsymbol{\Sigma}$ is invertible, its κατανομή πιθανότητας is fully characterized by the following συνάρτηση πυκνότητας πιθανότητας:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right].$$

Note that the συνάρτηση πυκνότητας πιθανότητας is only defined when $\boldsymbol{\Sigma}$ is invertible. More generally, any τυχαία μεταβλητή $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ admits the following innovation representation:

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a standard normal vector and $\mathbf{A} \in \mathbb{R}^{d \times d}$ satisfies $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$. This innovation representation is valid even when the πίνακας συνδιακύμανσης $\mathbf{\Sigma}$ is singular, in which case $\mathbf{A}$ is not necessarily full-rank [77, Ch. 23].

The family of multivariate normal distributions is exceptional among πιθανοτικό μοντέλος for numerical quantities at least for the following reasons. First, the family is closed under affine transformations, i.e.,

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma}) \text{ implies } \mathbf{B}\mathbf{x}+\mathbf{c} \sim \mathcal{N}\big(\mathbf{B}\boldsymbol{\mu} + \mathbf{c}, \mathbf{B}\mathbf{\Sigma}\mathbf{B}^{T}\big).$$

Second, the κατανομή πιθανότητας $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ maximizes the differential εντροπία among all distributions with the same πίνακας συνδιακύμανσης $\mathbf{\Sigma}$ [13].

Βλέπε επίσης: πιθανοτικό μοντέλο, διάνυσμα χαρακτηριστικών, κατανομή πιθανότητας, τυχαία μεταβλητή, πίνακας συνδιακύμανσης, συνάρτηση πυκνότητας πιθανότητας, standard normal vector, Gaussian RV, μέση τιμή, εντροπία.

πολυωνυμική παλινδρόμηση Polynomial regression aims at learning a polynomial υπόθεση map to predict a numeric ετικέτα based on the numeric features of a data point. For data points characterized by a single numeric feature, polynomial regression uses the χώρος υποθέσεων $\mathcal{H}_{d}^{(\mathrm{poly})} := \{h(x) = \sum_{j=0}^{d-1} x^{j} w_{j}\}$. The quality of a polynomial υπόθεση map is measured using the average απώλεια τετραγωνικού σφάλματος incurred on a set of σημείο δεδομένων με ετικέτας (which we refer to as the σύνολο εκπαίδευσης).

Βλέπε επίσης: regression, υπόθεση, map, ετικέτα, feature, data point, χώρος υποθέσεων, απώλεια τετραγωνικού σφάλματος, σημείο δεδομένων με ετικέτα, σύνολο εκπαίδευσης.

πραγμάτωση Consider an τυχαία μεταβλητή x which maps each element (i.e., outcome or elementary event) $\omega \in \mathcal{P}$ of a χώρος πιθανοτήτων $\mathcal{P}$ to an element a of a measurable space $\mathcal{N}$ [2], [6], [72]. A realization of x is any element $a' \in \mathcal{N}$ such that there is an element $\omega' \in \mathcal{P}$ with $x(\omega') = a'$.

Βλέπε επίσης: τυχαία μεταβλητή, χώρος πιθανοτήτων.

προβεβλημένη κάθοδος κλίσης Consider an ελαχιστοποίηση εμπειρικής διακινδύνευσης-based method that uses a parametrized model with χώρος παραμέτρων $\mathcal{W} \subseteq \mathbb{R}^d$. Even if the αντικειμενική συνάρτηση of ελαχιστοποίηση εμπειρικής διακινδύνευσης is λεία, we cannot use basic κάθοδος κλίσης, as it does not take into account contraints on the optimization variable (i.e., the παράμετροι μοντέλου). Projected κάθοδος κλίσης (projected gradient descent; projected GD) extends basic κάθοδος κλίσης to handle constraints on the optimization variable (i.e., the παράμετροι μοντέλου). A single iteration of projected κάθοδος κλίσης consists of first taking a βήμα κλίσης and then projecting the result back onto the χώρος παραμέτρων.

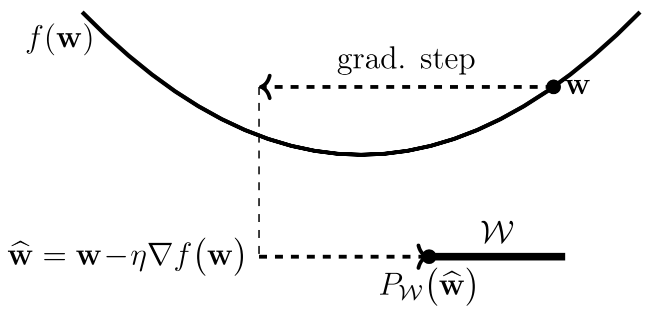


Fig. 23. Projected κάθοδος κλίσης augments a basic βήμα κλίσης with a προβολή back onto the constraint set $\mathcal{W}$.

Βλέπε επίσης: ελαχιστοποίηση εμπειρικής διακινδύνευσης, model, χώρος παραμέτρων, αντικειμενική συνάρτηση, λεία, κάθοδος κλίσης, παράμετροι μοντέλου, βήμα κλίσης, προβολή.

προβλέπουσα Μία προβλέπουσα είναι μία map υπόθεσης πραγματικής τιμής. Δεδομένου ενός σημείου δεδομένων με χαρακτηριστικά $\mathbf{x}$, η τιμή $h(\mathbf{x}) \in \mathbb{R}$ χρησιμοποιείται ως η πρόβλεψη για την αληθή αριθμητική ετικέτα $y \in \mathbb{R}$ του σημείου δεδομένων.

Βλέπε επίσης: υπόθεση, map, data point, feature, πρόβλεψη, ετικέτα.

πρόβλεψη Μία πρόβλεψη είναι μία εκτίμηση ή προσέγγιση για κάποια ποσότητα ενδιαφέροντος. Η μηχανική μάθηση περιστρέφεται γύρω από τη μάθηση ή εύρεση μίας map υπόθεσης h που διαβάζει τα χαρακτηριστικά $\mathbf{x}$ ενός σημείου δεδομένων και δίνει μία πρόβλεψη $\widehat{y} := h(\mathbf{x})$ για την ετικέτα του y.

Βλέπε επίσης: ml, υπόθεση, map, feature, data point, ετικέτα.

πρόβλημα βελτιστοποίησης An optimization problem is a mathematical structure consisting of an αντικειμενική συνάρτηση $f : \mathcal{U} \to \mathcal{V}$ defined over an optimization variable $\mathbf{w} \in \mathcal{U}$, together with a feasible set $\mathcal{W} \subseteq \mathcal{U}$. The co-domain $\mathcal{V}$ is assumed to be ordered, meaning that for any two elements $\mathbf{a}, \mathbf{b} \in \mathcal{V}$, we can determine whether $\mathbf{a} < \mathbf{b}$, $\mathbf{a} = \mathbf{b}$, or $\mathbf{a} > \mathbf{b}$. The goal of optimization is to find those values $\mathbf{w} \in \mathcal{W}$ for which the objective $f(\mathbf{w})$ is extremal—i.e., minimal or maximal [58], [59], [29].

Βλέπε επίσης: αντικειμενική συνάρτηση.

προβολή Θεωρούμε ένα υποσύνολο $\mathcal{W} \subseteq \mathbb{R}^d$ του d-διάστατου Ευκλείδειου χώρου. Ορίζουμε την προβολή $P_{\mathcal{W}}(\mathbf{w})$ ενός διανύσματος $\mathbf{w} \in \mathbb{R}^d$ στο $\mathcal{W}$ ως

$$P_{\mathcal{W}}(\mathbf{w}) = \operatorname*{arg\,min}_{\mathbf{w}' \in \mathcal{W}} \|\mathbf{w} - \mathbf{w}'\|_2 . \tag{9}$$

Με άλλα λόγια, η $P_{\mathcal{W}}(\mathbf{w})$ είναι το διάνυσμα στο $\mathcal{W}$ που είναι πιο κοντά στο $\mathbf{w}$. Η προβολή είναι καλά ορισμένη μόνο για υποσύνολα $\mathcal{W}$ για τα οποία υπάρχει το παραπάνω ελάχιστο [58].

Βλέπε επίσης: Ευκλείδειος χώρος, ελάχιστο.

προσδοκία Consider a numeric διάνυσμα χαρακτηριστικών $\mathbf{x} \in \mathbb{R}^d$ which we interpret as the πραγμάτωση of an τυχαία μεταβλητή with a κατανομή πιθανότητας $p(\mathbf{x})$. The expectation of $\mathbf{x}$ is defined as the integral $\mathbb{E}\{\mathbf{x}\} := \int \mathbf{x} p(\mathbf{x})$. Note that the expectation is only defined if this integral exists, i.e., if the τυχαία μεταβλητή is integrable [2], [6], [72]. Fig. 24 illustrates the expectation of a scalar discrete τυχαία μεταβλητή x which takes on values from a finite set only.

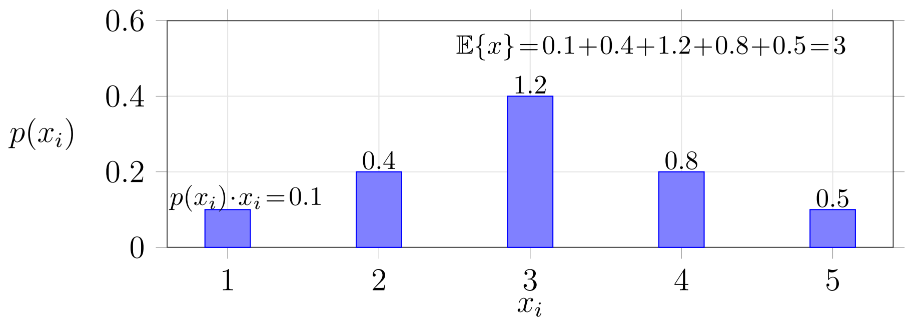


Fig. 24. The expectation of a discrete τυχαία μεταβλητή x is obtained by summing up its possible values x_i, weighted by the corresponding probability $p(x_i) = p(x = x_i)$.

Βλέπε επίσης: διάνυσμα χαρακτηριστικών, πραγμάτωση, τυχαία μεταβλητή, κατανομή πιθανότητας, probability.

προσεγγίσιμος Μία κυρτή συνάρτηση για την οποία ο εγγύς τελεστής μπορεί να υπολογιστεί αποτελεσματικά αναφέρεται μερικές φορές ως προσεγγίσιμη ή απλή [78].

Βλέπε επίσης: convex, συνάρτηση, εγγύς τελεστής.

προστασία της ιδιωτικότητας Θεωρούμε κάποια μέθοδο μηχανικής μάθησης $\mathcal{A}$ που διαβάζει ένα σύνολο δεδομένων $\mathcal{D}$ και δίνει κάποια έξοδο $\mathcal{A}(\mathcal{D})$. Η έξοδος θα μπορούσε να είναι οι παράμετροι μοντέλου $\widehat{\mathbf{w}}$ που μαθαίνονται ή η πρόβλεψη $\hat{h}(\mathbf{x})$ που προκύπτει για ένα συγκεκριμένο σημείο δεδομένων με χαρακτηριστικά $\mathbf{x}$. Πολλές σημαντικές εφαρμογές μηχανικής μάθησης περιλαμβάνουν σημεία δεδομένων που αντιπροσωπεύουν ανθρώπους. Κάθε σημείο δεδομένων χαρακτηρίζεται από χαρακτηριστικά $\mathbf{x}$, ενδεχομένως μία ετικέτα y, και ένα ευαίσθητο ιδιοχαρακτηριστικό s (π.χ. μία πρόσφατη ιατρική διάγνωση). Στο περίπου, προστασία της ιδιωτικότητας σημαίνει ότι θα έπρεπε να είναι αδύνατο να συμπεράνουμε, από την έξοδο $\mathcal{A}(\mathcal{D})$, οποιοδήποτε από τα ευαίσθητα ιδιοχαρακτηριστικά των σημείων δεδομένων στο $\mathcal{D}$. Από μαθηματικής άποψης, η προστασία της ιδιωτικότητας απαιτεί την μη αντιστρεψιμότητα της map $\mathcal{A}(\mathcal{D})$. Γενικά, το να κάνουμε απλώς το $\mathcal{A}(\mathcal{D})$ μη αντιστρέψιμο είναι συνήθως ανεπαρκές για την προστασία της ιδιωτικότητας. Χρειάζεται να κάνουμε το $\mathcal{A}(\mathcal{D})$ επαρκώς μη αντιστρέψιμο.

Βλέπε επίσης: ml, σύνολο δεδομένων, παράμετροι μοντέλου, πρόβλεψη, data point, feature, ετικέτα, ευαίσθητο ιδιοχαρακτηριστικό, map.

πυρήνας Θεωρούμε σημεία δεδομένων που χαρακτηρίζονται από ένα διάνυσμα χαρακτηριστικών $\mathbf{x} \in \mathcal{X}$ με έναν γενικό χώρο χαρακτηριστικών $\mathcal{X}$. Ένας πυρήνας (πραγματικής τιμής) $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ αποδίδει σε κάθε ζεύγος διανυσμάτων χαρακτηριστικών $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ έναν πραγματικό αριθμό $K(\mathbf{x}, \mathbf{x}')$. Η τιμή $K(\mathbf{x}, \mathbf{x}')$ συχνά ερμηνεύεται ως ένα μέτρο για την ομοιότητα μεταξύ $\mathbf{x}$ και $\mathbf{x}'$. Οι μέθοδοι πυρήνα χρησιμοποιούν έναν πυρήνα για να μετασχηματίσουν το διάνυσμα χαρακτηριστικών $\mathbf{x}$ σε ένα νέο διάνυσμα χαρακτηριστικών $\mathbf{z} = K(\mathbf{x}, \cdot)$. Αυτό το καινούριο διάνυσμα χαρακτηριστικών ανήκει σε έναν γραμμικό χώρο χαρακτηριστικών $\mathcal{X}'$ που είναι (γενικά) διαφορετικός από τον αρχικό χώρο χαρακτηριστικών $\mathcal{X}$. Ο χώρος χαρακτηριστικών $\mathcal{X}'$ έχει μία συγκεκριμένη μαθηματική δομή, δηλαδή είναι ένας αναπαραγωγός πυρήνας χώρος Hilbert [17], [55].

Βλέπε επίσης: data point, διάνυσμα χαρακτηριστικών, χώρος χαρακτηριστικών, kernel method, χώρος Hilbert.

ρυθμός μάθησης Θεωρούμε μία επαναληπτική μέθοδο μηχανικής μάθησης για την εύρεση ή μάθηση μίας χρήσιμης υπόθεσης $h \in \mathcal{H}$. Μία τέτοια επαναληπτική μέθοδος επαναλαμβάνει όμοια υπολογιστικά βήματα (ενημέρωσης) που προσαρμόζουν ή τροποποιούν την τρέχουσα υπόθεση για να προκύψει μία βελτιωμένη υπόθεση. Ένα καλά γνωστό παράδειγμα μίας τέτοιας επαναληπτικής μεθόδου μάθησης είναι η κάθοδος κλίσης και οι παραλλαγές της, στοχαστική κάθοδος κλίσης και προβεβλημένη κάθοδος κλίσης. Μία παράμετρος-κλειδί μίας επαναληπτικής μεθόδου είναι ο ρυθμός μάθησης. Ο ρυθμός μάθησης ελέγχει τον βαθμό που η τρέχουσα υπόθεση μπορεί να τροποποιηθεί κατά τη διάρκεια μίας μονής επανάληψης. Ένα καλά γνωστό παράδειγμα μίας τέτοιας παραμέτρου είναι το μέγεθος βήματος που χρησιμοποιείται στην καθόδο κλίσης [8, Κεφ. 5].

Βλέπε επίσης: ml, υπόθεση, κάθοδος κλίσης, στοχαστική κάθοδος κλίσης, προβεβλημένη κάθοδος κλίσης, παράμετρος, μέγεθος βήματος.

σημείο δεδομένων Ένα σημείο δεδομένων είναι οποιοδήποτε αντικείμενο που μεταφέρει πληροφορίες [13]. Σημεία δεδομένων μπορεί να είναι μαθητές, ραδιοσήματα, δέντρα, δάση, εικόνες, τυχαίες μεταβλητές, πραγματικοί αριθμοί, ή πρωτεΐνες. Χαρακτηρίζουμε σημεία δεδομένων χρησιμοποιώντας δύο τύπους ιδιοτήτων. Ένας τύπος ιδιότητας αναφέρεται ως ένα χαρακτηριστικό. Τα χαρακτηριστικά είναι ιδιότητες ενός σημείου δεδομένων που μπορούν να μετρηθούν ή να υπολογιστούν με έναν αυτόματο τρόπο. Ένα διαφορετικό είδος ιδιότητας αναφέρεται ως μία ετικέτα. Η ετικέτα ενός σημείου δεδομένων αναπαριστά κάποιο υψηλότερου επιπέδου γεγονός (ή ποσότητα ενδιαφέροντος). Σε αντίθεση με τα χαρακτηριστικά, ο προσδιορισμός της ετικέτας ενός σημείου δεδομένων συνήθως απαιτεί ανθρώπινους εμπειρογνώμονες (ή ειδικούς του τομέα). Στο περίπου, η μηχανική μάθηση στοχεύει στην πρόβλεψη της ετικέτας ενός σημείου δεδομένων βάσει μόνο των χαρακτηριστικών του.

Βλέπε επίσης: data, τυχαία μεταβλητή, feature, ετικέτα, ml.

σημείο δεδομένων με ετικέτα Ένα σημείο δεδομένων του οποίου η ετικέτα είναι γνωστή ή έχει προσδιοριστεί με κάποιον τρόπο που μπορεί να απαιτεί ανθρώπινη εργασία.

Βλέπε επίσης: data point, ετικέτα.

σκληρή συσταδοποίηση Η σκληρή συσταδοποίηση αναφέρεται στην εργασία χωρισμού ενός συγκεκριμένου συνόλου σημείων δεδομένων σε (μερικές) μη αλληλεπικαλυπτόμενες συστάδες. Η πιο ευρέως χρησιμοποιούμενη μέθοδος σκληρής συσταδοποίησης είναι ο αλγόριθμος k-μέσων.

Βλέπε επίσης: συσταδοποίηση, data point, συστάδα, αλγόριθμος k-μέσων.

στατιστικές διαστάσεις Ως στατιστικές διαστάσεις μίας μεθόδου μηχανικής μάθησης, αναφερόμαστε σε (ιδιότητες της) κατανομή πιθανότητας της εξόδου της κάτω από ένα πιθανοτικό μοντέλο για τα δεδομένα που τροφοδοτούνται στη μέθοδο.

Βλέπε επίσης: ml, κατανομή πιθανότητας, πιθανοτικό μοντέλο, data.

στοχαστική A process or method is called stochastic if it involves a random component or is governed by probabilistic laws. In ml, stochastic methods often incorporate randomness for reasons such as optimization (e.g., στοχαστική κάθοδος κλίσης) or αβεβαιότητα modeling (e.g., πιθανοτικό μοντέλος). A stochastic process is a collection of τυχαία μεταβλητής indexed by time or space, which are used to model random phenomena evolving over time (e.g., noise in sensors or financial time series).

Βλέπε επίσης: ml, στοχαστική κάθοδος κλίσης, αβεβαιότητα, πιθανοτικό μοντέλο, τυχαία μεταβλητή, μοντέλο στοχαστικής ομάδας.

στοχαστική κάθοδος κλίσης SGD (stochastic gradient descent; SGD) is obtained from κάθοδος κλίσης by replacing the gradient of the αντικειμενική συνάρτηση with a stochastic approximation. A main application of SGD is to train a parametrized model via ελαχιστοποίηση εμπειρικής διακινδύνευσης on a σύνολο εκπαίδευσης $\mathcal{D}$ that is either very large or not readily available (e.g., when data points are stored in a database distributed all over the planet). To evaluate the gradient of the empirical risk (as a συνάρτηση of the παράμετροι μοντέλου $\mathbf{w}$), we need to compute a sum $\sum_{r=1}^{m} \nabla_{\mathbf{w}} L\left(\mathbf{z}^{(r)}, \mathbf{w}\right)$ over all data points in the σύνολο εκπαίδευσης. We obtain a stochastic approximation to the gradient by replacing the sum $\sum_{r=1}^{m} \nabla_{\mathbf{w}} L\left(\mathbf{z}^{(r)}, \mathbf{w}\right)$ with a sum $\sum_{r \in \mathcal{B}} \nabla_{\mathbf{w}} L\left(\mathbf{z}^{(r)}, \mathbf{w}\right)$ over a randomly chosen subset $\mathcal{B} \subseteq \{1, \ldots, m\}$ (see Fig. 25). We often refer to these randomly chosen data points as a δέσμη. The δέσμη size $|\mathcal{B}|$ is an important παράμετρος of SGD. SGD with $|\mathcal{B}| > 1$ is referred to as mini-δέσμη SGD [79].

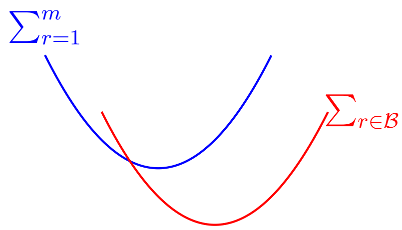


Fig. 25. SGD for ελαχιστοποίηση εμπειρικής διακινδύνευσης approximates the gradient $\sum_{r=1}^{m} \nabla_{\mathbf{w}} L\left(\mathbf{z}^{(r)}, \mathbf{w}\right)$ by replacing the sum over all data points in the σύνολο εκπαίδευσης (indexed by $r = 1, \ldots, m$) with a sum over a randomly chosen subset $\mathcal{B} \subseteq \{1, \ldots, m\}$.

Βλέπε επίσης: κάθοδος κλίσης, gradient, αντικειμενική συνάρτηση, stochastic, model, ελαχιστοποίηση εμπειρικής διακινδύνευσης, σύνολο εκπαίδευσης, data point, empirical risk, συνάρτηση, παράμετροι μοντέλου, δέσμη, παράμετρος.

στοχαστικός αλγόριθμος A stochastic αλγόριθμος uses a random mechanism during its execution. For example, στοχαστική κάθοδος κλίσης uses a randomly selected subset of data points to compute an approximation for the gradient of an αντικειμενική συνάρτηση. We can represent a stochastic αλγόριθμος by a stochastic process, i.e., the possible execution sequence is the possible outcomes of a random experiment [7], [121], [122].

Βλέπε επίσης: stochastic, αλγόριθμος, στοχαστική κάθοδος κλίσης, data point, gradient, αντικειμενική συνάρτηση, μέθοδος βελτιστοποίησης, μέθοδοι με βάση την κλίση.

συνάρτηση A function is a mathematical rule that assigns each element $u \in \mathcal{U}$ exactly one element $v \in \mathcal{V}$ [2]. We write this as $f : \mathcal{U} \to \mathcal{V}$, where $\mathcal{U}$ is the domain and $\mathcal{V}$ the co-domain of f. That is, a function f defines a unique output $f(u) \in \mathcal{V}$ for every input $u \in \mathcal{U}$.

συνάρτηση απώλειας A loss συνάρτηση is a map

$$L : \mathcal{X} \times \mathcal{Y} \times \mathcal{H} \rightarrow \mathbb{R}_{+} : \big((\mathbf{x}, y), h\big) \mapsto L\left((\mathbf{x}, y), h\right).$$

It assigns a non-negative real number (i.e., the loss) $L\left((\mathbf{x}, y), h\right)$ to a pair that consists of a data point, with features $\mathbf{x}$ and ετικέτα y, and a υπόθεση $h \in \mathcal{H}$. The value $L\left((\mathbf{x}, y), h\right)$ quantifies the discrepancy between the true ετικέτα y and the πρόβλεψη $h(\mathbf{x})$. Lower (closer to zero) values $L\left((\mathbf{x}, y), h\right)$ indicate a smaller discrepancy between πρόβλεψη $h(\mathbf{x})$ and ετικέτα y. Fig. 26 depicts a loss συνάρτηση for a given data point, with features $\mathbf{x}$ and ετικέτα y, as a συνάρτηση of the υπόθεση $h \in \mathcal{H}$.

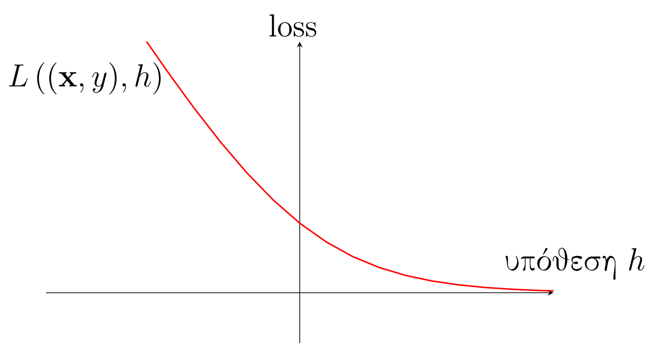


Fig. 26. Some loss συνάρτηση $L\left((\mathbf{x}, y), h\right)$ for a fixed data point, with διάνυσμα χαρακτηριστικών $\mathbf{x}$ and ετικέτα y, and a varying υπόθεση h. ml methods try to find (or learn) a υπόθεση that incurs minimal loss.

Βλέπε επίσης: loss, συνάρτηση, map, data point, feature, ετικέτα, υπόθεση, πρόβλεψη, διάνυσμα χαρακτηριστικών, ml.

συνάρτηση ενεργοποίησης Σε κάθε τεχνητό νευρώνα εντός ενός τεχνητού νευρωνικού δικτύου αποδίδεται μία συνάρτηση ενεργοποίησης (activation function) $\sigma(\cdot)$ που αντιστοιχίζει έναν σταθμισμένο συνδυασμό των εισόδων νευρώνα $x_1, \ldots, x_d$ σε μία μοναδική τιμή εξόδου $a = \sigma\big(w_1 x_1 + \ldots + w_d x_d\big)$. Σημείωση ότι κάθε νεωρώνας είναι παραμετροποιημένος από τα βάρη $w_1, \ldots, w_d$.

Βλέπε επίσης: ΤΝΔ, συνάρτηση, βάρη.

συνάρτηση πυκνότητας πιθανότητας The pdf $p(x)$ (probability density function; pdf) of a real-valued τυχαία μεταβλητή $x \in \mathbb{R}$ is a particular representation of its κατανομή πιθανότητας. If the pdf exists, it can be used to compute the probability that x takes on a value from a (measurable) set $\mathcal{B} \subseteq \mathbb{R}$ via $p(x \in \mathcal{B}) = \int_{\mathcal{B}} p(x')dx'$ [7, Ch. 3]. The pdf of a vector-valued τυχαία μεταβλητή $\mathbf{x} \in \mathbb{R}^d$ (if it exists) allows us to compute the probability of $\mathbf{x}$ belonging to a (measurable) region $\mathcal{R}$ via $p(\mathbf{x} \in \mathcal{R}) = \int_{\mathcal{R}} p(\mathbf{x}')dx'_1 \ldots dx'_d$ [7, Ch. 3].

Βλέπε επίσης: τυχαία μεταβλητή, κατανομή πιθανότητας, probability.

συνδεδεμένος γράφος Ένας μη κατευθυνόμενος γράφος $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ είναι συνδεδεμένος αν κάθε μη κενό υποσύνολο $\mathcal{V}' \subset \mathcal{V}$ έχει τουλάχιστον μία ακμή που το συνδέει με το $\mathcal{V} \setminus \mathcal{V}'$.

Βλέπε επίσης: graph.

συνθήκη μηδενικής κλίσης Consider the unconstrained optimization problem $\min_{\mathbf{w} \in \mathbb{R}^d} f(\mathbf{w})$ with a λεία and convex αντικειμενική συνάρτηση $f(\mathbf{w})$. A necessary and sufficient condition for a vector $\widehat{\mathbf{w}} \in \mathbb{R}^d$ to solve this problem is that the gradient $\nabla f\big(\widehat{\mathbf{w}}\big)$ is the zero vector,

$$\nabla f\big(\widehat{\mathbf{w}}\big) = \mathbf{0} \Leftrightarrow f\big(\widehat{\mathbf{w}}\big) = \min_{\mathbf{w} \in \mathbb{R}^d} f(\mathbf{w}).$$

Βλέπε επίσης: optimization problem, λεία, convex, αντικειμενική συνάρτηση, gradient.

σύνολο δεδομένων Ένα σύνολο δεδομένων αναφέρεται σε μία συλλογή σημείων δεδομένων. Αυτά τα σημεία δεδομένων φέρουν πληροφορίες σχετικά με κάποια ποσότητα ενδιαφέροντος (ή ετικέτα) εντός μίας εφαρμογής μηχανικής μάθησης. Οι μέθοδοι μηχανικής μάθησης χρησιμοποιούν σύνολα δεδομένων για την εκπαίδευση μοντέλων (π.χ. μέσω ελαχιστοποίησης εμπειρικής διακινδύνευσης) και την επικύρωση μοντέλων. Σημείωση ότι η έννοιά μας ενός συνόλου δεδομένων είναι πολύ ευέλικτη, καθώς επιτρέπει πολύ διαφορετικούς τύπους σημείων δεδομένων. Πράγματι, σημεία δεδομένων μπορεί να είναι συγκεκριμένα φυσικά αντικείμενα (όπως άνθρωποι ή ζώα) ή αφηρημένα αντικείμενα (όπως αριθμοί). Ως ένα χαρακτηριστικό παράδειγμα, το Σχ. 27 απεικονίζει ένα σύνολο δεδομένων που αποτελείται από αγελάδες ως σημεία δεδομένων.

Σχ. 27. "Cows in the Swiss Alps" από User:Huhu Uet αδειοδοτείται υπό την άδεια [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).

Αρκετά συχνά, ένας μηχανικός μηχανικής μάθησης δεν έχει άμεση πρόσβαση σε ένα σύνολο δεδομένων. Πράγματι, η πρόσβαση στο σύνολο δεδομένων στο Σχ. 27 θα απαιτούσε να επισκεφτούμε το κοπάδι αγελάδων στις Άλπεις. Αντ΄ αυτού, χρειάζεται να χρησιμοποιήσουμε μία προσέγγιση (ή αναπαράσταση) του συνόλου δεδομένων που είναι πιο βολική να χρησιμοποιηθεί. Διαφορετικά μαθηματικά μοντέλα έχουν αναπτυχθεί για την αναπαράσταση (ή προσέγγιση) συνόλων δεδομένων [80], [81], [82], [83]. Ένα από τα πιο εγκεκριμένα μοντέλα δεδομένων είναι το σχεσιακό μοντέλο, το οποίο οργανώνει δεδομένα ως έναν πίνακα (ή σχέση) [33], [80]. Ένας πίνακας αποτελείται από γραμμές και στήλες:

- Κάθε γραμμή του πίνακα αναπαριστά ένα μονό σημείο δεδομένων.
- Κάθε στήλη του πίνακα αντιστοιχεί σε ένα συγκεκριμένο ιδιοχαρακτηριστικό του σημείου δεδομένων. Οι μέθοδοι μηχανικής μάθησης μπορούν να χρησιμοποιήσουν ιδιοχαρακτηριστικά ως χαρακτηριστι-

κά και ετικέτες του σημείου δεδομένων.

Για παράδειγμα, ο Πίνακας 1 δείχνει μία αναπαράσταση του συνόλου δεδομένων στο Σχ. 27. Στο σχεσιακό μοντέλο, η σειρά των γραμμών δεν έχει σημασία, και κάθε ιδιοχαρακτηριστικό (δηλαδή στήλη) πρέπει να είναι ακριβώς ορισμένη με ένα πεδίο, το οποίο προσδιορίζει το σύνολο των πιθανών τιμών. Σε εφαρμογές μηχανικής μάθησης, αυτά τα πεδία ιδιοχαρακτηριστικών γίνονται ο χώρος χαρακτηριστικών και ο χώρος ετικετών.

Όνομα	**Βάρος**	**Ηλικία**	**Ύψος**	**Θερμοκρασία στομαχιού**
Zenzi	100	4	100	25
Berta	140	3	130	23
Resi	120	4	120	31

Πίνακας 1: Μία σχέση (ή πίνακας) που αναπαριστά το σύνολο δεδομένων στο Σχ. 27.

Ενώ το σχεσιακό μοντέλο είναι χρήσιμο για τη μελέτη πολλών εφαρμογών μηχανικής μάθησης, μπορεί να είναι ανεπαρκές όσον αφορά τις προϋποθέσεις για αξιόπιστη τεχνητή νοημοσύνη. Σύγχρονες προσεγγίσεις, όπως τα φύλλα δεδομένων για σύνολα δεδομένων, παρέχουν πιο περιεκτικά τεκμήρια, συμπεριλαμβανομένων λεπτομερειών για τη διαδικασία συλλογής του συνόλου δεδομένων, την επιθυμητή χρήση, και άλλες πληροφορίες σχετικές με τα συμφραζόμενα [39].

Βλέπε επίσης: data point, ετικέτα, ml, model, ελαχιστοποίηση εμπειρικής διακινδύνευσης, επικύρωση, data, feature, χώρος χαρακτηριστικών, χώρος ετικετών, αξιόπιστη ΤΝ.

σύνολο εκπαίδευσης Ένα σύνολο εκπαίδευσης είναι ένα σύνολο δεδομένων $\mathcal{D}$ που αποτελείται από κάποια σημεία δεδομένων που χρησιμοποιούνται στην ελαχιστοποίηση εμπειρικής διακινδύνευσης για τη μάθηση μίας υπόθεσης $\hat{h}$. Η μέση απώλεια της $\hat{h}$ στο σύνολο εκπαίδευσης αναφέρεται ως το σφάλμα εκπαίδευσης. Η σύγκριση του σφάλματος εκπαίδευσης με το σφάλματος επικύρωσης της $\hat{h}$ μας επιτρέπει να διαγνώσουμε τη μέθοδο μηχανικής μάθησης και ενημερώνει για το πώς να βελτιώσουμε το σφάλμα επικύρωσης (π.χ. χρησιμοποιώντας έναν διαφορετικό χώρο υποθέσεων ή συλλέγοντας περισσότερα σημεία δεδομένων) [8, Sec. 6.6].

Βλέπε επίσης: σύνολο δεδομένων, data point, ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπόθεση, loss, training error, σφάλμα επικύρωσης, ml, χώρος υποθέσεων.

σύνολο ελέγχου A set of data points that have been used neither to train a model (e.g., via ελαχιστοποίηση εμπειρικής διακινδύνευσης) nor in a σύνολο επικύρωσης to choose between different models.

Βλέπε επίσης: data point, model, ελαχιστοποίηση εμπειρικής διακινδύνευσης, σύνολο επικύρωσης.

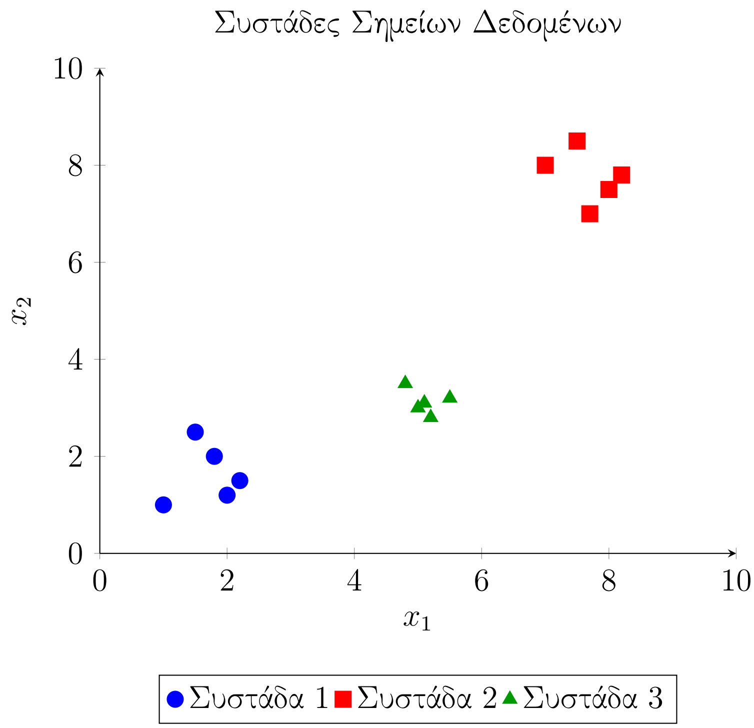
σύνολο επικύρωσης Ένα σύνολο σημείων δεδομένων που χρησιμοποιούνται για την εκτίμηση της διακινδύνευσης μίας υπόθεσης $\hat{h}$ που έχει μαθευτεί από κάποια μέθοδο μηχανικής μάθησης (π.χ. λύνοντας την ελαχιστοποίηση εμπειρικής διακινδύνευσης). Η μέση απώλεια της $\hat{h}$ στο σύνολο επικύρωσης αναφέρεται ως το σφάλμα επικύρωσης και μπορεί να χρησιμοποιηθεί για τη διάγνωση μίας μεθόδου μηχανικής μάθησης (βλέπε [8, Sec. 6.6]). Η σύγκριση μεταξύ σφάλματος εκπαίδευσης και σφάλματος επικύρωσης μπορεί να προσφέρει κατευθύνσεις για τη βελτίωση της μεθόδου μηχανικής μάθησης (όπως τη χρήση ενός διαφορετικού χώρου υποθέσεων).

Βλέπε επίσης: data point, διακινδύνευση, υπόθεση, ml, ελαχιστοποίηση εμπειρικής διακινδύνευσης, loss, επικύρωση, σφάλμα επικύρωσης, training error, χώρος υποθέσεων.

συσκευή Οποιοδήποτε φυσικό σύστημα που μπορεί να χρησιμοποιηθεί για την αποθήκευση και επεξεργασία δεδομένων. Στο πλαίσιο της μηχανικής μάθησης, συνήθως εννοούμε έναν υπολογιστή που έχει τη δυνατότητα να διαβάσει σημεία δεδομένων από διαφορετικές πηγές και, στη συνέχεια, να εκπαιδεύσει ένα μοντέλο μηχανικής μάθησης χρησιμοποιώντας αυτά τα σημεία δεδομένων.

Βλέπε επίσης: data, ml, data point, model.

συστάδα Μία συστάδα (cluster) είναι ένα υποσύνολο σημείων δεδομένων που είναι πιο όμοια μεταξύ τους παρά με τα σημεία δεδομένων εκτός της συστάδας. Το ποσοτικό μέτρο της ομοιότητας μεταξύ σημείων δεδομένων είναι μία επιλογή σχεδιασμού. Αν σημεία δεδομένων χαρακτηρίζονται από Ευκλείδεια διανύσματα χαρακτηριστικών $\mathbf{x} \in \mathbb{R}^d$, μπορούμε να ορίσουμε την ομοιότητα μεταξύ δύο σημείων δεδομένων μέσω της Ευκλείδειας απόστασης μεταξύ των διανυσμάτων χαρακτηριστικών τους. Ένα παράδειγμα τέτοιων συστάδων παρουσιάζεται στο Σχ. 28.



Σχ. 28. Εικονογράφηση τριών συστάδων σε έναν δισδιάστατο χώρο χαρακτηριστικών. Κάθε συστάδα ομαδοποιεί σημεία δεδομένων που είναι πιο όμοια μεταξύ τους παρά με αυτά σε άλλες συστάδες, με βάση την Ευκλείδεια απόσταση.

Βλέπε επίσης: data point, διάνυσμα χαρακτηριστικών, χώρος χαρακτηριστικών.

συσταδοποίηση Οι μέθοδοι συσταδοποίησης (clustering) διαμερίζουν ένα δεδομένο σύνολο σημείων δεδομένων σε λίγα υποσύνολα, τα οποία αναφέρονται ως συστάδες. Κάθε συστάδα αποτελείται από σημεία δεδομένων που είναι πιο όμοια μεταξύ τους παρά με σημεία δεδομένων εκτός της συστάδας. Διαφορετικές μέθοδοι συσταδοποίησης χρησιμοποιούν διαφορετικά μέτρα για την ομοιότητα μεταξύ σημείων δεδομένων και διαφορετικές μορφές αναπαράστασης συστάδων. Η μέθοδος συσταδοποίησης του αλγόριθμου k-μέσων χρησιμοποιεί το μέσο διάνυσμα χαρακτηριστικών μίας συστάδας (δηλαδή τη μέση τιμή της συστάδας) ως τον αντιπρόσωπό της. Μία δημοφιλής μέθοδος μαλακής συσταδοποίησης βασισμένη σε GMM αναπαριστά μία συστάδα από μία πολυμεταβλητή κανονική κατανομή.

Βλέπε επίσης: data point, συστάδα, αλγόριθμος k-μέσων, feature, μέση τιμή, soft clustering, GMM, πολυμεταβλητή κανονική κατανομή.

συσταδοποίηση γράφου Η συσταδοποίηση γράφου (graph clustering) στοχεύει στη συσταδοποίηση σημείων δεδομένων που αναπαριστώνται ως οι κόμβοι ενός γράφου $\mathcal{G}$. Οι ακμές του $\mathcal{G}$ αναπαριστούν κατά ζεύγη ομοιότητες μεταξύ σημείων δεδομένων. Κάποιες φορές μπορούμε να ποσοτικοποιήσουμε την έκταση αυτών των ομοιοτήτων με ένα βάρος ακμής [74], [84].

Βλέπε επίσης: graph, συσταδοποίηση, data point, βάρος ακμής.

συσταδοποίηση με βάση τη ροή Η συσταδοποίηση με βάση τη ροή ομαδοποιεί τους κόμβους ενός μη κατευθυνόμενου γράφου με την εφαρμογή συσταδοποίησης αλγόριθμου k-μέσων σε διανύσματα χαρακτηριστικών από θέμα κόμβων. Αυτά τα διανύσματα χαρακτηριστικών κατασκευάζονται από ροές δικτύου μεταξύ προσεκτικά επιλεγμένων πηγών και κόμβων προορισμού [84].

Βλέπε επίσης: συσταδοποίηση, graph, αλγόριθμος k-μέσων, διάνυσμα χαρακτηριστικών.

σφάλμα εκπαίδευσης Η μέση απώλεια μίας υπόθεσης όταν προβλέπει τις ετικέτες των σημείων δεδομένων σε ένα σύνολο εκπαίδευσης. Κάποιες φορές αναφερόμαστε ως σφάλμα εκπαίδευσης και στην ελάχιστη μέση απώλεια που επιτυγχάνεται από μία λύση της ελαχιστοποίησης εμπειρικής διακινδύνευσης.

Βλέπε επίσης: loss, υπόθεση, ετικέτα, data point, σύνολο εκπαίδευσης, ελαχιστοποίηση εμπειρικής διακινδύνευσης.

σφάλμα εκτίμησης Θεωρούμε σημεία δεδομένων, καθένα με διάνυσμα χαρακτηριστικών $\mathbf{x}$ και ετικέτα y. Σε κάποιες εφαρμογές, μπορούμε να μοντελοποιήσουμε τη σχέση μεταξύ του διανύσματος χαρακτηριστικών και της ετικέτας ενός σημείου δεδομένων ως $y = \bar{h}(\mathbf{x}) + \varepsilon$. Εδώ χρησιμοποιούμε κάποια αληθή υποκείμενη υπόθεση $\bar{h}$ και έναν όρο θορύβου ε που συνοψίζει οποιαδήποτε σφάλματα μοντελοποίησης ή ετικετοποίησης. Το σφάλμα εκτίμησης που προκαλείται από μία μέθοδο μηχανικής μάθησης που μαθαίνει μία υπόθεση $\widehat{h}$, π.χ. χρησιμοποιώντας την ελαχιστοποίηση εμπειρικής διακινδύνευσης, ορίζεται ως $\widehat{h}(\mathbf{x}) - \bar{h}(\mathbf{x})$, για κάποιο διάνυσμα χαρακτηριστικών. Για έναν παραμετρικό χώρο υποθέσεων, ο οποίος αποτελείται από maps υπόθεσης καθορισμένες από παραμέτρους μοντέλου $\mathbf{w}$, μπορούμε να ορίσουμε το σφάλμα εκτίμησης ως $\Delta\mathbf{w} = \widehat{\mathbf{w}} - \overline{\mathbf{w}}$ [63], [28].
Βλέπε επίσης: data point, διάνυσμα χαρακτηριστικών, ετικέτα, υπόθεση, ml, ελαχιστοποίηση εμπειρικής διακινδύνευσης, χώρος υποθέσεων, map, παράμετροι μοντέλου.

σφάλμα επικύρωσης Θεωρούμε μία υπόθεση $\hat{h}$ που προκύπτει από κάποια μέθοδο μηχανικής μάθησης, π.χ. χρησιμοποιώντας την ελαχιστοποίηση εμπειρικής διακινδύνευσης σε ένα σύνολο εκπαίδευσης. Η μέση απώλεια της $\hat{h}$ σε ένα σύνολο επικύρωσης, το οποίο είναι διαφορετικό από το σύνολο εκπαίδευσης, αναφέρεται ως το σφάλμα επικύρωσης.

Βλέπε επίσης: υπόθεση, ml, ελαχιστοποίηση εμπειρικής διακινδύνευσης, σύνολο εκπαίδευσης, loss, σύνολο επικύρωσης, επικύρωση.

ταξινόμηση Η ταξινόμηση είναι μία εργασία καθορισμού μίας ετικέτας διακριτής τιμής y για ένα δεδομένο σημείο δεδομένων, βασισμένη μόνο στα χαρακτηριστικά του $\mathbf{x}$. Η ετικέτα y ανήκει σε ένα πεπερασμένο σύνολο, όπως $y \in \{-1, 1\}$ ή $y \in \{1, \ldots, 19\}$, και αντιπροσωπεύει την κατηγορία στην οποία ανήκει το αντίστοιχο σημείο δεδομένων.

Βλέπε επίσης: ετικέτα, data point, feature.

ταξινομητής Ένας ταξινομητής είναι μία υπόθεση (δηλαδή μία map) $h(\mathbf{x})$ που χρησιμοποιείται για να προβλεφθεί μία ετικέτα που παίρνει τιμές από ένα πεπερασμένο χώρο ετικετών. Μπορεί να χρησιμοποιήσουμε την ίδια την τιμή συνάρτησης $h(\mathbf{x})$ ως μία πρόβλεψη $\hat{y}$ για την ετικέτα. Ωστόσο, είναι σύνηθες να χρησιμοποιούμε μία map $h(\cdot)$ που παραδίδει μία αριθμητική ποσότητα. Η πρόβλεψη έπειτα προκύπτει από ένα απλό βήμα κατωφλιού. Για παράδειγμα, σε ένα πρόβλημα δυαδικής ταξινόμησης με $\mathcal{Y} \in \{-1, 1\}$, μπορεί να χρησιμοποιήσουμε μία map υπόθεσης πραγματικής τιμής $h(\mathbf{x}) \in \mathbb{R}$ ως ταξινομητή. Μία πρόβλεψη $\hat{y}$ μπορεί έπειτα να προκύψει μέσω κατωφλιού,

$$\hat{y} = 1 \text{ για } h(\mathbf{x}) \geq 0 \text{ και } \hat{y} = -1 \text{ διαφορετικά.} \tag{10}$$

Μπορούμε να χαρακτηρίσουμε έναν ταξινομητή από τις περιοχές αποφάσεων $\mathcal{R}_a$, για κάθε πιθανή τιμή ετικέτας $a \in \mathcal{Y}$.

Βλέπε επίσης: υπόθεση, map, ετικέτα, χώρος ετικετών, συνάρτηση, πρόβλεψη, ταξινόμηση, περιοχή αποφάσεων.

τελεστής ελάχιστης απόλυτης συρρίκνωσης και επιλογής The Lasso (least absolute shrinkage and selection operator - Lasso) is an instance of δομημένη ελαχιστοποίηση διακινδύνευσης. It learns the βάρη $\mathbf{w}$ of a linear map $h(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$ based on a σύνολο εκπαίδευσης. Lasso is obtained from γραμμική παλινδρόμηση by adding the scaled ℓ_1-νόρμα $\alpha \|\mathbf{w}\|_1$ to the average απώλεια τετραγωνικού σφάλματος incurred on the σύνολο εκπαίδευσης.

Βλέπε επίσης: δομημένη ελαχιστοποίηση διακινδύνευσης, βάρη, linear map, σύνολο εκπαίδευσης, γραμμική παλινδρόμηση, νόρμα, απώλεια τετραγωνικού σφάλματος.

τεχνητή νοημοσύνη (ΤΝ) Η τεχνητή νοημοσύνη (artificial intelligence - AI) αναφέρεται σε συστήματα που συμπεριφέρονται λογικά με την έννοια της μεγιστοποίησης μίας μακροπρόθεσμης ανταμοιβής. Η προσέγγιση στην τεχνητή νοημοσύνη με βάση τη μηχανική μάθηση είναι να εκπαιδευτεί ένα μοντέλο για την πρόβλεψη βέλτιστων ενεργειών. Αυτές οι προβλέψεις υπολογίζονται από παρατηρήσεις σχετικά με την κατάσταση του περιβάλλοντος. Η επιλογή της συνάρτησης απώλειας διαφοροποιεί τις εφαρμογές της τεχνητής νοημοσύνης από πιο βασικές εφαρμογές της μηχανικής μάθησης. Τα συστήματα της τεχνητής νοημοσύνης σπάνια έχουν πρόσβαση σε ένα σύνολο εκπαίδευσης με ετικέτες που να επιτρέπει τη μέτρηση της μέσης απώλειας για οποιαδήποτε πιθανή επιλογή παραμέτρων μοντέλου. Αντίθετα, τα συστήματα της τεχνητής νοημοσύνης χρησιμοποιούν παρατηρούμενα σήματα ανταμοιβής για να αποκτηθεί μία (σημειακή) εκτίμηση για την απώλεια που προκαλείται από την τρέχουσα επιλογή παραμέτρων μοντέλου.

Βλέπε επίσης: ανταμοιβή, ml, model, συνάρτηση απώλειας, σύνολο εκπαίδευσης, loss, παράμετροι μοντέλου.

τεχνητό νευρωνικό δίκτυο (ΤΝΔ) Ένα τεχνητό νευρωνικό δίκτυο (artificial neural network - ANN) είναι μία γραφική (ροή σήματος) αναπαράσταση μίας συνάρτησης που αντιστοιχίζει τα χαρακτηριστικά ενός σημείου δεδομένων κατά την είσοδό του σε μία πρόβλεψη για την σχετική ετικέτα κατά την έξοδό του. Η βασική μονάδα ενός τεχνητού νευρωνικού δικτύου είναι ο τεχνητός νευρώνας, ο οποίος εφαρμόζει μία συνάρτηση ενεργοποίησης στις σταθμισμένες εισόδους του. Οι έξοδοι αυτών των νευρώνων χρησιμεύουν ως είσοδοι για άλλους νευρώνες, σχηματίζοντας διασυνδεδεμένα επίπεδα.

Βλέπε επίσης: συνάρτηση, feature, data point, πρόβλεψη, ετικέτα, συνάρτηση ενεργοποίησης.

τοπικό μοντέλο Θεωρούμε μία συλλογή συσκευών που αναπαριστώνται ως κόμβοι $\mathcal{V}$ ενός δικτύου ομοσπονδιακής μάθησης. Ένα τοπικό μοντέλο (local model) $\mathcal{H}^{(i)}$ είναι ένας χώρος υποθέσεων εκχωρημένος σε έναν κόμβο $i \in \mathcal{V}$. Σε διαφορετικούς κόμβους μπορεί να αποδίδονται διαφορετικοί χώροι υποθέσεων, δηλαδή, γενικά, $\mathcal{H}^{(i)} \neq \mathcal{H}^{(i')}$ για διαφορετικούς κόμβους $i, i' \in \mathcal{V}$.

Βλέπε επίσης: συσκευή, δίκτυο ομοσπονδιακής μάθησης, model, χώρος υποθέσεων.

τοπικό σύνολο δεδομένων Η έννοια του τοπικού συνόλου δεδομένων είναι μεταξύ της έννοιας ενός σημείου δεδομένων και ενός συνόλου δεδομένων. Ένα τοπικό σύνολο δεδομένων αποτελείται από αρκετά μεμονωμένα σημεία δεδομένων, τα οποία χαρακτηρίζονται από χαρακτηριστικά και ετικέτες. Σε αντίθεση με ένα μονό σύνολο δεδομένων που χρησιμοποιείται σε βασικές μεθόδους μηχανικής μάθησης, ένα τοπικό σύνολο δεδομένων σχετίζεται επίσης με άλλα τοπικά σύνολα δεδομένων μέσω διαφορετικών εννοιών ομοιότητας. Αυτές οι ομοιότητες μπορεί να ανακύψουν από πιθανοτικά μοντέλα ή υποδομές επικοινωνίας και είναι κωδικοποιημένες στις ακμές ενός δικτύου ομοσπονδιακής μάθησης.

Βλέπε επίσης: σύνολο δεδομένων, data point, feature, ετικέτα, ml, πιθανοτικό μοντέλο, δίκτυο ομοσπονδιακής μάθησης.

τυχαία μεταβλητή Μία τυχαία μεταβλητή (random variable - RV) είναι μία συνάρτηση που αντιστοιχίζει από έναν χώρο πιθανοτήτων $\mathcal{P}$ σε έναν χώρο τιμών [6], [24]. Ο χώρος πιθανοτήτων αποτελείται από στοιχειώδη γεγονότα και είναι εξοπλισμένος με ένα μέτρο πιθανότητας που αποδίδει πιθανότητες σε υποσύνολα του $\mathcal{P}$. Διαφορετικοί τύποι τυχαίων μεταβλητών περιλαμβάνουν

- δυαδικές τυχαίες μεταβλητές, οι οποίες αντιστοιχίζουν κάθε στοιχειώδες γεγονός σε ένα στοιχείο ενός δυαδικού συνόλου (π.χ. $\{-1, 1\}$ ή $\{\text{γάτα}, \text{όχι γάτα}\}$)·
- τυχαίες μεταβλητές πραγματικής τιμής, οι οποίες παίρνουν τιμές στους πραγματικούς αριθμούς $\mathbb{R}$·
- τυχαίες μεταβλητές διανυσματικής τιμής, οι οποίες αντιστοιχίζουν στοιχειώδη γεγονότα στον Ευκλείδειο χώρο $\mathbb{R}^d$.

Η θεωρία πιθανοτήτων χρησιμοποιεί την έννοια των μετρήσιμων χώρων για να ορίσει ενδελεχώς και να μελετήσει τις ιδιότητες (μεγάλων) συλλογών τυχαίων μεταβλητών [6].

Βλέπε επίσης: συνάρτηση, χώρος πιθανοτήτων, probability, Ευκλείδειος χώρος.

τυχαίο δάσος Ένα τυχαίο δάσος (random forest) είναι ένα σύνολο διαφορετικών δέντρων αποφάσεων. Καθένα από αυτά τα δέντρα αποφάσεων προκύπτει από την προσαρμογή ενός διαταραγμένου αντιγράφου του αρχικού συνόλου δεδομένων.

Βλέπε επίσης: decision tree, σύνολο δεδομένων.

υπερπροσαρμογή Θεωρούμε μία μέθοδο μηχανικής μάθησης που χρησιμοποιεί ελαχιστοποίηση εμπειρικής διακινδύνευσης για να μάθει μία υπόθεση με την ελάχιστη εμπειρική διακινδύνευση σε ένα δεδομένο σύνολο εκπαίδευσης. Μία τέτοια μέθοδος υπερπροσαρμόζει το σύνολο εκπαίδευσης αν μάθει μία υπόθεση με μία μικρή εμπειρική διακινδύνευση στο σύνολο εκπαίδευσης αλλά με μία σημαντικά μεγαλύτερη απώλεια έξω από το σύνολο εκπαίδευσης.

Βλέπε επίσης: ml, ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπόθεση, ελάχιστο, empirical risk, σύνολο εκπαίδευσης, loss.

υπόθεση Μία υπόθεση (hypothesis) αναφέρεται σε μία map (ή συνάρτηση) $h : \mathcal{X} \to \mathcal{Y}$ από τον χώρο χαρακτηριστικών $\mathcal{X}$ στον χώρο ετικετών $\mathcal{Y}$. Δεδομένου ενός σημείου δεδομένων με χαρακτηριστικά $\mathbf{x}$, χρησιμοποιούμε μία map υπόθεσης h για να εκτιμήσουμε (ή να προσεγγίσουμε) την ετικέτα y χρησιμοποιώντας την πρόβλεψη $\hat{y} = h(\mathbf{x})$. Η μηχανική μάθηση έχει σχέση με τη μάθηση (ή εύρεση) μίας map υπόθεσης h, έτσι ώστε $y \approx h(\mathbf{x})$ για οποιοδήποτε σημείο δεδομένων (που έχει χαρακτηριστικά $\mathbf{x}$ και ετικέτα y).

Βλέπε επίσης: map, συνάρτηση, χώρος χαρακτηριστικών, χώρος ετικετών, data point, feature, ετικέτα, πρόβλεψη, ml.

υποκλίση For a real-valued συνάρτηση $f : \mathbb{R}^d \rightarrow \mathbb{R} : \mathbf{w} \mapsto f(\mathbf{w})$, a vector $\mathbf{a}$ such that $f(\mathbf{w}) \geq f(\mathbf{w}') + \left(\mathbf{w} - \mathbf{w}'\right)^T \mathbf{a}$ is referred to as a subgradient of f at $\mathbf{w}'$ [105], [29].

Βλέπε επίσης: συνάρτηση.

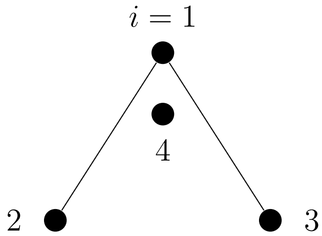
υπολογιστικές διαστάσεις Με τις υπολογιστικές διαστάσεις (computational aspects) μίας μεθόδου μηχανικής μάθησης, αναφερόμαστε κυρίως στους υπολογιστικούς πόρους που απαιτούνται για την εκτέλεσή της. Για παράδειγμα, αν μία μέθοδος μηχανικής μάθησης χρησιμοποιεί επαναληπτικές τεχνικές βελτιστοποίησης για να λύσει την ελαχιστοποίηση εμπειρικής διακινδύνευσης, τότε οι υπολογιστικές διαστάσεις της περιλαμβάνουν: 1) πόσες αριθμητικές πράξεις χρειάζονται για να εκτελεστεί μία μονή επανάληψη (δηλαδή ένα βήμα κλίσης)· και 2) πόσες επαναλήψεις χρειάζονται για να προκύψουν χρήσιμες παράμετροι μοντέλου. Ένα σημαντικό παράδειγμα μίας επαναληπτικής τεχνικής βελτιστοποίησης είναι η κάθοδος κλίσης.

Βλέπε επίσης: ml, ελαχιστοποίηση εμπειρικής διακινδύνευσης, βήμα κλίσης, παράμετροι μοντέλου, κάθοδος κλίσης.

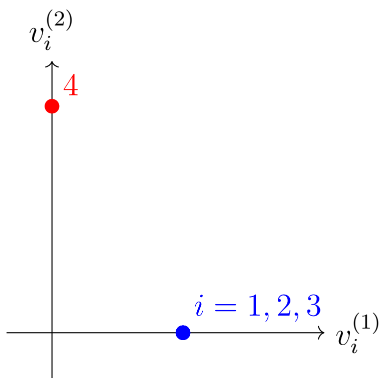
υποπροσαρμογή Consider an ml method that uses ελαχιστοποίηση εμπειρικής διακινδύνευσης to learn a υπόθεση with the ελάχιστο empirical risk on a given σύνολο εκπαίδευσης. Such a method is underfitting the σύνολο εκπαίδευσης if it is not able to learn a υπόθεση with a sufficiently small empirical risk on the σύνολο εκπαίδευσης. If a method is underfitting, it will typically also not be able to learn a υπόθεση with a small διακινδύνευση.

Βλέπε επίσης: ml, ελαχιστοποίηση εμπειρικής διακινδύνευσης, υπόθεση, ελάχιστο, empirical risk, σύνολο εκπαίδευσης, διακινδύνευση.

φασματική συσταδοποίηση Spectral συσταδοποίηση is a particular instance of συσταδοποίηση γράφου, i.e., it clusters data points represented as the nodes $i = 1, \ldots, n$ of a graph $\mathcal{G}$. Spectral συσταδοποίηση uses the ιδιοδιάνυσμαs of the πίνακας Laplace $\mathbf{L}^{(\mathcal{G})}$ to construct διάνυσμα χαρακτηριστικώνs $\mathbf{x}^{(i)} \in \mathbb{R}^d$ for each node (i.e., for each data point) $i = 1, \ldots, n$. We can feed these διάνυσμα χαρακτηριστικώνs into Ευκλείδειος χώρος-based συσταδοποίηση methods, such as αλγόριθμος k-μέσων or soft clustering via GMM. Roughly speaking, the διάνυσμα χαρακτηριστικώνs of nodes belonging to a well-connected subset (or συστάδα) of nodes in $\mathcal{G}$ are located nearby in the Ευκλείδειος χώρος $\mathbb{R}^d$ (see Fig. 29).



$$\mathbf{L}^{(\mathcal{G})} = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$$



$$\mathbf{V} = \left(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}, \mathbf{v}^{(4)}\right)$$

$$\mathbf{v}^{(1)} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Fig. 29. **Top.** Left: An undirected graph $\mathcal{G}$ with four nodes $i = 1, 2, 3, 4$, each representing a data point. Right: The πίνακας Laplace $\mathbf{L}^{(\mathcal{G})} \in \mathbb{R}^{4 \times 4}$ and its ανάλυση ιδιοτιμών. **Bottom.** Left: A διάγραμμα διασποράς of data points using the διάνυσμα χαρακτηριστικώνs $\mathbf{x}^{(i)} = \left(v_i^{(1)}, v_i^{(2)}\right)^T$. Right: Two ιδιοδιάνυσμας $\mathbf{v}^{(1)}, \mathbf{v}^{(2)} \in \mathbb{R}^d$ corresponding to the ιδιοτιμή $\lambda = 0$ of the πίνακας Laplace $\mathbf{L}^{(\mathcal{G})}$.

Βλέπε επίσης: συσταδοποίηση, συσταδοποίηση γράφου, data point, graph, ιδιοδιάνυσμα, πίνακας Laplace, διάνυσμα χαρακτηριστικών, Ευκλείδειος χώρος, αλγόριθμος k-μέσων, soft clustering, GMM, συστάδα, ανάλυση ιδιοτιμών, διάγραμμα διασποράς, ιδιοτιμή.

Φινλανδικό Μετεωρολογικό Ινστιτούτο The FMI (Finnish Meteorological Institute; FMI) is a government agency responsible for gathering and reporting weather data in Finland.

Βλέπε επίσης: data.

χαρακτηριστικό Ένα χαρακτηριστικό ενός σημείου δεδομένων είναι μία από τις ιδιότητες που μπορούν να μετρηθούν ή να υπολογιστούν εύκολα χωρίς την ανάγκη ανθρώπινης εποπτείας. Για παράδειγμα, αν ένα σημείο δεδομένων είναι μία ψηφιακή εικόνα (π.χ. αποθηκευμένη ως ένα αρχείο `.jpeg`), τότε θα μποούσαμε να χρησιμοποιήσουμε τις εντάσεις κόκκινου-πράσινου-μπλε των εικονοστοιχείων της ως χαρακτηριστικά. Συνώνυμα του όρου χαρακτηριστικό που χρησιμοποιούνται στον τομέα είναι «συμμεταβλητή,» «επεξηγηματική μεταβλητή,» «ανεξάρτητη μεταβλητή,» «είσοδος (μεταβλητή),» «προβλέπουσα (μεταβλητή),» ή «παλινδρομούσα μεταβλητή» [52], [53], [54].

Βλέπε επίσης: data point.

χάρτης χαρακτηριστικών Feature map refers to a map that transforms the original features of a data point into new features. The so-obtained new features might be preferable over the original features for several reasons. For example, the arrangement of data points might become simpler (or more linear) in the new χώρος χαρακτηριστικών, allowing the use of γραμμικό μοντέλοs in the new features. This idea is a main driver for the development of kernel methods [55]. Moreover, the hidden layers of a βαθύ δίκτυο can be interpreted as a trainable feature map followed by a γραμμικό μοντέλο in the form of the output layer. Another reason for learning a feature map could be that learning a small number of new features helps to avoid υπερπροσαρμογή and ensures ερμηνευσιμότητα [27]. The special case of a feature map delivering two numeric features is particularly useful for data visualization. Indeed, we can depict data points in a διάγραμμα διασποράς by using two features as the coordinates of a data point.

Βλέπε επίσης: feature, map, data point, χώρος χαρακτηριστικών, γραμμικό μοντέλο, kernel method, βαθύ δίκτυο, υπερπροσαρμογή, ερμηνευσιμότητα, data, διάγραμμα διασποράς.

χωρική συσταδοποίηση εφαρμογών με θόρυβο με βάση την πυκνότητα

DBSCAN (density-based spatial clustering of applications with noise - DBSCAN) refers to a συσταδοποίηση αλγόριθμος for data points that are characterized by numeric διάνυσμα χαρακτηριστικώνs. Like αλγόριθμος k-μέσων and soft clustering via GMM, also DBSCAN uses the Euclidean distances between διάνυσμα χαρακτηριστικώνs to determine the συστάδαs. However, in contrast to αλγόριθμος k-μέσων and GMM, DBSCAN uses a different notion of similarity between data points. DBSCAN considers two data points as similar if they are connected via a sequence (path) of close-by intermediate data points. Thus, DBSCAN might consider two data points as similar (and therefore belonging to the same cluster) even if their διάνυσμα χαρακτηριστικώνs have a large

Euclidean distance.

Βλέπε επίσης: συσταδοποίηση, αλγόριθμος, data point, διάνυσμα χαρακτηριστικών, αλγόριθμος k-μέσων, soft clustering, GMM, συστάδα.

χώρος ετικετών Θεωρούμε μία εφαρμογή μηχανικής μάθησης που περιλαμβάνει σημεία δεδομένων που χαρακτηρίζονται από χαρακτηριστικά και ετικέτες. Ο χώρος ετικετών αποτελείται από όλες τις πιθανές τιμές που η ετικέτα ενός σημείου δεδομένων μπορεί να πάρει. Μέθοδοι παλινδρόμησης, που στοχεύουν στην πρόβλεψη αριθμητικών ετικετών, συχνά χρησιμοποιούν τον χώρο ετικετών $\mathcal{Y} = \mathbb{R}$. Μέθοδοι δυαδικής ταξινόμησης χρησιμοποιούν έναν χώρο ετικετών που αποτελείται από δύο διαφορετικά στοιχεία, π.χ.

- $\mathcal{Y} = \{-1, 1\}$·
- $\mathcal{Y} = \{0, 1\}$·
- $\mathcal{Y} = \{$«εικόνα γάτας», «όχι εικόνα γάτας»$\}$.

Βλέπε επίσης: ml, data point, feature, ετικέτα, regression, ταξινόμηση.

χώρος παραμέτρων The παράμετρος space $\mathcal{W}$ of an ml model $\mathcal{H}$ is the set of all feasible choices for the παράμετροι μοντέλου (see Fig. 30). Many important ml methods use a model that is parametrized by vectors of the Ευκλείδειος χώρος $\mathbb{R}^d$. Two widely used examples of parametrized models are γραμμικό μοντέλος and βαθύ δίκτυος. The παράμετρος space is then often a subset $\mathcal{W} \subseteq \mathbb{R}^d$, e.g., all vectors $\mathbf{w} \in \mathbb{R}^d$ with a νόρμα smaller than one.

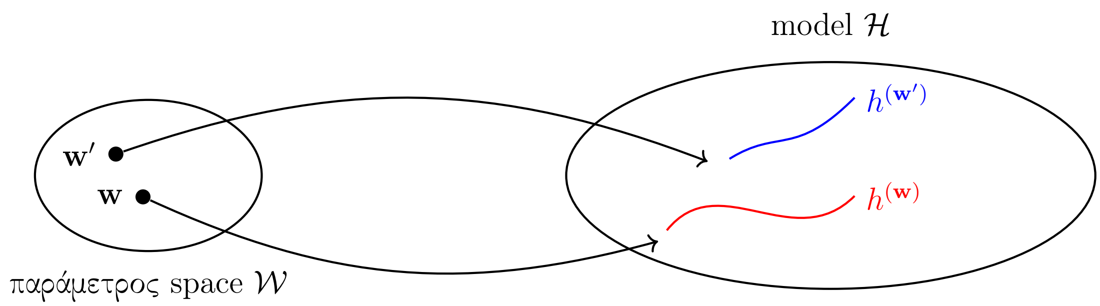


Fig. 30. The παράμετρος space $\mathcal{W}$ of an ml model $\mathcal{H}$ consists of all feasible choices for the παράμετροι μοντέλου. Each choice $\mathbf{w}$ for the παράμετροι μοντέλου selects a υπόθεση map $h^{(\mathbf{w})} \in \mathcal{H}$.

Βλέπε επίσης: παράμετρος, ml, model, παράμετροι μοντέλου, Ευκλείδειος χώρος, γραμμικό μοντέλο, βαθύ δίκτυο, νόρμα, υπόθεση, map.

χώρος πιθανοτήτων A probability space is a mathematical model of a physical process (a random experiment) with an uncertain outcome. Formally, a probability space $\mathcal{P}$ is a triplet $(\Omega, \mathcal{F}, P)$ where

- Ω is a δείγμα space containing all possible elementary outcomes of a random experiment;
- $\mathcal{F}$ is a sigma-algebra, a collection of subsets of Ω (called events) that satisfies certain closure properties under set operations;
- P is a probability measure, a συνάρτηση that assigns a probability $P(\mathcal{A}) \in [0, 1]$ to each event $\mathcal{A} \in \mathcal{F}$. The συνάρτηση must satisfy $P(\Omega) = 1$ and $P\left(\bigcup_{i=1}^{\infty} \mathcal{A}_i\right) = \sum_{i=1}^{\infty} P(\mathcal{A}_i)$ for any countable sequence of pairwise disjoint events $\mathcal{A}_1, \mathcal{A}_2, \ldots$ in $\mathcal{F}$.

Probability spaces provide the foundation for defining τυχαία μεταβλητής and to reason about αβεβαιότητα in ml applications [24], [6], [85].

Βλέπε επίσης: probability, model, δείγμα, συνάρτηση, τυχαία μεταβλητή, αβεβαιότητα, ml.

χώρος υποθέσεων Every practical ml method uses a υπόθεση space (or model) $\mathcal{H}$. The υπόθεση space of an ml method is a subset of all possible maps from the χώρος χαρακτηριστικών to the χώρος ετικετών. The design choice of the υπόθεση space should take into account available computational resources and στατιστικές διαστάσεις. If the computational infrastructure allows for efficient matrix operations, and there is an (approximately) linear relation between a set of features and a ετικέτα, a useful choice for the υπόθεση space might be the γραμμικό μοντέλο.

Βλέπε επίσης: ml, υπόθεση, model, map, χώρος χαρακτηριστικών, χώρος ετικετών, στατιστικές διαστάσεις, feature, ετικέτα, γραμμικό μοντέλο.

χώρος χαρακτηριστικών Ο χώρος χαρακτηριστικών μία δεδομένης εφαρμογής ή μεθόδου μηχανικής μάθησης αποτελείται από όλες τις πιθανές τιμές που μπορεί να πάρει το διάνυσμα χαρακτηριστικών ενός σημείου δεδομένων. Μία ευρέως χρησιμοποιούμενη επιλογή για τον χώρο χαρακτηριστικών είναι ο Ευκλείδειος χώρος $\mathbb{R}^d$, με τη διάσταση d να είναι ο αριθμός ξεχωριστών χαρακτηριστικών ενός σημείου δεδομένων.

Βλέπε επίσης: feature, ml, διάνυσμα χαρακτηριστικών, data point, feature, Ευκλείδειος χώρος.

χώρος Hilbert Ένας χώρος Hilbert είναι ένας πλήρης χώρος με εσωτερικό γινόμενο [86]. Για την ακρίβεια, είναι ένας διανυσματικός χώρος εξοπλισμένος με ένα εσωτερικό γινόμενο μεταξύ ζευγών διανυσμάτων, και πληροί την πρόσθετη προϋπόθεση της πληρότητας, δηλαδή κάθε ακολουθία Cauchy διανυσμάτων συγκλίνει σε ένα όριο εντός του χώρου. Ένα κανονικό παράδειγμα ενός χώρου Hilbert είναι ο Ευκλείδειος χώρος $\mathbb{R}^d$, για κάποια διάσταση d, που αποτελείται από διανύσματα $\mathbf{u} = \big(u_1, \ldots, u_d\big)^T$ και το τυπικό εσωτερικό γινόμενο $\mathbf{u}^T\mathbf{v}$.

Βλέπε επίσης: Ευκλείδειος χώρος.

0/1 απώλεια The 0/1 loss $L^{(0/1)}\left((\mathbf{x}, y), h\right)$ measures the quality of a ταξινομητής $h(\mathbf{x})$ that delivers a πρόβλεψη $\hat{y}$ (e.g., via thresholding (10)) for the ετικέτα y of a data point with features $\mathbf{x}$. It is equal to 0 if the πρόβλεψη is correct, i.e., $L^{(0/1)}\left((\mathbf{x}, y), h\right) = 0$ when $\hat{y} = y$. It is equal to 1 if the πρόβλεψη is wrong, i.e., $L^{(0/1)}\left((\mathbf{x}, y), h\right) = 1$ when $\hat{y} \neq y$.

Βλέπε επίσης: loss, ταξινομητής, πρόβλεψη, ετικέτα, data point, feature.

vertical federated learning (VFL) VFL refers to FL applications where συσκευήs have access to different features of the same set of data points [87]. Formally, the underlying global σύνολο δεδομένων is

$$\mathcal{D}^{(\mathrm{global})} := \left\{ \left(\mathbf{x}^{(1)}, y^{(1)}\right), \ldots, \left(\mathbf{x}^{(m)}, y^{(m)}\right) \right\}.$$

We denote by $\mathbf{x}^{(r)} = \left(x_1^{(r)}, \ldots, x_{d'}^{(r)}\right)^T$, for $r = 1, \ldots, m$, the complete διάνυσμα χαρακτηριστικώνs for the data points. Each συσκευή $i \in \mathcal{V}$ observes only a subset $\mathcal{F}^{(i)} \subseteq \{1, \ldots, d'\}$ of features, resulting in a τοπικό σύνολο δεδομένων $\mathcal{D}^{(i)}$ with διάνυσμα χαρακτηριστικώνs

$$\mathbf{x}^{(i,r)} = \left(x_{j_1}^{(r)}, \ldots, x_{j_d}^{(r)}\right)^T.$$

Some of the συσκευής might also have access to the ετικέτας $y^{(r)}$, for $r = 1, \ldots, m$, of the global σύνολο δεδομένων. One potential application of VFL is to enable collaboration between different healthcare providers. Each provider collects distinct types of measurements—such as blood values, electrocardiography, and lung X-rays—for the same patients. Another application is a national social insurance system, where health records, financial indicators, consumer behavior, and mobility data are collected by different institutions. VFL enables joint learning across these parties while allowing well-defined levels of προστασία της ιδιωτικότητας.

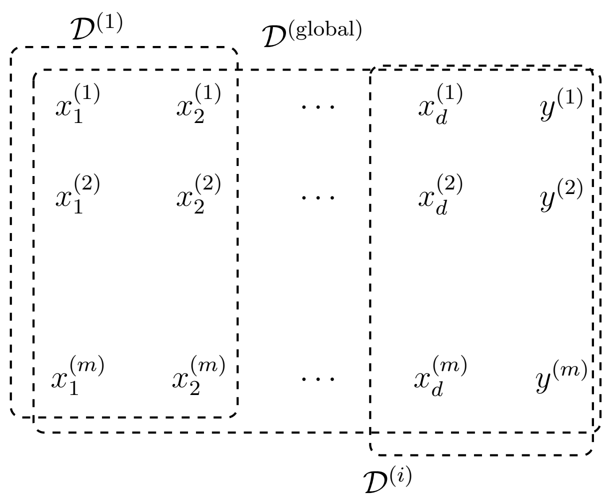


Fig. 31. VFL uses τοπικό σύνολο δεδομένωνs that are derived from the data points of a common global σύνολο δεδομένων. The τοπικό σύνολο δεδομένωνs differ in the choice of features used to characterize the data points.

Βλέπε επίσης: FL, συσκευή, feature, data point, σύνολο δεδομένων, διάνυσμα χαρακτηριστικών, τοπικό σύνολο δεδομένων, ετικέτα, data, προστασία της ιδιωτικότητας.

local interpretable model-agnostic explanations (LIME) Consider a trained model (or learned υπόθεση) $\widehat{h} \in \mathcal{H}$, which maps the διάνυσμα χαρακτηριστικών of a data point to the πρόβλεψη $\widehat{y} = \widehat{h}$. LIME is a technique for explaining the behavior of $\widehat{h}$, locally around a data point with διάνυσμα χαρακτηριστικών $\mathbf{x}^{(0)}$ [27]. The εξήγηση is given in the form of a local approximation $g \in \mathcal{H}'$ of $\widehat{h}$ (see Fig. 32). This approximation can be obtained by an instance of ελαχιστοποίηση εμπειρικής διακινδύνευσης with carefully designed σύνολο εκπαίδευσης. In particular, the σύνολο εκπαίδευσης consists of data points with διάνυσμα χαρακτηριστικών $\mathbf{x}$ close to $\mathbf{x}^{(0)}$ and the (pseudo-)ετικέτα $\widehat{h}(\mathbf{x})$. Note that we can use a different model $\mathcal{H}'$ for the approximation from the original model $\mathcal{H}$. For example, we can use a decision tree to locally approximate a βαθύ δίκτυο. Another widely-used choice for $\mathcal{H}'$ is the γραμμικό μοντέλο.

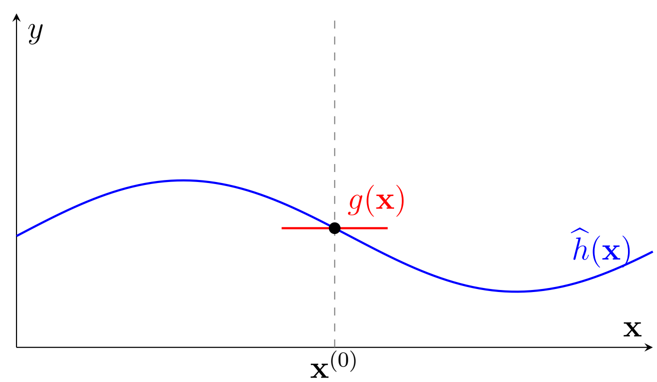


Fig. 32. To explain a trained model $\widehat{h} \in \mathcal{H}$, around a given διάνυσμα χαρακτηριστικών $\mathbf{x}^{(0)}$, we can use a local approximation $g \in \mathcal{H}'$.

Βλέπε επίσης: model, υπόθεση, διάνυσμα χαρακτηριστικών, data point, πρόβλεψη, εξήγηση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, σύνολο εκπαίδευσης, ετικέτα, decision tree, βαθύ δίκτυο, γραμμικό μοντέλο.

Gaussian random variable (Gaussian RV) A standard Gaussian τυχαία μεταβλητή is a real-valued τυχαία μεταβλητή x with συνάρτηση πυκνότητας πιθανότητας [7], [24], [68]

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp^{-x^2/2}.$$

Given a standard Gaussian τυχαία μεταβλητή x, we can construct a general Gaussian τυχαία μεταβλητή x' with μέση τιμή μ and διακύμανση σ^2 via $x' := \sigma x + \mu$. The κατανομή πιθανότητας of a Gaussian τυχαία μεταβλητή is referred to as normal distribution, denoted $\mathcal{N}(\mu, \sigma^2)$.

A Gaussian random vector $\mathbf{x} \in \mathbb{R}^d$ with πίνακας συνδιακύμανσης $\mathbf{C}$ and μέση τιμή $\boldsymbol{\mu}$ can be constructed as [24], [68], [76]

$$\mathbf{x} := \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$$

where $\mathbf{z} := \big(z_1, \ldots, z_d\big)^T$ is a vector of ανεξάρτητες και ταυτόσημα κατανεμημένες standard Gaussian τυχαία μεταβλητής, and $\mathbf{A} \in \mathbb{R}^{d \times d}$ is any matrix satisfying $\mathbf{A}\mathbf{A}^T = \mathbf{C}$. The κατανομή πιθανότητας of a Gaussian random vector is referred to as the πολυμεταβλητή κανονική κατανομή, denoted $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$.

Gaussian random vectors arise as finite-dimensional marginals of Gaussian processs, which define consistent joint Gaussian distributions over arbitrary (potentially infinite) index sets [88].

Gaussian τυχαία μεταβλητήs are widely used πιθανοτικό μοντέλοs in the statistical analysis of ml methods. Their significance arises partly from the central limit theorem (CLT), which is a mathematically precise formulation of the following rule-of-thumb: The average of a large number of independent τυχαία μεταβλητήs (not necessarily Gaussian themselves) tends towards a Gaussian τυχαία μεταβλητή [85].

Compared to other κατανομή πιθανότηταςs, the πολυμεταβλητή κανονική κατανομή is also distinct in that—in a mathematically precise sense—represents maximum uncertainty. Among all continuous random vectors with a given covariance matrix $\mathbf{C}$, the Gaussian random vector $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ maximizes differential εντροπία [13, Th. 8.6.5]. This makes Gaussian distributions a natural choice for capturing uncertainty (or lack of knowledge) in the absence of additional structural information.

Βλέπε επίσης: τυχαία μεταβλητή, συνάρτηση πυκνότητας πιθανότητας, μέση τιμή, διακύμανση, κατανομή πιθανότητας, πίνακας συνδιακύμανσης, ανεξάρτητες και ταυτόσημα κατανεμημένες, πολυμεταβλητή κανονική κατανομή, GP, πιθανοτικό μοντέλο, ml, CLT, εντροπία.

central limit theorem (CLT) The CLT refers to mathematically precise statements about the tendency of an average of a large number of independent τυχαία μεταβλητήs to tend towards a Gaussian RV.

Βλέπε επίσης: τυχαία μεταβλητή, Gaussian RV.

Gaussian process (GP) A GP is a collection of τυχαία μεταβλητήs $\{f(\mathbf{x})\}_{\mathbf{x} \in \mathcal{X}}$

indexed by input values $\mathbf{x}$ from some input space $\mathcal{X}$, such that, for any finite subset $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathcal{X}$, the corresponding τυχαία μεταβλητήs $f(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ have a joint multivariate Gaussian distribution:

$$\big(f(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\big) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}).$$

For a fixed input space $\mathcal{X}$, a GP is fully specified (or parametrized) by

- a μέση τιμή συνάρτηση $\mu(\mathbf{x}) = \mathbb{E}\{f(\mathbf{x})\}$
- and a covariance συνάρτηση $K\big(\mathbf{x}, \mathbf{x}'\big) = \mathbb{E}\{\big(f(\mathbf{x}) - \mu(\mathbf{x})\big)\big(f(\mathbf{x}') - \mu(\mathbf{x}')\big)\big\}$.

Example: We can interpret the temperature distribution across Finland (at a specific point in time) as the πραγμάτωση of a GP $f(\mathbf{x})$, where each input $\mathbf{x} = (\text{lat}, \text{lon})$ denotes a geographic location. Temperature observations from Φινλανδικό Μετεωρολογικό Ινστιτούτο weather stations provide δείγμαs of $f(\mathbf{x})$ at specific locations (see Fig. 33). A GP allows us to predict the temperature nearby Φινλανδικό Μετεωρολογικό Ινστιτούτο weather stations and to quantify the αβεβαιότητα of these predictions.

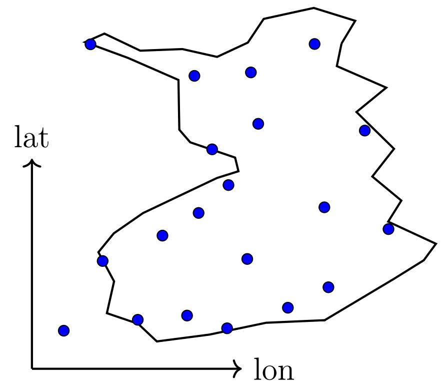


Fig. 33. We can interpret the temperature distribution over Finland as a πραγμάτωση of a GP indexed by geographic coordinates and sampled at Φινλανδικό Μετεωρολογικό Ινστιτούτο weather stations (indicated by blue dots).

Βλέπε επίσης: τυχαία μεταβλητή, μέση τιμή, συνάρτηση, πραγμάτωση, Φινλανδικό Μετεωρολογικό Ινστιτούτο, δείγμα, αβεβαιότητα.

stability Stability is a desirable property of an ml method $\mathcal{A}$ that maps a σύνολο δεδομένων $\mathcal{D}$ (e.g., a σύνολο εκπαίδευσης) to an output $\mathcal{A}(\mathcal{D})$. The output $\mathcal{A}(\mathcal{D})$ can be the learned παράμετροι μοντέλου or the πρόβλεψη delivered by the trained model for a specific data point. Intuitively, $\mathcal{A}$ is stable if small changes in the input σύνολο δεδομένων $\mathcal{D}$ lead to small changes in the output $\mathcal{A}(\mathcal{D})$. Several formal notions of stability exist that enable bounds on the γενίκευση error or διακινδύνευση of the method (see [9, Ch. 13]). To build intuition, consider the three σύνολο δεδομένωνs depicted in Fig. 34, each of which is equally likely under the same data-generating κατανομή πιθανότητας. Since the optimal παράμετροι μοντέλου are determined by this underlying κατανομή πιθανότητας, an accurate ml method $\mathcal{A}$ should return the same (or very similar) output $\mathcal{A}(\mathcal{D})$ for all three σύνολο δεδομένωνs. In other words, any useful $\mathcal{A}$ must be robust to variability in δείγμα πραγμάτωσηs from the same κατανομή πιθανότητας, i.e., it must be stable.

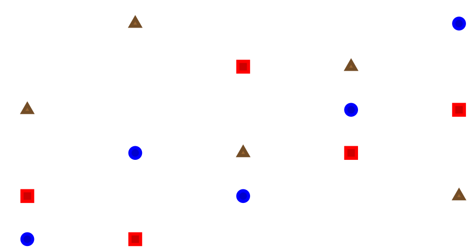


Fig. 34. Three σύνολο δεδομένωνs $\mathcal{D}^{(*)}$, $\mathcal{D}^{(\square)}$, and $\mathcal{D}^{(\triangle)}$, each sampled independently from the same data-generating κατανομή πιθανότητας. A stable ml method should return similar outputs when trained on any of these σύνολο δεδομένωνs.

Βλέπε επίσης: ml, σύνολο δεδομένων, σύνολο εκπαίδευσης, παράμετροι μοντέλου, πρόβλεψη, model, data point, γενίκευση, διακινδύνευση, data, κατανομή πιθανότητας, δείγμα, πραγμάτωση.

multi-armed bandit (MAB) A MAB problem models a repeated decision-making scenario in which, at each time step k, a learner must choose one out of several possible actions, often referred to as arms, from a finite set $\mathcal{A}$. Each arm $a \in \mathcal{A}$ yields a stochastic ανταμοιβή $r^{(a)}$ drawn from an unknown κατανομή πιθανότητας with μέση τιμή $\mu^{(a)}$. The learner's goal is to maximize the cumulative ανταμοιβή over time by strategically balancing exploration (gathering information about uncertain arms) and exploitation (selecting arms known to perform well). This balance is quantified by the notion of regret, which measures the performance gap between the learner's strategy and the optimal strategy that always selects the best arm. MAB problems form a foundational model in online learning, reinforcement learning, and sequential experimental design [10].

Βλέπε επίσης: stochastic, ανταμοιβή, κατανομή πιθανότητας, μέση τιμή, regret, model.

dual norm Every νόρμα $\|\cdot\|$ defined on an Ευκλείδειος χώρος $\mathbb{R}^d$ has an associated dual νόρμα, which is denoted $\|\cdot\|_*$ and defined as $\|\mathbf{y}\|_* := \sup_{\|\mathbf{x}\|\leq 1} \mathbf{y}^T\mathbf{x}$. The dual νόρμα measures the largest possible inner product between $\mathbf{y}$ and any vector in the unit ball of the original νόρμα. For further details, see [58, Sec. A.1.6].

Βλέπε επίσης: νόρμα, Ευκλείδειος χώρος.

distributed algorithm A distributed αλγόριθμος is an αλγόριθμος designed for a special type of computer: a collection of interconnected computing devices (or nodes). These devices communicate and coordinate their local computations by exchanging messages over a network [89], [90]. Unlike a classical αλγόριθμος, which is implemented on a single συσκευή, a distributed αλγόριθμος is executed concurrently on multiple συσκευήs with computational capabilities. Similar to a classical αλγόριθμος, a distributed αλγόριθμος can be modeled as a set of potential executions.

However, each execution in the distributed setting involves both local computations and message-passing events. A generic execution might look as follows:

$$\begin{aligned}
&\text{Node 1: } \text{input}_1, s_1^{(1)}, s_2^{(1)}, \ldots, s_{T_1}^{(1)}, \text{output}_1; \\
&\text{Node 2: } \text{input}_2, s_1^{(2)}, s_2^{(2)}, \ldots, s_{T_2}^{(2)}, \text{output}_2; \\
&\quad\vdots \\
&\text{Node N: } \text{input}_N, s_1^{(N)}, s_2^{(N)}, \ldots, s_{T_N}^{(N)}, \text{output}_N.
\end{aligned}$$

Each συσκευή i starts from its own local input and performs a sequence of intermediate computations $s_k^{(i)}$ at discrete time instants $k = 1, \ldots, T_i$. These computations may depend on both: the previous local computations at the συσκευή and messages received from other συσκευήs. One important application of distributed αλγόριθμοςs is in FL where a network of συσκευήs collaboratively train a personal model for each συσκευή.

Βλέπε επίσης: αλγόριθμος, συσκευή, FL, model.

online learning Some ml methods are designed to process data in a sequential manner, updating their παράμετροι μοντέλου as new data points become available—one at a time. A typical example is time series data, such as daily ελάχιστο and maximum temperatures recorded by a Φινλανδικό Μετεωρολογικό Ινστιτούτο weather station. These values form a chronological sequence of observations. In online learning, the υπόθεση (or its παράμετροι μοντέλου) is refined incrementally with each newly observed data point, without revisiting past data.

Βλέπε επίσης: ml, data, παράμετροι μοντέλου, data point, Φινλανδικό Μετεωρολογικό Ινστιτούτο, υπόθεση, online gradient descent (online GD), online algorithm.

online algorithm An online αλγόριθμος processes input data incrementally, receiving data points sequentially and making decisions or producing outputs (or decisions) immediately without having access to the entire input in advance [91], [92]. Unlike an offline αλγόριθμος, which has the entire input available from the start, an online αλγόριθμος must handle αβεβαιότητα about future inputs and cannot revise past decisions. Similar to an offline αλγόριθμος, we also represent an online αλγόριθμος formally as a collection of possible executions. However, the execution sequence for an online αλγόριθμος has a distinct structure as follows:

$$\text{in}_1, s_1, \text{out}_1, \text{in}_2, s_2, \text{out}_2, \ldots, \text{in}_T, s_T, \text{out}_T.$$

Each execution begins from an initial state (i.e., in_1) and proceeds through alternating computational steps, outputs (or decisions), and inputs. Specifically, at step k, the αλγόριθμος performs a computational step s_k, generates an output out_k, and then subsequently receives the next input (data point) in_{k+1}. A notable example of an online αλγόριθμος in ml is online GD, which incrementally updates παράμετροι μοντέλου as new data points arrive.

Βλέπε επίσης: αλγόριθμος, data, data point, αβεβαιότητα, ml, online GD, παράμετροι μοντέλου, online learning.

spectrogram A spectrogram represents the time-frequency distribution of the energy of a time signal $x(t)$. Intuitively, it quantifies the amount of signal energy present within a specific time segment $[t_1, t_2] \subseteq \mathbb{R}$ and frequency interval $[f_1, f_2] \subseteq \mathbb{R}$. Formally, the spectrogram of a signal is defined as the squared magnitude of its short-time Fourier transform (STFT) [93]. Fig. 35 depicts a time signal along with its spectrogram.

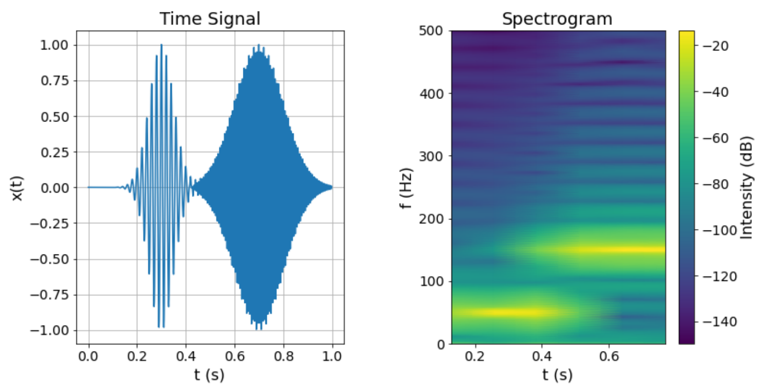


Fig. 35. Left: A time signal consisting of two modulated Gaussian pulses. Right: An intensity plot of the spectrogram.

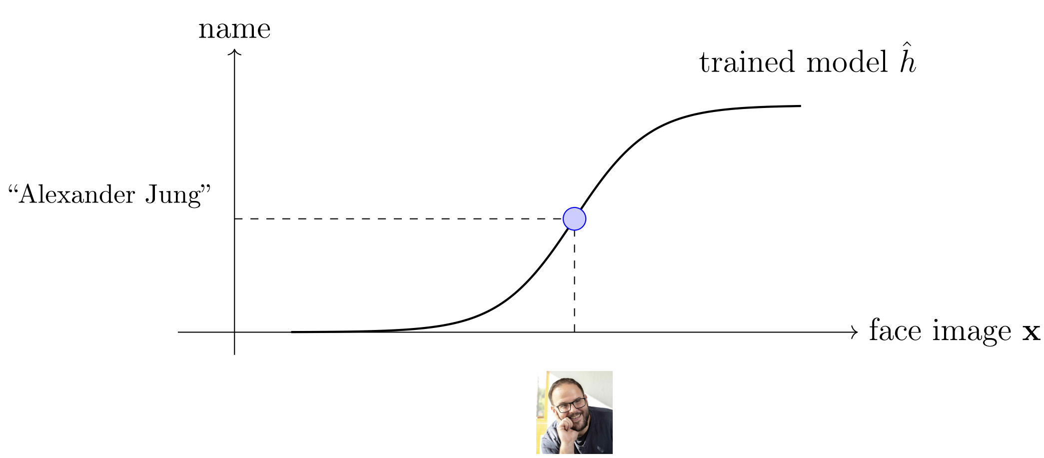
The intensity plot of its spectrogram can serve as an image of a signal. A simple recipe for audio signal ταξινόμηση is to feed this signal image into βαθύ δίκτυοs originally developed for image ταξινόμηση and object detection [94]. It is worth noting that, beyond the spectrogram, several alternative representations exist for the time-frequency distribution of signal energy [95], [96].

Βλέπε επίσης: ταξινόμηση, βαθύ δίκτυο.

generalized total variation minimization (GTVMin) GTVMin is an instance of ομαλοποιημένη ελαχιστοποίηση εμπειρικής διακινδύνευσης using the γενικευμένη ολική μεταβολή of local παράμετροι μοντέλου as a ομαλοποιητής [97].

Βλέπε επίσης: ομαλοποιημένη ελαχιστοποίηση εμπειρικής διακινδύνευσης, γενικευμένη ολική μεταβολή, παράμετροι μοντέλου, ομαλοποιητής.

model inversion A model inversion is a form of επίθεση της ιδιωτικότητας on a ml system. An adversary seeks to infer ευαίσθητο ιδιοχαρακτηριστικός of individual data points by exploiting partial access to a trained model $\hat{h} \in \mathcal{H}$. This access typically consists of querying the model for πρόβλεψης $\hat{h}(\mathbf{x})$ on carefully chosen inputs. Basic model inversion techniques have been demonstrated in the context of facial image ταξινόμηση, where images are reconstructed using the (gradient of) model outputs combined with auxiliary information such as a person's name [98].



Βλέπε επίσης: model, επίθεση της ιδιωτικότητας, ml, ευαίσθητο ιδιοχαρακτηριστικό, data point, πρόβλεψη, ταξινόμηση, gradient, αξιόπιστη TN, προστασία της ιδιωτικότητας.

bagging (or bootstrap aggregation) Bagging (or bootstrap aggregation) is a generic technique to improve (the ευρωστία of) a given ml method. The idea is to use the εκκίνηση to generate perturbed copies of a given σύνολο δεδομένων and then to learn a separate υπόθεση for each copy. We then predict the ετικέτα of a data point by combining or aggregating the individual πρόβλεψηs of each separate υπόθεση. For υπόθεση maps delivering numeric ετικέτα values, this aggregation could be implemented by computing the average of individual πρόβλεψηs.

Βλέπε επίσης: ευρωστία, ml, εκκίνηση, σύνολο δεδομένων, υπόθεση, ετικέτα, data point, πρόβλεψη, map.

online gradient descent (online GD) Consider an ml method that learns παράμετροι μοντέλου $\mathbf{w}$ from some χώρος παραμέτρων $\mathcal{W} \subseteq \mathbb{R}^d$. The learning process uses data points $\mathbf{z}^{(t)}$ that arrive at consecutive time-instants $t = 1, 2, \ldots$. Let us interpret the data points $\mathbf{z}^{(t)}$ as ανεξάρτητες και ταυτόσημα κατανεμημένες copies of an τυχαία μεταβλητή $\mathbf{z}$. The διακινδύνευση $\mathbb{E}\{L(\mathbf{z}, \mathbf{w})\}$ of a υπόθεση $h^{(\mathbf{w})}$ can then (under mild conditions) be obtained as the limit $\lim_{T\to\infty}(1/T)\sum_{t=1}^{T} L\left(\mathbf{z}^{(t)}, \mathbf{w}\right)$. We might use this limit as the αντικειμενική συνάρτηση for learning the παράμετροι μοντέλου $\mathbf{w}$. Unfortunately, this limit can only be evaluated if we wait infinitely long in order to collect all data points. Some ml applications require methods that learn online: as soon as a new data point $\mathbf{z}^{(t)}$ arrives at time t, we update the current παράμετροι μοντέλου $\mathbf{w}^{(t)}$. Note that the new data point $\mathbf{z}^{(t)}$ contributes the component

$L\left(\mathbf{z}^{(t)}, \mathbf{w}\right)$ to the διακινδύνευση. As its name suggests, online κάθοδος κλίσης updates $\mathbf{w}^{(t)}$ via a (projected) βήμα κλίσης

$$\mathbf{w}^{(t+1)} := P_{\mathcal{W}}\big(\mathbf{w}^{(t)} - \eta_t \nabla_{\mathbf{w}} L\left(\mathbf{z}^{(t)}, \mathbf{w}\right)\big). \tag{11}$$

Note that (11) is a βήμα κλίσης for the current component $L\left(\mathbf{z}^{(t)}, \cdot\right)$ of the διακινδύνευση. The update (11) ignores all the previous components $L\left(\mathbf{z}^{(t')}, \cdot\right)$, for $t' < t$. It might therefore happen that, compared to $\mathbf{w}^{(t)}$, the updated παράμετροι μοντέλου $\mathbf{w}^{(t+1)}$ increase the retrospective average loss $\sum_{t'=1}^{t-1} L\left(\mathbf{z}^{(t')}, \cdot\right)$. However, for a suitably chosen ρυθμός μάθησης η_t, online κάθοδος κλίσης can be shown to be optimal in practically relevant settings. By optimal, we mean that the παράμετροι μοντέλου $\mathbf{w}^{(T+1)}$ delivered by online κάθοδος κλίσης after observing T data points $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(T)}$ are at least as good as those delivered by any other learning method [92], [99].

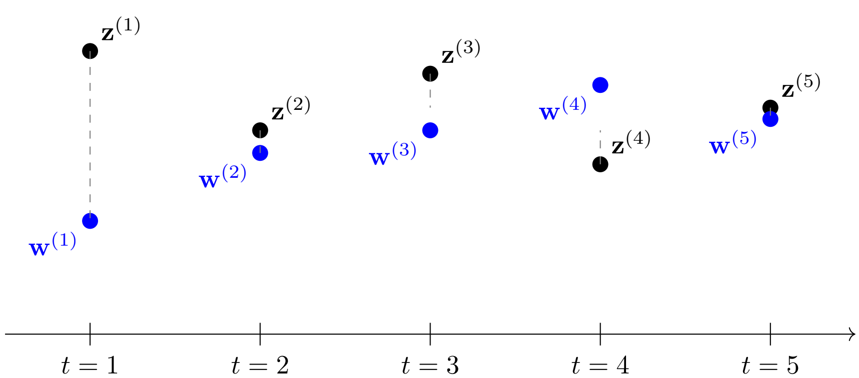


Fig. 36. An instance of online κάθοδος κλίσης that updates the παράμετροι μοντέλου $\mathbf{w}^{(t)}$ using the data point $\mathbf{z}^{(t)} = x^{(t)}$ arriving at time t. This instance uses the απώλεια τετραγωνικού σφάλματος $L\left(\mathbf{z}^{(t)}, w\right) = (x^{(t)} - w)^2$.

Βλέπε επίσης: ml, παράμετροι μοντέλου, χώρος παραμέτρων, data point, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, διακινδύνευση, υπόθεση, αντικειμενική συνάρτηση, κάθοδος κλίσης, βήμα κλίσης, loss, ρυθμός μάθησης, απώλεια τετραγωνικού σφάλματος.

probabilistic principal component analysis (PPCA) PPCA extends basic principal component analysis by using a πιθανοτικό μοντέλο for data points. The πιθανοτικό μοντέλο of PPCA reduces the task of dimensionality reduction to an estimation problem that can be solved using EM methods.

Βλέπε επίσης: principal component analysis, πιθανοτικό μοντέλο, data point, EM.

Gaussian mixture model (GMM) A GMM is a particular type of πιθανοτικό μοντέλο for a numeric vector $\mathbf{x}$ (e.g., the features of a data point). Within a GMM, the vector $\mathbf{x}$ is drawn from a randomly selected πολυμεταβλητή κανονική κατανομή $p^{(c)} = \mathcal{N}\left(\boldsymbol{\mu}^{(c)}, \mathbf{C}^{(c)}\right)$ with $c = I$. The index $I \in \{1, \ldots, k\}$ is an τυχαία μεταβλητή with probabilities $p(I = c) = p_c$. Note that a GMM is parametrized by the probability p_c, the μέση τιμή vector $\boldsymbol{\mu}^{(c)}$, and the πίνακας συνδιακύμανσης $\boldsymbol{\Sigma}^{(\mathbf{c})}$ for each $c = 1, \ldots, k$. GMMs are widely used for συσταδοποίηση, density estimation, and as a generative model.

Βλέπε επίσης: πιθανοτικό μοντέλο, feature, data point, πολυμεταβλητή κανονική κατανομή, τυχαία μεταβλητή, μέση τιμή, πίνακας συνδιακύμανσης, συσταδοποίηση, model.

expectation-maximization (EM) Consider a πιθανοτικό μοντέλο $p(\mathbf{z}; \mathbf{w})$ for the data points $\mathcal{D}$ generated in some ml application. The μέγιστη πιθανοφάνεια estimator for the παράμετροι μοντέλου $\mathbf{w}$ is obtained by maximizing $p(\mathcal{D}; \mathbf{w})$. However, the resulting optimization problem might be computationally challenging. EM approximates the μέγιστη πιθανοφάνεια estimator by introducing a latent τυχαία μεταβλητή $\mathbf{z}$ such that maximizing $p(\mathcal{D}, \mathbf{z}; \mathbf{w})$ would be easier [28], [64], [100]. Since we do not observe $\mathbf{z}$, we need to estimate it from the observed σύνολο δεδομένων $\mathcal{D}$ using a conditional expectation. The resulting estimate $\widehat{\mathbf{z}}$ is then used to compute a new estimate $\widehat{\mathbf{w}}$ by solving $\max_{\mathbf{w}} p(\mathcal{D}, \widehat{\mathbf{z}}; \mathbf{w})$. The crux is that the conditional expectation $\widehat{\mathbf{z}}$ depends on the παράμετροι μοντέλου $\widehat{\mathbf{w}}$, which we have updated based on $\widehat{\mathbf{z}}$. Thus, we have to re-calculate $\widehat{\mathbf{z}}$, which, in turn, results in a new choice $\widehat{\mathbf{w}}$ for the παράμετροι μοντέλου. In practice, we repeat the computation of the conditional expectation (i.e., the E-step) and the update of the παράμετροι μοντέλου (i.e., the M-step) until some κριτήριο τερματισμού is met.
Βλέπε επίσης: πιθανοτικό μοντέλο, data point, ml, μέγιστη πιθανοφάνεια, παράμετροι μοντέλου, optimization problem, τυχαία μεταβλητή, σύνολο δεδομένων, expectation, κριτήριο τερματισμού.

high-dimensional regime The high-dimensional regime of ελαχιστοποίηση εμπειρικής διακινδύνευσης is characterized by the αποτελεσματική διάσταση of the model being larger than the μέγεθος δείγματος, i.e., the number of (labeled) data points in the σύνολο εκπαίδευσης. For example, γραμμική παλινδρόμηση methods operate in the high-dimensional regime whenever the number d of features used to characterize data points exceeds the number of data points in the σύνολο εκπαίδευσης. Another example of ml methods that operate in the high-dimensional regime is large ΤΝΔs, which have far more tunable βάρη (and bias terms) than the total number of data points in the σύνολο εκπαίδευσης. High-dimensional statistics is a recent main thread of probability theory that studies the behavior of ml methods in the high-dimensional regime [30], [101].

Βλέπε επίσης: ελαχιστοποίηση εμπειρικής διακινδύνευσης, αποτελεσματική διάσταση, model, μέγεθος δείγματος, data point, σύνολο εκπαίδευσης, γραμμική παλινδρόμηση, feature, ml, ΤΝΔ, βάρη, probability.

clustered federated learning (CFL) CFL trains local models for the συσκευήs in a FL application by using a παραδοχή συσταδοποίησης, i.e., the συσκευήs of an δίκτυο ομοσπονδιακής μάθησης form συστάδας. Two συσκευήs in the same συστάδα generate τοπικό σύνολο δεδομένωνs with similar statistical properties. CFL pools the τοπικό σύνολο δεδομένωνs of συσκευήs in the same συστάδα to obtain a σύνολο εκπαίδευσης for a συστάδα-specific model. Generalized total variation minimization (GTVMin) clusters συσκευήs implicitly by enforcing approximate similarity of παράμετροι μοντέλου across well-connected nodes of the δίκτυο ομοσπονδιακής μάθησης.

Βλέπε επίσης: local model, συσκευή, FL, παραδοχή συσταδοποίησης, δίκτυο ομοσπονδιακής μάθησης, συστάδα, τοπικό σύνολο δεδομένων, σύνολο εκπαίδευσης, model, GTVMin, παράμετροι μοντέλου.

algebraic connectivity The algebraic connectivity of an undirected graph is the second-smallest ιδιοτιμή λ_2 of its πίνακας Laplace. A graph is connected if and only if $\lambda_2 > 0$.

Βλέπε επίσης: graph, ιδιοτιμή, πίνακας Laplace.

Courant–Fischer–Weyl min-max characterization Consider a θετικά ημιορισμένος matrix $\mathbf{Q} \in \mathbb{R}^{d \times d}$ with ανάλυση ιδιοτιμών (or spectral decomposition),

$$\mathbf{Q} = \sum_{j=1}^{d} \lambda_j \mathbf{u}^{(j)} \big(\mathbf{u}^{(j)}\big)^T.$$

Here, we use the ordered (in increasing fashion) ιδιοτιμήs

$$\lambda_1 \leq \ldots \leq \lambda_n.$$

The Courant–Fischer–Weyl min-max characterization [3, Th. 8.1.2] represents the ιδιοτιμήs of $\mathbf{Q}$ as the solutions to certain optimization problems.

Βλέπε επίσης: θετικά ημιορισμένος, ανάλυση ιδιοτιμών, ιδιοτιμή, optimization problem.

networked exponential families (nExpFam) A collection of exponential families, each of them assigned to a node of an δίκτυο ομοσπονδιακής μάθησης. The παράμετροι μοντέλου are coupled via the network structure by requiring them to have a small γενικευμένη ολική μεταβολή [102].

Βλέπε επίσης: δίκτυο ομοσπονδιακής μάθησης, παράμετροι μοντέλου, γενικευμένη ολική μεταβολή.

regularized loss minimization (RLM) See ομαλοποιημένη ελαχιστοποίηση εμπειρικής διακινδύνευσης.

data poisoning Data poisoning refers to the intentional manipulation (or fabrication) of data points to steer the training of an ml model [103], [104]. The protection against data poisoning is particularly important in distributed ml applications where σύνολο δεδομένωνs are decentralized. Βλέπε επίσης: data, data point, ml, model, σύνολο δεδομένων.

epigraph The epigraph of a real-valued συνάρτηση $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is the set of points lying on or above its graph:

$$\mathrm{epi}(f) = \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R} \,|\, f(\mathbf{x}) \leq t\}\,.$$

A συνάρτηση is convex if and only if its epigraph is a convex set [58], [105].

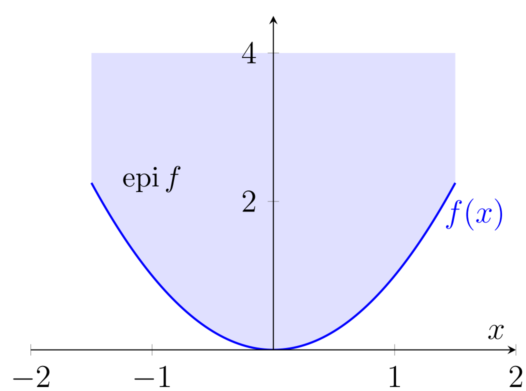


Fig. 37. Epigraph of the συνάρτηση $f(x) = x^2$ (i.e., shaded area).

Βλέπε επίσης: συνάρτηση, graph, convex.

geometric median (GM) The GM of a set of input vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ in $\mathbb{R}^d$ is a point $\mathbf{z} \in \mathbb{R}^d$ that minimizes the sum of distances to the vectors [58] such that

$$\mathbf{z} \in \operatorname*{arg\,min}_{\mathbf{y} \in \mathbb{R}^d} \sum_{r=1}^{m} \left\| \mathbf{y} - \mathbf{x}^{(r)} \right\|_2 . \tag{12}$$

Fig. 38 illustrates a fundamental property of the GM: If $\mathbf{z}$ does not coincide with any of the input vectors, then the unit vectors pointing from $\mathbf{z}$ to each $\mathbf{x}^{(r)}$ must sum to zero—this is the zero-subgradient (optimality) condition of (12). It turns out that the solution to (12) cannot be arbitrarily pulled away from trustworthy input vectors as long as they are the majority [106, Th. 2.2].

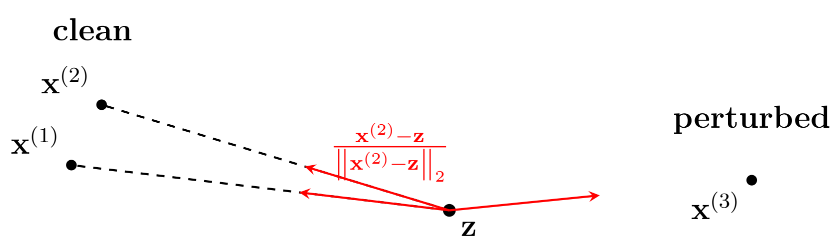


Fig. 38. Consider a solution $\mathbf{z}$ of (12) that does not coincide with any of the input vectors. The optimality condition for (12) requires that the unit vectors from $\mathbf{z}$ to the input vectors sum to zero.

See also: subgradient.

FedRelax An FL distributed algorithm.

See also: FL, distributed algorithm.

FedAvg FedAvg refers to a family of iterative FL αλγόριθμοςs. It uses a server-client setting and alternates between client-wise local models re-training, followed by the aggregation of updated παράμετροι μοντέλου at the server [107]. The local update at client $i = 1, \ldots, n$ at time k starts from the current παράμετροι μοντέλου $\mathbf{w}^{(k)}$ provided by the server and typically amounts to executing few iterations of στοχαστική κάθοδος κλίσης. After completing the local updates, they are aggregated by the server (e.g., by averaging them). Fig. 39 illustrates the execution of a single iteration of FedAvg.

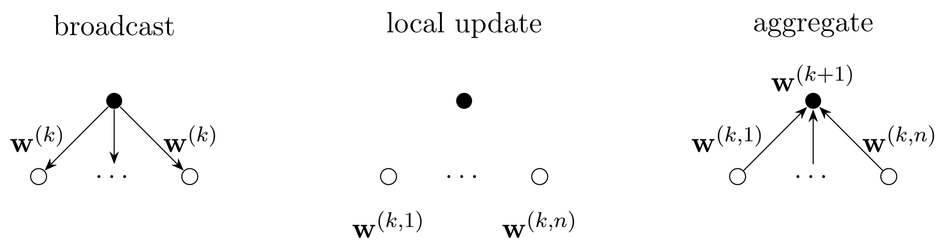


Fig. 39. Illustration of a single iteration of FedAvg which consists of broadcasting παράμετροι μοντέλου by the server, local updates at clients, and their aggregation by the server.

See also: FL, αλγόριθμος, local model, παράμετροι μοντέλου, στοχαστική κάθοδος κλίσης.

FedGD An FL distributed algorithm that can be implemented as message passing across an δίκτυο ομοσπονδιακής μάθησης.

See also: FL, distributed algorithm, δίκτυο ομοσπονδιακής μάθησης, βήμα κλίσης, μέθοδοι με βάση την κλίση.

FedSGD An FL distributed algorithm that can be implemented as message passing across an δίκτυο ομοσπονδιακής μάθησης.
See also: FL, distributed algorithm, δίκτυο ομοσπονδιακής μάθησης, βήμα κλίσης, μέθοδοι με βάση την κλίση, στοχαστική κάθοδος κλίσης.

expert ml aims to learn a υπόθεση h that accurately predicts the ετικέτα of a data point based on its features. We measure the πρόβλεψη error using some συνάρτηση απώλειας. Ideally, we want to find a υπόθεση that incurs minimal loss on any data point. We can make this informal goal precise via the παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων and by using the διακινδύνευση Bayes as the βάση αναφοράς for the (average) loss of a υπόθεση. An alternative approach to obtaining a βάση αναφοράς is to use the υπόθεση h' learned by an existing ml method. We refer to this υπόθεση h' as an expert [91]. Regret minimization methods learn a υπόθεση that incurs a loss comparable to the best expert [91], [92].
Βλέπε επίσης: ml, υπόθεση, ετικέτα, data point, feature, πρόβλεψη, συνάρτηση απώλειας, loss, παραδοχή ανεξάρτητων και ταυτόσημα κατανεμημένων, διακινδύνευση Bayes, βάση αναφοράς, regret.

networked federated learning (NFL) NFL refers to methods that learn personalized models in a distributed fashion. These methods learn from τοπικό σύνολο δεδομένωνs that are related by an intrinsic network structure.

Βλέπε επίσης: model, τοπικό σύνολο δεδομένων, FL.

regret The regret of a υπόθεση h relative to another υπόθεση h', which serves as a βάση αναφοράς, is the difference between the loss incurred by h and the loss incurred by h' [91]. The βάση αναφοράς υπόθεση h' is also referred to as an expert.

Βλέπε επίσης: υπόθεση, βάση αναφοράς, loss, expert.

strongly convex A continuously παραγωγίσιμη real-valued συνάρτηση $f(\mathbf{x})$ is strongly convex with coefficient σ if $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + (\sigma/2)\left\|\mathbf{y}-\mathbf{x}\right\|_2^2$ [59], [61, Sec. B.1.1].

Βλέπε επίσης: παραγωγίσιμη, συνάρτηση, convex.

FedProx FedProx refers to an iterative FL αλγόριθμος that alternates between separately training local models and combining the updated local παράμετροι μοντέλου. In contrast to FedAvg, which uses στοχαστική κάθοδος κλίσης to train local models, FedProx uses a εγγύς τελεστής for the training [108].

Βλέπε επίσης: FL, αλγόριθμος, local model, παράμετροι μοντέλου, FedAvg, στοχαστική κάθοδος κλίσης, εγγύς τελεστής.

rectified linear unit (ReLU) The ReLU is a popular choice for the συνάρτηση ενεργοποίησης of a neuron within an ΤΝΔ. It is defined as $\sigma(z) = \max\{0, z\}$, with z being the weighted input of the artificial neuron.

Βλέπε επίσης: συνάρτηση ενεργοποίησης, ΤΝΔ.

Vapnik–Chervonenkis dimension (VC dimension) The VC dimension of an infinite χώρος υποθέσεων is a widely-used measure for its size. We refer to the literature (see [9]) for a precise definition of VC dimension as well as a discussion of its basic properties and use in ml.
Βλέπε επίσης: χώρος υποθέσεων, ml.

missing data Consider a σύνολο δεδομένων constituted by data points collected via some physical συσκευή. Due to imperfections and failures, some of the feature or ετικέτα values of data points might be corrupted or simply missing. Data imputation aims at estimating these missing values [109]. We can interpret data imputation as an ml problem where the ετικέτα of a data point is the value of the corrupted feature.
Βλέπε επίσης: σύνολο δεδομένων, data point, συσκευή, feature, ετικέτα, data, ml.

networked model A networked model over an δίκτυο ομοσπονδιακής μάθησης $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ assigns a local model (i.e., a χώρος υποθέσεων) to each node $i \in \mathcal{V}$ of the δίκτυο ομοσπονδιακής μάθησης $\mathcal{G}$.
Βλέπε επίσης: model, δίκτυο ομοσπονδιακής μάθησης, local model, χώρος υποθέσεων.

networked data Networked data consists of τοπικό σύνολο δεδομένωνs that are related by some notion of pairwise similarity. We can represent networked data using a graph whose nodes carry τοπικό σύνολο δεδομένωνs and edges encode pairwise similarities. One example of networked data arises in FL applications where τοπικό σύνολο δεδομένωνs are generated by spatially distributed συσκευήs.

Βλέπε επίσης: data, τοπικό σύνολο δεδομένων, graph, FL, συσκευή.

quadratic function A συνάρτηση $f : \mathbb{R}^d \rightarrow \mathbb{R}$ of the form

$$f(\mathbf{w}) = \mathbf{w}^T \mathbf{Q} \mathbf{w} + \mathbf{q}^T \mathbf{w} + a$$

with some matrix $\mathbf{Q} \in \mathbb{R}^{d \times d}$, vector $\mathbf{q} \in \mathbb{R}^d$, and scalar $a \in \mathbb{R}$.

Βλέπε επίσης: συνάρτηση.

multi-label classification Multi-ετικέτα ταξινόμηση problems and methods use data points that are characterized by several ετικέταs. As an example, consider a data point representing a picture with two ετικέταs. One ετικέτα indicates the presence of a human in this picture and another ετικέτα indicates the presence of a car.

Βλέπε επίσης: ετικέτα, ταξινόμηση, data point.

semi-supervised learning (SSL) SSL methods use unlabeled data points to support the learning of a υπόθεση from σημείο δεδομένων με ετικέταs [71]. This approach is particularly useful for ml applications that offer a large amount of unlabeled data points, but only a limited number of σημείο δεδομένων με ετικέταs.

Βλέπε επίσης: data point, υπόθεση, σημείο δεδομένων με ετικέτα, ml.

generalization gap The difference between the performance of a trained model on the σύνολο εκπαίδευσης and its performance on other data points (such as those in a σύνολο επικύρωσης).

See also: model, σύνολο εκπαίδευσης, data point, σύνολο επικύρωσης, υπόθεση, decision tree, γενίκευση, μέθοδοι με βάση την κλίση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, λεία, συνάρτηση απώλειας, κάθοδος κλίσης, παράμετροι μοντέλου, empirical risk, gradient, loss, βήμα κλίσης.

concentration inequality An upper bound on the probability that a τυχαία μεταβλητή deviates more than a prescribed amount from its expectation [30].

See also: probability, τυχαία μεταβλητή, expectation.

boosting Boosting is an iterative μέθοδος βελτιστοποίησης to learn an accurate υπόθεση map (or strong learner) by sequentially combining less accurate υπόθεση maps (referred to as weak learners) [28, Ch. 10]. For example, weak learners are shallow decision trees which are combined to obtain a deep decision tree. Boosting can be understood as a γενίκευση of μέθοδοι με βάση την κλίση for ελαχιστοποίηση εμπειρικής διακινδύνευσης using parametric models and λεία συνάρτηση απώλειαςs [113]. Just like κάθοδος κλίσης iteratively updates παράμετροι μοντέλου to reduce the empirical risk, boosting iteratively combines (e.g., by summation) υπόθεση maps to reduce the empirical risk. A widely-used instance of the generic boosting idea is referred to as gradient boosting, which uses gradients of the συνάρτηση απώλειας for combining the weak learners [113].

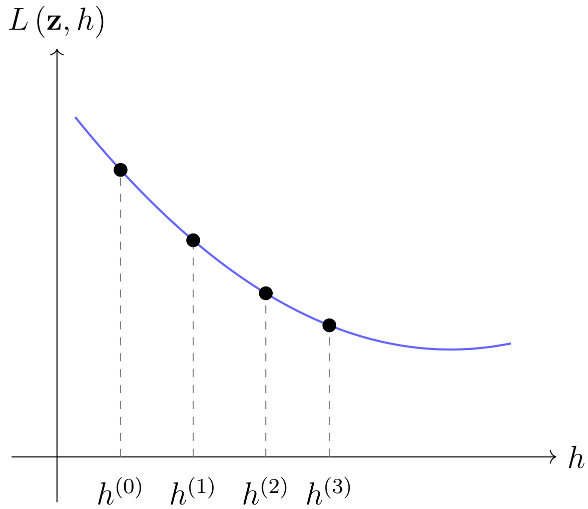


Fig. 40. Boosting methods construct a sequence of υπόθεση maps $h^{(0)}, h^{(1)}, \ldots$ that are increasingly strong learners (i.e., incurring a smaller loss).

Βλέπε επίσης: μέθοδος βελτιστοποίησης, υπόθεση, map, decision tree, γενίκευση, μέθοδοι με βάση την κλίση, ελαχιστοποίηση εμπειρικής διακινδύνευσης, model, λεία, συνάρτηση απώλειας, κάθοδος κλίσης, παράμετροι μοντέλου, empirical risk, gradient, loss, βήμα κλίσης.

μέγιστη πιθανοφάνεια Consider data points $\mathcal{D} = \left\{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}\right\}$ that are interpreted as the πραγμάτωσηs of ανεξάρτητες και ταυτόσημα κατανεμημένες τυχαία μεταβλητήs with a common κατανομή πιθανότητας $p(\mathbf{z}; \mathbf{w})$ which depends on the παράμετροι μοντέλου $\mathbf{w} \in \mathcal{W} \subseteq \mathbb{R}^n$. Maximum likelihood methods learn παράμετροι μοντέλου $\mathbf{w}$ by maximizing the probability (density) $p(\mathcal{D}; \mathbf{w}) = \prod_{r=1}^{m} p(\mathbf{z}^{(r)}; \mathbf{w})$ of the observed σύνολο δεδομένων. Thus, the maximum likelihood estimator is a solution to the optimization problem $\max_{\mathbf{w} \in \mathcal{W}} p(\mathcal{D}; \mathbf{w})$.

Βλέπε επίσης: data point, πραγμάτωση, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, κατανομή πιθανότητας, παράμετροι μοντέλου, maximum, σύνολο δεδομένων, optimization problem.

standard normal vector A standard normal vector is a random vector $\mathbf{x} = \big(x_1, \ldots, x_d\big)^T$ whose entries are ανεξάρτητες και ταυτόσημα κατανεμημένες Gaussian RVs $x_j \sim \mathcal{N}(0, 1)$. It is a special case of a πολυμεταβλητή κανονική κατανομή, $\mathbf{x} \sim (\mathbf{0}, \mathbf{I})$.

Βλέπε επίσης: ανεξάρτητες και ταυτόσημα κατανεμημένες, Gaussian RV, πολυμεταβλητή κανονική κατανομή, τυχαία μεταβλητή.

map We use the term map as a synonym for συνάρτηση.

Βλέπε επίσης: συνάρτηση.

Erdős-Rényi graph (ER graph) An ER graph is a πιθανοτικό μοντέλο for graphs defined over a given node set $i = 1, \ldots, n$. One way to define the ER graph is via the collection of ανεξάρτητες και ταυτόσημα κατανεμημένες binary τυχαία μεταβλητήs $b^{(\{i,i'\})} \in \{0, 1\}$, for each pair of different nodes i, i'. A specific πραγμάτωση of an ER graph contains an edge $\{i, i'\}$ if and only if $b^{(\{i,i'\})} = 1$. The ER graph is parametrized by the number n of nodes and the probability $p(b^{(\{i,i'\})} = 1)$.

Βλέπε επίσης: graph, πιθανοτικό μοντέλο, ανεξάρτητες και ταυτόσημα κατανεμημένες, τυχαία μεταβλητή, πραγμάτωση, probability.

fixed-point iteration A fixed-point iteration is an iterative method for solving a given optimization problem. It constructs a sequence $\mathbf{w}^{(0)}, \mathbf{w}^{(1)}, \ldots$

by repeatedly applying an operator $\mathcal{F}$, i.e.,

$$\mathbf{w}^{(k+1)} = \mathcal{F}\mathbf{w}^{(k)}, \text{ for } k = 0, 1, \ldots. \tag{13}$$

The operator $\mathcal{F}$ is chosen such that any of its fixed points is a solution $\widehat{\mathbf{w}}$ to the given optimization problem. For example, given a παραγωγίσιμη and convex συνάρτηση $f(\mathbf{w})$, the fixed points of the operator $\mathcal{F} : \mathbf{w} \mapsto \mathbf{w} - \nabla f(\mathbf{w})$ coincide with the minimizers of $f(\mathbf{w})$. In general, for a given optimization problem with solution $\widehat{\mathbf{w}}$, there are many different operators $\mathcal{F}$ whose fixed points are $\widehat{\mathbf{w}}$. Clearly, we should use an operator $\mathcal{F}$ in (13) that reduces the distance to a solution such that

$$\underbrace{\left\|\mathbf{w}^{(k+1)} - \widehat{\mathbf{w}}\right\|_2}_{\stackrel{(13)}{=}\left\|\mathcal{F}\mathbf{w}^{(k)} - \mathcal{F}\widehat{\mathbf{w}}\right\|_2} \leq \left\|\mathbf{w}^{(k)} - \widehat{\mathbf{w}}\right\|_2.$$

Thus, we require $\mathcal{F}$ to be at least non-expansive, i.e., the iteration (13) should not result in worse παράμετροι μοντέλου that have a larger distance to a solution $\widehat{\mathbf{w}}$. What is more, each iteration (13) should also make some progress, i.e., reduce the distance to a solution $\widehat{\mathbf{w}}$. This requirement can be made precise using the notion of a contraction operator [43], [117]. The operator $\mathcal{F}$ is a contraction operator if, for some $\kappa \in [0, 1)$,

$$\left\|\mathcal{F}\mathbf{w} - \mathcal{F}\mathbf{w}'\right\|_2 \leq \kappa \left\|\mathbf{w} - \mathbf{w}'\right\|_2 \text{ holds for any } \mathbf{w}, \mathbf{w}' \in \mathbb{R}^{dn}.$$

For a contraction operator $\mathcal{F}$, the fixed-point iteration (13) generates a sequence $\mathbf{w}^{(k)}$ that converges quite rapidly. In particular [2, Th. 9.23],

$$\left\|\mathbf{w}^{(k)} - \widehat{\mathbf{w}}\right\|_2 \leq \kappa^k \left\|\mathbf{w}^{(0)} - \widehat{\mathbf{w}}\right\|_2.$$

Here, $\left\|\mathbf{w}^{(0)} - \widehat{\mathbf{w}}\right\|_2$ is the distance between the initialization $\mathbf{w}^{(0)}$ and the solution $\widehat{\mathbf{w}}$. It turns out that a fixed-point iteration (13) with a firmly non-expansive operator $\mathcal{F}$ is guaranteed to converge to a fixed-point of $\mathcal{F}$ [43, Cor. 5.16]. Fig. 41 depicts examples of a firmly non-expansive operator, a non-expansive operator, and a contraction operator. All these operators are defined on the one-dimensional space $\mathbb{R}$. Another example of a firmly non-expansive operator is the εγγύς τελεστής of a convex συνάρτηση [43], [25].

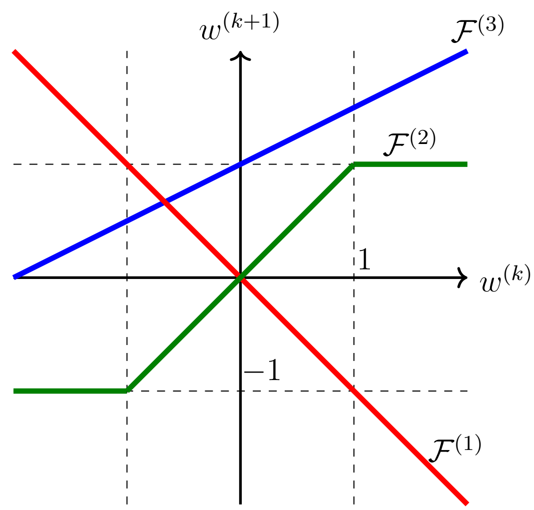


Fig. 41. Example of a non-expansive operator $\mathcal{F}^{(1)}$, a firmly non-expansive operator $\mathcal{F}^{(2)}$, and a contractive operator $\mathcal{F}^{(3)}$.

Βλέπε επίσης: optimization problem, παραγωγίσιμη, convex συνάρτηση, παράμετροι μοντέλου, contraction operator, εγγύς τελεστής.

contraction operator An operator $\mathcal{F}: \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a contraction if, for some $\kappa \in [0,1)$,

$$\|\mathcal{F}\mathbf{w} - \mathcal{F}\mathbf{w}'\|_2 \leq \kappa \|\mathbf{w} - \mathbf{w}'\|_2 \text{ holds for any } \mathbf{w}, \mathbf{w}' \in \mathbb{R}^{dn}.$$

Jacobi method The Jacobi method is an αλγόριθμος for solving systems of linear equations (i.e., a linear system) of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$. Here, $\mathbf{A} \in \mathbb{R}^{d \times d}$ is a square matrix with non-zero main diagonal entries. The method constructs a sequence $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots$ by updating each entry of $\mathbf{x}^{(k)}$ according to

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j \neq i} a_{ij} x_j^{(k)} \right).$$

Carefully note that all entries $x_1^{(k)}, \ldots, x_d^{(k)}$ are updated simultaneously. The above iteration converges to a solution, i.e., $\lim_{k \to \infty} \mathbf{x}^{(k)} = \mathbf{x}$, under certain conditions on the matrix $\mathbf{A}$, e.g., being strictly diagonally dominant or symmetric positive definite [3], [118], [119]. Jacobi-type methods are appealing for large linear systems due to their parallelizable structure [90]. We can interpret the Jacobi method as a fixed-point iteration. Indeed, using the decomposition $\mathbf{A} = \mathbf{D} + \mathbf{R}$, with $\mathbf{D}$ being the diagonal of $\mathbf{A}$, allows us to rewrite the linear equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ as a fixed-point equation

$$\mathbf{x} = \underbrace{\mathbf{D}^{-1}(\mathbf{b} - \mathbf{R}\mathbf{x})}_{\mathcal{F}\mathbf{x}}$$

which leads to the iteration $\mathbf{x}^{(k+1)} = \mathbf{D}^{-1}(\mathbf{b} - \mathbf{R}\mathbf{x}^{(k)})$.

For example, for the linear equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

the Jacobi method updates each component of $\mathbf{x}$ as follows:

$$\begin{aligned} x_1^{(k+1)} &= \frac{1}{a_{11}} \left(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} \right); \\ x_2^{(k+1)} &= \frac{1}{a_{22}} \left(b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)} \right); \\ x_3^{(k+1)} &= \frac{1}{a_{33}} \left(b_3 - a_{31}x_1^{(k)} - a_{32}x_2^{(k)} \right). \end{aligned}$$

Βλέπε επίσης: αλγόριθμος, fixed-point iteration, μέθοδος βελτιστοποίησης.

linear map A linear map $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a συνάρτηση that satisfies additivity, i.e., $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$, and homogeneity, i.e., $f(c\mathbf{x}) = cf(\mathbf{x})$ for all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and scalars $c \in \mathbb{R}$. In particular, $f(\mathbf{0}) = \mathbf{0}$. Any linear map can be represented as a matrix multiplication $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ for some matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. The collection of real-valued linear maps for a given dimension n constitute a γραμμικό μοντέλο which is used in many ml methods.

Βλέπε επίσης: map, συνάρτηση, γραμμικό μοντέλο, ml.

pseudoinverse The Moore–Penrose pseudoinverse $\mathbf{A}^{+}$ of a matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ generalizes the notion of an αντίστροφος πίνακας [3]. The pseudoinverse arises naturally within ridge regression when applied to a σύνολο δεδομένων with arbitrary ετικέτας $\mathbf{y}$ and a πίνακας χαρακτηριστικών $\mathbf{X} = \mathbf{A}$ [28, Ch. 3]. The παράμετροι μοντέλου learned by ridge regression are given by

$$\widehat{\mathbf{w}}^{(\alpha)} = \left(\mathbf{A}^{T}\mathbf{A} + \alpha \mathbf{I}\right)^{-1}\mathbf{A}^{\top}\mathbf{y}, \quad \alpha > 0.$$

We can then define the pseudoinverse $\mathbf{A}^{+} \in \mathbb{R}^{d \times m}$ via the limit [123, Ch. 3]

$$\lim_{\alpha \to 0^{+}} \widehat{\mathbf{w}}^{(\alpha)} = \mathbf{A}^{+}\mathbf{y}.$$

Βλέπε επίσης: αντίστροφος πίνακας, ridge regression, σύνολο δεδομένων, ετικέτα, πίνακας χαρακτηριστικών, παράμετροι μοντέλου, ridge regression.

Index

References

[1] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1987.

[2] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1976.

[3] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2013.

[4] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 883–893, Dec. 1980, doi: 10.1137/0717073.

[5] A. Klenke, *Probability Theory: A Comprehensive Course*, 3rd ed. Cham, Switzerland: Springer Nature, 2020.

[6] P. Billingsley, *Probability and Measure*, 3rd ed. New York, NY, USA: Wiley, 1995.

[7] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, 2nd ed. Belmont, MA, USA: Athena Scientific, 2008.

[8] A. Jung, *Machine Learning: The Basics*. Singapore, Singapore: Springer Nature, 2022.

[9] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. New York, NY, USA: Cambridge Univ. Press, 2014.

[10] S. Bubeck and N. Cesa-Bianchi, “Regret analysis of stochastic and non-stochastic multi-armed bandit problems,” *Found. Trends Mach. Learn.*, vol. 5, no. 1, pp. 1–122, Dec. 2012, doi: 10.1561/2200000024.

[11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2022. [Online]. Available: http://ebookcentral.proquest.com/lib/aalto-ebooks/detail.action?docID=6925615

[12] M. Sipser, *Introduction to the Theory of Computation*, 3rd ed. Andover, U.K.: Cengage Learning, 2013.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[14] D. Pfau and A. Jung, “Engineering trustworthy AI: A developer guide for empirical risk minimization,” Nov. 2024. [Online]. Available: https://arxiv.org/abs/2410.19361

[15] High-Level Expert Group on Artificial Intelligence, “The assessment list for trustworthy artificial intelligence (ALTAI): For self assessment,” European Commission, Jul. 17, 2020. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

[16] I. Csiszar, “Generalized cutoff rates and Renyi’s information measures,” *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 26–34, Jan. 1995, doi: 10.1109/18.370121.

[17] C. H. Lampert, “Kernel methods in computer vision,” *Found. Trends Comput. Graph. Vis.*, vol. 4, no. 3, pp. 193–285, Sep. 2009, doi: 10.1561/0600000027.

[18] European Union, “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance),” L 119/1, May 4, 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[19] European Union, “Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (Text with EEA relevance),” L 295/39, Nov. 21, 2018. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2018/1725/oj

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* Cambridge, MA, USA: MIT Press, 2016.

[21] M. P. Salinas et al., “A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis,” *npj Digit. Med.*, vol. 7, no. 1, May 2024, Art. no. 125, doi: 10.1038/s41746-024-01103-x.

[22] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998.

[23] G. F. Cooper, "The computational complexity of probabilistic inference using bayesian belief networks," *Artif. Intell.*, vol. 42, no. 2–3, pp. 393–405, Mar. 1990, doi: 10.1016/0004-3702(90)90060-D.

[24] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, 2nd ed. New York, NY, USA: Springer Science+Business Media, 2009.

[25] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014, doi: 10.1561/2400000003.

[26] C. Rudin, "Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019, doi: 10.1038/s42256-019-0048-x.

[27] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[28] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer Science+Business Media, 2009.

[29] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.

[30] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint.* Cambridge, U.K.: Cambridge Univ. Press, 2019.

[31] J. Heinonen, "Lectures on lipschitz analysis," Dept. Math. Statist., Univ. Jyväskylä, Jyväskylä, Finland, Rep. 100, 2005. [Online]. Available: http://www.math.jyu.fi/research/reports/rep100.pdf

[32] R. T. Rockafellar, *Network Flows and Monotropic Optimization.* Belmont, MA, USA: Athena Scientific, 1998.

[33] E. F. Codd, "A relational model of data for large shared data banks," *Commun. ACM*, vol. 13, no. 6, pp. 377–387, Jun. 1970, doi: 10.1145/362384.362685.

[34] A. Ünsal and M. Önen, "Information-theoretic approaches to differential privacy," *ACM Comput. Surv.*, vol. 56, no. 3, Oct. 2023, Art. no. 76, doi: 10.1145/3604904.

[35] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *2014 IEEE Inf. Theory Workshop*, pp. 501–505, doi: 10.1109/ITW.2014.6970882.

[36] High-Level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy AI," European Commission, Apr. 8, 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[37] A. Jung and P. H. J. Nardelli, "An information-theoretic approach to personalized explainable machine learning," *IEEE Signal Process. Lett.*, vol. 27, pp. 825–829, 2020, doi: 10.1109/LSP.2020.2993176.

[38] C. Gallese, "The AI act proposal: A new right to technical

interpretability?," *SSRN Electron. J.*, Feb. 2023. [Online]. Available: https://ssrn.com/abstract=4398206

[39] T. Gebru et al., "Datasheets for datasets," *Commun. ACM*, vol. 64, no. 12, pp. 86–92, Nov. 2021, doi: 10.1145/3458723.

[40] M. Mitchell et al., "Model cards for model reporting," in *Proc. Conf. Fairness, Accountability, Transparency*, 2019, pp. 220–229, doi: 10.1145/3287560.3287596.

[41] K. Shahriari and M. Shahriari, "IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems," in *2017 IEEE Canada Int. Humanitarian Technol. Conf.*, pp. 197–201, doi: 10.1109/IHTC.2017.8058187.

[42] L. Richardson and M. Amundsen, *RESTful Web APIs.* Sebastopol, CA, USA: O'Reilly Media, 2013.

[43] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. New York, NY, USA: Springer Science+Business Media, 2017.

[44] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 3rd ed., 2025. [Online]. Available: https://christophm.github.io/interpretable-ml-book/

[45] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE Int. Conf. Comput. Vis.*, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[46] J. Colin, T. Fel, R. Cadène, and T. Serre, "What I cannot predict, I do not understand: A human-centered evaluation framework for explainability methods," in *Adv. Neural Inf. Process. Syst.*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds. vol. 35, 2022, pp. 2832–2845. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/hash/13113e938f2957891c0c5e8df811dd01-Abstract-Conference.html

[47] L. Zhang, G. Karakasidis, A. Odnoblyudova, L. Dogruel, Y. Tian, and A. Jung, "Explainable empirical risk minimization," *Neural Comput. Appl.*, vol. 36, no. 8, pp. 3983–3996, Mar. 2024, doi: 10.1007/s00521-023-09269-3.

[48] J. Chen, L. Song, M. J. Wainwright, and M. I. Jordan, "Learning to explain: An information-theoretic perspective on model interpretation," in *Proc. 35th Int. Conf. Mach. Learn.*, J. Dy and A. Krause, Eds. vol. 80, 2018, pp. 883–892. [Online]. Available: https://proceedings.mlr.press/v80/chen18j.html

[49] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, pp. 41–75, Jul. 1997, doi: 10.1023/A:1007379606734.

[50] A. Jung, G. Hannak, and N. Goertz, "Graphical lasso based model selection for time series," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1781–1785, Oct. 2015, doi: 10.1109/LSP.2015.2425434.

[51] A. Jung, "Learning the conditional independence structure of stationary

time series: A multitask learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 21, Nov. 2015, doi: 10.1109/TSP.2015.2460219.

[52] D. N. Gujarati and D. C. Porter, *Basic Econometrics*, 5th ed. New York, NY, USA: McGraw-Hill/Irwin, 2009.

[53] Y. Dodge, Ed. *The Oxford Dictionary of Statistical Terms*. New York, NY, USA: Oxford Univ. Press, 2003.

[54] B. S. Everitt, *The Cambridge Dictionary of Statistics*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[55] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* Cambridge, MA, USA: MIT Press, 2002.

[56] D. Sun, K.-C. Toh, and Y. Yuan, "Convex clustering: Model, theoretical guarantee and efficient algorithm," *J. Mach. Learn. Res.*, vol. 22, no. 9, pp. 1–32, Jan. 2021. [Online]. Available: http://jmlr.org/papers/v22/18-694.html

[57] K. Pelckmans, J. De Brabanter, J. A. K. Suykens, and B. De Moor, "Convex clustering shrinkage," presented at the PASCAL Workshop Statist. Optim. Clustering Workshop, 2005.

[58] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

[59] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course.* Boston, MA, USA: Kluwer Academic, 2004.

[60] S. Bubeck, "Convex optimization: Algorithms and complexity," *Found. Trends Mach. Learn.*, vol. 8, no. 3–4, pp. 231–357, Nov. 2015, 10.1561/2200000050.

[61] D. P. Bertsekas, *Convex Optimization Algorithms.* Belmont, MA, USA: Athena Scientific, 2015.

[62] A. Vaswani et al., "Attention is all you need," in *Adv. Neural Inf. Process. Syst.*, I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. vol. 30, 2017, pp. 5998–6008. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[63] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice Hall PTR, 1993.

[64] C. M. Bishop, *Pattern Recognition and Machine Learning.* New York, NY, USA: Springer Science+Business Media, 2006.

[65] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* New York, NY, USA: Cambridge Univ. Press, 2000.

[66] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations.* Boca Raton, FL, USA: CRC Press, 2015.

[67] E. Abbe, "Community detection and stochastic block models: Recent developments," *J. Mach. Learn. Res.*, vol. 18, no. 177, pp. 1–86, Apr. 2018. [Online]. Available: http://jmlr.org/papers/v18/16-480.html

[68] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill Higher Education, 2002.

[69] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2013.

[70] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Horizontal federated learning," in *Federated Learning.* Cham, Switzerland: Springer Nature, 2020, ch. 4, pp. 49–67.

[71] O. Chapelle, B. Schölkopf, and A. Zien, Eds. *Semi-Supervised Learning*. Cambridge, MA, USA: MIT Press, 2006.

[72] P. R. Halmos, *Measure Theory.* New York, NY, USA: Springer-Verlag, 1974.

[73] O. Kallenberg, *Foundations of Modern Probability.* New York, NY, USA: Springer-Verlag, 1997.

[74] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007, doi: 10.1007/s11222-007-9033-z.

[75] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Adv. Neural Inf. Process. Syst.*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds. vol. 14, 2001, pp. 849–856. [Online]. Available: https://papers.nips.cc/paper_files/paper/2001/hash/801272ee79cfde7fa5960571fee36b9b-Abstract.html

[76] A. Lapidoth, *A Foundation in Digital Communication.* Cambridge, U.K.: Cambridge Univ. Press, 2009.

[77] A. Lapidoth, *A Foundation in Digital Communication*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[78] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Aug. 2013, doi: 10.1007/s10957-012-0245-9.

[79] L. Bottou, "On-line learning and stochastic approximations," in *On-Line Learning in Neural Networks*, D. Saad, Ed. New York, NY, USA: Cambridge Univ. Press, 1999, ch. 2, pp. 9–42.

[80] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database System Concepts*, 7th ed. New York, NY, USA: McGraw-Hill Education, 2019. [Online]. Available: https://db-book.com/

[81] S. Abiteboul, R. Hull, and V. Vianu, *Foundations of Databases.* Reading, MA, USA: Addison-Wesley, 1995.

[82] S. Hoberman, *Data Modeling Made Simple: A Practical Guide for Business and IT Professionals*, 2nd ed. Basking Ridge, NJ, USA: Technics Publications, 2009.

[83] R. Ramakrishnan and J. Gehrke, *Database Management Systems*, 3rd ed. New York, NY, USA: McGraw-Hill, 2002.

[84] Y. SarcheshmehPour, Y. Tian, L. Zhang, and A. Jung, "Flow-based clustering and spectral clustering: A comparison," in *2021 55th Asilomar Conf. Signals, Syst., Comput.*, M. B. Matthews, Ed. pp. 1292–1296, doi: 10.1109/IEEECONF53345.2021.9723162.

[85] S. Ross, *A First Course in Probability*, 9th ed. Boston, MA, USA: Pearson Education, 2014.

[86] N. Young, *An Introduction to Hilbert Space.* New York, NY, USA: Cambridge Univ. Press, 1988.

[87] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Vertical federated learning," in *Federated Learning.* Cham, Switzerland: Springer Nature, 2020, ch. 5, pp. 69–81.

[88] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning.* Cambridge, MA, USA: MIT Press, 2006.

[89] G. Tel, *Introduction to Distributed Algorithms*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[90] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Belmont, MA, USA: Athena Scientific, 2015.

[91] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games.* New York, NY, USA: Cambridge Univ. Press, 2006.

[92] E. Hazan, "Introduction to online convex optimization," *Found. Trends Optim.*, vol. 2, no. 3–4, pp. 157–325, Aug. 2016, doi: 10.1561/2400000013.

[93] L. Cohen, *Time-Frequency Analysis.* Upper Saddle River, NJ, USA: Prentice Hall PTR, 1995.

[94] J. Li, L. Han, X. Li, J. Zhu, B. Yuan, and Z. Gou, “An evaluation of deep neural network models for music classification using spectrograms,” *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 4621–4647, Feb. 2022, doi: 10.1007/s11042-020-10465-9.

[95] B. Boashash, Ed. *Time Frequency Signal Analysis and Processing: A Comprehensive Reference*. Oxford, U.K.: Elsevier, 2003.

[96] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Burlington, MA, USA: Academic, 2009.

[97] Y. SarcheshmehPour, Y. Tian, L. Zhang, and A. Jung, “Clustered federated learning via generalized total variation minimization,” *IEEE Trans. Signal Process.*, vol. 71, pp. 4240–4256, 2023, doi: 10.1109/TSP.2023.3322848.

[98] M. Fredrikson, S. Jha, and T. Ristenpart, “Model inversion attacks that exploit confidence information and basic countermeasures,” in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2015, pp. 1322–1333, doi: 10.1145/2810103.2813677.

[99] A. Rakhlin, O. Shamir, and K. Sridharan, “Making gradient descent optimal for strongly convex stochastic optimization,” in *Proc. 29th Int. Conf. Mach. Learn.*, J. Langford and J. Pineau, Eds. 2012, pp. 449–456. [Online]. Available: https://icml.cc/Conferences/2012/papers/261.pdf

[100] M. J. Wainwright and M. I. Jordan, “Graphical models, exponential families, and variational inference,” *Found. Trends Mach. Learn.*, vol. 1, no. 1–2, pp. 1–305, Nov. 2008, doi: 10.1561/2200000001.

[101] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications.* Berlin, Germany: Springer-Verlag, 2011.

[102] A. Jung, “Networked exponential families for big data over networks,” *IEEE Access*, vol. 8, pp. 202 897–202 909, Nov. 2020, doi: 10.1109/ACCESS.2020.3033817.

[103] X. Liu, H. Li, G. Xu, Z. Chen, X. Huang, and R. Lu, “Privacy-enhanced federated learning against poisoning adversaries,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4574–4588, 2021, doi: 10.1109/TIFS.2021.3108434.

[104] J. Zhang, B. Chen, X. Cheng, H. T. T. Binh, and S. Yu, “PoisonGAN: Generative poisoning attacks against federated learning in edge computing systems,” *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3310–3322, Mar. 2021, doi: 10.1109/JIOT.2020.3023126.

[105] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization.* Belmont, MA, USA: Athena Scientific, 2003.

[106] H. P. Lopuhaä and P. J. Rousseeuw, “Breakdown points of affine equivariant estimators of multivariate location and covariance matrices,” *Ann. Statist.*, vol. 19, no. 1, pp. 229–248, Mar. 1991, doi: 10.1214/aos/1176347978.

[107] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, A. Singh and J. Zhu, Eds. vol. 54, 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[108] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, I. Dhillon, D. Papailiopoulos, and V. Sze, Eds. vol. 2, 2020. [Online]. Available: https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html

[109] K. Abayomi, A. Gelman, and M. Levy, "Diagnostics for multivariate imputations," *J. Roy. Statist. Soc.: Ser. C (Appl. Statist.)*, vol. 57, no. 3, pp. 273–291, Jun. 2008, doi: 10.1111/j.1467-9876.2007.00613.x.

[110] S. Shalev-Shwartz and A. Tewari, "Stochastic methods for ℓ_1 regularized loss minimization," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, L. Bottou and M. Littman, Eds. Jun. 2009, pp. 929–936.

[111] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Trans. Evol. Comput.*, vol. 23, no. 5, pp. 828–841, Oct. 2019, doi: 10.1109/TEVC.2019.2890858.

[112] S. Mallat, "Understanding deep convolutional networks," *Philos. Trans. Roy. Soc. A*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150203, doi: 10.1098/rsta.2015.0203.

[113] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[114] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York, NY, USA: Springer-Verlag, 1991.

[115] M. Kearns and M. Li, “Learning in the presence of malicious errors,” *SIAM J. Comput.*, vol. 22, no. 4, pp. 807–837, Aug. 1993, doi: 10.1137/0222052.

[116] G. Lugosi and S. Mendelson, “Robust multivariate mean estimation: The optimality of trimmed mean,” *Ann. Statist.*, vol. 49, no. 1, pp. 393–410, Feb. 2021, doi: 10.1214/20-AOS1961.

[117] V. I. Istrăţescu, *Fixed Point Theory: An Introduction.* Dordrecht, The Netherlands: D. Reidel, 1981.

[118] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis.* Cambridge, UK: Cambridge Univ. Press, 1991.

[119] G. Strang, *Introduction to Linear Algebra*, 5th ed. Wellesley-Cambridge Press, MA, 2016.

[120] G. Strang, *Computational Science and Engineering.* Wellesley, MA, USA: Wellesley-Cambridge Press, 2007.

[121] R. Motwani and P. Raghavan, *Randomized Algorithms.* Cambridge, U.K.: Cambridge Univ. Press, 1995.

[122] R. G. Gallager, *Stochastic Processes: Theory for Applications*. New York, NY, USA: Cambridge Univ. Press, 2013.

[123] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed. New York, NY, USA: Springer-Verlag, 2003.