ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

7ο εξάμηνο — Ακαδημαϊκό έτος 2022-2023

1η Σειρά Ασκήσεων

Ημερ. Παράδ.: 14.12.2022

Γενικές Οδηγίες: Οι αναλυτικές σειρές ασκήσεων είναι ατομικές, και οι λύσεις που θα δώσετε πρέπει να αντιπροσωπεύουν μόνο την προσωπική σας εργασία. Εξηγήστε επαρκώς την εργασία σας. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός των σημειώσεων για την λύση σας, πρέπει να το αναφέρετε. Η παράδοση των λύσεων των αναλυτικών ασκήσεων της σειράς αυτής θα γίνει ηλεκτρονικά στην HELIOS ιστοσελίδα του μαθήματος και θα πρέπει να την υποβάλετε ως ένα ενιαίο αρχείο PDF με το εξής filename format χρησιμοποιώντας μόνο λατινικούς χαρακτήρες: ML22_hwk1_AM_LastnameFirstname.pdf, όπου AM ειναι ο 8-ψήφιος αριθμός μητρώου σας. Σκαναριμένες χειρόγραφες λύσεις επιτρέπονται αρκεί να είναι καθαρογραμμένες και ευανάγνωστες. Επίσης στην 1η σελίδα των λύσεων θα αναγράφετε το ονοματεπώνυμο, Α.Μ., και email address σας. Συμπεριλάβετε και τον κώδικα προγραμμάτων, π.χ. Matlab ή Python, που χρησιμοποιήσατε για αριθμητική επίλυση.

Άσκηση 1.1 (Linear and Ridge Regression)

Θεωρήστε το ακόλουθο μοντέλο παλινδρόμησης πάνω σε 2 ανεξάρτητες μεταβλητές (χαρακτηριστικά) x_1 και x_2:

$$y = w_0 + w_1 x_1 + w_2 x_2$$

Το σύνολο δεδομένων που θα χρησιμοποιηθούν στη συγκεκριμένη άσκηση περιλαμβάνεται στο αρχείο ML2022-23_hwk1_olympic_teams.zip, το οποίο είναι διαθέσιμο στην ιστοσελίδα του μαθήματος. Τα δεδομένα σε αυτό το αρχείο αφορούν τα μετάλλια που έχουν κερδίσει αθλητές από διάφορες χώρες στους Ολυμπιακούς Αγώνες τα τελευταία χρόνια, καθώς και το πλήθος των αθλητών, των αγωνισμάτων, το ύψος και βάρος αθλητών κλπ.

Από το σύνολο των χαρακτηριστικών που σας δίνονται, θεωρήστε ως ανεξάρτητες μεταβλητές (χαρακτηριστικά) x_1 και x_2 το πλήθος των αθλητών/athletes (A) και το πλήθος των αγωνισμάτων/events (N), αντίστοιχα, ενώ ως εξαρτημένη μεταβλητή y τον αριθμό των κερδισμένων μεταλλίων/medals (M). Πριν χρησιμοποιήσετε τις αριθμητικές τιμές των δεδομένων για τα ανωτέρω μεγέθη, θα σας βοηθήσει πρώτα να τα κανονικοποιήσετε (δηλ. μέση τιμή 0, και τυπική απόκλιση 1).

(α) Υπολογίστε τον κανονικοποιημένο συντελεστή συσχέτισης r_{12} μεταξύ των ανεξάρτητων μεταβλητών αθλητών (A) και αγωνισμάτων (N). Τι παρατηρείτε σχεδιάζοντας σε ένα γράφημα διασποράς (scatterplot) τις δύο αυτές μεταβλητές?

(β) Υπολογίστε τα βάρη w_0, w_1, w_2 εφαρμόζοντας τον αλγόριθμο/εξισώσεις για γραμμική παλινδρόμηση (Linear Regression).

(γ) Υπολογίστε τα βάρη w_0, w_1, w_2 εφαρμόζοντας τον αλγόριθμο/εξισώσεις για Ridge Regression για τιμές της παραμέτρου $\lambda = 1, 10, 100$.

(δ) Υπολογίστε τα σφάλματα RMSE (τετραγωνική ρίζα των μέσων τετραγωνικών σφαλμάτων) εκπαίδευσης και επαλήθευσης που αντιστοιχούν στα παραπάνω ερωτήματα (β) και (γ), και έπειτα συμπληρώστε τα σε έναν πίνακα της παρακάτω μορφής, μαζί με τις εξισώσεις που προέκυψαν σε κάθε περίπτωση. Ποια τιμή της παραμέτρου λ θα επιλέγατε και με βάση ποιο κριτήριο? Ποια συμπεράσματα προκύπτουν?

Σημείωση: Επεξηγήστε αναλυτικά τη διαδικασία που ακολουθήσατε για να φθάσετε στις λύσεις σας. Για την επίλυση μπορείτε να χρησιμοποιήσετε προγραμματιστικά εργαλεία (π.χ. Python, Matlab) που διευκολύνουν λειτουργίες γραμμικής άλγεβρας (όπως πολλαπλασιασμό ή αντιστροφή πινάκων), αλλά όχι τις έτοιμες υλοποιήσεις για Linear Regression, Ridge Regression που περιέχονται σε βιβλιοθήκες (όπως scikit-learn κλπ). Προαιρετικά, θα μπορούσατε να τις χρησιμοποιήσετε για επαλήθευση (μόνο) των ανωτέρω αποτελεσμάτων/εξισώσεων.

	Linear Regression ($\lambda = 0$)	Ridge Regression ($\lambda = 1$)	Ridge Regression ($\lambda = 10$)	Ridge Regression ($\lambda = 100$)
RMSE (Train set)				
RMSE (Test set)				
Εξίσωση $y = \dots$				

Άσκηση 1.2 (Gaussian distribution)

(α) Αποδείξτε ότι αν η τυχαία μεταβλητή (τ.μ.) x ακολουθεί την Gaussian κατανομή με παραμέτρους μ και σ^2, τότε ισχύει $\sigma_x^2 = \sigma^2$, δηλαδή η μεταβλητότητα της x ισούται με την παράμετρο σ^2.

(β) Έστω το τυχαίο διάνυσμα (τ.δ.) $\mathrm{x} \in \mathbb{R}^2$ που ακολουθεί την κανονική κατανομή, $\mathrm{x} \sim \mathcal{N}(\mathrm{x}|\boldsymbol{\mu}, \Sigma)$, με

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Να βρείτε αναλυτικά την έκφραση που δίνει τις ισοσταθμικές καμπύλες της κατανομής του x και να τις σχεδιάσετε προσεγγιστικά στο επίπεδο-x_1x_2 που ορίζουν οι συνιστώσες του x.

(γ) Θεωρούμε ότι τα $\mathrm{x}_1, \mathrm{x}_1, \ldots, \mathrm{x}_N \in \mathbb{R}^l$ είναι ανεξάρτητες παρατηρήσεις που ακολουθούν την κανονική κατανομή, $\mathcal{N}(\mathrm{x}_n|, \Sigma)$. Να δειχτεί ότι ο εκτιμητής μεγίστης πιθανοφάνειας $\boldsymbol{\mu}_{\mathrm{ML}}$ της παραμέτρου $\boldsymbol{\mu}$ ισούται με τον δειγματικό μέσο:

$$\boldsymbol{\mu}_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{x}_n \ .$$

Άσκηση 1.3 (Bayes classifier)

Η κατανομή Laplace με παραμέτρους μ και a ορίζεται ως εξής:

$$f(x|\mu, a) = \frac{1}{2a} e^{-\frac{|x-\mu|}{a}}, \quad -\infty < x < \infty.$$

Για την κατανομή αυτή μπορεί να δειχτεί ότι $\mathbb{E}[x] = \mu$ και $\sigma_x^2 = 2a^2$.

(α) Θεωρούμε ένα πρόβλημα ταξινόμησης σε δύο κλάσεις ω_1 και ω_2, οι οποίες είναι ισοπίθανες, δηλαδή $P(\omega_1) = P(\omega_2) = \frac{1}{2}$. Έστω, επίσης, ότι τα δεδομένα στις δύο κλάσεις ακολουθούν την κατανομή Laplace με παραμετρους $(\mu = 0, a = 1)$ και $(\mu = 3, a = 1)$, αντίστοιχα, δηλαδή, $p(x|\omega_1) = f(x|0, 1)$ και $p(x|\omega_2) = f(x|3, 1)$. Να βρεθεί η τιμή κατωφλίου x_t που διαχωρίζει τις δύο κλάσεις και ελαχιστοποιεί το σφάλμα ταξινόμησης.
(β) Αν στο προηγούμενο ερώτημα τα λάθη που σχετίζονται με τις δύο κλάσεις δεν έχουν την ίδια βαρύτητα, αλλά τα αντίστοιχα βάρη είναι $\lambda_{12} = \frac{1}{2}$ και $\lambda_{21} = 1$, να βρεθεί η τιμή κατωφλίου x_r που διαχωρίζει τις δύο κλάσεις και ελαχιστοποιεί το μέσο ρίσκο. Συγκρίνετε και σχολιάστε τα αποτελέσματα που πήρατε στα (α) και (β).

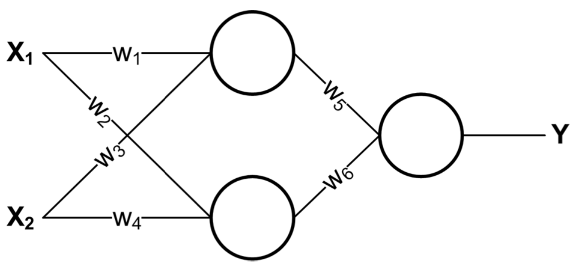
Έστω, τώρα, το πρόβλημα ταξινόμησης σε δύο ισοπίθανες κλάσεις στο $\mathbb{R}^2$, όπου τα $p(\mathrm{x}|\omega_1), p(\mathrm{x}|\omega_2)$ είναι Gaussian κατανομές με μέσα διανύσματα $\boldsymbol{\mu}_1 = [2, -2]^T$, $\boldsymbol{\mu}_2 = [-1, 2]^T$ και κοινό πίνακα συμμεταβλητότητας (δύο περιπτώσεις):

$$\text{(i)}\ \ \Sigma = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}, \quad \text{(ii)}\ \ \Sigma = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}.$$

(γ) Για κάθε μία από τις περιπτώσεις (i) και (ii), προσδιορίστε την εξίσωση της ευθείας που διαχωρίζει τις δύο κλάσεις κατά βέλτιστο τρόπο, ελαχιστοποιώντας το σφάλμα ταξινόμησης. Δώστε μια ενιαία γεωμετρική απεικόνιση των παραπάνω ευθειών σε σχέση με τα κέντρα των δύο κλάσεων και σχολιάστε.
(δ) Προσδιορίστε, αλγεβρικά και γεωμετρικά, σε ποια κλάση θα ταξινομηθεί το πρότυπο $\hat{\mathrm{x}} = [4, 3]^T$, για κάθε μία από τις περιπτώσεις (i) και (ii).

Άσκηση 1.4 (Perceptron - MultiLayer Perceptron)

Έστω νευρωνικό δίκτυο με δύο εισόδους, ένα κρυμμένο στρώμα με δύο νευρώνες και έναν νευρώνα στο στρώμα εξόδου, όπως απεικονίζεται στην εικόνα.



(α) Υποθέστε αρχικά ότι χρησιμοποιούμε γραμμική συνάρτηση ενεργοποίησης $f(u) = cu$, όπου c σταθερά. Σχεδιάστε ένα νέο νευρωνικό δίκτυο χωρίς κρυμμένο στρώμα, το οποίο να είναι ισοδύναμο με το παραπάνω MLP. Προσδιορίστε τη συνάρτηση ενεργοποίησης και εκφράστε τα βάρη του νέου δικτύου συναρτήσει των w_i και της σταθεράς c.
(β) Ισχύει ή όχι ότι για οποιοδήποτε MultiLayer Perceptron με γραμμικές συναρτήσεις ενεργοποίησης μπορούμε να βρούμε ένα ισοδύναμο νευρωνικό δίκτυο χωρίς κρυμμένα στρώματα; Εξηγήστε.
(γ) Υποθέστε τώρα ότι στο νευρωνικό δίκτυο της εικόνας χρησιμοποιούμε ως συναρτήσεις ενεργοποίησης: τη σιγμοειδή για τους νευρώνες του κρυμμένου στρώματος και τη βηματική ($g(u) = 1$ αν $u > 0$ και $g(u) = 0$ αν $u < 0$) για τον νευρώνα εξόδου. Προσδιορίστε τις τιμές των βαρών w_i ώστε το νευρωνικό δίκτυο να υπολογίζει το αποτέλεσμα της λογικής πράξης XOR (ΑΠΟΚΛΕΙΣΤΙΚΟ Ή) των $X1$ και $X2$. Θεωρούμε ότι στο νευρωνικό δίκτυο δεν υπάρχουν όροι πόλωσης (bias).

Άσκηση 1.5 (Support Vector Machines - Kernels)

Δίνονται τα ακόλουθα έξι (6) σημεία στον μονοδιάστατο χώρο $\mathbb{R}^1$: τρία με αρνητική (-) ετικέτα (label) $x_1 = -1, x_2 = 0, x_3 = 1$ και τρία με θετική (+) ετικέτα $x_4 = -3, x_5 = -2, x_6 = 3$. Τα σημεία αυτά είναι προφανώς μη γραμμικά διαχωρίσιμα στο χώρο εισόδου (input space) $\mathbb{R}^1$. Ωστόσο, χρησιμοποιώντας τον μετασχηματισμό $\phi(u) = (u, u^2)$, τα σημεία καθίστανται γραμμικά διαχωρίσιμα στον χώρο χαρακτηριστικών (feature space) $\mathbb{R}^2$.
(α) Δώστε την αναλυτική μορφή του πυρήνα (kernel) $k(X_1, X_1')$ που αντιστοιχεί στον μετασηχματισμό ϕ.
(β) Προσδιορίστε το υπερεπίπεδο μέγιστου περιθωρίου $w_1Y_1+w_2Y_2+c = 0$, όπου $(Y_1, Y_2) = \phi(X_1)$ και υπολογίστε το πλάτος γ του περιθωρίου. Σημειώστε ότι δεν χρειάζεται να λύσετε πρόβλημα τετραγωνικού προγραμματισμού, αλλά μπορείτε να προσδιορίσετε γεωμετρικά τις τιμές w_1, w_2, c, παρατηρώντας ότι η ευθεία διαχωρισμού πρέπει να περνάει μεταξύ των σημείων (-2,4) και (-1,1) και να είναι κάθετη στην ευθεία που ορίζουν τα σημεία αυτά (εξηγήστε γιατί ισχύει αυτό). Χρησιμοποιήστε μόνο δύο διανύσματα υποστήριξης (support vectors) για την επίλυση.
(γ) Αναπαραστήστε γραφικά τα έξι σημεία μετά τον μετασχηματισμό τους στο χώρο χαρακτηριστικών $\mathbb{R}^2$. Στο ίδιο διάγραμμα σχεδιάστε το υπερεπίπεδο και το περιθώριο διαχωρισμού και επισημάνετε τα διανύσματα υποστήριξης.
(δ) Σχεδιάστε (σε νέο διάγραμμα) το όριο απόφασης του υπερεπιπέδου διαχωρισμού στον αρχικό χώρο εισόδου $\mathbb{R}^1$.
(ε) Υπολογίστε τους συντελεστές α_n και τη σταθερά b της εξίσωσης (1) για τον πυρήνα k και τα διανύσματα υποστήριξης $SV = \{u_1, u_2\}$ που χρησιμοποιήσατε παραπάνω. Προς τούτο, λάβετε υπόψη το δυϊκό πρόβλημα τετραγωνικού προγραμματισμού και τους περιορισμούς που τίθενται για τους συντελεστές α_n.

$$y(x) = \text{sign}\left(\sum_{n=1}^{|SV|} \alpha_n y_n k(x, u_n) + b\right) \tag{1}$$

(στ) Εάν προσθέσουμε στο αρχικό σύνολο δεδομένων ένα επιπλέον σημείο $x_7 = 5$ με θετική (+) ετικέτα, θα αλλάξει ή όχι το υπερεπίπεδο διαχωρισμού ή/και το περιθώριο; Αιτιολογήστε την απάντησή σας.
(ζ) Θεωρήστε τώρα έναν νέο μετασχηματισμό $\phi_n : \mathbb{R}^1 \to \mathbb{R}^n$:

$$\phi_n(x) = \left\{e^{-x^2/2}, e^{-x^2/2}x, \frac{e^{-x^2/2}x^2}{\sqrt{2}}, \ldots, \frac{e^{-x^2/2}x^i}{\sqrt{i!}} \ldots, \frac{e^{-x^2/2}x^n}{\sqrt{n!}}\right\} \tag{2}$$

Εάν $n \to \infty$, τότε παίρνουμε τον μετασχηματισμό ϕ_∞:

$$\phi_\infty(x) = \left\{e^{-x^2/2}, e^{-x^2/2}x, \frac{e^{-x^2/2}x^2}{\sqrt{2}}, \ldots, \frac{e^{-x^2/2}x^i}{\sqrt{i!}} \ldots\right\} \tag{3}$$

Έστω $\mu = \{\mu_1, \ldots, \mu_i, \ldots\}$ και $\nu = \{\nu_1, \ldots, \nu_i, \ldots\}$ δύο διανύσματα άπειρων διαστάσεων του χώρου χαρακτηριστικών που λαμβάνουμε από τον μετασχηματισμό ϕ_∞. Υπολογίστε το εσωτερικό γινόμενο:

$$k(\mu, \nu) = \mu \cdot \nu = \sum_{i=1}^{\infty} \mu_i \nu_i. \tag{4}$$

Υπόδειξη: Ίσως σας φανεί χρήσιμο το ανάπτυγμα Taylor της εκθετικής συνάρτησης e^x.
(η) Με δεδομένη την υψηλή διάσταση του χώρου χαρακτηριστικών του προηγούμενου ερωτήματος, κατά τη γνώμη σας πόσο θα πρέπει να μας προβληματίζει το ζήτημα της υπερπροσαρμογής (overfitting) στη συγκεκριμένη περίπτωση;

Άσκηση 1.6 (Decision Trees)

(α) Έστω T ένα δέντρο απόφασης που κατασκευάζεται με τον αλγόριθμο DecisionTree.ID3($\mathbb{D}$) από ένα σύνολο δεδομένων $\mathbb{D}$ και έστω $p = t_1 t_2 \cdots t_n$ ένα τυχαίο μονοπάτι του δέντρου, όπου $t_i, i \in \mathbb{N}_n$ ένας κόμβος του T, με t_i πρόγονο του t_j για $i < j$. Έστω επίσης $\text{ig}(t)$ η συνάρτηση που υπολογίζει το κέρδος πληροφορίας του κόμβου t του δέντρου, με βάση την εντροπία.

1. Να ελέγξετε αν ισχύει $\text{ig}(t) \geq 0$ για κάθε κόμβο t του δέντρου T.

2. Έστω t_i, t_j με $i < j$ δύο κόμβοι του T με την ίδια επιλογή χαρακτηριστικού, η οποία στηρίζεται σε κριτήριο ισότητας (η διακλάδωση γίνεται με έλεγχο κριτηρίου ισότητας για την τιμή του χαρακτηριστικού). Να ελέγξετε αν ισχύει $\text{ig}(t_j) > 0$.

(β) Δίνονται οι παρατηρήσεις που απεικονίζονται στο παρακάτω Σχήμα. Να υπολογίσετε δύο δέντρα απόφασης, χρησιμοποιώντας το κριτήριο gini για τον υπολογισμό του κέρδους πληροφορίας, αντιμετωπίζοντας το χαρακτηριστικό Temperature εναλλακτικά ως αριθμητικό ή ως κατηγορικό (εφαρμόζοντας κριτήριο ανισότητας με μία τιμή ή κριτήριο ισότητας αντίστοιχα). Ποιο από τα δύο δέντρα θα επιλέγατε για την ταξινόμηση και γιατί;

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	40	High	FALSE	No
Sunny	40	High	TRUE	No
Overcast	36	High	FALSE	Yes
Rainy	22	High	FALSE	Yes
Rainy	8	Normal	FALSE	Yes
Rainy	8	Normal	TRUE	No
Overcast	4	Normal	TRUE	Yes
Sunny	22	High	FALSE	No
Sunny	4	Normal	TRUE	Yes
Rainy	22	Normal	FALSE	Yes
Sunny	22	Normal	TRUE	Yes
Overcast	22	High	TRUE	Yes
Overcast	36	Normal	FALSE	Yes
Rainy	22	High	TRUE	No