
Οι αναλυτικές σειρές ασκήσεων είναι ατομικές, και οι λύσεις που θα δώσετε πρέπει να αντιπροσωπεύουν μόνο την προσωπική σας εργασία. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός βιβλίου για τη λύση σας, πρέπει να το αναφέρετε. Παραδίδονται γραπτώς και προσωπικώς στη Γραμματεία Εργ. Ρομποτικής (Αίθ. 2.1.12, παλαιό Κτ. Ηλεκτρ.) 9.00-14.30.

Υλικό για Ανάγνωση:
Βιβλία: [1], [2], [3] και [4]
Διαφάνειες διαλέξεων μαθήματος

Αναλυτικές Ασκήσεις

Άσκηση 2.1: (Hidden Markov Models)
Ζητείται να χρησιμοποιηθεί ένα HMM για να αποκωδικοποιηθεί μια απλή ακολουθία DNA. Είναι γνωστό ότι μια ακολουθία DNA είναι μια σειρά από στοιχεία του συνόλου $\{A, C, G, T\}$. Ας υποθέσουμε ότι υπάρχει μία κρυμμένη κατάσταση S που ελέγχει τη δημιουργία της ακολουθίας DNA και έχει 2 πιθανές καταστάσεις $\{S_1, S_2\}$. Επίσης, δίνονται οι ακόλουθες πιθανότητες μετάβασης για το HMM λ:

$$P(S_1|S_1) = 0.8 \quad P(S_2|S_1) = 0.2 \quad P(S_1|S_2) = 0.2 \quad P(S_2|S_2) = 0.8$$

οι ακόλουθες πιθανότητες των παρατηρήσεων:

$$P(A|S_1) = 0.4 \quad P(C|S_1) = 0.1 \quad P(G|S_1) = 0.4 \quad P(T|S_1) = 0.1$$

$$P(A|S_2) = 0.1 \quad P(C|S_2) = 0.4 \quad P(G|S_2) = 0.1 \quad P(T|S_2) = 0.4$$

και οι ακόλουθες a-priori πιθανότητες:

$$P(S_1) = 0.5 \quad P(S_2) = 0.5$$

Έστω ότι η παρατηρούμενη ακολουθία είναι $x = CGTCAG$. Υπολογίστε:
(α) Την πιθανότητα $P(x|\lambda)$ χρησιμοποιώντας τον forward αλγόριθμο.
(β) Τις εκ των υστέρων πιθανότητες $P(\pi_i = S_1|x, \lambda)$ για $i = 1, \ldots, 6$.
(γ) Το πιο πιθανό μονοπάτι κρυμμένων καταστάσεων χρησιμοποιώντας τον αλγόριθμο Viterbi.

Άσκηση 2.2: (Principal Component Analysis)
Ζητείται η εφαρμογή της Principal Component Analysis (PCA) πάνω στο ευρέως διαδεδομένο σύνολο δεδομένων κρίνων του Fisher, προκειμένου να μετασχηματιστούν τα δεδομένα σε ένα χώρο χαμηλότερων διαστάσεων. Τα δεδομένα αποτελούνται από 3 κλάσεις (για τους 3 διαφορετικούς τύπους κρίνου), καθεμιά από τις οποίες περιλαμβάνει 50 δείγματα. Τα δεδομένα περιγράφονται από 4 διαφορετικά χαρακτηριστικά:

- μήκος σεπάλων σε εκ.
- πλάτος σεπάλων σε εκ.

- μήκος πετάλων σε εκ.
- πλάτος πετάλων σε εκ.
- τύπος κρίνου (Iris Setosa/Iris Versicolour/Iris Virginica)

(α) Κατεβάστε το σύνολο δεδομένων που έχει ανεβεί στο mycourses (αρχείο `PCA.data`).
(β) Προεπεξεργαστείτε τα δεδομένα αφαιρώντας τη μέση τιμή και διαιρώντας με την τυπική απόκλιση του κάθε χαρακτηριστικού ξεχωριστά. Τα προκύπτοντα δεδομένα θα πρέπει να έχουν μέση τιμή 0 και διασπορά 1.
(γ) Υπολογίστε τον δειγματικό πίνακα συνδιασπορών $\Sigma = \frac{1}{m-1} \sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^\top$, όπου $\mathbf{x}_i$ είναι το i-στό δείγμα και m το πλήθος των δειγμάτων.
(δ) Παραγοντοποιήστε τον πίνακα συνδιασπορών κάνοντας χρήση του Singular Value Decomposition (SVD) και βρείτε τις αντίστοιχες ιδιοτιμές και ιδιοδιανύσματα. Προσέξτε εάν η συγκεκριμένη υλοποίηση του SVD δίνει τα αποτελέσματα με φθίνουσα ή αύξουσα σειρά ιδιοτιμών. Ο μετασχηματισμός SVD ενός πίνακα Σ είναι μια παραγοντοποίηση της μορφής $\Sigma = UDV^\top$. Για έναν συμμετρικό, θετικά ορισμένο πίνακα Σ, οι $U = V$ περιέχουν τα ιδιοδιανύσματα και D είναι ένας διαγώνιος πίνακας με τις αντίστοιχες ιδιοτιμές.
(ε) Προβάλετε τα δεδομένα πάνω στις δύο πρώτες κύριες συνιστώσες και σχεδιάστε τα αποτελέσματα που προκύπτουν.
(στ) Ποιος είναι ο ελάχιστος απαιτούμενος αριθμός από κύριες συνιστώσες ώστε να “ερμηνεύεται” το 95% της διασποράς των τιμών;

Άσκηση 2.3: (Linear Discriminant Analysis)
Στο μάθημα είδαμε ότι η Linear Discriminant Analysis (LDA) βασίζεται στην ανάστροφη σχέση των μητρών (πινάκων) S_W και S_B:

$$S_W = \sum_{i=1}^{|Classes|} \mathbb{E}_{x|x\in\omega_i} \left[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T \right]$$

$$S_B = \sum_{i=1}^{|Classes|} P(\omega_i)(\vec{\mu_i} - \vec{\mu})(\vec{\mu_i} - \vec{\mu})^T$$

όπου το ω_i αναπαριστά μια κλάση με μέση τιμή $\vec{\mu_i}$, $|Classes|$ είναι το πλήθος των κλάσεων και $\vec{\mu}$ είναι η μέση τιμή όλων των δειγμάτων.
(α) Δείξτε ότι στην περίπτωση διαχωρισμού δύο κλάσεων ω_1 και ω_2, ο πίνακας S_B μπορεί να γραφτεί στη μορφή $S_B = P(\omega_1)P(\omega_2)(\vec{\mu_2} - \vec{\mu_1})(\vec{\mu_2} - \vec{\mu_1})^T$.
(β) Βασιζόμενοι στο υποερώτημα **(α)**, να βρείτε το ιδιοδιάνυσμα του πίνακα $S_W^{-1} S_B$ και την ιδιοτιμή του.

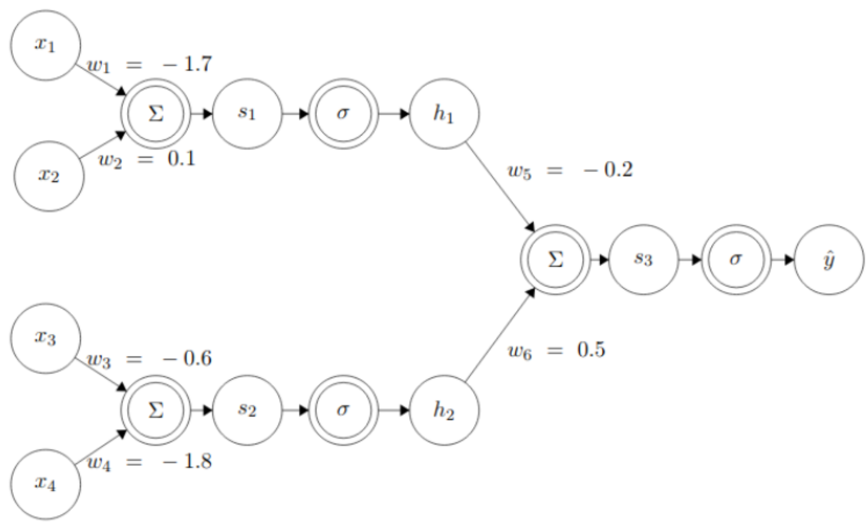
Άσκηση 2.4: (Multilayer Perceptron)
Μας δίνονται 10 διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις ω_1 και ω_2:

$$\omega_1 : [4.5, 1]^\top, [5, 2]^\top, [5, 1]^\top, [0, 4.5]^\top, [0.5, 4]^\top$$

$$\omega_2 : [0, 1]^\top, [0.5, 2]^\top, [5, 4]^\top, [4.5, 4]^\top, [1, 1]^\top$$

Ελέγξτε αν οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες, και αν όχι, σχεδιάστε ένα κατάλληλο multilayer perceptron με τους κόμβους να έχουν βηματική συνάρτηση ενεργοποίησης (step) για να ταξινομήσετε τα διανύσματα στις δύο κλάσεις.

Άσκηση 2.5: (Backpropagation)
Υποθέστε ότι έχουμε το ακόλουθο νευρωνικό δίκτυο. Οι κόμβοι που βρίσκονται μέσα σε μονό κύκλο υποδηλώνουν μεταβλητές (για παράδειγμα η x_1 είναι μια μεταβλητή εισόδου, h_1 είναι μια ενδιάμεση μεταβλητή, και $\hat{y}$ είναι μια μεταβλητή εξόδου). Οι κόμβοι που βρίσκονται μέσα σε διπλό κύκλο υποδηλώνουν συναρτήσεις (για παράδειγμα το $\sum$ υπολογίζει το άθροισμα των εισόδων του και η σ αναπαριστά τη συνάρτηση logistic $\sigma(x) = \frac{1}{1+e^{-x}}$.



Θεωρήστε ότι η συνάρτηση για το L2 loss δίνεται από τη σχέση $L(y, \hat{y}) = ||y - \hat{y}||_2^2$. Επίσης, μας δίνονται τα δεδομένα ενός δείγματος $(x_1, x_2, x_3, x_4) = (-0.7, 1.2, 1.1, -2)$ με τιμή για το πραγματικό του label ίση με 0.5. Χρησιμοποιήστε τον αλγόριθμο backpropagation για να υπολογίσετε τη μερική παράγωγο $\frac{\partial L}{\partial w_1}$.
Σημείωση: Το gradient για μια συνάρτηση L2 loss είναι ίσο με $2||y - \hat{y}||$.

Άσκηση 2.6: (Support Vector Machine)
Θεωρήστε το πρόβλημα του διαχωρισμού ενός συνόλου από διανύσματα εκπαίδευσης δύο κλάσεων. Τα δεδομένα εκπαίδευσης είναι της μορφής $\{(\mathbf{x}_i, y_i)\}$, όπου τα διανύσματα χαρακτηριστικών $\mathbf{x}_i \in \mathbb{R}^m$ και τα labels των κλάσεων $y_i \in \{-1, 1\}$.
Όπως είναι γνωστό, στην περίπτωση όπου τα δεδομένα εκπαίδευσης δεν είναι γραμμικώς διαχωρίσιμα (για παράδειγμα μέσω ενός κανόνα απόφασης $\text{sign}(\mathbf{w} \cdot \mathbf{x} + b)$ για κάποια $\mathbf{w}, b$), τότε το πρόβλημα χρειάζεται να διατυπωθεί με χρήση slack variables $\{\xi_i\}$, $1 \leq i \leq n$. Έτσι, ο ταξινομητής SVM με το μεγαλύτερο περιθώριο αποκτάται μέσω της επίλυσης του δυϊκού προβλήματος:

$$L(\mathbf{w}, b, \alpha, \xi, \beta) = \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i[(\mathbf{w} \cdot \mathbf{x}_i + b)y_i - (1 - \xi_i)] - \sum_{i=1}^{n} \beta_i \xi_i$$

όπου το C είναι μια σταθερά, $\alpha_i, \beta_i \geq 0$, $\forall i$ είναι οι πολλαπλασιαστές Lagrange, και $\xi_i \geq 0$ είναι οι slack variables.
(α) Υποθέστε ότι $n = 4$ και ότι το $\mathbf{x}$ είναι δύο διαστάσεων $< x_i^1, x_i^2 >:< 2, 2 >, < 2.5, 2.5 >, < 5, 5 >, < 7, 7 >$. Τώρα το SVM εκπαιδεύεται με βάση την παραπάνω εξίσωση. Να δείξετε ότι οποιαδήποτε labels y και αν έχουν τα τέσσερα δείγματα εκπαίδευσης, το βέλτιστο διάνυσμα

παραμέτρων $\hat{\mathbf{w}} = (\hat{w}^1, \hat{w}^2)$ έχει την ιδιότητα ότι $\hat{w}^1 = \hat{w}^2$.
(β) Θεωρήστε την εκπαίδευση ενός SVM με slack variables, αλλά δίχως την ύπαρξη του bias όρου ($b = 0$). Θα χρησιμοποιήσουμε έναν kernel $\mathbf{K}(\mathbf{u}, \mathbf{v})$ με την ιδιότητα ότι για δύο οποιαδήποτε σημεία $\mathbf{u}$ και $\mathbf{v}$ που ανήκουν στο σύνολο εκπαίδευσης ισχύει ότι $-1 < \mathbf{K}(\mathbf{u}, \mathbf{v}) < 1$. Επιπρόσθετα, $\mathbf{K}(\mathbf{u}, \mathbf{u}) < 1$. Να δείξετε ότι αν υπάρχουν n δείγματα στο σύνολο εκπαίδευσης και η σταθερά C επιλέγεται ούτως ώστε $C < \frac{1}{n-1}$, τότε όλες οι δυϊκές μεταβλητές α_i είναι μη μηδενικές (και άρα όλα τα δείγματα του συνόλου εκπαίδευσης αποτελούν support vectors).
(γ) Θεωρήστε τον εξής kernel:

$$\mathbf{K}(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} + 4(\mathbf{u} \cdot \mathbf{v})^2$$

όπου τα διανύσματα $\mathbf{u}$ και $\mathbf{v}$ είναι δύο διαστάσεων. Ο kernel αυτός είναι ίσος με το εσωτερικό γινόμενο $\phi(\mathbf{u}) \cdot \phi(\mathbf{v})$ για κάποιο ορισμό της συνάρτησης ϕ. Ποια είναι η συνάρτηση αυτή ϕ;

Άσκηση 2.7: (Logistic Regression)
Θεωρήστε το πρόβλημα logistic regression για ένα σύνολο δεδομένων $\{\boldsymbol{\phi}_n, t_n\}$, όπου $t_n \in \{0, 1\}$ και $\boldsymbol{\phi}_n = \boldsymbol{\phi}(\mathbf{x}_n)$ είναι οι κατηγορίες και οι συναρτήσεις βάσης, αντίστοιχα, για δείγματα $n = \{1, 2, \ldots, N\}$. Η συνάρτηση σφάλματος $E(\mathbf{w})$, η οποία αναφέρεται συνήθως και ως cross-entropy, ορίζεται ως:

$$E(\mathbf{w}) = -\sum_{n=1}^{N} \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$

όπου $\mathbf{w}$ είναι το διάνυσμα βαρών, $y_n = \sigma(\mathbf{w}^\top \boldsymbol{\phi}_n)$ η έξοδος του μοντέλου logistic regression στο διάνυσμα εισόδου $\mathbf{x}_n$, και $\sigma(a) = \frac{1}{1+\exp(-a)}$ η logistic sigmoid συνάρτηση.
(α) Να δείξετε ότι για ένα γραμμικώς διαχωρίσιμο σύνολο δεδομένων, η λύση μέγιστης πιθανοφάνειας για το μοντέλο logistic regression αντιστοιχεί στην εύρεση ενός διανύσματος $\mathbf{w}$, για το οποίο η επιφάνεια απόφασης $\mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x}) = 0$ διαχωρίζει τις κλάσεις, απειρίζοντας ταυτόχρονα το μέτρο του διανύσματος $\mathbf{w}$.
(β) Η Hessian μήτρα για το logistic regression δίνεται από τη σχέση:

$$\mathbf{H} = [\boldsymbol{\Phi}^\top \mathbf{R} \boldsymbol{\Phi}]$$

όπου $\boldsymbol{\Phi}$ είναι ο πίνακας των χαρακτηριστικών και $\mathbf{R}$ είναι ένας διαγώνιος πίνακας με στοιχεία $y_n(1 - y_n)$.
Να δείξετε ότι η Hessian μήτρα $\mathbf{H}$ είναι θετικώς ορισμένη. Ως εκ τούτου, δείξτε ότι η συνάρτηση σφάλματος είναι κυρτή συνάρτηση του $\mathbf{w}$ και ότι έχει μοναδικό ελάχιστο.
(γ) Να γράψετε κώδικα που θα υλοποιεί τον iterative reweighted least squares (IWLS) αλγόριθμο για logistic regression. Χρησιμοποιώντας τον αλγόριθμο αυτό, να υπολογίσετε και να σχεδιάσετε τα διαχωριστικά επίπεδα απόφασης που αντιστοιχούν στο σύνολο δεδομένων του προβλήματος τριών κλάσεων που έχει ανεβεί στο mycourses (αρχείο `MLR.data`). Οι δύο πρώτες στήλες περιλαμβάνουν τα διανύσματα χαρακτηριστικών, ενώ η τρίτη την κλάση. Συγκρίνετε τα αποτελέσματα με εκείνα που θα προέκυπταν εάν εφαρμοζόταν ταξινόμηση με βάση τα ελάχιστα τετράγωνα, σχεδιάζοντας τα αντίστοιχα διαχωριστικά επίπεδα απόφασης.
Σημείωση: Λεπτομέρειες από θεωρία του logistic regression και IWLS μπορούν να βρεθούν στις σχετικές διαφάνειες του μαθήματος και τις ενότητες 4.3.2, 4.3.3, 4.3.4 από [3].

ΒΙΒΛΙΟΓΡΑΦΙΑ:

[1] Γ. Καραγιάννης και Γ. Σταϊνχάουερ, *Αναγνώριση Προτύπων και Μάθηση Μηχανών*, ΕΜΠ, 2001.

[2] R. O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, Wiley, 2001.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[4] S. Theodoridis and K. Koutroumbas, Pattern Recognition, 4th Edition Academic Pres, Elsevier, 2009. *Ελληνική μετάφραση*: απόδοση-επιμέλεια-πρόλογος ελληνικής έκδοσης Α. Πικράκης, Κ. Κουτρουμπάς, Θ. Γιαννακόπουλος, Επιστημονικές Εκδόσεις Π.Χ. Πασχαλίδης-Broken Hill Publishers LTD, 2012.