
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 την λύση σας, πρέπει να το αναφέρετε. Παραδίδονται γραπτώς και προσωπικώς στην Γραμματεία Εργ. Ρομποτικής (Αιθ. 2.1.12, παλαιό Κτ.Ηλεκτρ.) 9.00-14.30.

Κεφάλαια για Ανάγνωση:
Βιβλίο [1]: I, II, III, IV.
Βιβλίο [2]: Κεφ. 1, 2, 3, 5.
Βιβλίο [3]: Κεφ. 1, 2, 9.
Βιβλίο [4]: Κεφ. 1, 2, 3.

Αναλυτικές Ασκήσεις

Ασκηση 1.1: Θεωρήστε ένα πρόβλημα ταξινόμησης d-διάστατων προτύπων $\boldsymbol{x} \in \mathbb{R}^d$ σε δύο τάξεις ω_1 και ω_2 με πρότερες πιθανότητες (priors) $P(\omega_1)$ και $P(\omega_2)$, αντίστοιχα, βάσει του κανόνα Bayes με ελάχιστη πιθανότητα λάθους:

$$\text{Decide } \omega_1 \text{ if } P(\omega_1|\boldsymbol{x}) > P(\omega_2|\boldsymbol{x}); \text{ otherwise decide } \omega_2$$

Θεωρήστε ότι οι πυκνότητες πιθανότητας για κάθε τάξη είναι d-διάστατες Gaussian

$$p(\boldsymbol{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}\sqrt{|\det(\boldsymbol{C})|}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_i)^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{m}_i)\right], \quad i = 1, 2$$

όπου $\boldsymbol{m}_i$ είναι τα μέσα διανύσματα και $\boldsymbol{C}$ ένας κοινός πίνακας συμμεταβλητότητας (covariance). Ο ανωτέρω ταξινομητής Bayes μπορεί να εκφρασθεί και βάσει της διαχωριστικής επιφάνειας

$$g(\boldsymbol{x}) = g_1(\boldsymbol{x}) - g_2(\boldsymbol{x}) = 0, \quad g_i(\boldsymbol{x}) = \ln P(\omega_i|\boldsymbol{x})$$

(a) Να βρείτε αναλυτικά την συνάρτηση της επιφάνειας $g(\boldsymbol{x}) = 0$ και να περιγράψετε συνοπτικά την γεωμετρία της.
(b) Να βρείτε την αναλυτική σχέση που πρέπει να ικανοποιούν οι δύο πρότερες πιθανότητες $P(\omega_1)$ και $P(\omega_2)$ ώστε η επιφάνεια $g(\boldsymbol{x}) = 0$ να μην τέμνει την ευθεία που ενώνει τα μέσα διανύσματα σε σημείο που κείται ενδιάμεσα των δύο μέσων $\boldsymbol{m}_1$,$\boldsymbol{m}_2$.

Ασκηση 1.2: Στις Ενότητες 2.3.1 και 2.3.2 του βιβλίου [3] παρουσιάστηκαν οι υπο-συνθήκη (conditional) και marginal κατανομές για μια πολυμεταβλητή γκαουσιανή κατανομή. Γενικότερα, μπορούμε να θεωρήσουμε τον διαχωρισμό των συνιστωσών του $\boldsymbol{x}$ σε τρεις ομάδες $\boldsymbol{x}_a$, $\boldsymbol{x}_b$ και $\boldsymbol{x}_c$ μαζί με έναν αντίστοιχο διαχωρισμό του διανύσματος μέσης τιμής $\boldsymbol{\mu}$ και της μήτρας συμμεταβλητότητας (covariance) $\boldsymbol{\Sigma}$ στη μορφή:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \\ \boldsymbol{\mu}_c \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} & \boldsymbol{\Sigma}_{ac} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} & \boldsymbol{\Sigma}_{bc} \\ \boldsymbol{\Sigma}_{ca} & \boldsymbol{\Sigma}_{cb} & \boldsymbol{\Sigma}_{cc} \end{pmatrix}$$

Χρησιμοποιώντας τα αποτελέσματα της Ενότητας 2.3 του [3], βρείτε μια έκφραση για την υποσυνθήκη κατανομή $p(\boldsymbol{x}_a|\boldsymbol{x}_b)$.

Ασκηση 1.3: Για το ακόλουθο πρόβλημα χρησιμοποιήστε το τελικό αποτέλεσμα του Προβλήματος 19 από το Κεφάλαιο 2 του βιβλίου [2], το οποίο επισυνάπτεται στο Παράρτημα Α.
(a) Υποθέστε πως γνωρίζομε μόνο ότι μια κατανομή είναι μη-μηδενική στο εύρος $x_l \leq x \leq x_u$. Αποδείξετε ότι η κατανομή μέγιστης εντροπίας είναι ομοιόμορφη σε αυτό το εύρος, δηλαδή

$$p(x) \sim U(x_l, x_u) = \begin{cases} 1/|x_u - x_l| & x_l \leq x \leq x_u \\ 0 & \text{αλλού} \end{cases}$$

(b) Υποθέστε πως γνωρίζομε μόνο ότι μια κατανομή είναι μη-μηδενική για $x \geq 0$ και ότι έχει μέση τιμή μ. Αποδείξετε ότι η κατανομή μέγιστης εντροπίας είναι:

$$p(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu} & \text{για } x \geq 0 \\ 0 & \text{αλλού} \end{cases}$$

(c) Τώρα υποθέστε πως γνωρίζομε μόνο ότι η κατανομή είναι κανονικοποιημένη, έχει μέση τιμή μ, τυπική απόκλιση σ, και επομένως με βάση το Πρόβλημα 19 η κατανομή μεγίστης εντροπίας πρέπει να έχει τη μορφή

$$p(x) = \exp\left[\lambda_0 - 1 + \lambda_1 x + \lambda_2 x^2\right]$$

Γράψετε τους τρεις περιορισμούς και λύστε ως προς τα λ_0, λ_1 και λ_2 και έτσι αποδείξετε ότι η λύση μεγίστης εντροπίας είναι μία γκαουσιανή, δηλαδή:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

Ασκηση 1.4: Στόχος αυτού του προβλήματος είναι η εύρεση του ταξινομητή Bayes για την περίπτωση της d-διάστατης πολυμεταβλητής κατανομής Bernoulli. Ως σύνηθες, εργαστείτε για κάθε τάξη ξεχωριστά, ερμηνεύοντας την $P(\boldsymbol{x}|D)$ ως $P(\boldsymbol{x}|D_i, \omega_i)$. Έστω ότι η υπό-συνθήκη πιθανότητα για μια δοθείσα κατηγορία δίνεται από την

$$P(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i}$$

και έστω $D = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ ένα σύνολο n δειγμάτων που έχουν ληφθεί ανεξάρτητα, με βάση αυτή την πυκνότητα πιθανότητας.
(a) Αν $\boldsymbol{s} = (s_1, ..., s_d)^T$ είναι το άθροισμα των n δειγμάτων, δείξετε ότι

$$P(D|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}$$

(b) Υποθέτοντας μία ομοιόμορφη a priori κατανομή για την $\boldsymbol{\theta}$ και χρησιμοποιώντας την ταυτότητα

$$\int_0^1 \theta^m (1-\theta)^n d\theta = \frac{m!n!}{(m+n+1)!}$$

δείξετε ότι

$$p(\boldsymbol{\theta}|D) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}$$

(c) Σχεδιάστε την πυκνότητα αυτή για την περίπτωση $d = 1, n = 1$, και για τις δύο προκύπτουσες πιθανότητες για την s_1. **(d)** Ολοκληρώστε το γινόμενο $P(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|D)$ ως προς $\boldsymbol{\theta}$ για να λάβετε την επιθυμητή υπό-συνθήκη πιθανότητα

$$P(\boldsymbol{x}|D) = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}$$

(e) Αν σκεφτούμε να βρούμε την $P(\boldsymbol{x}|D)$ αντικαθιστώντας μια εκτίμηση $\hat{\boldsymbol{\theta}}$ αντί για $\boldsymbol{\theta}$ στην $P(\boldsymbol{x}|\boldsymbol{\theta})$, ποια είναι η εκτίμηση Bayes για την $\boldsymbol{\theta}$;

Ασκηση 1.5: **(a)**. Θεωρήστε μία γραμμική μηχανή με συναρτήσεις διαχωρισμού $g_i(\boldsymbol{x}) = \boldsymbol{w}_i^T\boldsymbol{x} + w_{i0}, i = 1, ..., c$. Δείξετε πως οι περιοχές απόφασης R_i είναι κυρτές, δείχνοντας ότι αν $\boldsymbol{x}_1 \in R_i$ και $\boldsymbol{x}_2 \in R_i$ τότε $\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in R_i$ αν $0 \leq \lambda \leq 1$.
(b). Η κυρτή θήκη (convex hull) ενός συνόλου διανυσμάτων $\boldsymbol{x}_i$, $i = 1, ..., n$, είναι το σύνολο όλων των διανυσμάτων της μορφής

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i$$

όπου οι συντελεστές α_i είναι μη-αρνητικοί και έχουν άθροισμα 1. Δοθέντων δύο συνόλων διανυσμάτων, δείξετε ότι είτε είναι γραμμικώς διαχωρίσιμα, είτε οι convex hulls τους τέμνονται. (Υπόδειξη: Υποθέστε ότι και οι δύο προτάσεις ισχύουν, και θεωρήστε την ταξινόμηση ενός σημείου στην τομή των convex hulls).

Ασκηση 1.6: Βασιζόμενοι στην ενότητα 5.8.4 του κεφαλαίου 5 του βιβλίου Duda, Hart & Stork (2001), αποδείξετε ότι το $\boldsymbol{a}$ που ελαχιστοποιεί το κόστος

$$J_s(\boldsymbol{a}) = ||\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{b}||^2 \tag{1}$$

προκύπτει από την παρακάτω επαναληπτική μέθοδο gradient descent:

$$\boldsymbol{a}(1) \qquad arbitrary$$

$$\boldsymbol{a}_{k+1} = \boldsymbol{a}_k - \eta_k \boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{a}_k - \boldsymbol{b}) \tag{2}$$

Συγκεκριμένα, εάν χρησιμοποιήσομε βήματα $\eta_k = 1/k$, αποδείξετε οτι η παραπάνω επαναληπτική μέθοδος στο όριο ($\lim_{k\to\infty} \boldsymbol{a}_k \to \boldsymbol{a}$) δίνει διάνυσμα $\boldsymbol{a}$ το οποίο ικανοποιεί την παρακάτω εξίσωση :

$$\boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{b}) = 0 \tag{3}$$

Οπότε, ο ανωτέρω αλγόριθμος δίνει λύση ανεξάρτητα από το εάν ο πίνακας $\boldsymbol{Y}^T\boldsymbol{Y}$ έχει μηδενική ορίζουσα.
Σημείωση: Γνωρίζομε ότι αν ο πίνακας $\boldsymbol{Y}^T\boldsymbol{Y}$ δεν έχει μηδενική ορίζουσα, τότε το διάνυσμα που ελαχιστοποιεί το κριτήριο J_s προκύπτει από τον ψευδο-αντίστροφο: $\boldsymbol{a} = \boldsymbol{Y}^\dagger\boldsymbol{b}$, $\boldsymbol{Y}^\dagger = (\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T$.

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.

Παράρτημα Α

Κεφάλαιο 2, Άσκηση 19 από [2]. (Για την άσκηση 1.3)

19. Starting from the definition of entropy (Eq. 37), derive the general equation for the maximum-entropy distribution given constraints expressed in the general form

$$\int b_k(x)p(x)\,dx = a_k, \quad k = 1, 2, ..., q$$

as follows:

(a) Use Lagrange undetermined multipliers $\lambda_1, \lambda_2, ..., \lambda_q$ and derive the synthetic function:

$$H_s = -\int p(x)\left[\ln p(x) - \sum_{k=0}^{q} \lambda_k b_k(x)\right]\,dx - \sum_{k=0}^{q} \lambda_k a_k.$$

State why we know $a_0 = 1$ and $b_0(x) = 1$ for all x.

(b) Take the derivative of H_s with respect to $p(x)$. Equate the integrand to zero, and thereby prove that the maximum entropy distribution obeys

$$p(x) = \exp\left[\sum_{k=0}^{q} \lambda_k b_k(x) - 1\right],$$

where the $q+1$ parameters are determined by the constraint equation.