
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 τη λύση σας, πρέπει να το αναφέρετε. Η παράδοση των λύσεων των αναλυτικών ασκήσεων της σειράς αυτής θα γίνει ηλεκτρονικά στο helios και θα πρέπει να την υποβάλετε ως ένα ενιαίο αρχείο PDF με το εξής filename format, χρησιμοποιώντας μόνο λατινικούς χαρακτήρες: pr22_hwk1_AM_FirstnameLastname.pdf, όπου AM είναι ο 8-ψήφιος αριθμός μητρώου σας. Σκαναρισμένες χειρόγραφες λύσεις επιτρέπονται, αρκεί να είναι καθαρογραμμένες και ευανάγνωστες. Επίσης, στην πρώτη σελίδα των λύσεων θα αναγράφετε το ονοματεπώνυμο, AM, και email address σας.

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

Αναλυτικές Ασκήσεις

Άσκηση 1.1: (Maximum Likelihood Estimation)
Έστω ότι η τυχαία διακριτή μεταβλητή X ακολουθεί την κατανομή Poisson:

$$X \sim \frac{\theta^x}{x!} e^{-\theta}, \quad x \in \mathbb{Z}_+ \cup \{0\}$$

(α) Να σχεδιάσετε την σ.π.π. $P(x|\theta)$ συναρτήσει του x για $\theta = 2$ και για $\theta = 10$.
(β) Εάν υποθέσετε ότι γνωρίζετε N ανεξάρτητα δείγματα $\{x_1, x_2, \dots, x_N\}$ τα οποία προέρχονται από μία κατανομή Poisson, να βρείτε τον εκτιμητή μέγιστης πιθανοφάνειας για το θ

Άσκηση 1.2: (Gradient Descent for MSE Perceptron Training)
Βασιζόμενοι στην Ενότητα 5.8.4 του βιβλίου [2], να αποδείξετε ότι το $\mathbf{a}$ που ελαχιστοποιεί το κόστος εκπαίδευσης ενός Perceptron με τη μέθοδο ελαχίστων τετραγώνων

$$J_s(\mathbf{a}) = ||\mathbf{Y}\mathbf{a} - \mathbf{b}||^2$$

προκύπτει από την παρακάτω επαναληπτική μέθοδο gradient descent:

$$\mathbf{a}(1) \qquad \textit{arbitrary}$$

$$\mathbf{a}_{k+1} = \mathbf{a}_k - \eta_k \mathbf{Y}^\top (\mathbf{Y}\mathbf{a}_k - \mathbf{b})$$

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

$$\mathbf{Y}^\top (\mathbf{Y}\mathbf{a} - \mathbf{b}) = 0$$

Οπότε, ο ανωτέρω αλγόριθμος δίνει λύση ανεξάρτητα από το εάν ο πίνακας $\mathbf{Y}^\top\mathbf{Y}$ έχει μηδενική ορίζουσα.
Σημείωση: Γνωρίζουμε ότι αν ο πίνακας $\mathbf{Y}^\top\mathbf{Y}$ δεν έχει μηδενική ορίζουσα, τότε το διάνυσμα που

ελαχιστοποιεί το κριτήριο J_s προκύπτει από τον ψευδο-αντίστροφο: $\mathbf{a} = \mathbf{Y}^{\dagger}\mathbf{b}$, $\mathbf{Y}^{\dagger} = (\mathbf{Y}^{\top}\mathbf{Y})^{-1}\mathbf{Y}^{\top}$.

***Άσκηση* 1.3**: (EM on GMMs)
Θεωρήστε δύο Γκαουσιανές συναρτήσεις πυκνότητας πιθανότητας $\mathcal{N}(1.5, 0.1)$ και $\mathcal{N}(2.5, 0.2)$. Δημιουργήστε 15 δείγματα (με μία γεννήτρια τυχαίων αριθμών) σύμφωνα με τον εξής κανόνα: τα πρώτα δύο δείγματα να προέρχονται από την πρώτη Γκαουσιανή και το τρίτο δείγμα από τη δεύτερη. Ο κανόνας αυτός επαναλαμβάνεται μέχρις ότου δημιουργηθούν και τα 15 δείγματα. Η υποκείμενη συνάρτηση πυκνότητας πιθανότητας των τυχαίων δειγμάτων μπορεί να μοντελοποιηθεί ως ένα μείγμα Γκαουσιανών:

$$\sum_{i=1}^{2} \mathcal{N}(\mu_i, \sigma_i^2) P_i$$

Να χρησιμοποιήσετε τον αλγόριθμο Expectation-Maximization (EM) και τα παραχθέντα δείγματα προκειμένου να εκτιμήσετε τις άγνωστες παραμέτρους μ_i, σ_i^2, P_i. Να εξηγήσετε οποιεσδήποτε παραδοχές χρειαστεί να κάνετε. Τέλος, να δώσετε έναν σύντομο σχολιασμό για τα αριθμητικά αποτελέσματα που προκύπτουν.
Σημείωση: Για να λύσετε την άσκηση να πραγματοποιήσετε όλες τις πράξεις που απαιτούνται *χειροκίνητα*. Δεν ζητείται προγραμματιστική υλοποίηση.

***Άσκηση* 1.4:** (Bayes Classifier for multivariate Bernoulli) Αυτή η άσκηση μελετά το πρόβλημα της εύρεσης του ταξινομητή Bayes για την περίπτωση τηε d-διάστατης πολυμεταβλητής κατανομής Bernoulli. Θεωρήστε κάθε κλάση ανεξάρτητα, ερμηνεύοντας δηλαδή την $P(\mathbf{x}|D)$ ως $P(\mathbf{x}|D_i, \omega_i)$. Έστω το ότι η υποσυνθήκη πιθανότητα για μια δοθείσα κατηγορία δίνεται από την

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

και έστω $D = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ ένα σύνολο n δειγμάτων που έχουν ληφθεί ανεξάρτητα, με βάση αυτή την πυκνότητα πιθανότητας.

(a) Αν $\mathbf{s} = (s_1, \cdots, s_d)^T$ είναι το άθροισμα των n δειγμάτων, δείξτε ότι

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

(b) Υποθέτωντας μια ομοιόμορφη (uniform) apriori κατανομή για τη μεταβλητή $\boldsymbol{\theta}$ και χρησιμοποιώντας την ταυτότητα

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

να 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 $P(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|D)$ ως προς θ για να λάβετε την επιθυμητή υπό-συνθήκη πιθανότητα

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

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

***Άσκηση* 1.5:** (Bayesian parameter estimation)
Έστω ότι η κατανομή $p(\mathbf{x}|\boldsymbol{\Sigma})$ ακολουθεί την κανονική κατανομή $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, όπου το διάνυσμα μέσης τιμής $\boldsymbol{\mu}$ θεωρείται γνωστό και ο πίνακας συνδιασποράς $\boldsymbol{\Sigma}$) άγνωστος. Εάν έχουμε n δεδομένα $x_1, \cdots, x_n$, να δείξετε ότι ο εκτιμητής μέγιστης πιθανοφάνειας του $mtahbf\Sigma$ δίνεται από τη σχέση:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^T$$

Για να το αποδείηετε ακολουθήστε τα παρακάτω βήματα:

(**a**) Αποδείξτε την ιδιότητα $\mathbf{a}^T\mathbf{A}\mathbf{a} = trace(\mathbf{A}\mathbf{a}\mathbf{a}^T)$, όπου το ίχνος (trace) είναι το άθροισμα των διαγώνιων στοιχείων του πίνακα $\mathbf{A}$.

(**b**) Δείξτε οτι η συνάρτηση πιθανοφάνειας μπορεί να γραφεί στη μορφή:

$$p(\mathbf{x}_1, \cdots, \mathbf{x}_n|\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{nd/2}|\hat{\boldsymbol{\Sigma}}|^{n/2}} \exp\Big(- \frac{1}{2}trace\big(\boldsymbol{\Sigma}^{-1}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^T\big)\Big)$$

(**c**) Έστω ότι $\mathbf{A} = \boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}$ και $\lambda_1, \cdots, \lambda_n$ οι ιδιοτιμές του $\mathbf{A}$. Δείξτε ότι το παραπάνω αποτέλεσμα οδηγεί στη σχέση

$$p(\mathbf{x}_1, \cdots, \mathbf{x}_n|\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{nd/2}|\hat{\boldsymbol{\Sigma}}|^{n/2}}(\lambda_1 \cdots \lambda_d)^{n/2} \exp\big(- \frac{n}{2}\sum_{i=1}^{d}\lambda_i\big)$$

(**d**) ΟΛοκληρώστε την απόδειξη δείχνοντας ότι η πιθανοφάνεια μεγιστοποιείται με την επιλογή $\lambda_1 = \cdots \lambda_d = 1$. Εξηγήστε αναλυτικά.

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.