Λύσεις 1ου Σύν. Αν. Ασκήσεων

Οι ακόλουθες λύσεις είναι συνοπτικές. Οι κανονικές λύσεις που παραδίδονται θα πρέπει να είναι πιο λεπτομερείς.

Ασκηση 1.1: a) Η διαχωριστική επιφάνεια είναι:

$$(\vec{\mu}_1 - \vec{\mu}_2)^T C^{-1}(\vec{x} - \vec{x}_0) = 0$$

όπου

$$\vec{x}_0 = \frac{1}{2}(\vec{\mu}_1 + \vec{\mu}_2) - \frac{\ln P(\omega_1)/\ln P(\omega_2)}{(\vec{\mu}_1 - \vec{\mu}_2)^T C^{-1}(\vec{\mu}_1 - \vec{\mu}_2)}(\vec{\mu}_1 - \vec{\mu}_2)$$

Η διαχωριστική επιφάνεια είναι υπερεπίπεδο κάθετο στη διεύθυνση $(\vec{\mu}_1 - \vec{\mu}_2)^T C^{-1}$ και διέρχεται από το σημείο $\vec{x}_0$ το οποίο βρίσκεται στην ευθεία που ενώνει τα σημεία $\vec{\mu}_1$ και $\vec{\mu}_2$.
b) Θα πρέπει να ισχύει:

$$P(\omega_1) < P(\omega_2)\exp\left[-\frac{1}{2}(\vec{\mu}_1 - \vec{\mu}_2)^T C^{-1}(\vec{\mu}_1 - \vec{\mu}_2)\right]$$

Ασκηση 1.2: Περιθωριοποιώντας ως προς το $\boldsymbol{x}_c$ παίρνουμε τη νέα κατανομή:

$$\boldsymbol{\mu}_{ab} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$$

Επομένως, οι παράμετροι της κατανομής $p(\boldsymbol{x}_a|\boldsymbol{x}_b)$ είναι:

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b)$$
$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}$$

Ασκηση 1.3: a) Ο περιορισμός εδώ είναι $\int_{x_u}^{x_l} p(x)dx = 1$ και με $p(x) = e^{\lambda_0 - 1}$ παίρνουμε

$$p(x) = \begin{cases} \frac{1}{|x_u - x_l|} & x_l \le x \le x_u \\ 0 & \text{αλλού} \end{cases}$$

b) Εδώ, οι δύο περιορισμοί που έχουμε είναι: $\int_0^\infty p(x)dx = 1$ και $\int_0^\infty xp(x)dx = \mu$. Επίσης έχουμε $p(x) = e^{\lambda_0 - 1 + \lambda_1 x}$. Λύνοντας ως προς τα λ_0 και λ_1 παίρνουμε

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

c) Εδώ, έχουμε τους τρεις περιορισμούς: $\int_{-\infty}^\infty p(x)dx = 1$, $\int_{-\infty}^\infty xp(x)dx = \mu$ και $\int_{-\infty}^\infty x^2 p(x)dx = \sigma^2$. Επίσης, $p(x) = e^{\lambda_0 - 1 + \lambda_1 x - \lambda_2 x^2}$. Λύνοντας ως προς τα λ_0, λ_1 και λ_2 παίρνουμε:

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

Ασκηση 1.4: a) Έχουμε

$$P(D|\boldsymbol{\theta}) = P(\boldsymbol{x}_1, ..., \boldsymbol{x}_n|\boldsymbol{\theta}) = \prod_{k=1}^{n} P(\boldsymbol{x}_k|\boldsymbol{\theta})$$

αφού τα $\boldsymbol{x}_k$ είναι ανεξάρτητα. Μετά από πράξεις παίρνουμε το τελικό αποτέλεσμα:

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

b) Ξέρουμε τον κανόνα Bayes:

$$p(\boldsymbol{\theta}|D) = \frac{p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(D)}$$

όπου

$$p(D) = \int P(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$$

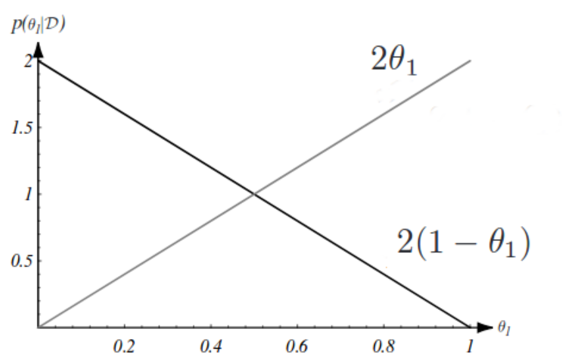
Επίσης έχουμε

$$p(\theta_i) = \begin{cases} 1 & 0 \le \theta_i \le 1 \\ 0 & \text{αλλού} \end{cases}$$

Χρησιμοποιώντας το αποτέλεσμα του μέρους a), βρίσκουμε το $p(D)$ και στη συνέχεια από κανόνα Bayes:

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

c)



e)

$$\hat{\theta}_i = \frac{s_i}{n}$$

Ασκηση 1.5: a) Αν τα $\boldsymbol{x}_1$ και $\boldsymbol{x}_2$ ανήκουν στο ω_i έχουμε:

$$\max_j g_j(\boldsymbol{x}_1) = g_i(\boldsymbol{x}_1) = \boldsymbol{w}_i^t\boldsymbol{x}_1 + w_{i0}$$

και

$$\max_j g_j(\boldsymbol{x}_0) = g_i(\boldsymbol{x}_0) = \boldsymbol{w}_i^t\boldsymbol{x}_0 + w_{i0}$$

Για $0 \le \lambda \le 1$ και για κάθε j έχουμε επομένως:

$$\boldsymbol{w}_j^t\left[\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_0\right]+w_{j0}=\lambda\left[\boldsymbol{w}_j^t\boldsymbol{x}_1+w_{j0}\right]+(1-\lambda)\left[\boldsymbol{w}_j^t\boldsymbol{x}_0+w_{j0}\right]$$
$$\leq\lambda\left[\boldsymbol{w}_i^t\boldsymbol{x}_1+w_{i0}\right]+(1-\lambda)\left[\boldsymbol{w}_i^t\boldsymbol{x}_0+w_{i0}\right]$$

Άρα τώρα βρίσκοντας την μέγιστη συνάρτηση ταξινόμησης:

$$\max_j\left[\boldsymbol{w}_j^t\left[\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_0\right]+w_{j0}\right]=\boldsymbol{w}_i^t\left[\lambda\boldsymbol{x}_1+(1-\lambda)\boldsymbol{x}_0\right]+w_{i0}$$

που μας δείχνει ότι και οποιοδήποτε σημείο μεταξύ των $\boldsymbol{x}_0$ και $\boldsymbol{x}_1$ ανήκει στο ω_i. Επομένως, η περιοχή ω_i είναι κυρτή και το ίδιο ισχύει για κάθε κατηγορία ω_j (αρκεί κατά την απόδειξη μας η κατηγορία να είναι η ω_j).
b) Έστω δύο σύνολα διανυσμάτων $S_1=\boldsymbol{x}_1,\boldsymbol{x}_2,...\boldsymbol{x}_n$ και $S_2=\boldsymbol{y}_1,\boldsymbol{y}_2,...\boldsymbol{y}_n$. Αν υποθέσουμε ότι είναι γραμμικώς διαχωρίσιμα τότε υπάρχει γραμμική συνάρτηση διαχωρισμού $g(\boldsymbol{x})$ έτσι ώστε:

$$g(\boldsymbol{x})>0\Rightarrow\boldsymbol{x}\in S_1$$
$$g(\boldsymbol{x})<0\Rightarrow\boldsymbol{x}\in S_2$$

Αν παράλληλα υποθέσουμε ότι οι κυρτές θήκες των δύο συνόλων τέμνονται, υπάρχει σημείο $\boldsymbol{x}$ που ανήκει στην κυρτή θήκη και των δύο συνόλων. Αν υπολογίσουμε την συνάρτηση διαχωρισμού $g(\boldsymbol{x})$ για τις δύο διαφορετικές κυρτές θήκες θα βρούμε μετά από πράξεις ότι $g(\boldsymbol{x})>0$ και $g(\boldsymbol{x})<0$ που είναι άτοπο. Άρα είτε οι κυρτές θήκες δύο συνόλων διανυσμάτων θα τέμνονται είτε θα είναι γραμμικά διαχωρίσιμες.

Ασκηση 1.6: Βρίσκουμε την παράγωγο του κόστους ως προς α:

$$\nabla J_s=\nabla||\boldsymbol{Y}\boldsymbol{\alpha}-\boldsymbol{b}||^2=2\boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{\alpha}-\boldsymbol{b})$$

Άρα με βάση την μέθοδο gradient descent, και για $\eta_k=1/k$ έχουμε:

$$\boldsymbol{\alpha}_{k+1}=\boldsymbol{\alpha}_k-\frac{1}{k}\boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{\alpha}-\boldsymbol{b}) \tag{1}$$

Ομοίως μπορούμε να πάρουμε:

$$\boldsymbol{\alpha}_k=\boldsymbol{\alpha}_{k-1}-\frac{1}{k-1}\boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{\alpha}-\boldsymbol{b}) \tag{2}$$

Αντικαθιστώντας παίρνουμε:

$$\boldsymbol{\alpha}_{k+1}=\boldsymbol{\alpha}_{k-1}+\left[\frac{1}{k-1}\boldsymbol{Y}^T\boldsymbol{Y}+\frac{1}{k}\boldsymbol{Y}^T\boldsymbol{Y}+\frac{1}{k(k-1)}(\boldsymbol{Y}^T\boldsymbol{Y})^2\right]\boldsymbol{\alpha}_{k-1}$$
$$-\left[\frac{1}{k-1}\boldsymbol{Y}^T+\frac{1}{k(k-1)}\boldsymbol{Y}^T\boldsymbol{Y}\boldsymbol{Y}^T\right]\boldsymbol{b}$$

όπου μπορούμε να δούμε εύκολα πως για $k\to\infty$ οι συντελεστές των $\boldsymbol{\alpha}_{k-1}$ και $\boldsymbol{b}$ μηδενίζονται και επομένως η μέθοδος συγκλίνει σε διάνυσμα $\boldsymbol{\alpha}$ για το οποίο ισχύει:

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