Μηχανική Μάθηση

1η Σειρά Ασκήσεων

Έτος 2023-2024

Χαρίλαος Κουκουλάρης
03118137
Νοέμβριος-Δεκέμβριος 2023

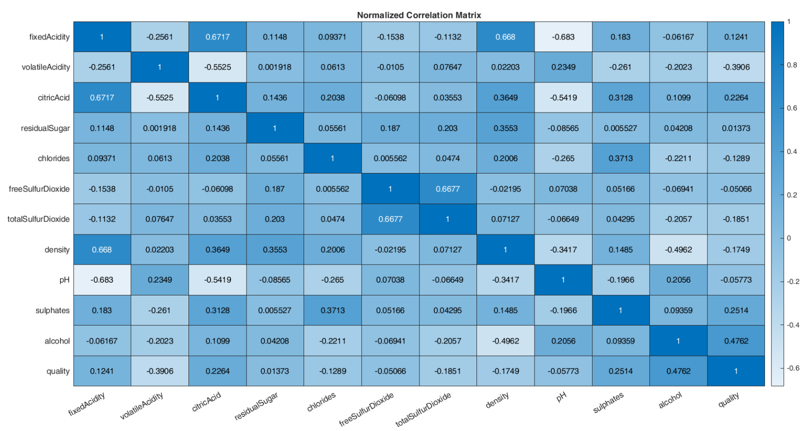
Άσκηση 1.1

Ερώτημα α.

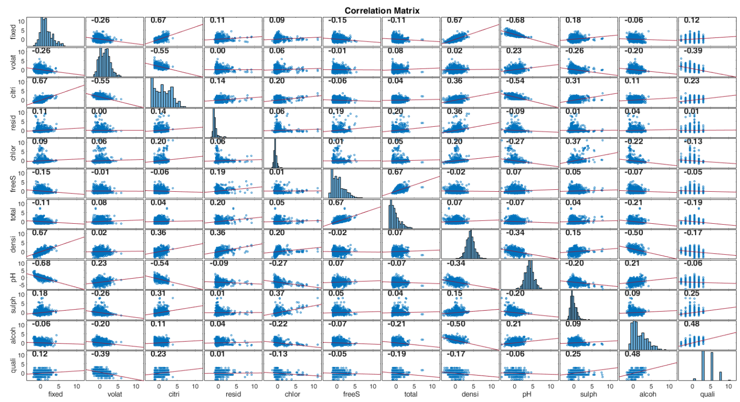
Ο κανονικοποιημένος συντελεστής συσχέτισης $r_{(9,10)}$ μεταξύ της ένατης (pH) και δέκατης (sulphates) ανεξάρτητης μεταβλητής (στήλης) είναι ίσος με:

$$r_{(9,10)} = -0.1966$$

Ακολουθούν πίνακες συσχέτισης μεταξύ των διαφόρων χαρακτηριστικών.



Σχήμα 1: Κανονικοποιημένος Πίνακας Συσχέτισης



Σχήμα 2: Πίνακας Συσχέτισης με Γραφήματα Διασποράς

Ερώτημα β.

Οι εξισώσεις της Γραμμικής Παλινδρόμησης (Linear Regression) σε μορφή πινάκων δίνουν τα βάρη ως εξής:

$$w^* = X^+ y = (X^T X)^{-1} X^T y$$

$$w^* = \begin{bmatrix} 0.2026 \\ -0.4274 \\ -0.3318 \\ -0.0067 \\ -0.0082 \\ 0.2699 \\ -0.3115 \\ -0.0863 \\ -0.1934 \\ 0.0199 \\ 0.4241 \end{bmatrix}$$

Ερώτημα γ.

Οι εξισώσεις της Παλινδρόμησης Ridge (Ridge Regression) σε μορφή πινάκων δίνουν τα βάρη ως εξής:

$$w^* = (X^T X + \lambda I)^{-1} X^T y$$

Παλινδρόμηση Ridge με $\lambda = 10$

$$w^* = \begin{bmatrix} 0.1810 \\ -0.3287 \\ -0.2114 \\ 0.0032 \\ -0.0201 \\ 0.2002 \\ -0.2654 \\ -0.0712 \\ -0.1640 \\ 0.0149 \\ 0.3576 \end{bmatrix}$$

Παλινδρόμηση Ridge με $\lambda = 100$

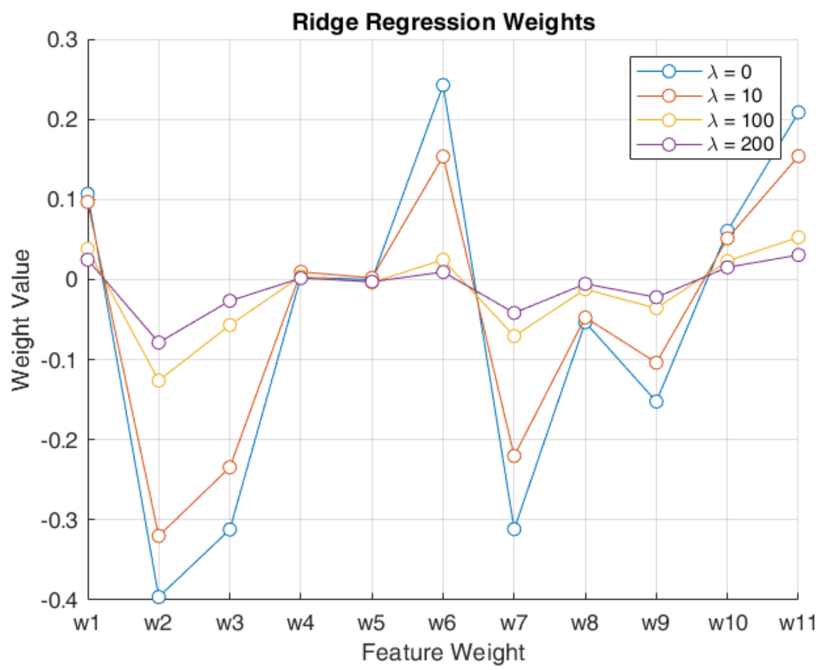
$$w^* = \begin{bmatrix} 0.1066 \\ -0.1352 \\ -0.0126 \\ 0.0097 \\ -0.0310 \\ 0.0447 \\ -0.1228 \\ -0.0177 \\ -0.0754 \\ -0.0001 \\ 0.1682 \end{bmatrix}$$

Παλινδρόμηση Ridge με $\lambda = 200$

$$w^* = \begin{bmatrix} 0.0747 \\ -0.0896 \\ 0.0083 \\ 0.0074 \\ -0.0266 \\ 0.0190 \\ -0.0808 \\ -0.0076 \\ -0.0492 \\ -0.0028 \\ 0.1095 \end{bmatrix}$$

Ερώτημα δ.

Τα διανύσματα των βαρών που προέκυψαν στις παραπάνω περιπτώσεις φαίνονται από κοινού στο επόμενο γράφημα.



Σχήμα 3: Βάρη χαρακτηριστικών συναρτήσει της υπερπαραμέτρου λ

Από το γράφημα φαίνεται πως αυξάνοντας την τιμή της παραμέτρου λ, τα βάρη καταλήγουν να έχουν μικρότερη απόλυτη τιμή. Αυτό είναι αναμενόμενο καθώς η παράμετρος λ καθορίζει το μέγεθος της συνεισφοράς του μέτρου του διανύσματος των βαρών στη συνάρτηση κόστους

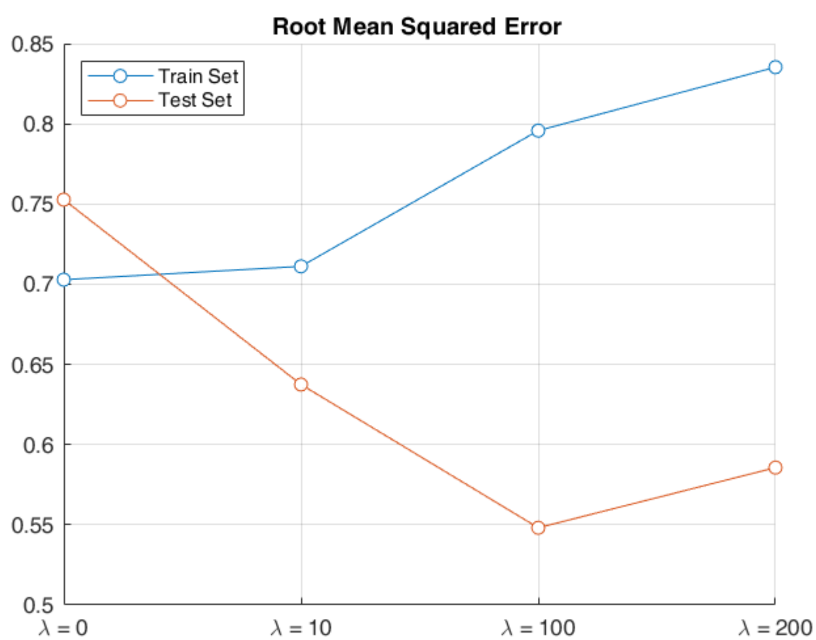
$$L(\boldsymbol{w}) = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} + \lambda \boldsymbol{w}^T \boldsymbol{w}$$

Ερώτημα ε.

	Linear Regression ($\lambda = 0$)	Ridge Regression ($\lambda = 10$)	Ridge Regression ($\lambda = 100$)	Ridge Regression ($\lambda = 200$)
RMSE (Training Set)	0.70281	0.71112	0.79577	0.83533
RMSE (Test Set)	0.75272	0.63745	0.54812	0.58565

Πίνακας 1: Μέσα τετραγωνικά σφάλματα σετ εκπαίδευσης και επαλήθευσης για διαφορετικές τιμές της υπερπαραμέτρου λ

Η βέλτιστη τιμή, βάσει των δεδομένων του πίνακα, φαίνεται να είναι η $\lambda = 100$ καθώς ελαχιστοποιεί τη ρίζα του μέσου τετραγωνικού σφάλματος τόσο του συνόλου εκπαίδευσης όσο και του συνόλου επαλήθευσης.



Σχήμα 4: Μέσο τετραγωνικό σφάλμα εκπαίδευσης και επαλήθευσης για διάφορες τιμές του λ

Άσκηση 1.2

Ερώτημα α.

Πολυμεταβλητή Κανονική Κατανομή δύο μεταβλητών $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$:

$$p(\boldsymbol{x}) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} exp\Bigg(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Bigg)$$

με μέση τιμή και συνδιασπορά: $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}^T, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$

Η ορίζουσα και ο αντίστροφος του πίνακα συνδιασποράς υπολογίζονται παρακάτω:

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \sigma_{12}^2$$

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{21} & \sigma_1^2 \end{bmatrix}$$

Θεωρώντας το x_2 σταθερό και το x_1 ως τη μόνη μεταβλητή, ο εκθέτης μπορεί να γραφεί σε πιο βολική μορφή.

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix} \frac{1}{|\Sigma|}\begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{21} & \sigma_1^2 \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} =$$
$$= \frac{1}{|\Sigma|}\Bigg((x_1-\mu_1)^2\sigma_2^2 - 2(x_1-\mu_1)(x_2-\mu_2)\sigma_{12} + (x_2-\mu_2)^2\sigma_1^2 \Bigg) =$$
$$= \frac{1}{|\Sigma|}\Bigg((x_1^2 - 2x_1\mu_1 + \mu_1^2)\sigma_2^2 - 2x_1(x_2-\mu_2)\sigma_{12} + 2\mu_1(x_2-\mu_2)\sigma_{12} + (x_2-\mu_2)^2\sigma_1^2 \Bigg) =$$
$$= \frac{1}{|\Sigma|}\Bigg((x_1^2 - 2x_1\mu_1)\sigma_2^2 - 2x_1(x_2-\mu_2)\sigma_{12} + C(x_2) \Bigg) =$$
$$= \frac{\sigma_2^2}{|\Sigma|}\Bigg[x_1^2 - 2x_1\Bigg(\mu_1 + (x_2-\mu_2)\frac{\sigma_{12}}{\sigma_2^2}\Bigg) + \Bigg(\mu_1 + (x_2-\mu_2)\frac{\sigma_{12}}{\sigma_2^2}\Bigg)^2 + C'(x_2) \Bigg] =$$
$$= \frac{\sigma_2^2}{|\Sigma|}\Bigg[x_1 - \Bigg(\mu_1 + (x_2-\mu_2)\frac{\sigma_{12}}{\sigma_2^2}\Bigg) \Bigg]^2 + C''(x_2)$$

Όπου,

$$C(x_2) = \mu_1^2\sigma_2^2 + 2\mu_1(x_2-\mu_2)\sigma_{12} + (x_2-\mu_2)^2\sigma_1^2$$

$$\begin{aligned}
C'(x_2) &= \mu_1^2 + \frac{1}{\sigma_2^2}\Bigg((x_2-\mu_2)^2\sigma_1^2 + 2\mu_1(x_2-\mu_2)\sigma_{12}\Bigg) - \Bigg(\mu_1 + (x_2-\mu_2)\frac{\sigma_{12}}{\sigma_2^2}\Bigg)^2 \\
&= \mu_1^2 + (x_2-\mu_2)^2\frac{\sigma_1^2}{\sigma_2^2} + 2\mu_1(x_2-\mu_2)\frac{\sigma_{12}}{\sigma_2^2} - \mu_1^2 - 2\mu_1(x_2-\mu_2)\frac{\sigma_{12}}{\sigma_2^2} - (x_2-\mu_2)^2\Bigg(\frac{\sigma_{12}}{\sigma_2^2}\Bigg)^2 \\
&= (x_2-\mu_2)^2\frac{\sigma_1^2}{\sigma_2^2} - (x_2-\mu_2)^2\Bigg(\frac{\sigma_{12}}{\sigma_2^2}\Bigg)^2 \\
&= \frac{(x_2-\mu_2)^2}{\sigma_2^2}\Bigg(\sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\Bigg)
\end{aligned}$$

$$\begin{aligned}
C''(x_2) &= \frac{\sigma_2^2}{|\Sigma|}C'(x_2) \\
&= \frac{1}{|\Sigma|}(x_2-\mu_2)^2\Bigg(\sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\Bigg)
\end{aligned}$$

Θέτοντας:

$$\mu = \mu_1 + \frac{\sigma_{12}(x_2-\mu_2)}{\sigma_2^2}$$

και

$$\sigma^2 = \frac{|\Sigma|}{\sigma_2^2} = \frac{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}{\sigma_2^2}$$

και αντικαθιστώντας τη μετασχηματισμένη μορφή του εκθέτη στον τύπο της κατανομής προκύπτει:

$$\begin{aligned}
p(x_1,x_2) &= \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}exp\Bigg(-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2} - \frac{1}{2}C''(x_2)\Bigg) \\
&= \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}exp\Bigg(-\frac{1}{2}C''(x_2)\Bigg)exp\Bigg(-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2}\Bigg) =
\end{aligned}$$

Από αυτό συμπεραίνεται πως η υπό συνθήκη κατανομή $p(x_1|x_2=\alpha)$ είναι Γκαουσιανή και έχει τις ακόλουθες παραμέτρους:

$$\mu = \mu_1 + \frac{\sigma_{12}(\alpha-\mu_2)}{\sigma_2^2}$$

και

$$\sigma^2 = \frac{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}{\sigma_2^2}$$

Πολυμεταβλητή κανονική κατανομή μεταβλητών $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ και μέσης τιμής και συνδιασποράς:

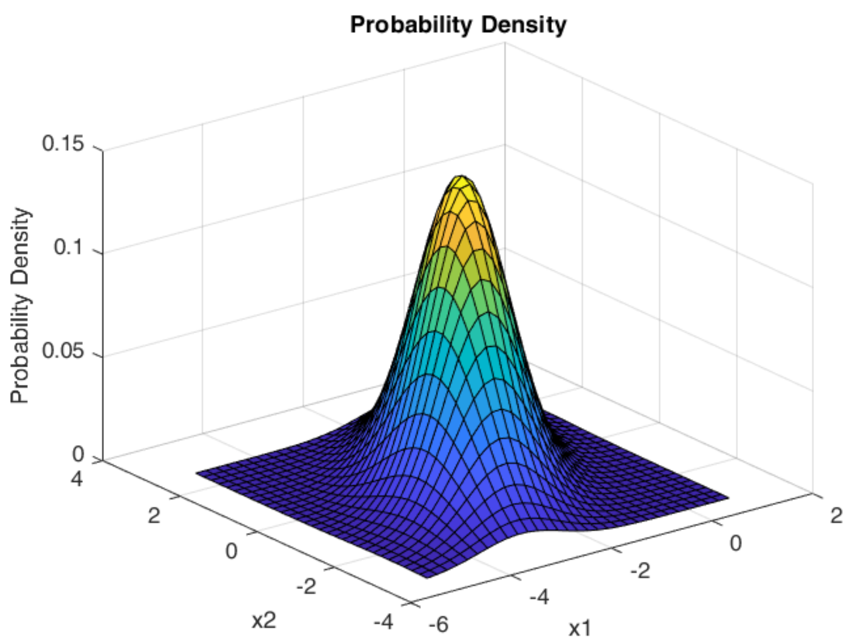
$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 2 & 1 \\ 0.5 & 1 & 3 \end{bmatrix}$$

Για τα επόμενα ερωτήματα χρησιμοποιούνται οι τύποι της σελίδας 106 του βιβλίου "Αναγνώριση Προτύπων και Μηχανική Μάθηση" του Christopher M. Bishop:

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

Ερώτημα β.

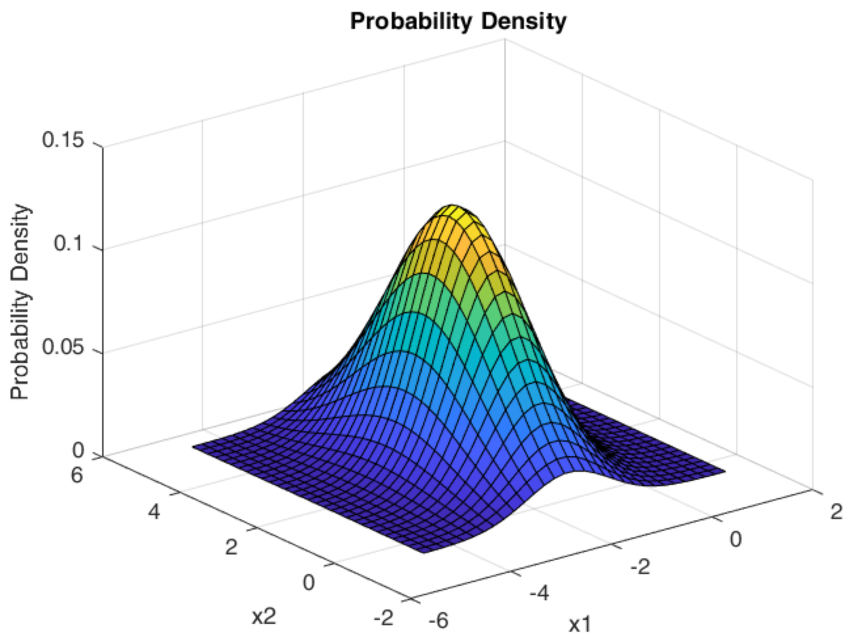
$$\boldsymbol{\mu}_{a|b} = \begin{bmatrix} -2.1667 \\ -0.3333 \end{bmatrix}, \quad \Sigma_{a|b} = \begin{bmatrix} 0.9167 & 0.6333 \\ 0.6333 & 1.6667 \end{bmatrix}$$



Σχήμα 5: Κατανομή Υπο Συνθήκη Συνάρτησης Πυκνότητας Πιθανότητας $p(x_1, x_2|x_3 = 1)$

Ερώτημα γ.

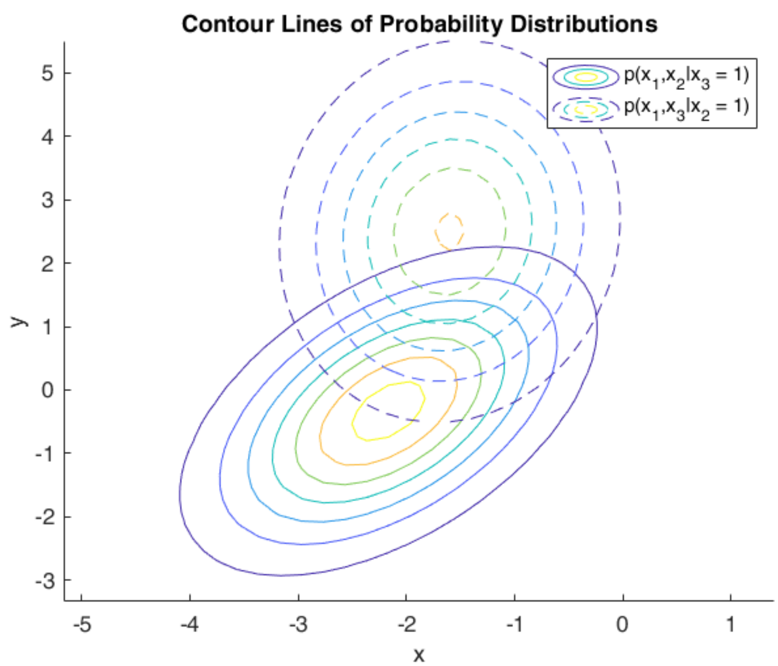
$$\boldsymbol{\mu}_{a|b} = \begin{bmatrix} -1.6 \\ 2.5 \end{bmatrix}, \quad \Sigma_{a|b} = \begin{bmatrix} 0.68 & 0.1 \\ 0.1 & 2.5 \end{bmatrix}$$



Σχήμα 6: Κατανομή Υπο Συνθήκη Συνάρτηση Πυκνότητας Πιθανότητας $p(x_1, x_3|x_2 = 1)$

Ερώτημα δ.

Οι ισοσταθμικές καμπύλες των κατανομών των ερωτημάτων (β) και (γ) παρουσιάζονται σε κοινό σχήμα παρακάτω με κοινούς άξονες παρακάτω.



Σχήμα 7: Ισοσταθμικές καμπύλες των κατανομών των ερωτημάτων (β) και (γ)

Άσκηση 1.3

Ισοπίθανες κλάσεις ω_1 και ω_2. Δηλαδή $P(\omega_1) = P(\omega_2) = \frac{1}{2}$.

Η βέλτιστη ευθεία απόφασης δίνεται από τον τύπο:

$$g(\boldsymbol{x}) = \boldsymbol{\theta^T(x - x_0) = 0}$$

όπου

$$\boldsymbol{\theta} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$\boldsymbol{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) - ln\frac{P(\omega_1)}{P(\omega_2)}\frac{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2}{||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||_{\Sigma^{-1}}}$$

Απόσταση Mahalanobis των $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$:

$$||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||_{\Sigma^{-1}} = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}$$

Ερώτημα α.

$$\boldsymbol{\mu}_1 = \begin{bmatrix}-2 & 0\end{bmatrix}^T, \quad \boldsymbol{\mu}_2 = \begin{bmatrix}2 & 1\end{bmatrix}^T, \quad \Sigma = I = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

$$\boldsymbol{x}_0 = \begin{bmatrix}0 & 0.5\end{bmatrix}^T, \ \boldsymbol{\theta} = \begin{bmatrix}-4 & -1\end{bmatrix}^T$$

Βέλτιστη Ευθεία Απόφασης: $g(\boldsymbol{x}) = -4x_1 - x_2 + 0.5 = 0$

Ερώτημα β.

$$\boldsymbol{\mu}_1 = \begin{bmatrix}-2 & 0\end{bmatrix}^T, \ \boldsymbol{\mu}_2 = \begin{bmatrix}2 & 1\end{bmatrix}^T, \ \hat{\Sigma} = \begin{bmatrix}1 & -0.6\\ -0.6 & 1\end{bmatrix}$$

$$\boldsymbol{x_0} = \begin{bmatrix}0 & 0.5\end{bmatrix}^T, \ \boldsymbol{\theta} = \begin{bmatrix}-7.1875 & -5.3125\end{bmatrix}^T$$

Βέλτιστη Ευθεία Απόφασης: $g(\boldsymbol{x}) = -7.1875x_1 - 5.3125x_2 + 2.6562 = 0$

Ερώτημα γ.

$$\boldsymbol{\mu}_1 = \begin{bmatrix}-2 & 0\end{bmatrix}^T, \ \boldsymbol{\mu}_2 = \begin{bmatrix}2 & 1\end{bmatrix}^T, \ \hat{\Sigma} = \begin{bmatrix}1 & -0.6\\ -0.6 & 1\end{bmatrix}$$

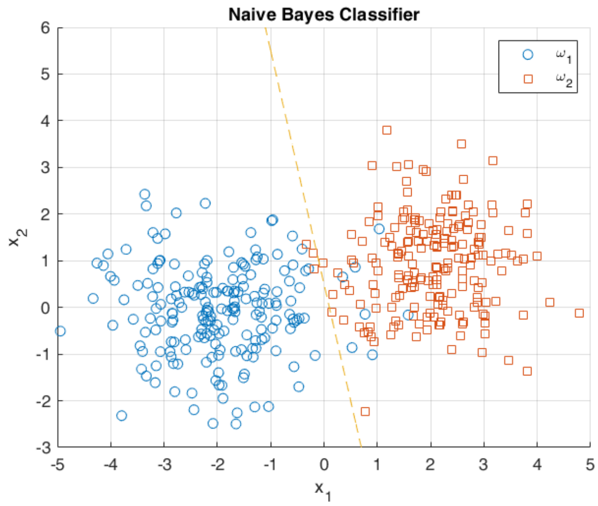
Συντελεστές Βαρύτητας Σφαλμάτων Ταξινόμησης: $\lambda_{12} = 1, \ \lambda_{21} = 0.5$

Χρησιμοποιούνται οι ίδιοι τύποι με $P'(\omega_1) = \lambda_{12}P(\omega_1), \ P'(\omega_2) = \lambda_{21}P(\omega_2)$ οπότε προκύπτουν:

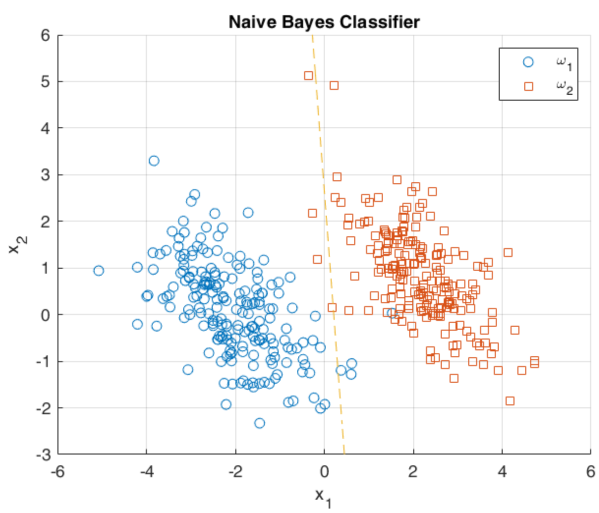
$$\boldsymbol{x}_0 = \begin{bmatrix}0.4751 & 0.6188\end{bmatrix}^T, \ \boldsymbol{\theta} = \begin{bmatrix}-7.1875 & -5.3125\end{bmatrix}^T$$

Βέλτιστη Ευθεία Απόφασης: $g(\boldsymbol{x}) = -7.1875x_1 - 5.3125x_2 + 6.7017 = 0$

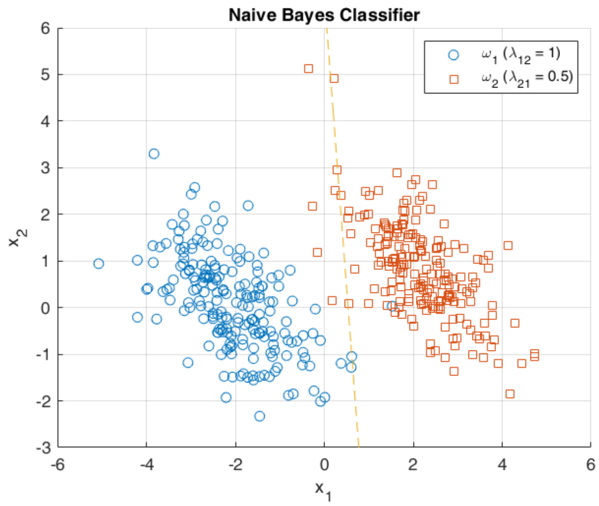
Ερώτημα δ.



Σχήμα 8: Σημεία από τις κατανομές του ερωτήματος α και ευθεία απόφασης του ταξινομητή Naive Bayes



Σχήμα 9: Σημεία από τις κατανομές του ερωτήματος β και ευθεία απόφασης του ταξινομητή Naive Bayes



Σχήμα 10: Ευθεία απόφασης του ταξινομητή Naive Bayes για συντελεστές βαρύτητας σφαλμάτων ταξινόμησης $\lambda_{12} = 1$ και $\lambda_{21} = 0.5$ και σημεία από τις κατανομές του ερωτήματος γ.

Όταν το σφάλμα ταξινόμησης στην κλάση ω_1 θεωρείται σοβαρότερο από το σφάλμα ταξινόμησης στην κλάση ω_2 ο ταξινομητής μετακινεί την ευθεία απόφασης προς την περιοχή της κλάσης ω_2, αυξάνοντας την περιοχή της κλάσης ω_1.

Άσκηση 1.4

Ερώτημα α.

Παρακάτω φαίνονται τα αποτελέσματα της εκπαίδευσης του αλγορίθμου μάθησης perceptron με τιμή βήματος εκπαίδευσης $\beta = 0.1$ στο επαυξημένο σύνολο δειγμάτων:

$$\begin{aligned}
\boldsymbol{x}_1 &= (1,\ 4,\ 3,\ 6)^T \in C_N \\
\boldsymbol{x}_2 &= (1,\ 2, -2,\ 3)^T \in C_P \\
\boldsymbol{x}_3 &= (1,\ 1,\ 0, -3)^T \in C_P \\
\boldsymbol{x}_4 &= (1,\ 4,\ 2,\ 3)^T \in C_N
\end{aligned}$$

και για αρχική τιμή βαρών $\boldsymbol{w} = (w_0, w_1, w_2, w_3)^T = (1, 0, 0, 0)^T$

Epoch 1

Data Sample 1:

$$\begin{aligned}
\boldsymbol{x}_1 &: \begin{bmatrix}1.00 & 4.00 & 3.00 & 6.00\end{bmatrix}^T \\
t_1 &: -1 \quad \rightarrow \quad Class: \quad C_N \\
y_1 &: +1 \quad \rightarrow \quad Label: \quad False\ Positive \\
\boldsymbol{\Delta w}_1 &: \begin{bmatrix}-0.20 & -0.80 & -0.60 & -1.20\end{bmatrix}^T \\
\boldsymbol{w}_1 &: \begin{bmatrix}0.80 & -0.80 & -0.60 & -1.20\end{bmatrix}^T
\end{aligned}$$

Data Sample 2:

$$\begin{aligned}
\boldsymbol{x}_2 &: \begin{bmatrix}1.00 & 2.00 & -2.00 & 3.00\end{bmatrix}^T \\
t_2 &: +1 \quad \rightarrow \quad Class: \quad C_P \\
y_2 &: -1 \quad \rightarrow \quad Label: \quad False\ Negative \\
\boldsymbol{\Delta w}_2 &: \begin{bmatrix}0.20 & 0.40 & -0.40 & 0.60\end{bmatrix}^T \\
\boldsymbol{w}_2 &: \begin{bmatrix}1.00 & -0.40 & -1.00 & -0.60\end{bmatrix}^T
\end{aligned}$$

Data Sample 3:

$$\begin{aligned}
\boldsymbol{x}_3 &: \begin{bmatrix}1.00 & 1.00 & 0.00 & -3.00\end{bmatrix}^T \\
t_3 &: +1 \quad \rightarrow \quad Class: \quad C_P \\
y_3 &: +1 \quad \rightarrow \quad Label: \quad True\ Positive \\
\boldsymbol{\Delta w}_3 &: \begin{bmatrix}0.00 & 0.00 & 0.00 & -0.00\end{bmatrix}^T \\
\boldsymbol{w}_3 &: \begin{bmatrix}1.00 & -0.40 & -1.00 & -0.60\end{bmatrix}^T
\end{aligned}$$

Data Sample 4:

$$\begin{aligned}
\boldsymbol{x}_4 &: \begin{bmatrix}1.00 & 4.00 & 2.00 & 3.00\end{bmatrix}^T \\
t_4 &: -1 \quad \rightarrow \quad Class: \quad C_N \\
y_4 &: -1 \quad \rightarrow \quad Label: \quad True\ Negative \\
\boldsymbol{\Delta w}_4 &: \begin{bmatrix}0.00 & 0.00 & 0.00 & 0.00\end{bmatrix}^T \\
\boldsymbol{w}_4 &: \begin{bmatrix}1.00 & -0.40 & -1.00 & -0.60\end{bmatrix}^T
\end{aligned}$$

Epoch 2

Data Sample 1:

$$\begin{aligned}
\boldsymbol{x}_1 &: \begin{bmatrix}1.00 & 4.00 & 3.00 & 6.00\end{bmatrix}^T \\
t_1 &: -1 \quad \rightarrow \quad Class: \quad C_N \\
y_1 &: -1 \quad \rightarrow \quad Label: \quad True\ Negative \\
\boldsymbol{\Delta w}_1 &: \begin{bmatrix}0.00 & 0.00 & 0.00 & 0.00\end{bmatrix}^T \\
\boldsymbol{w}_1 &: \begin{bmatrix}1.00 & -0.40 & -1.00 & -0.60\end{bmatrix}^T
\end{aligned}$$

Data Sample 2:

$$\begin{aligned}
\boldsymbol{x}_2 &: \begin{bmatrix}1.00 & 2.00 & -2.00 & 3.00\end{bmatrix}^T \\
t_2 &: +1 \quad \rightarrow \quad Class: \quad C_P \\
y_2 &: +1 \quad \rightarrow \quad Label: \quad True\ Positive \\
\boldsymbol{\Delta w}_2 &: \begin{bmatrix}0.00 & 0.00 & -0.00 & 0.00\end{bmatrix}^T \\
\boldsymbol{w}_2 &: \begin{bmatrix}1.00 & -0.40 & -1.00 & -0.60\end{bmatrix}^T
\end{aligned}$$

Data Sample 3:

$$\begin{aligned}
\boldsymbol{x}_3 &: \begin{bmatrix}1.00 & 1.00 & 0.00 & -3.00\end{bmatrix}^T \\
t_3 &: +1 \quad \rightarrow \quad Class: \quad C_P \\
y_3 &: +1 \quad \rightarrow \quad Label: \quad True\ Positive \\
\boldsymbol{\Delta w}_3 &: \begin{bmatrix}0.00 & 0.00 & 0.00 & -0.00\end{bmatrix}^T \\
\boldsymbol{w}_3 &: \begin{bmatrix}1.00 & -0.40 & -1.00 & -0.60\end{bmatrix}^T
\end{aligned}$$

Data Sample 4:

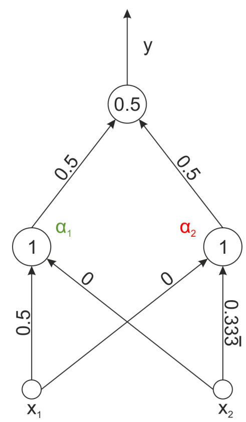
$$\begin{aligned}
\boldsymbol{x}_4 &: \begin{bmatrix}1.00 & 4.00 & 2.00 & 3.00\end{bmatrix}^T \\
t_4 &: -1 \quad \rightarrow \quad Class: \quad C_N \\
y_4 &: -1 \quad \rightarrow \quad Label: \quad True\ Negative \\
\boldsymbol{\Delta w}_4 &: \begin{bmatrix}0.00 & 0.00 & 0.00 & 0.00\end{bmatrix}^T \\
\boldsymbol{w}_4 &: \begin{bmatrix}1.00 & -0.40 & -1.00 & -0.60\end{bmatrix}^T
\end{aligned}$$

Ερώτημα β.

Παρατηρείται οτι τα δείγματα $\boldsymbol{x} = (x_1,\ x_2)^T$ ταξινομούνται στην αρνητική κλάση (λευκή περιοχή) όταν $max\{\frac{1}{2}x_1,\ \frac{1}{3}x_2\} < 1$ ενώ θετική κλάση (γκρίζα περιοχή και όριο) στην αντίθετη περίπτωση. Ισοδύναμα, μια είσοδος $\boldsymbol{x}$ ταξινομείται στη θετική κλάση αν

$$\boldsymbol{x} \in \left\{\boldsymbol{x} = (x_1, x_2)^T \mid \{x_1 \geq 2 \Leftrightarrow \frac{1}{2}x_1 \geq 1\} \vee \{x_2 \geq 3 \Leftrightarrow \frac{1}{3}x_2 \geq 1\}\right\}$$

και στην αρνητική όταν το $\boldsymbol{x}$ δεν ανήκει στο παραπάνω σύνολο.

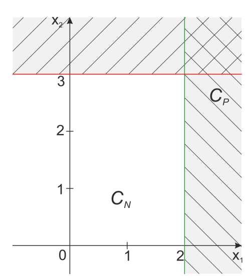


Σχήμα 11: Πολυστρωματικό δίκτυο Perceptron τριών νευρώνων

Ως συνάρτηση ενεργοποίησης των νευρώνων χρησιμοποιείται η κλασσική βηματική

$$f(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases}$$

και θεωρείται πως η τιμή 1 αντιστοιχεί στη θετική κλάση ενώ η τιμή 0 στην αρνητική.



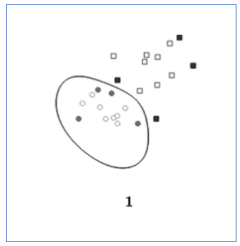
Σχήμα 12: Περιοχές απόφασης που προκύπτουν από το προηγούμενο μοντέλο

Βάσει της προηγούμενης ανάλυσης επιδιώκεται να ενεργοποιείται η έξοδος όταν υπάρχει $\boldsymbol{x}$ με $\frac{1}{2}x_1 \geq 1$ και $x_2 \in R$. Δηλαδή, αρκούν βάρη εισόδων ίσα με $\frac{1}{2}$ και 0 για x_1 και x_2 αντίστοιχα και κατώφλι ενεργοποίησης ίσο με 1 ώστε να τεθεί στο 1 η έξοδος α_1 του κρυφού στρώματος.

Ομοίως, όταν υπάρχει $\boldsymbol{x}$ με $\frac{1}{3}x_2 \geq 1$ επιλέγονται βάρη εισόδων ίσα με 0 και $\frac{1}{3}$ για x_1 και x_2 αντίστοιχα και κατώφλι ίσο με 1 ώστε να τεθεί στο 1 η έξοδος α_2.

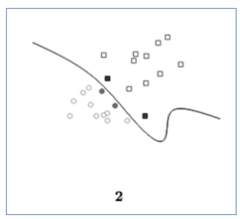
Τέλος, οι έξοδοι α_1 και α_2 του κρυφού στρώματος αθροίζονται με ίσα βάρη τιμής $\frac{1}{2}$ και με κατώφλι $\frac{1}{2}$ ώστε να παράγουν την έξοδο y. Η πράξη αυτή αποτελεί μια απλή διάζευξη που ενεργοργοποιεί το νευρώνα εξόδου όταν τουλάχιστον μία από τις εξόδους του προηγούμενου στρώματος (α_1, α_2) είναι ενεργή.

Άσκηση 1.5



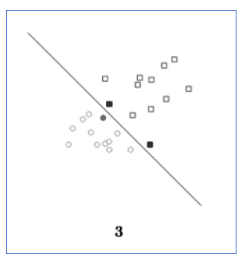
Σχήμα 13: (δ) $k(u,v) = exp\big(-0.25||u{-}v||_2^2\big)$

Γκαουσιανός πυρήνας με διασπορά: $\frac{1}{2\sigma^2} = 0.25 \Longrightarrow \sigma^2 = 2$ δηλαδή μεγαλύτερη διασπορά από ότι στην περίπτωση της εικόνας 6



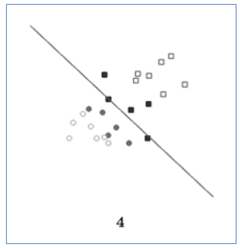
Σχήμα 14: (γ) Μη γραμμικό SVM με $k\,(u,\ v) =\ u^T v\ +\ \left(u^T v\right)^2$

Πολυωνυμικός πυρήνας



Σχήμα 15: (β) Γραμμικό SVM με C = 10

Η ποινή χαλαρότητας είναι μεγάλη (σε σύγκριση με το σχήμα 4) κι επομένως δίνεται μεγάλη έμφαση τα δείγματα εκπαίδευσης να ταξινομούνται στη σωστή κλάση.



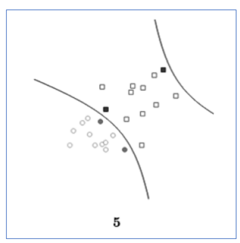
Σχήμα 16: (α) Γραμμικό SVM με C = 0.1

Η ποινή χαλαρότητας είναι μικρή (σε σύγκριση με το σχήμα 3) οπότε δε δίνεται τόση βαρύτητα στη σωστή ταξινόμηση των δειγμάτων εκπαίδευσης όσο στην ελαχιστοποίηση της πολυπλοκότητας του μοντέλου.

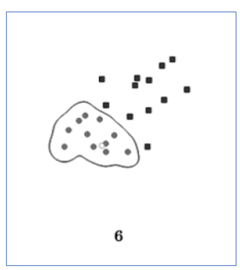
Για τα σχήματα 3 και 4 έχει υποτεθεί πως η συνάρτηση κόστους προς ελαχιστοποίηση είναι της μορφής

$$J(\boldsymbol{w}, w_0) = \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{P}\xi_i$$

στην οποία ο πρώτος όρος αντιστοιχεί στην πολυπλοκότητα του μοντέλου και ο δεύτερος στην σωστή ταξινόμηση των δειγμάτων εκπαίδευσης.



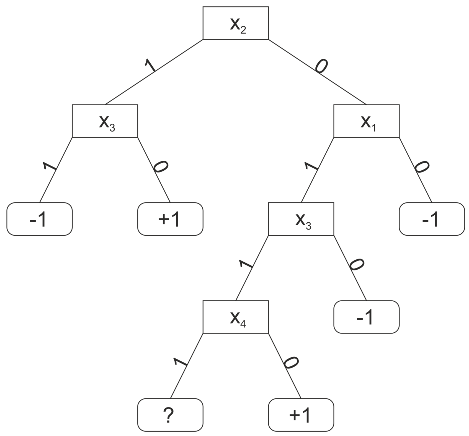
Σχήμα 17: (στ) Κανένα από τα παραπάνω



Σχήμα 18: (ε) Μη γραμμικό SVM με $k(u, v) = exp\big(-4||u{-}v||_2^2\big)$

Γκαουσιανός πυρήνας με διασπορά: $\frac{1}{2\sigma^2} = 4 \Longrightarrow \sigma^2 = 0.125$ δηλαδή μικρότερη διασπορά από ότι στην περίπτωση της εικόνας 1

Άσκηση 1.6



Σχήμα 19: Δέντρο Απόφασης με βάση την εντροπία και το κριτήριο gini