

Σχολή Ηλεκτρολογών Μηχανικών και Μηχανικών
Υπολογιστών

ΤΗΛ311-Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων

2η Σειρά Ασκήσεων

Μαραγκάκη Μαρία
Α.Μ.:2015030153

Πολυτεχνείο Κρήτης
Ημερομηνία παράδοσης: 06/05/2020

Contents

1	Λογιστική Παλινδρόμηση: Αναλυτική Εύρεση Κλίσης (Greadient)	2
2	Λογιστική Παλινδρόμηση με Ομαλοποίηση	4
3	Εκτίμηση Παραμέτρων (Maximum Likelihood)	8
4	Bayesian εκτίμηση παραμέτρου (Bayesian Inference)	8
5	Support Vector Machines (Αναλυτική βελτιστοποίηση με KKT)	11
6	Support Vector Machines (Εφαρμογή σε τεχνητό σύνολο δεδομένων)	17
7	Παραπομπές	20

1 Λογιστική Παλινδρόμηση: Αναλυτική Εύρεση Κλίσης (Gradient)

Σκοπός της συγκεκριμένης άσκησης είναι η μελέτη της εύρεσης κλίσης μέσω της μεθόδου της λογιστικής παλινδρόμησης. Δηλαδή, δεδομένου ενός συνόλου δεδομένων $\{(x^1, y^1), \dots, (x^m, y^m)\}$ σκοπός είναι η πρόβλεψη του label της κλάσης y^i από της αντίστοιχες τιμές x^i , $i \in \{1, 2, \dots, m\}$.

1. Σε πρώτη φάση υπολογίστηκε η κλίση του σφάλματος δεδομένων, της συνάρτησης λογιστικής παλινδρόμησης, τη λογιστική συνάρτηση και τη συνάρτηση κόστους σφάλματος. Ο υπολογισμός έγινε ως εξής:

- Έστω $a = h_\theta(x^i) = f(\theta^T x)$ τότε $\log a = \ln(1/(1 + e^{-\theta^T x})) = 0 - \ln(1 + e^{-\theta^T x}) = -\ln(1 + e^{-\theta^T x})$
- $\partial \ln a / \partial \theta = -1/(1 + e^{-\theta^T x}) * (-x^i e^{-\theta^T x}) = (x^i e^{-\theta^T x}) / (1 + e^{-\theta^T x}) = x_j(1 - a)$
- $\ln(1 - a) = \log(e^{-\theta^T x} / (1 + e^{-\theta^T x})) = -\theta^T x - \ln(1 + e^{-\theta^T x})$
- $\partial \ln(1 - a) / \partial \theta = -x_j - 1/(1 + e^{-\theta^T x}) * (-x^i e^{-\theta^T x}) = -x_j + (x^i e^{-\theta^T x}) / (1 + e^{-\theta^T x}) = -x_j + x_j(1 - a) = -ax_j$
- Συνεπώς τελικά:

$$\begin{aligned} \partial J(\theta) / \partial \theta &= -1/m \sum (y^i \ln a + (1 - y^i) \ln(1 - a)) = -1/m \sum (y^i x_j (1 - a) + (1 - y^i)(-ax_j)) = \\ &= -1/m \sum (y^i x_j - y^i ax_j - ax_j + y^i ax_j) = -1/m \sum (y^i x_j - h_\theta(x^i) x_j) = 1/m \sum (h_\theta(x^i) - y^i) x_j \end{aligned}$$

2. Στη συνέχεια χρησιμοποιήθηκε η λογιστική παλινδρόμηση για την πρόβλεψη αν ένας φοιτητής θα γίνει δεκτός σε ένα πανεπιστήμιο ή όχι με βάση τους βαθμούς σε δύο εξετάσεις
 - (a) Αρχικά σχεδιάστηκαν τα δεδομένα μέσω της συνάρτησης `plotData.m` και το αποτέλεσμα ήταν το εξής:

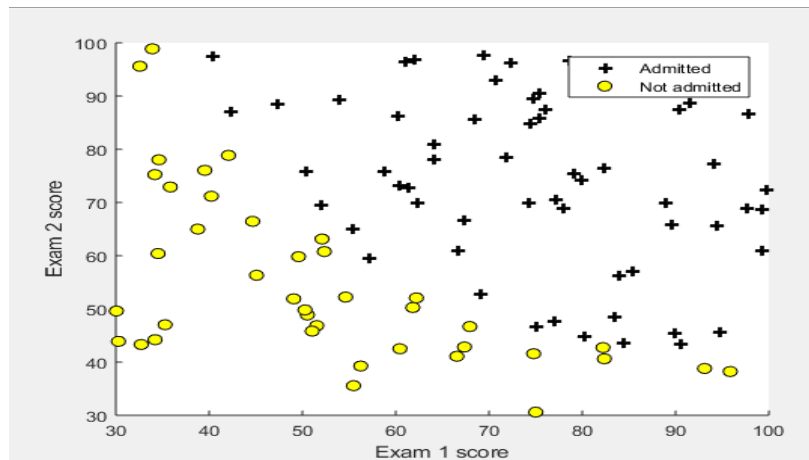


Figure 1: Αποτύπωση δειγμάτων στο χώρο

- (b) Στη συνέχεια, υλοποιήθηκε η συνάρτηση *sigmoid.m* με βάση τη δοσμένη συνάρτηση $f(z)$.
- (c) Έπειτα υλοποιήθηκε η συνάρτηση *costFunction.m* η οποία επιστρέφει το κόστος $J(\theta)$ και την κλίση $\nabla J(\theta)$ όπως υπολογίστηκε παραπάνω. Συνεπώς βρέθηκε ότι το κόστος θα είναι $J = 0.693147$ και η κλίση θα είναι $\nabla J(\theta) = [-0.1000000, -12.009217, -11.262842]$
- (d) Τέλος πραγματοποιήθηκε βελτιστοποίηση των παραμέτρων και σχεδιάστηκε το όριο απόφασης. Ακόμη μέσω τη συνάρτησης *predict.m* υπολογίστηκε η πρόβλεψη αν ο φοιτητής θα γίνει δεκτός με βάση διάφορες τιμές βαθμών, έτσι το $TrainAccuracy = 89$ που σημαίνει ότι το συγκριμένο μοντέλο κάνει σωστές προβλέψεις κατά 89%. Τα αποτελέσματα παρουσιάζονται παρακάτω.

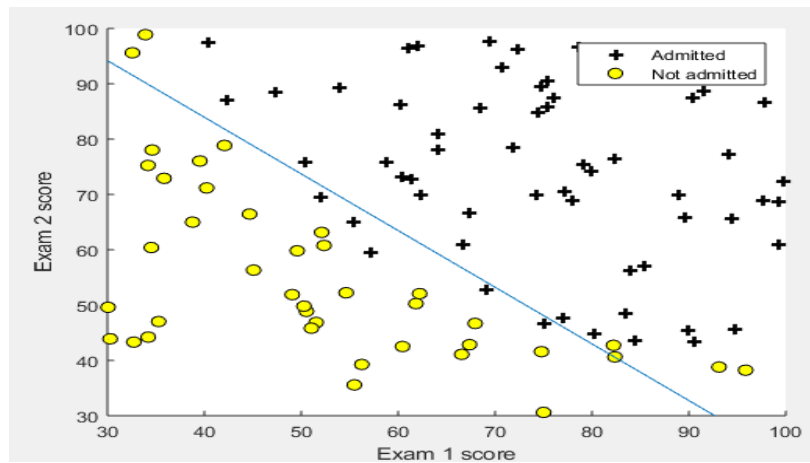


Figure 2: Αποτύπωση ορίου απόφασης

2 Λογιστική Παλινδρόμηση με Ομαλοποίηση

Στην συγκεκριμένη άσκηση εφαρμόστηκε ομαλοποιημένη λογιστική παλινδρόμηση με σκοπό την πρόβλεψη αν τα μικροσίπ από μια μονάδα κατασκευής περνούν τον έλεγχο ποιότητας.

1. Αρχικά σχεδιάστηκαν τα δεδομένα. Το αποτέλεσμα παρουσιάζεται παρακάτω.

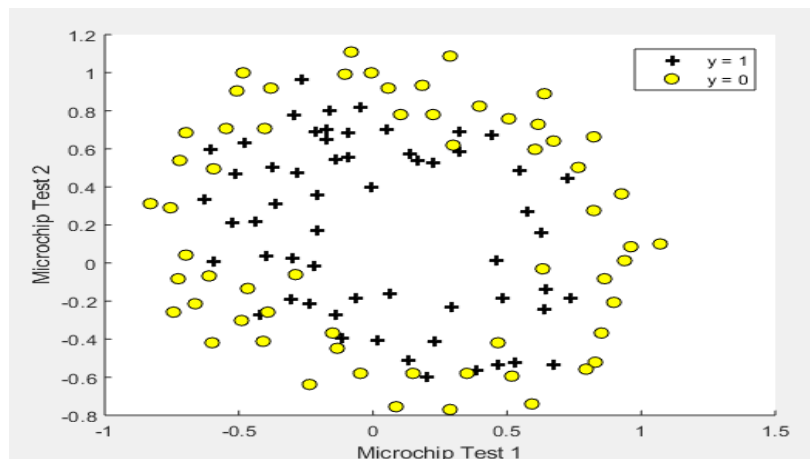


Figure 3: Αποτύπωση δειγμάτων στο χώρο

2. Στην συνέχεια, συμπληρώθηκε η συνάρτηση *mapFeature.m* όπου επικονίζει τα χαρακτηριστικά σε όρους πολυωνύμων x_1 και x_2 βαθμού μέχρι 6. Οι συναρτήσεις που χρησιμοποιήθηκαν ήταν οι εξής:

$$P(x_1, x_2) = \sum_{i=0}^6 \sum_{j=i}^i x_1^j x_2^{i-j}$$

$$\text{mapFeature}(x) = [1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, \dots, x_1^5x_2^5, x_2^6]$$

Συνεπώς απο αυτήν την συνάρτηση επιστρέφεται ένα διάνυσμα 118*28 δηλαδή για κάθε στοιχείο X_i γίνεται *map*.

3. Η ομαλοποιημένη συνάρτηση κόστους δίνεται απο τη σχέση:

$$\partial J(\theta)/\partial \theta = 1/m \sum_{i=1}^m (-y^i \ln(h_\theta(x^i)) - (1 - y^i) \ln(1 - h_\theta(x^i))) + \lambda/2m \sum_{j=1}^n \theta_j^2$$

4.

- Έστω $a = h_\theta(x^i) = f(\theta^T x)$ τότε $\log a = \ln(1/(1 + e^{-\theta^T x})) = 0 - \ln(1 + e^{-\theta^T x}) = -\ln(1 + e^{-\theta^T x})$
- $\partial \ln a / \partial \theta = -1/(1 + e^{-\theta^T x}) * (-x^i e^{-\theta^T x}) = (x^i e^{-\theta^T x})/(1 + e^{-\theta^T x}) = x_j(1 - a)$
- $\ln(1 - a) = \log(e^{-\theta^T x}/(1 + e^{-\theta^T x})) = -\theta^T x - \ln(1 + e^{-\theta^T x})$
- $\partial \ln(1 - a) / \partial \theta = -x_j - 1/(1 + e^{-\theta^T x}) * (-x^i e^{-\theta^T x}) = -x_j + (x^i e^{-\theta^T x})/(1 + e^{-\theta^T x}) = -x_j + x_j(1 - a) = -ax_j$
- Συνεπώς τελικά:

$$\begin{aligned} \partial J(\theta)/\partial \theta &= -1/m \sum (y^i \ln a + (1 - y^i) \ln(1 - a)) + \lambda/2m \sum_{j=1}^n \theta_j^2 \\ &= -1/m \sum (y^i x_j(1 - a) + (1 - y^i)(-ax_j)) + \lambda/m * \theta_j = \\ &\quad -1/m \sum (y^i x_j - y^i ax_j - ax_j + y^i ax_j) + \lambda/m * \theta_j \\ &= -1/m \sum (y^i x_j - h_\theta(x^i) x_j) + \lambda/m * \theta_j = 1/m \sum (h_\theta(x^i) - y^i) x_j + \lambda/m * \theta_j \end{aligned}$$

5. Στη συνέχεια με βάση τον τύπο που υπολογίστηκε συμπληρώθηκε η συνάρτηση *costFunction.m* όπου επιστρέφει το κόστος και την κλίση. Το κόστος για $\lambda = 1$ είναι $J = 0.693147$.

6. Τέλος σχεδιάστηκαν τα διαφορετικά όρια απόφαση για διαφορετικές τιμές του λ . Οι τιμές που πήρε το λ ήταν οι: 0,1,10,100 και τα αποτελέσματα φαίνονται παρακάτω.

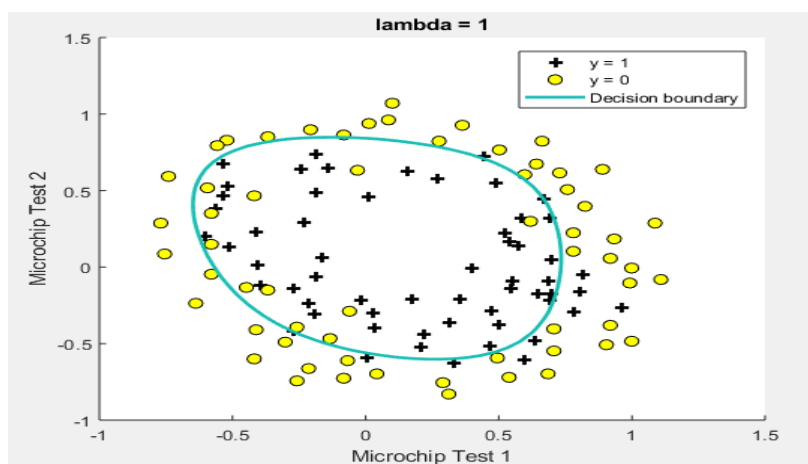


Figure 4: Όριο απόφασης για $\lambda=1$

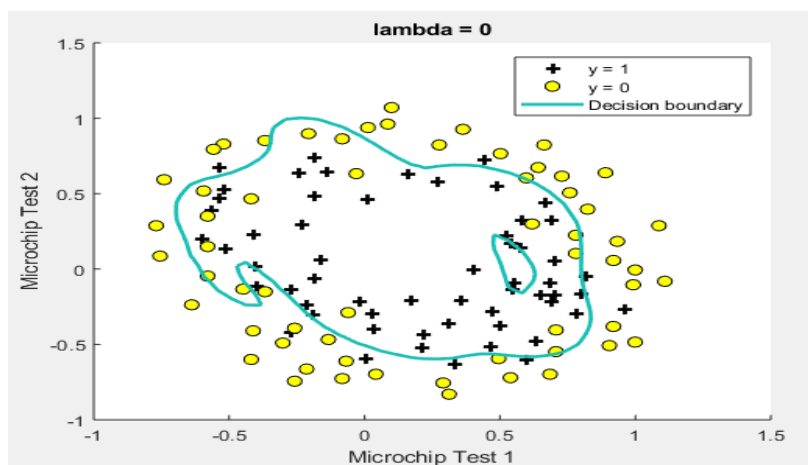


Figure 5: Όριο απόφασης για $\lambda=0$

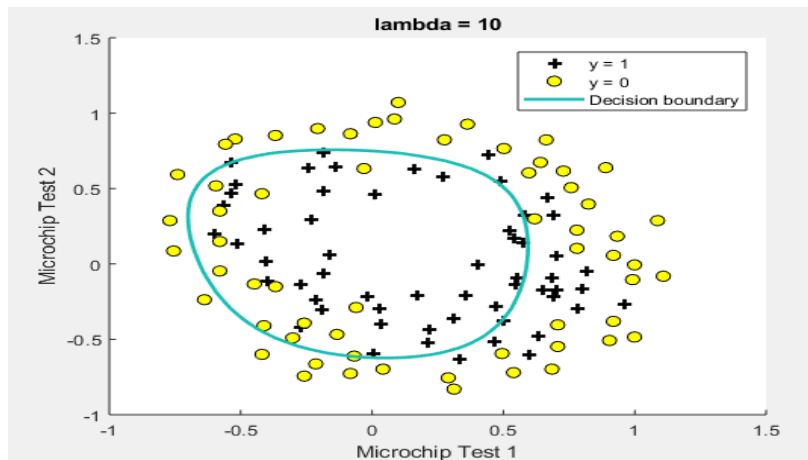


Figure 6: Όριο απόφασης για $\lambda=10$

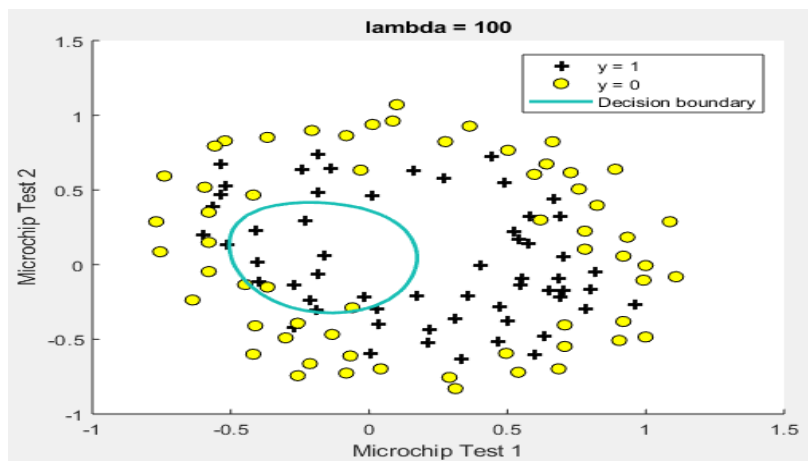


Figure 7: Όριο απόφασης για $\lambda=100$

Παρατηρείται ότι όσο το λ αυξάνεται τόσο μειώνεται το όριο απόφασης αλλά και το train accuracy. Όσον το Train Accuracy για την κάθε περίπτωση έχουμε: 82.2 για $\lambda = 1$, 83.33 για $\lambda = 0$, 74.57 για $\lambda = 10$, 60.16 για $\lambda = 100$. Παρατηρείται ότι όσο αυξάνεται το λ μειώνεται το Train Accuracy.

3 Εκτίμηση Παραμέτρων (Maximum Likelihood)

Η εκτίμηση μεγίστης πιθανοφάνειας της παραμέτρου λ υπολογίζεται ως εξής:

$$l(\lambda) = \ln\left(\prod_{i=1}^n P(x_i|\lambda)\right) = \sum_{i=1}^n \ln((e^{-\lambda}\lambda^{x_i})/x_i!) = \sum_{i=1}^n (\ln(e^{-\lambda}\lambda^{x_i}) - \ln x_i!) = \sum_{i=1}^n (\ln e^{-\lambda} + \ln \lambda^{x_i} - \ln x_i!) =$$

$$\sum_{i=1}^n \ln e^{-\lambda} + \sum_{i=1}^n \ln x_i \lambda - \sum_{i=1}^n \ln x_i! = \sum_{i=1}^n -\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!$$

Συνεπώς παραγωγίζοντας έχουμε:

$$\partial l(\lambda)/\partial \lambda = \partial \left(\sum_{i=1}^n -\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i! \right) / \partial \lambda = 0 \Rightarrow -n + 1/\lambda \sum_{i=1}^n x_i = 0 \Rightarrow \hat{\lambda} = 1/n \sum_{i=1}^n x_i$$

4 Bayesian εκτίμηση παραμέτρου (Bayesian Inference)

Στη συγκεκριμένη άσκηση μελετήθηκε η Bayesian εκτίμηση της μέση τιμής (μ) δεδομένου ενός συνόλου 25 μετρήσεων. Η συνάρτηση πυκνότητας πιθανότητας των μετρήσεων είναι Γκαουσιανή με μέση τιμή μ και τυπική απόκλιση $\sigma=1.25$, $p(x) \sim N(\mu, \sigma^2)$. Ακόμη, η κατανομή της μέσης τιμής είναι Γκαουσιανή $p(\mu) \sim N(\mu_0, \sigma_0^2)$.

1. Αφού η συνάρτηση πυκνότητας πιθανότητας και η κατανομή της μέσης τιμής είναι Γκαουσιανές, τότε και η δεσμευμένη πυκνότητα πιθανότητας θα είναι Γκαουσιανή με $\mu_0 = 0$ και $\sigma_0^2 = 10\sigma^2$. Επίσης θα ισχύει:

$$p(\mu|X) = 1/(\sqrt{2\pi\sigma_n}) \exp(-(\mu - \mu_N)^2/2\sigma_N^2)$$

Όπου,

$$\mu_N = (\sigma_0^2 \hat{x}_N + \sigma^2 \mu) / (\sigma_0^2 + \sigma^2)$$

$$\sigma_N^2 = (\sigma_0^2 \sigma^2) / (\sigma_0^2 + \sigma^2)$$

$$\hat{x}_N = 1/N \sum_{k=1}^N x_k$$

Συνεπώς, η δεσμευμένη πυκνότητα πιθανότητας καθώς το n μεταβάλλεται από 1 έως 25 θα έχει την παρακάτω μορφή.

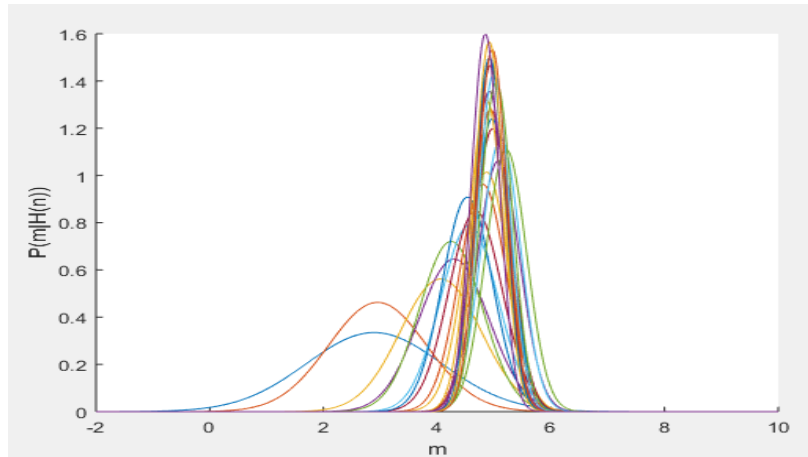


Figure 8: Δεσμευμένη πυκνότητα πιθανότητας $P(m|H(n))$

2. Τέλος σχεδιάστηκαν οι δεσμευμένες πυκνότητες $P(x|H(n))$ για διαφορετικές τιμές του σ_0^2 τα αποτελέσματα παρουσιάζονται παρακάτω.

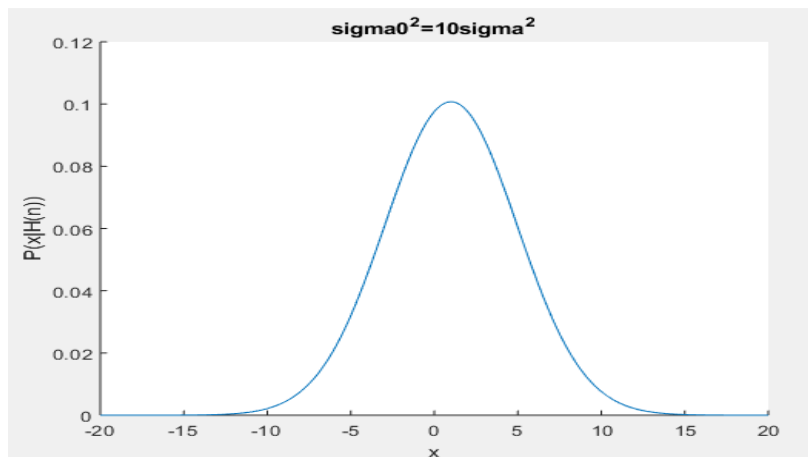


Figure 9: Δεσμευμένη πυκνότητα πιθανότητας $P(x|H(n))$ για $\sigma_0^2 = 10\sigma^2$

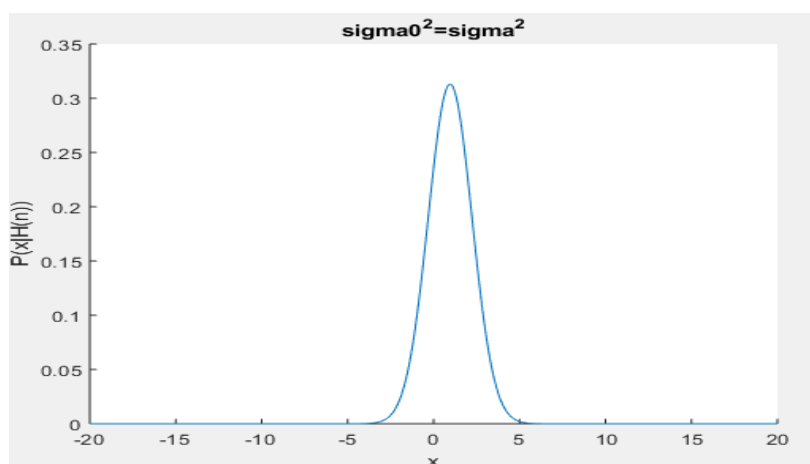


Figure 10: Δεσμευμένη πυκνότητα πιθανότητας $P(x|H(n))$ για $\sigma_0^2 = \sigma^2$

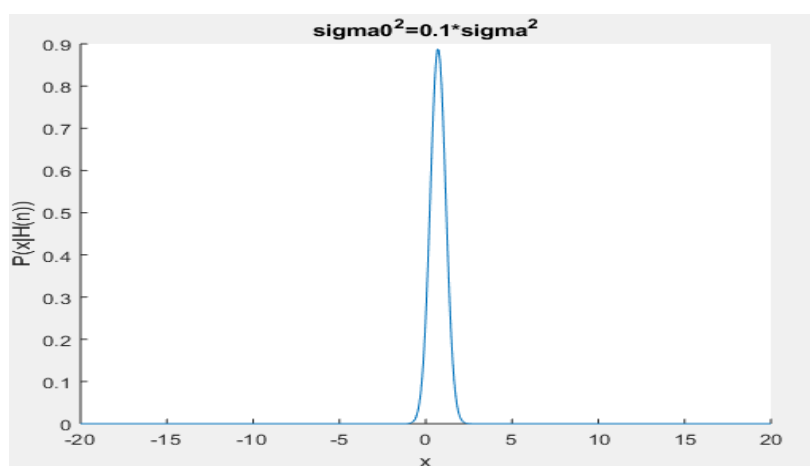


Figure 11: Δεσμευμένη πυκνότητα πιθανότητας $P(x|H(n))$ για $\sigma_0^2 = 0.1\sigma^2$

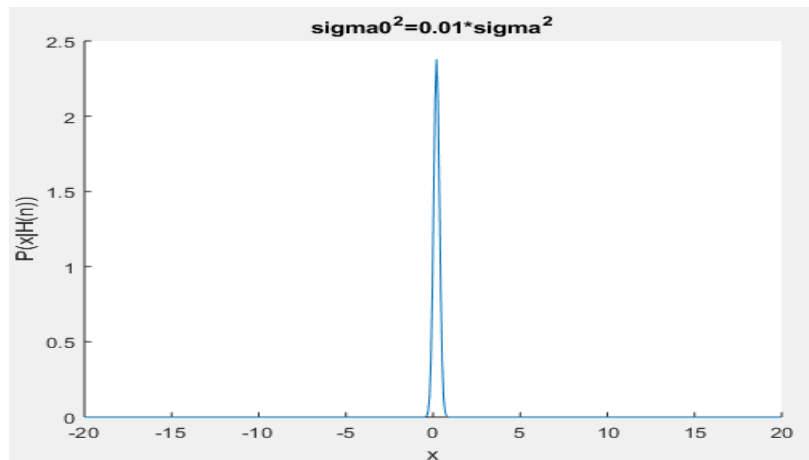


Figure 12: Δεσμευμένη πυκνότητα πιθανότητας $P(x|H(n))$ για $\sigma_0^2 = 0.01\sigma^2$

Παρατηρείται ότι όσο το σ_0^2 μειώνεται τόσο η καμπύλη γίνεται πιο στενή και ψηλή πλησιάζει δηλαδή περισσότερο στον δειγματικό μέσο.

5 Support Vector Machines (Αναλυτική βελτιστοποίηση με KKT)

Στη συγκεκριμένη άσκηση χρησιμοποιήθηκαν τα Support Vector Machines σε ένα πρόβλημα ταξινόμησης σε δύο κατηγορίες ω_1 και ω_2 .

1. Τα δείγματα $\mathbf{x} = [x_1, x_2]$ για τις δύο κλάσεις είναι τα εξής:

- $\omega_1 = \mathbf{x}^+ = \{[3, 1]^T, [3, -1]^T, [5, 1]^T, [5, -1]^T\}$
- $\omega_1 = \mathbf{x}^- = \{[1, 0]^T, [0, 1]^T, [0, -1]^T, [-1, 0]^T\}$

(a) Αρχικά, σχεδιάστηκαν τα δείγματα και το αποτέλεσμα παρουσιάζεται στο παρακάτω figure. Τα support vectors στη συγκεκριμένη περίπτωση θα είναι από την \mathbf{x}^- το $[1, 0]$ και από την \mathbf{x}^+ τα $[3, 1]$ και $[3, -1]$. Διαισθητικά το γραμμικό επίπεδο διαχωρισμού θα είναι η ευθεία $g(x) = 2$. Θα είναι αυτή η ευθεία διότι είναι η μοναδική ευθεία που χωρίζει τα δείγματα με τέτοιο τρόπο ώστε να ισαπέχουν από αυτή. Μεγιστοποιεί δηλαδή το *margin* των δειγμάτων.

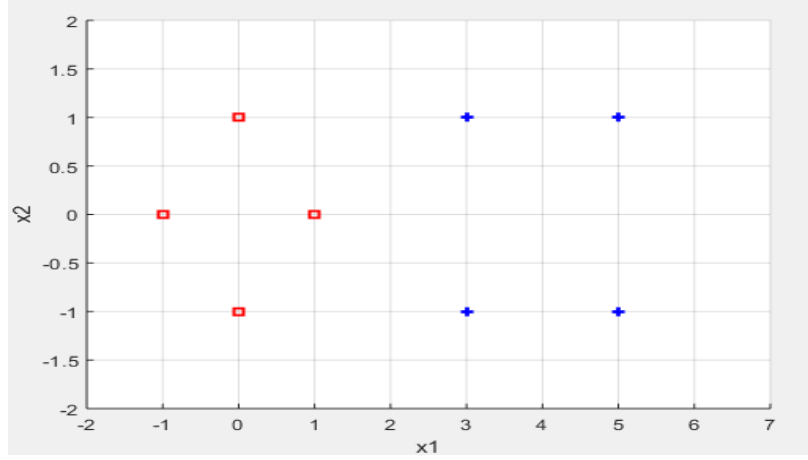


Figure 13: Δείγματα κλάσεων

- (b) Έπειτα, υπολογίστηκε η βέλτιστη γραμμή διαχωρισμού με τη χρήση πολλαπλασιαστών Lagrange και τις συνθήκες Karush-Khun-Tucker. Επειδή τα Support Vectors θα είναι 3 τότε και οι πολλαπλασιαστές Lagrange θα είναι 3. Συνεπώς, θα υπάρχουν 4 περιπτώσεις, για $\lambda_1 = 0, \lambda_2, \lambda_3 \neq 0$, για $\lambda_2 = 0, \lambda_1, \lambda_3 \neq 0$, για $\lambda_3 = 0, \lambda_1, \lambda_2 \neq 0$ και για $\lambda_1, \lambda_2, \lambda_3 \neq 0$. Θα αναλυθεί μόνο η περίπτωση όπου $\lambda_1, \lambda_2, \lambda_3 \neq 0$ καθώς είναι και η περίπτωση από την οποία προκύπτει η $g(x)$. Το υπερεπίπεδο διαχωρισμού θα έχει τη μορφή $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$. Πρέπει να ισχύουν οι εξής περιορισμοί

- $\mathbf{w}^T \mathbf{x} + w_0 \geq 0, \forall \mathbf{x}_+ \in \omega_1$
- $\mathbf{w}^T \mathbf{x} + w_0 \leq 0, \forall \mathbf{x}_- \in \omega_2$

Σκοπός είναι η ελαχιστοποίηση του $J(\mathbf{w}^T, w_0) = 1/2 \|\mathbf{w}\|_2^2 = 1/2(w_1^2 + w_2^2)$. Υποθέτοντας ότι $f_i(\mathbf{w}^T, w_0) = y_i(\mathbf{w}^T \mathbf{x} + w_0) - 1 \geq 0, \forall \mathbf{x}_i \in \omega_1, \omega_2$, με $y_i = 1$ για την ω_1 και $y_i = -1$ για την ω_2 . Συνεπώς προκύπτουν οι παρακάτω εξισώσεις για τα Support Vectors:

- $[3, 1]^T : 3w_1 + w_2 + w_0 - 1 \geq 0$
- $[3, -1]^T : 3w_1 + w_2 - w_0 - 1 \geq 0$
- $[1, 0]^T : -w_1 - w_0 - 1 \geq 0$

Η βοηθητική συνάρτηση Lagrange την οποία θα ελαχιστοποιήσουμε είναι η εξής:
 $L(\mathbf{w}, w_0, \lambda) = J(\mathbf{w}, w_0) - \sum_{i=1}^N \lambda_i [y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1]$. Συνεπώς αντικαθιστώντας έχουμε.

$$L(\mathbf{w}, \lambda_1, \lambda_2, \lambda_3) = (w_1^2 + w_2^2)/2 - \lambda_1(3w_1 + w_2 + w_0 - 1) - \lambda_2(3w_1 + w_2 - w_0 - 1) - \lambda_3(-w_1 - w_0 - 1)$$

Παραγωγίζοντας θα πάρουμε τις εξής εξισώσεις σύμφωνα με τις συνθήκες KKT.

- $\partial(\mathbf{w}, \lambda_1, \lambda_2, \lambda_3)/\partial w_1 = w_1 - 3\lambda_1 - 3\lambda_2 + \lambda_3 = 0 \Rightarrow w_1 = 3(\lambda_1 + \lambda_2) - \lambda_3(1)$
- $\partial(\mathbf{w}, \lambda_1, \lambda_2, \lambda_3)/\partial w_2 = w_2 - \lambda_1 + \lambda_2 = 0 \Rightarrow w_2 = \lambda_1 - \lambda_2(2)$
- $\partial(\mathbf{w}, \lambda_1, \lambda_2, \lambda_3)/\partial w_0 = -\lambda_1 - \lambda_2 + \lambda_3 = 0 \Rightarrow \lambda_3 = \lambda_1 + \lambda_2(3)$
- $(1) \Rightarrow w_1 = 2\lambda_3(4)$
- $\lambda_1(3w_1 + w_2 + w_0 - 1) = 0(5)$
- $\lambda_2(3w_1 - w_2 + w_0 - 1) = 0(6)$
- $\lambda_3(-w_1 - w_0 - 1) = 0(7)$
- $\lambda_1, \lambda_2, \lambda_3 \geq 0$

Περίπτωση όπου $\lambda_1, \lambda_2, \lambda_3 \neq 0$

- $(2) \Rightarrow w_2 = \lambda_1 - \lambda_2$
- $(1) \Rightarrow w_1 = 3(\lambda_1 + \lambda_2) - \lambda_3 = 2\lambda_3$
- $(5) \Rightarrow 3w_1 + w_2 + w_0 - 1 = 0 \Rightarrow 6\lambda_3 + \lambda_1 - \lambda_2 + w_0 - 1 = 0 \Rightarrow w_0 = 1 - 7\lambda_3 + 2\lambda_2(8)$
- $(6) \Rightarrow 3w_1 - w_2 + w_0 - 1 = 0 \Rightarrow 6\lambda_3 - (\lambda_1 - \lambda_2) + w_0 - 1 = 0 \Rightarrow w_0 = 1 - 5\lambda_3 - 2\lambda_2(9)$
- $(7) \Rightarrow -w_1 - w_0 - 1 = 0 \Rightarrow -2\lambda_3 - w_0 - 1 = 0 \Rightarrow w_0 = -2\lambda_3 - 1(10)$
- $(8) = (10) \Rightarrow 1 - 7\lambda_3 + 2\lambda_2 = -2\lambda_3 - 1(10) \Rightarrow \lambda_3 = (2\lambda_2 + 2)/5$
- $(9) = (10) \Rightarrow 1 - 5\lambda_3 - 2\lambda_2 = -2\lambda_3 - 1(10) \Rightarrow \lambda_3 = (2 - 2\lambda_2)/3$
- $(2\lambda_2 + 2)/5 = (2 - 2\lambda_2)/3 \Rightarrow 6\lambda_2 + 6 = 10 - 10\lambda_2 \Rightarrow \lambda_2 = 1/4\lambda_3 = 1/2\lambda_1 = 1/4$
- Άρα τελικά $w_0 = -2, w_1 = 1, w_2 = 0$ Οπότε για την $g(x)$ έχουμε.

$$g(x) = \mathbf{w}^T \mathbf{x} + w_0 = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} + w_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} + -2 = x_1 - 2 = 0 \Rightarrow x_1 = 2$$

Άρα τελικά όπως προβλέφθηκε στην αρχή $g(x) = 2$

Συνεπώς το υπερεπίπεδο διαχωρισμού θα είναι αυτό που παρουσιάζεται παρακάτω.

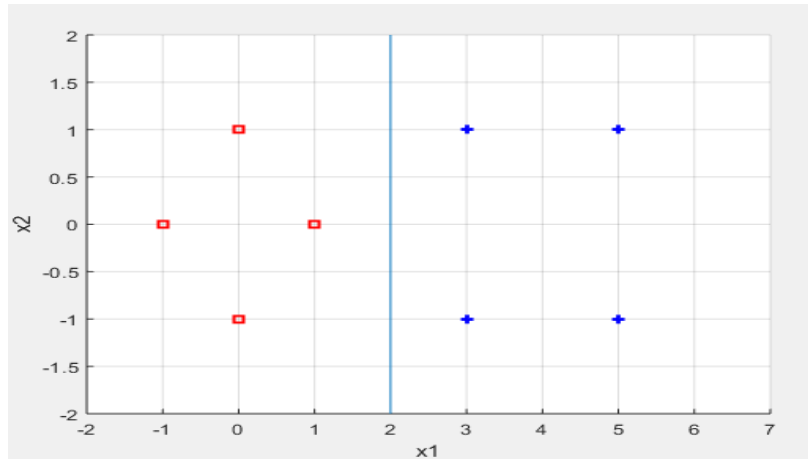


Figure 14: Support Vectors με το υπερεπίπεδο διαχωρισμού

2. Στη συνέχεια εφαρμόστηκε η ίδια διαδικασία για τις ακόλουθες κλάσεις:

- $\omega_1 = \mathbf{x}^+ = \{[2, 2]^T, [2, -2]^T, [-2, -2]^T, [-2, 2]^T\}$
- $\omega_1 = \mathbf{x}^- = \{[1, 1]^T, [1, -1]^T, [-1, -1]^T, [-1, 1]^T\}$

(a) Αρχικά, σχεδιάστηκαν τα δείγματα στον χώρο και το αποτέλεσμα φαίνεται παρακάτω.

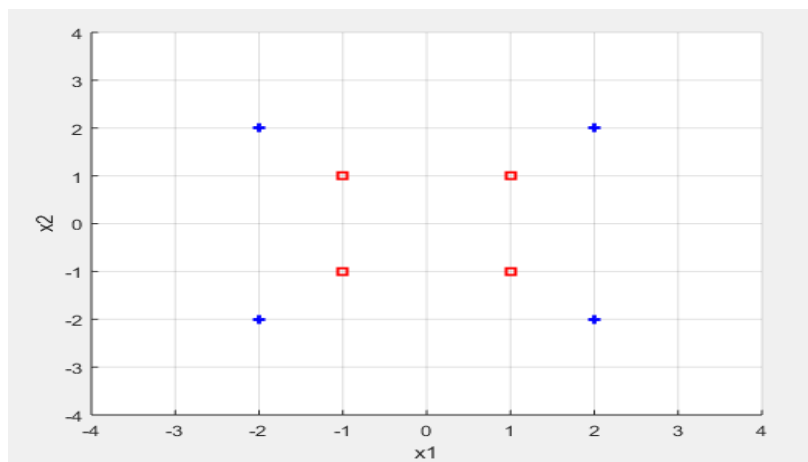


Figure 15: Δείγματα κλάσεων

Παρατηρείται ότι τα δείγματα δεν είναι γραμμικά διαχωρίσιμα καθώς δεν υπάρχει κάποιο υπερπίπεδο που να μπορεί να μεγιστοποιήσει το *margin*. Συνεπώς θα πρέπει να εφαρμοστεί κάποιος μετασχηματισμός ώστε τα δείγματα να μεταφερθούν στο χωρό και να είναι γραμμικά διαχωρίσιμα.

- (b) Στα δείγματα των κλάσεων ω_1 και ω_2 εφαρμόστηκε ο παρακάτω μετασχηματισμός:

$$\Phi(\mathbf{x}) = \mathbf{x} - \|\mathbf{x}\|_2^2 - 4v \|\mathbf{x}\|_2^2 = x_1^2 + x_2^2$$

Συνεπώς εφαρμόζοντας τον μετασχηματισμό προκύπτουν τα εξής δείγματα:

- $\omega_1 = \mathbf{x}^+ = \{[-10, -10]^T, [-10, -14]^T, [-14, -14]^T, [-14, -10]^T\}$
- $\omega_2 = \mathbf{x}^- = \{[-5, -5]^T, [-5, -7]^T, [-7, -7]^T, [-7, -5]^T\}$

Οπότε τα δείγματα που θα προκύψουν θα είναι τα παρακάτω:

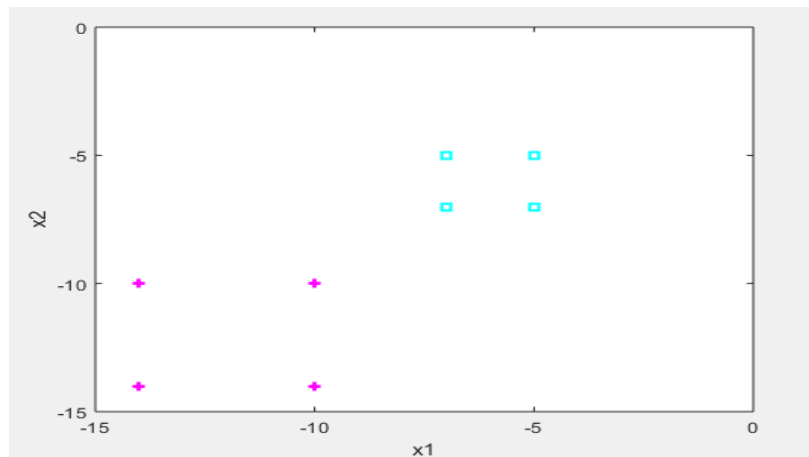


Figure 16: Δείγματα κλάσεων μετά την εφαρμογή του μετασχηματισμού

Συνεπώς παρατηρείται ότι στη συγκεκριμένη περίπτωση τα δείγματα είναι γραμμικά διαχωρίσιμα με Support Vectors να είναι τα δείγματα $[-10, -10]$, $[-7, -7]$. Διαισθητικά το υπερπίπεδο διαχωρισμού θα είναι η ευθεία της μορφής $y = ax + b$ που θα περνάει ανάμεσα από τα Support Vectors και θα μεγιστοποιεί το *margin*.

- (c) Τέλος, υπολογίστηκε η βέλτιστη γραμμή διαχωρισμού με τη χρήση πολλαπλασιαστών Lagrange και τις συνθήκες Karush-Khun-Tucker. Επειδή τα Support Vectors θα είναι 2 τότε και οι πολλαπλασιαστές Lagrange θα είναι 2. Συνεπώς, θα υπάρχουν

3 περιπτώσεις, για $\lambda_1 = 0, \lambda_2 \neq 0$, για $\lambda_2 = 0, \lambda_1 \neq 0$ και για $\lambda_1, \lambda_2 \neq 0$. Θα αναλυθεί μόνο η περίπτωση όπου $\lambda_1, \lambda_2 \neq 0$ καθώς είναι και η περίπτωση από την οποία προκύπτει η $g(x)$. Το υπερεπίπεδο διαχωρισμού θα έχει τη μορφή $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$. Πρέπει να ισχύουν οι εξής περιορισμοί

- $\mathbf{w}^T \mathbf{x} + w_0 \geq 0, \forall \mathbf{x}_+ \in \omega_1$
- $\mathbf{w}^T \mathbf{x} + w_0 \geq 0, \forall \mathbf{x}_- \in \omega_2$

Σκοπός είναι η ελαχιστοποίηση του $J(\mathbf{w}^T, w_0) = 1/2 \|\mathbf{w}\|_2^2 = 1/2(w_1^2 + w_2^2)$. Υποθέτοντας ότι $f_i(\mathbf{w}^T, w_0) = y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1 \geq 0, \forall \mathbf{x}_i \in \omega_1, \omega_2$, με $y_i = 1$ για την ω_1 και $y_i = -1$ για την ω_2 . Συνεπώς προκύπτουν οι παρακάτω εξισώσεις για τα Support Vectors:

- $[-10, -10]^T : -10w_1 - 10w_2 + w_0 - 1 \geq 0$
- $[-7, -7]^T : 7w_1 + 7w_2 - w_0 - 1 \geq 0$

Η βοηθητική συνάρτηση Lagrange την οποία θα ελαχιστοποιήσουμε είναι η εξής: $L(\mathbf{w}, w_0, \lambda) = J(\mathbf{w}, w_0) - \sum_{i=1}^N \lambda_i [y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1]$. Συνεπώς αντικαθιστώντας έχουμε.

$$L(\mathbf{w}, \lambda_1, \lambda_2, \lambda_3) = (w_1^2 + w_2^2)/2 - \lambda_1(-10w_1 - 10w_2 + w_0 - 1) - \lambda_2(7w_1 + 7w_2 - w_0 - 1)$$

Παραγωγίζοντας θα πάρουμε τις εξής εξισώσεις σύμφωνα με τις συνθήκες KKT.

- $\partial(\mathbf{w}, \lambda_1, \lambda_2)/\partial w_1 = w_1 + 10\lambda_1 + 7\lambda_2 = 0 \Rightarrow w_1 = 7\lambda_2 - 10\lambda_1$ (1)
- $\partial(\mathbf{w}, \lambda_1, \lambda_2)/\partial w_2 = w_2 + 10\lambda_1 - 7\lambda_2 = 0 \Rightarrow w_2 = 7\lambda_2 - 10\lambda_1$ (2)
- $\partial(\mathbf{w}, \lambda_1, \lambda_2)/\partial w_0 = -\lambda_1 + \lambda_2 \Rightarrow \lambda_2 = \lambda_1$ (3)
- (1) $\Rightarrow w_1 = 2\lambda_3$ (4)
- $\lambda_1(-10w_1 - 10w_2 + w_0 - 1) = 0$ (5)
- $\lambda_2(7w_1 + 7w_2 - w_0 - 1) = 0$ (6)
- $\lambda_1, \lambda_2 \geq 0$

Περίπτωση όπου $\lambda_1, \lambda_2 \neq 0$

- (1) $\Rightarrow w_0 = 10w_1 + 10w_2 + 1 = -30\lambda_1 - 30\lambda_1 + 1 = -60\lambda_1 + 1$
- (1) $\Rightarrow w_0 = 7w_1 + 7w_2 - 1 = -21\lambda_1 - 21\lambda_1 - 1 = -42\lambda_1 - 1$
- Από τις παραπάνω συνεπάγεται ότι $-60\lambda_1 + 1 = -42\lambda_1 - 1 \Rightarrow \lambda_1 = 1/9$
- (1) $\Rightarrow w_1 = -1/3 = w_2$ και $w_0 = -17/3$
- Άρα τελικά $w_0 = -17/3, w_1 = -1/3, w_2 = -1/3$ Οπότε για την $g(x)$ έχουμε.

$$g(x) = \mathbf{w}^T \mathbf{x} + w_0 = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} + w_0 = \begin{bmatrix} -1/3 \\ -1/3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} - 17/3$$

$$= -1/3x_1 - 1/3x_2 - 17/3 = 0$$

Συνεπώς το επίπεδο διαχωρισμού που προκύπτει παρουσιάζεται παρακάτω.

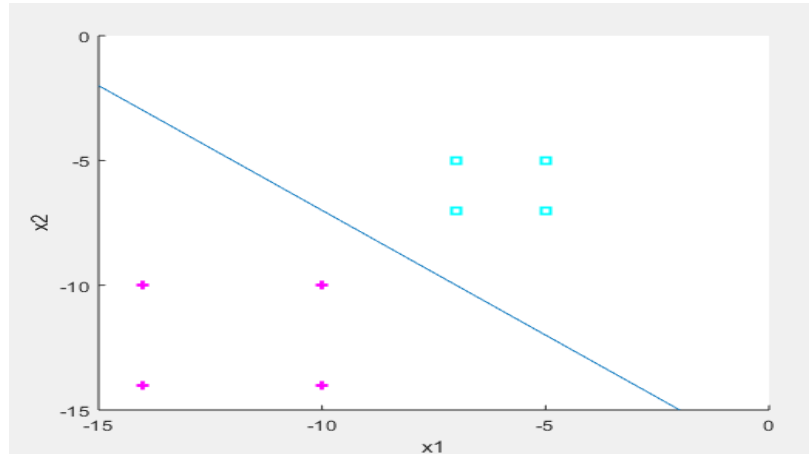


Figure 17: Δείγματα κλάσεων μετά την εφαρμογή του μετασχηματισμού και το υπερεπίπεδο διαχωρισμού

6 Support Vector Machines (Εφαρμογή σε τεχνητό σύνολο δεδομένων)

Στη συγκεκριμένη άσκηση εφαρμόστηκε ο αλγόριθμος των *Support Vector Machines* σε τεχνητό σύνολο δεδομένων. Συγκεκριμένα τα δεδομένα στα οποία θα εφαρμοστεί ο αλγόριθμος βρίσκονται στον πίνακα X όπου οι γραμμές του πίνακα είναι τα δείγματα και οι στήλες τα χαρακτηριστικά. Σκοπός είναι η πρόβλεψη των τιμών y^i από τις αντίστοιχες τιμές x^i .

Μέρος 1

Δεδομένης της Λαγκρανσιανής του δυικού προβλήματος

$$L(\lambda) = \sum_{i=1}^n \lambda^i - 1/2 \sum_{i=1}^n \sum_{j=1}^n \lambda^i \lambda^j y^i y^j K(\mathbf{x}^i, \mathbf{x}^j)$$

όπου $K(\mathbf{x}^i, \mathbf{x}^j) = \mathbf{x}^i * (\mathbf{x}^j)$

Η οποία μεγιστοποιείται κάτω από τους περιορισμούς

$$\lambda^i \geq 0, i = 1, \dots, n$$

$$\sum_{i=1}^n \lambda^i y^i = 0$$

1. Αρχικά, ελαχιστοποιήθηκε η κατάλληλη συνάρτηση υπολογίστηκε και υπολογίστηκαν τα λ^i .
2. Στη συνέχεια υπολογίστηκαν τα βάρη w , το bias w_0 και το margin των support vectors. Βρέθηκε ότι
 - $w = [2, 5414, 3, 0694]$
 - $w_0 = -16,3162$
 - $width = 0,5019$
3. Τέλος σχεδιάστηκαν τα δείγματα τα Support Vectors καθώς και το υπερεπίπεδο διαχωρισμού και το αποτέλεσμα παρουσιάζεται παρακάτω.

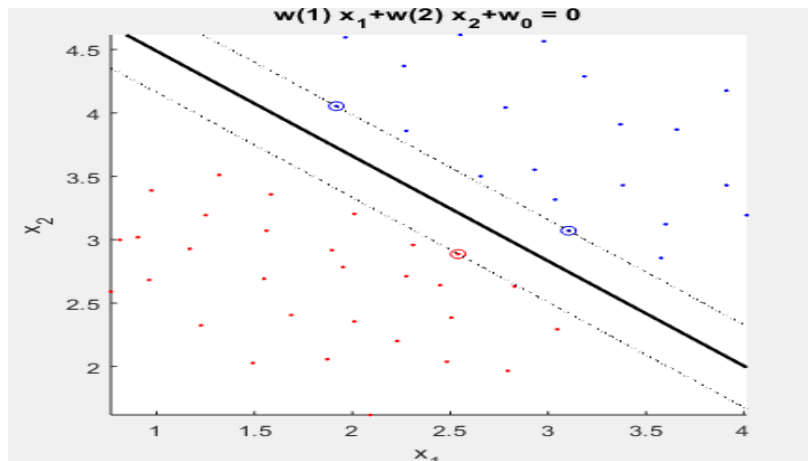


Figure 18: Δείγματα κλάσεων μετά την εφαρμογή του μετασχηματισμού και το υπερεπίπεδο διαχωρισμού

Μέρος 2

Έπειτα εφαρμόστηκε η ίδια διαδικασία με τη διαφορά ότι προστέθηκε ο περιορισμός

$$0 \leq \lambda^i \leq C, i = 1, \dots, n$$

Συνεπώς τα αποτελέσματα παρουσιάζονται παρακάτω.

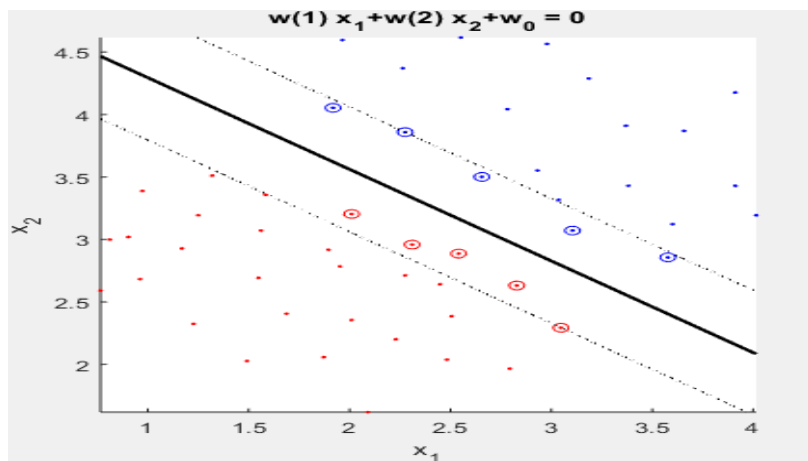


Figure 19: Support Vectors και υπερεπίπεδο διαχωρισμού για $C=1$

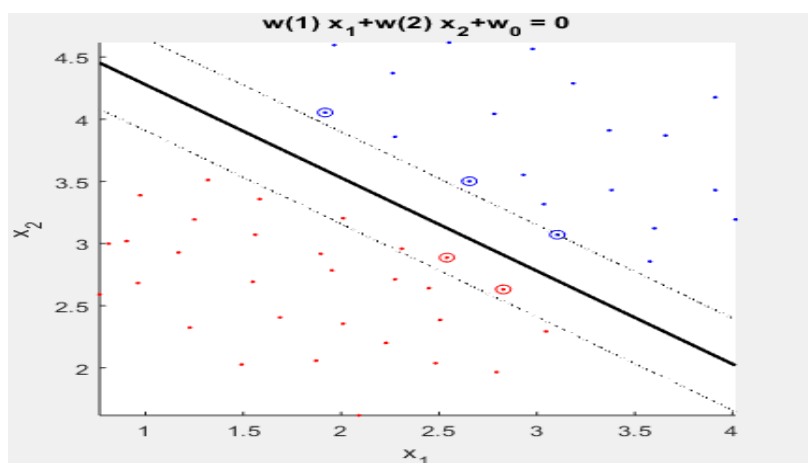


Figure 20: Support Vectors και υπερεπίπεδο διαχωρισμού για $C=5$

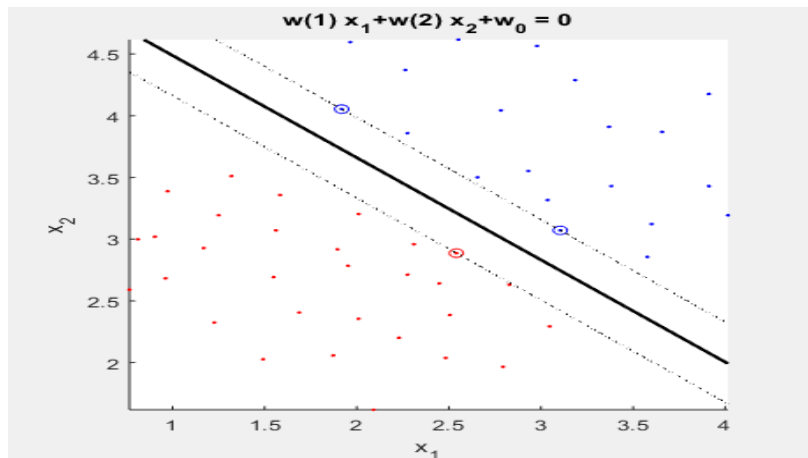


Figure 21: Support Vectors και υπερεπίπεδο διαχωρισμού για $C=100$

Παρατηρείται ότι αυξάνοντας το όριο C μειώνεται ο αριθμός των Support Vectors καθώς και αυξάνεται η ακρίβεια στον υπολογισμό τους.

7 Παραπομπές

- Διαλέξεις μαθήματος και διαφάνειες φροντιστηρίου.
- <https://www.mathworks.com/matlabcentral/fileexchange/68218-svm-basic-support-vector-machines-using-quadprog>
- <https://www.youtube.com/watch?v=hWLdFMccpTY>