**Μηχανική Μάθηση**

1η Σειρά Ασκήσεων

*Ιωάννης Τσαντήλας, 03120883*

[*el20883@mail.ntua.gr*](mailto:el20883@mail.ntua.gr)

*Σημείωση: Όπου χρησιμοποιήθηκε python, υπάρχει το αντίστοιχο footnote.*

Contents

[Άσκηση 1.1 (Linear and Ridge Regression) 1](#_Toc152415501)

[Άσκηση 1.2 (Multivariate Gausian distribution) 4](#_Toc152415502)

[Άσκηση 1.3 (Bayes Classifier) 6](#_Toc152415503)

[Άσκηση 1.4 (Perceptron – Multilayer Perceptron) 9](#_Toc152415504)

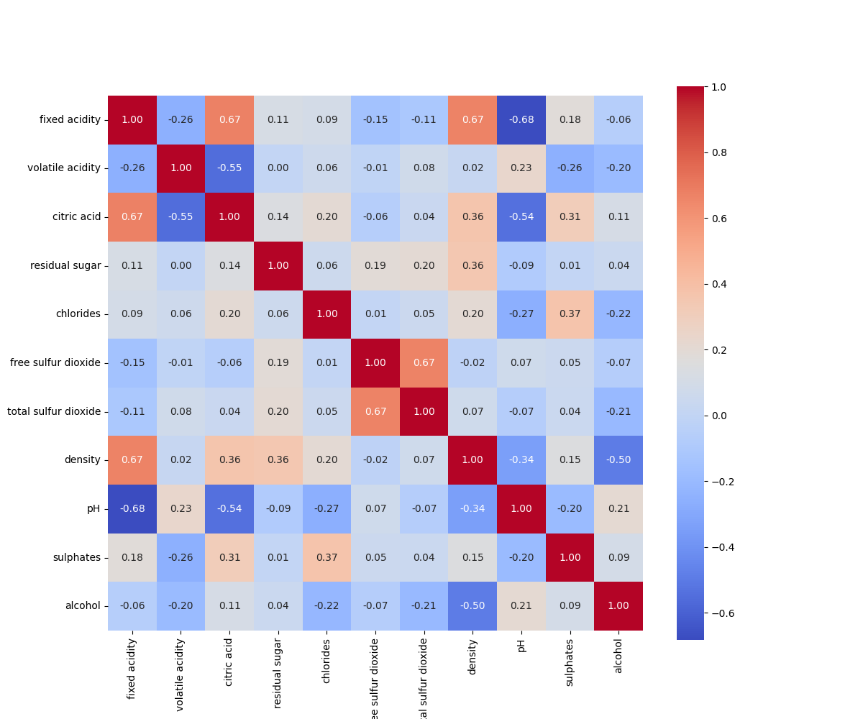
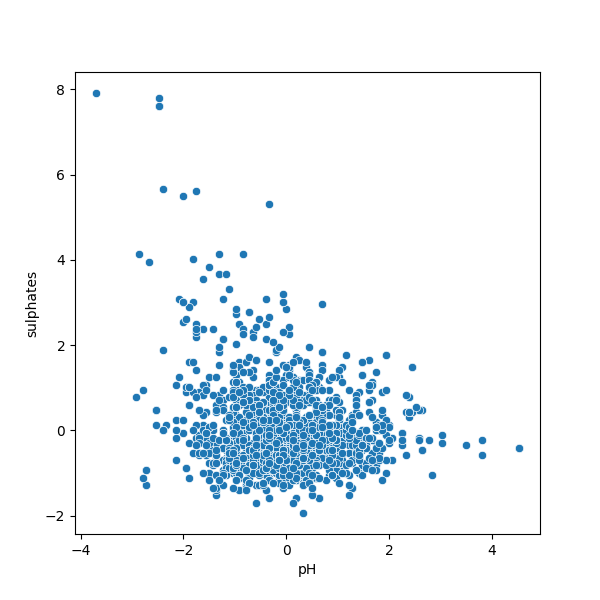
[Άσκηση 1.5 Support Vector Machines – Kernels 10](#_Toc152415505)

[Άσκηση 1.6 (Decision Trees) 11](#_Toc152415506)

# ***Άσκηση 1.1 (Linear and Ridge Regression)[[1]](#footnote-1)***

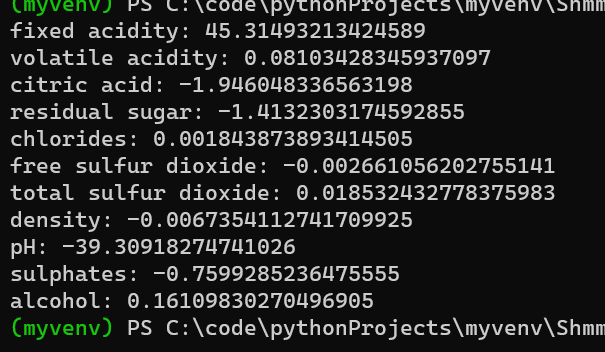
Ερώτημα (α)

Ακολουθούμε τα εξής βήματα: φορτώνουμε τα δεδομένα, αλλάζουμε το σωστό dellimeter (από ; σε ,), κανονικοποιούμε τα δεδομένα, υπολογίζουμε τον κανονικοποιημένο συντελεστή συσχέτισης και τον πίνακα συσχέτισης και τυπώνουμε τα heatmap και scatter plot:

Ερώτημα (β)

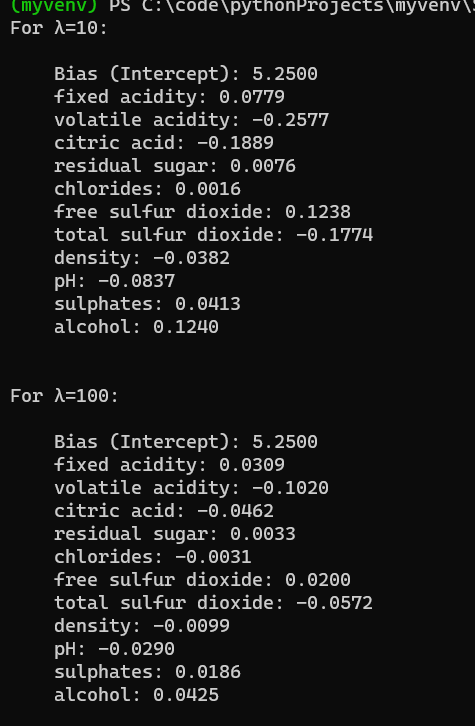
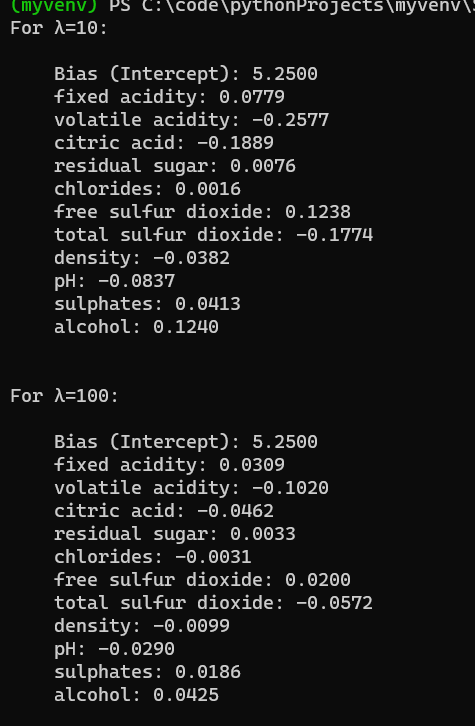
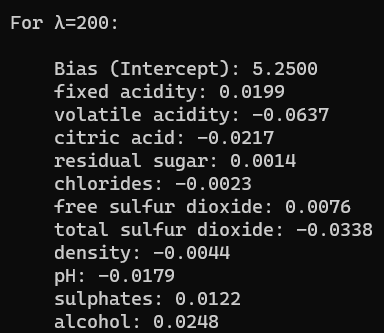
Χωρίζουμε τα δεδομένα σε εκπαίδευση και επαλήθευση, και εφαρμόζουμε την εξίσωση για γραμμική παλινδρόμηση, με Χ να είναι το input και y το output:



Ερώτημα (γ)

Η παλινδρόμηση Ridge εισάγει έναν όρο κανονικοποίησης στη συνάρτηση κόστους, ο οποίος συμβάλλει στην αποφυγή υπερβολικής προσαρμογής. Η εξίσωση για τον υπολογισμό των βαρών στην παλινδρόμηση Ridge, με Χ το input, y το output, λ παράμετρος κανονικοποίησης και Ι ο πίνακας ταυτότητας, είναι:

Παρατηρούμε πως όσο αυξάνετε το *λ*, η απόλυτη τιμή των βαρών μειώνεται, το οποίο αιτιολογείται από την Ridge, αφού μειώνει τα βάρη για να αποφύγει την υπερβολική προσαρμογή.

Ερώτημα (δ)

A graph with red line

Description automatically generated

Ερώτημα (ε)

Θυμίζουμε πως:

Με yi η πραγματική τιμή, η προβλεπόμενη τιμή από το μοντέλο και n το πλήθος των παρατηρήσεων.

A computer screen shot of numbers and letters

Description automatically generated

Παρατηρούμε πως για λ=100 έχω το χαμηλότερο RMSE για την εκπαίδευση, το οποίο υποδηλώνει πως δεν πρέπει να το αυξάνουμε άσκοπα, αλλά να βρούμε μία χρυσή τομή για να πετύχουμε το καλύτερο RMSE.

# ***Άσκηση 1.2 (Multivariate Gaussian distribution)***

Ερώτημα (α)

Η μέση τιμή της υπό συνθήκης μx1|x2 είναι η αναμενόμενη τιμή της x1, θεωρώντας πως x2 = α. Γνωρίζουμε από την κοινή κατανομή Gauss ότι οι x1 και x2 είναι γραμμικά εξαρτημένες. Η γραμμική διακύμανση της x1 στη x2 μας δίνει την καλύτερη γραμμική πρόβλεψη του x1 με βάση το x2. Σε όρους της αναδρομικής εξίσωσης, έχουμε:

Όπου το ϵ αντιπροσωπεύει τον όρο σφάλματος ο οποίος κατανέμεται κανονικά με μηδενική μέση τιμή και διακύμανση σ2x1|x2. Θέτοντας x2 = α, εξαλείφουμε τον όρο σφάλματος για να λάβουμε τον υπό συνθήκη μέσο όρο:

Η διακύμανση του x1 που δεν εξηγείται από το x2, ή η υπό συνθήκη διακύμανση σ2x1|x2, είναι το μέρος του σ21 που δεν εξηγείται από τη συνδιακύμανση σ212. Αυτό μπορεί να προκύψει από τον τύπο για τη διακύμανση μιας υπό συνθήκη κατανομής σε μια διμεταβλητή κανονική:

Αυτός ο τύπος αντιπροσωπεύει τη διακύμανση του όρου σφάλματος ϵ στην παραπάνω αναδρομική εξίσωση.

Ερώτημα (β)

Για να βρούμε την υπό συνθήκη κατανομή του (x1,x2) δεδομένου x3=1, χωρίζουμε το μέσο διάνυσμα και τον πίνακα συνδιακύμανσης σε δύο μέρη: ένα για το (x1,x2) και ένα για το x3:

Όπου ο Σ11 είναι ο πίνακας συνδιακύμανσης των (x1,x2), Σ22 είναι η διασπορά της x3 και ο Σ21 = Σ12Τ είναι ο πίνακας συνδιακύμανσης των (x1,x2) με την x3. Το υπό συνθήκη μέσο διάνυσμα μ(x1,x2)|x3 και ο υπό συνθήκη πίνακας συνδιακύμανσης Σ(x1,x2)|x3 δίνονται από:

Για x3 = 1:

Ερώτημα (γ)

Ομοίως:

Όπου ο Σ11 είναι ο πίνακας συνδιακύμανσης των (x1,x3), Σ22 είναι η διασπορά της x2 και ο Σ21 = Σ12Τ είναι ο πίνακας συνδιακύμανσης των (x1,x3) με την x2. Το υπό συνθήκη μέσο διάνυσμα μ(x1,x3)|x2 και ο υπό συνθήκη πίνακας συνδιακύμανσης Σ(x1,x3)|x2 δίνονται από:

Για x2 = 1:

Ερώτημα (δ)**[[2]](#footnote-2)**

A graph of an isostatic curves

Description automatically generated

# ***Άσκηση 1.3 (Bayes Classifier)***

Ερώτημα (α)

Αφού Σ = Ι, αυτό σημαίνει ότι οι μεταβλητές είναι ανεξάρτητες και έχουν την ίδια διακύμανση. Σε έναν Bayesian ταξινομητή δύο κλάσεων, το όριο απόφασης μεταξύ των δύο κλάσεων μπορεί να βρεθεί θέτοντας τις εκ των υστέρων πιθανότητές τους ίσες μεταξύ τους, πράγμα που απλοποιείται στη σύγκριση των διακριτικών συναρτήσεων τους. Για δύο Gausian κατανομές με ίσους πίνακες συνδιακύμανσης, η συνάρτηση διάκρισης δίνεται από:

Όπου μi είναι ο μέσος όρος της i-οστής κλάσης, Σ είναι ο πίνακας συνδιακύμανσης και P(ωi) είναι η εκ των προτέρων πιθανότητα της κλάσης ωi. Δεδομένου ότι δεν έχουμε πληροφορίες σχετικά με τις προκαταλήψεις, θα υποθέσουμε ότι είναι ίσες, επομένως ο όρος lnP(ωi) μπορεί να αγνοηθεί για το όριο απόφασης. Το όριο απόφασης βρίσκεται τότε όταν g1(x)=g2(x), το οποίο απλοποιείται, δεδομένου ότι ο πίνακας συνδιακύμανσης είναι ο πίνακας ταυτότητας (άρα Σ-1 = Ι), σε:

Και από τα δεδομένα του προβλήματος:

Εδώ, x1 και x2 είναι οι συντεταγμένες στο δισδιάστατο χώρο χαρακτηριστικών R2. Αυτή η γραμμή θα διαχωρίσει τις δύο κλάσεις με βάση τον κανόνα απόφασης του Bayes.

Ερώτημα (β)

Τώρα έχουμε τον κοινό πίνακα συνδιακύμανσης και για τις δύο κλάσεις, ο οποίος δεν είναι ο πίνακας ταυτότητας. Ο πίνακας συνδιακύμανσης είναι:

Για δύο κλάσεις με κοινό πίνακα συνδιακύμανσης, το όριο απόφασης μπορεί ακόμα να βρεθεί θέτοντας τις συναρτήσεις διάκρισης τους ίσες μεταξύ τους. Οι συναρτήσεις διάκρισης για Gausian κατανομές με κοινό αλλά μη ταυτόσημο πίνακα συνδιακύμανσης είναι:

Δεδομένου ότι οι προκαταλήψεις P(ωi) είναι ίσες, ο όρος lnP(ωi) δεν επηρεάζει το όριο της απόφασης και μπορεί να αγνοηθεί. Το όριο απόφασης βρίσκεται τότε όταν g1(x)=g2(x), το οποίο απλοποιείται ως εξής:

Και από τα δεδομένα του προβλήματος**[[3]](#footnote-3)**:

Ερώτημα (γ)

Εδώ, το λij αντιπροσωπεύει το κόστος λανθασμένης ταξινόμησης ενός αντικειμένου από την κλάση i ως κλάση j. Στο πλαίσιο της Bayes θεωρίας αποφάσεων, αυτό σημαίνει ότι εξετάζουμε τώρα μια συνάρτηση απωλειών στη διαδικασία λήψης αποφάσεων. Για δύο κλάσεις, ο κανόνας απόφασης γίνεται:

* Ανάθεσε x στην κλάση ω1, αν .
* Διαφορετικά, αναθέστε x στην κλάση ω2.

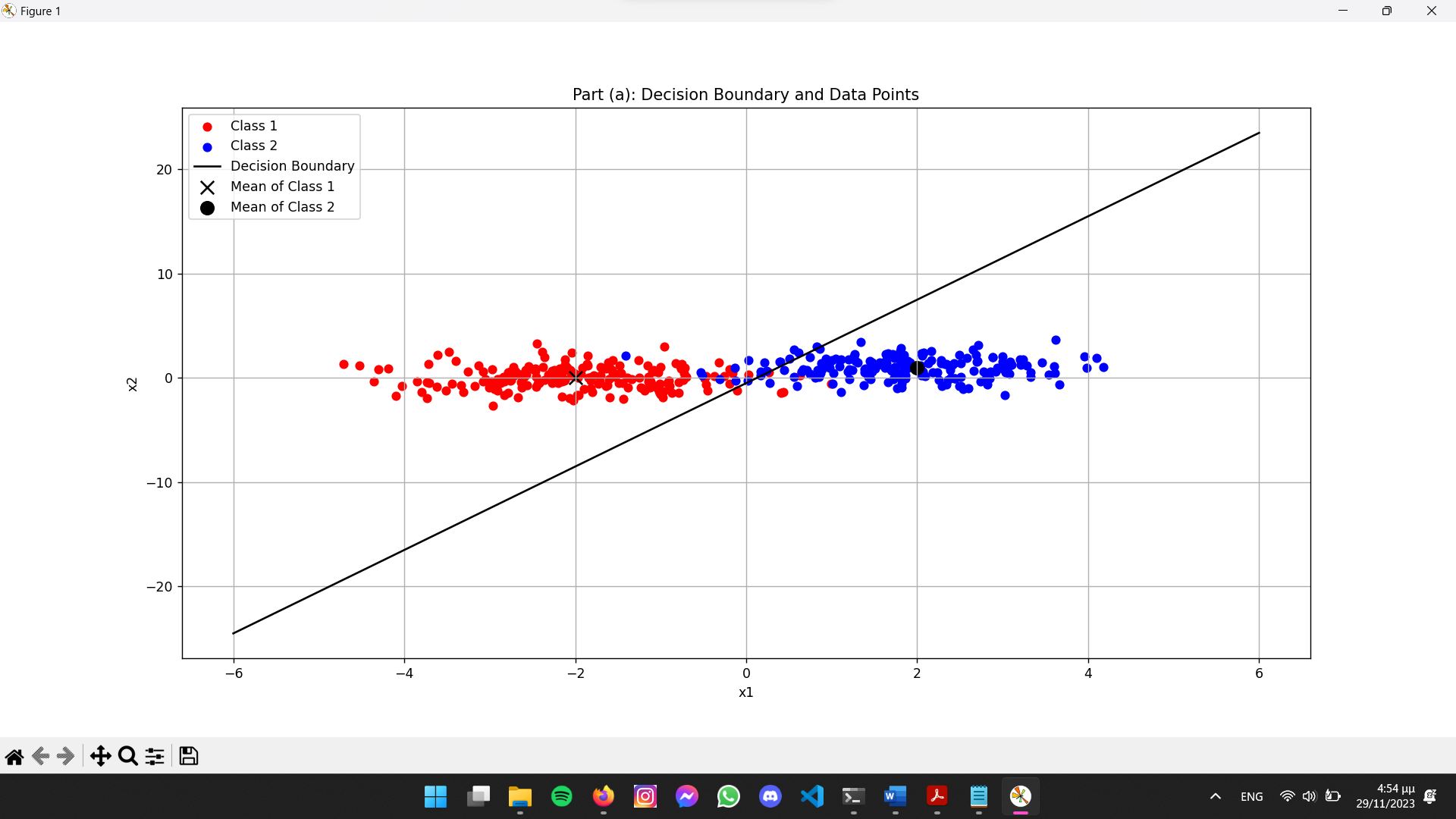
Δεδομένων των εκ των προτέρων πιθανοτήτων των κλάσεων Ρ(ω1) και Ρ(ω2) και των υπό συνθήκη πιθανοτήτων p(x|ω1) και p(x|ω2), ο κανόνας αυτός μπορεί να ξαναγραφεί χρησιμοποιώντας το θεώρημα του Bayes:

* Ανάθεσε x στην κλάση ω1, αν .
* Διαφορετικά, αναθέστε x στην κλάση ω2.

Αφού οι εκ των προτέρων πιθανότητες δεν δίνονται, υποθέτουμε ότι είναι ίσες, δηλαδή Ρ(ω1) = Ρ(ω2). Άρα, η συνθήκη γίνεται:

Και από τα δεδομένα του προβλήματος**[[4]](#footnote-4)**:

Ερώτημα (δ)**[[5]](#footnote-5)**



A screen shot of a computer screen

Description automatically generated

A screen shot of a computer screen

Description automatically generated

Σχόλια:

* Σε όλα τα διαγράμματα, όσο μικρότερη είναι η επικάλυψη, τόσο πιο αποτελεσματικό είναι το όριο απόφασης στην ορθή ταξινόμηση των σημείων. Οι θέσεις των κέντρων των κλάσεων (μέσων) σε σχέση με τα όρια απόφασης παρέχουν πληροφορίες για το πόσο καλά αντιπροσωπεύεται κάθε κλάση από το μέσο της και πώς οι εν λόγω μέσοι επηρεάζουν τα όρια απόφασης.
* Συγκρίνοντας τα (α) και (β) δείχνει πώς το σχήμα και ο προσανατολισμός του ορίου απόφασης μπορεί να αλλάξει με βάση τον πίνακα συνδιακύμανσης των κατανομών, ακόμη και όταν οι μέσοι όροι των κλάσεων παραμένουν αμετάβλητοι.
* Το (γ) δείχνει πώς η ενσωμάτωση διαφορετικού κόστους για την εσφαλμένη ταξινόμηση μπορεί να αλλάξει σημαντικά τη φύση του ορίου απόφασης.

# ***Άσκηση 1.4 (Perceptron – Multilayer Perceptron)***

Ερώτημα (α)**[[6]](#footnote-6)**

A screenshot of a computer

Description automatically generated

Ερώτημα (β)**[[7]](#footnote-7)**

1. ***Νεύρων 1 (Ν1)***: Αυτός ο νευρώνας θα δημιουργήσει ένα κατακόρυφο όριο στο x1 = 0. Έτσι, θέτουμε w1 = 1, w2 = 0 και μια προκατάληψη 0.
2. ***Νεύρων 2 (Ν2)***: Αυτός ο νευρώνας θα δημιουργήσει ένα οριζόντιο όριο στο σημείο x2 = 0. Έτσι, θέτουμε w1 = 0, w2 = 1 και μια προκατάληψη 0.
3. ***Νεύρων Εξόδου 3 (Ν3)***: Αυτός ο νευρώνας θα λάβει τις εξόδους από τα Ν1 και Ν2. Εάν Ν1 = 1 (x1 > 0) και η Ν2 = 0 (x2 < 0), τότε η περιοχή είναι γκρίζα και θα πρέπει να χαρακτηριστεί ως θετική. Τα βάρη από το N1 και το N2 στο N3 θα πρέπει να ρυθμιστούν έτσι ώστε η έξοδος να είναι 1 μόνο όταν N1 = 1 και το N2 = 0. Αυτό θα μπορούσε να επιτευχθεί με ένα βάρος 1 από το N1, ένα βάρος -1 από το N2 και μια μεροληψία 0,5 (για να εξασφαλιστεί ότι το ίδιο το όριο ταξινομείται ως θετικό).

A screenshot of a computer

Description automatically generated

# ***Άσκηση 1.5 Support Vector Machines – Kernels***

|  |  |  |
| --- | --- | --- |
|  | **SVM** | **Περιγραφή** |
| **α)** |  | Σχετικά ευθεία γραμμή με ανεκτές κάποιες λανθασμένες ταξινομήσεις (δεδομένου ότι το C είναι χαμηλό, πράγμα που σημαίνει ότι η ποινή για λανθασμένη ταξινόμηση είναι χαμηλή). |
| **β)** |  | Ευθεία γραμμή με λιγότερες λανθασμένες ταξινομήσεις, επειδή ένα υψηλότερο C σημαίνει υψηλότερη ποινή για λανθασμένες ταξινομήσεις. |
| **γ)** |  | Πολυωνυμικός πυρήνας βαθμού 2, ο οποίος θα καμπυλώσει το όριο απόφασης. Το όριο θα είναι μια καμπύλη που επιτρέπει κάποιο διαχωρισμό μεταξύ των κλάσεων. |
| **δ)** |  | Gausian πυρήνας (RBF) με ορισμένη παράμετρο διασποράς (σχετίζεται με το 0,25 στον εκθέτη). Το όριο περικλείει μια κλάση εξ ολοκλήρου ή καμπυλώνει γύρω από μια κλάση στενά. |
| **ε)** |  | Ομοίως με (δ), αλλά με πιο στενή εξάπλωση λόγω του μεγαλύτερου αρνητικού εκθέτη, που οδηγεί σε ένα όριο απόφασης που αγκαλιάζει στενά τα σημεία δεδομένων μιας κλάσης. |

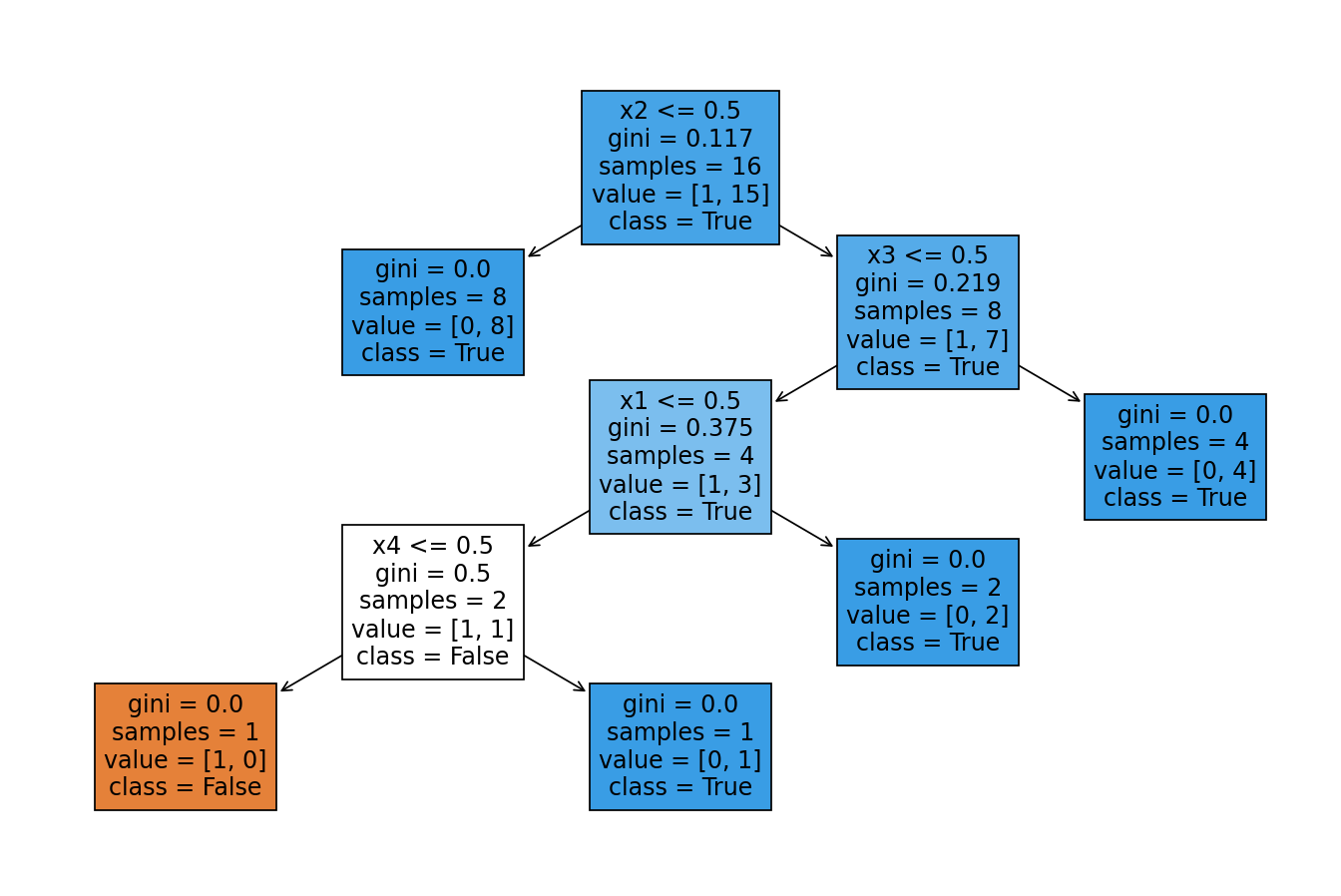
|  |  |  |
| --- | --- | --- |
| **Σχήμα** | **Περιγραφή** | **SVM** |
| **1)** | Μη γραμμικό όριο απόφασης που δεν φαίνεται να ταιριάζει πολύ στενά με τα δεδομένα. Αυτό είναι πιο πιθανό να είναι ένας πολυωνυμικός πυρήνας. | **(γ)** |
| **2)** | Καμπύλο όριο που ταιριάζει αρκετά σφιχτά στα δεδομένα, πιθανώς ενδεικτικό ενός πυρήνα RBF με μικρότερη διασπορά (πιο σφιχτή προσαρμογή). | **(ε)** |
| **3)** | Ευθεία γραμμή με πολύ λίγες λανθασμένες ταξινομήσεις, γεγονός που υποδηλώνει υψηλότερη τιμή C για ένα γραμμικό SVM. | **(β)** |
| **4)** | Ευθεία γραμμή, αλλά με μερικές λανθασμένες ταξινομήσεις, γεγονός που υποδηλώνει χαμηλότερη τιμή C για ένα γραμμικό SVM. | **(α)** |
| **5)** | Μη γραμμικό όριο με μέτρια προσαρμογή, το οποίο θα μπορούσε να υποδηλώνει έναν πυρήνα RBF με μεγαλύτερη διασπορά (πιο χαλαρή προσαρμογή). | **(δ)** |
| **6)** | Καμπύλη με στενή προσαρμογή γύρω από μια κλάση, γεγονός που υποδηλώνει έναν πυρήνα RBF με στενή εξάπλωση. | **(ε)** |

# ***Άσκηση 1.6 (Decision Trees)***

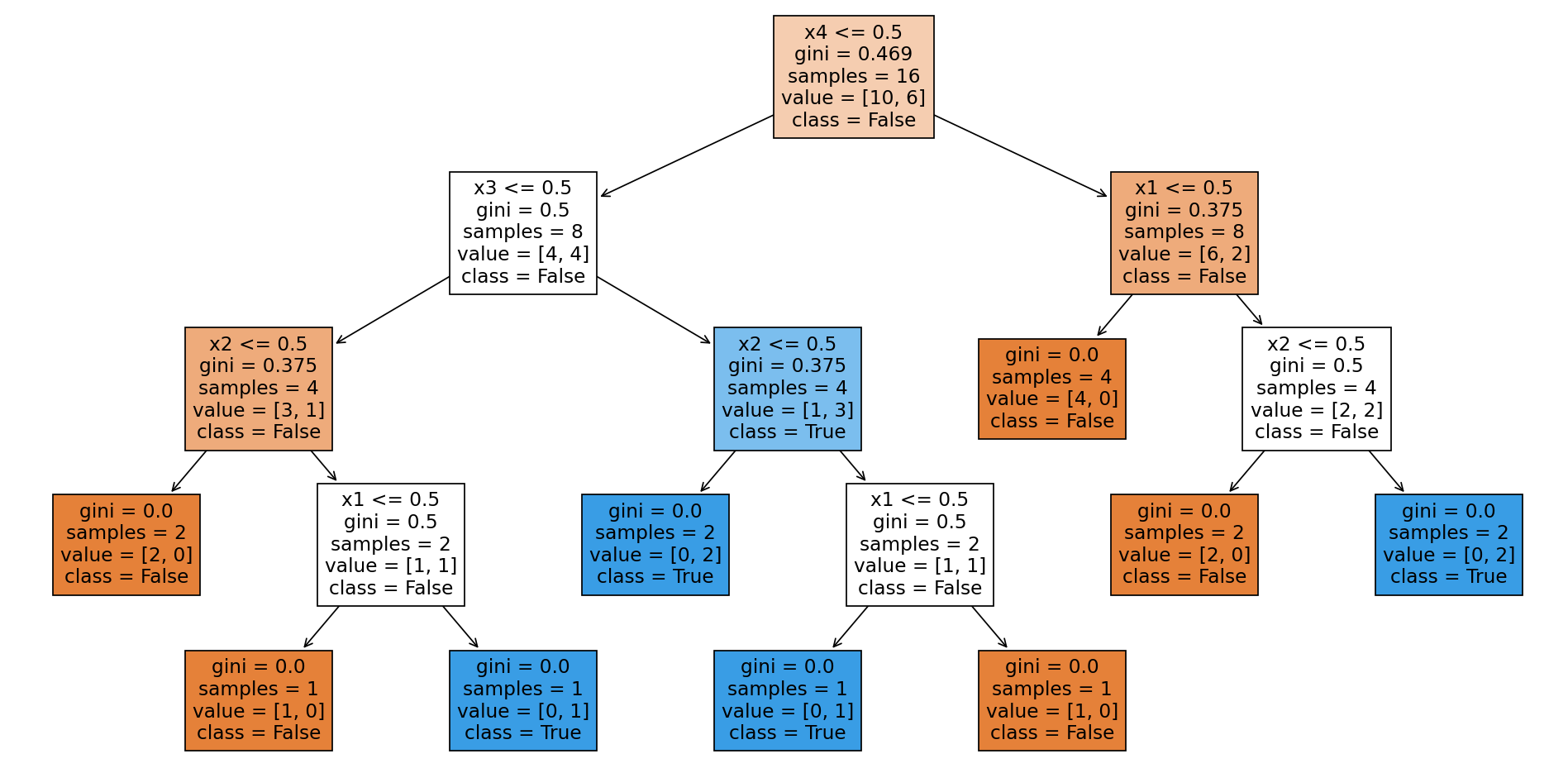
Ερώτημα (α)

**Υποερώτημα (1)**: Για το συγκεκριμένο δέντρο, κάθε κόμβος αντιπροσωπεύει ένα από τα x1, x2, ..., xk και έχει δύο διακλαδώσεις, που αντιπροσωπεύουν το αληθές και το ψευδές του xi. Αφού το fn εκφράζεται με n διαζεύξεις, το βάθος του δέντρου στην απλούστερη περίπτωση, θα είναι n, με κάθε επίπεδο να αντιπροσωπεύει μία από τις διαζεύξεις. Τέλος, τα φύλλα αντιπροσωπεύουν την έξοδο του ταξινομητή (αληθής/ ψευδής). Η προσέγγιση στο 1 σημαίνει πως πρέπει να ταξινομεί σωστά τις περιπτώσεις με περιθώριο σφάλματος 1. Επομένως, το δέντρο πρέπει να καλύπτει τα περισσότερα, αλλά όχι απαραίτητα όλα, τα σενάρια που περιγράφονται από τις n διαζεύξεις. Ένα παράδειγμα δέντρου απόφασης, n=2:

Εδώ, πρώτα ελέγχουμε x1 ή ¬x2, και στη συνέχεια, με βάση αυτό, ελέγχουμε x3 ή x4. Η απεικόνιση του**[[8]](#footnote-8)**:

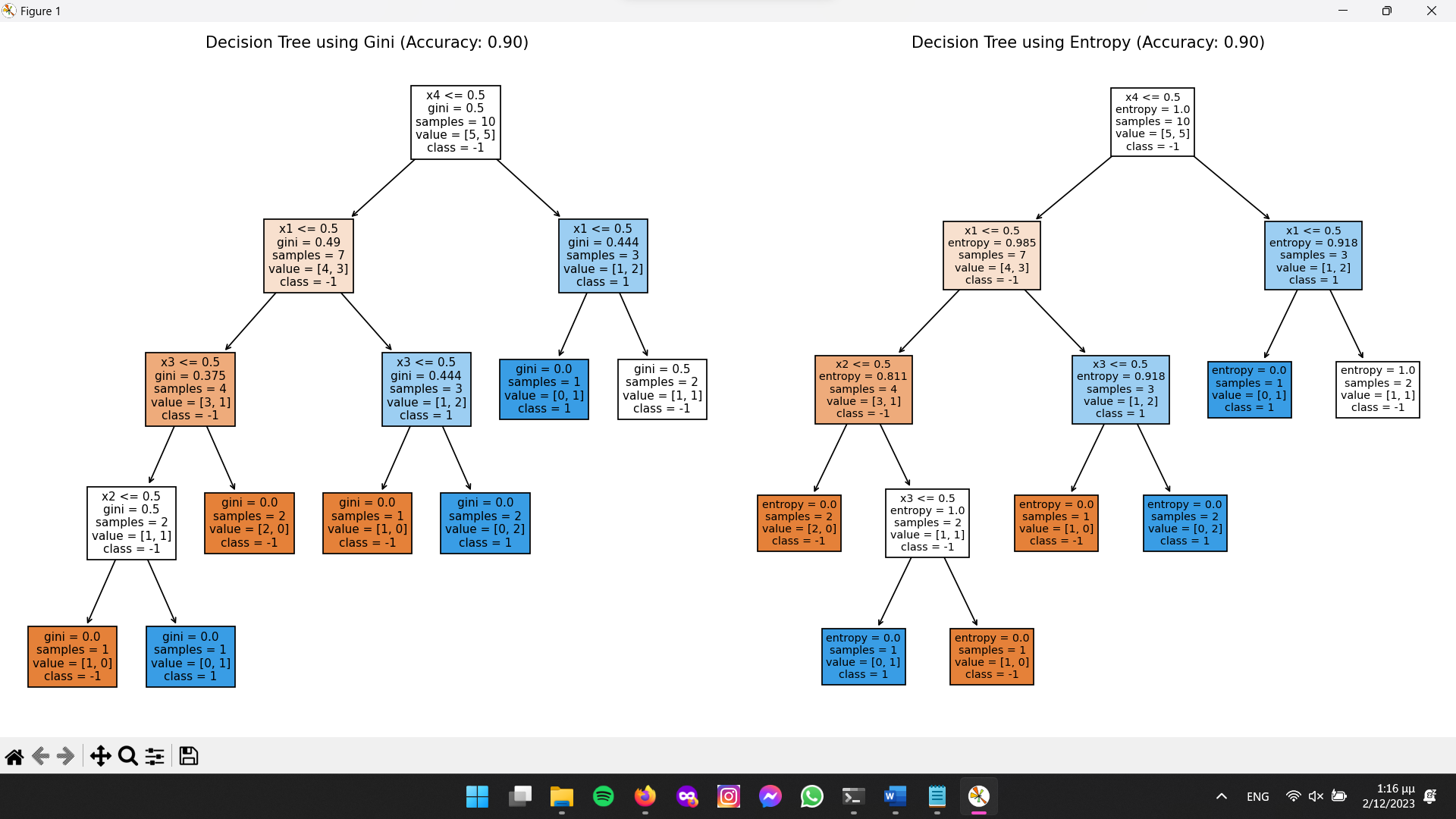


**Υποερώτημα (2)**: Για το συγκεκριμένο δέντρο, κάθε κόμβος θα πρέπει να αναπαριστά μια σύζευξη (πράξη AND) δύο χαρακτηριστικών ή τις αρνήσεις τους. Δεδομένου ότι έχουμε να κάνουμε με αποκλίσεις (πράξεις XOR), πρέπει να διασφαλίσουμε ότι οι διακλαδώσεις του δέντρου αντιπροσωπεύουν σενάρια όπου ακριβώς μία από τις συζυγίες είναι αληθής, αλλά όχι και οι δύο. Ομοίως, τα φύλλα θα αντιπροσώπευαν την τελική ταξινόμηση (αληθής/ ψευδής). Τέλος, η ακρίβεια 1 σημαίνει πως το δέντρο πρέπει να ταξινομεί τέλεια όλες τις πιθανές περιπτώσεις σύμφωνα με την fn. Ένα παράδειγμα δέντρου απόφασης, n=2[[9]](#footnote-9):

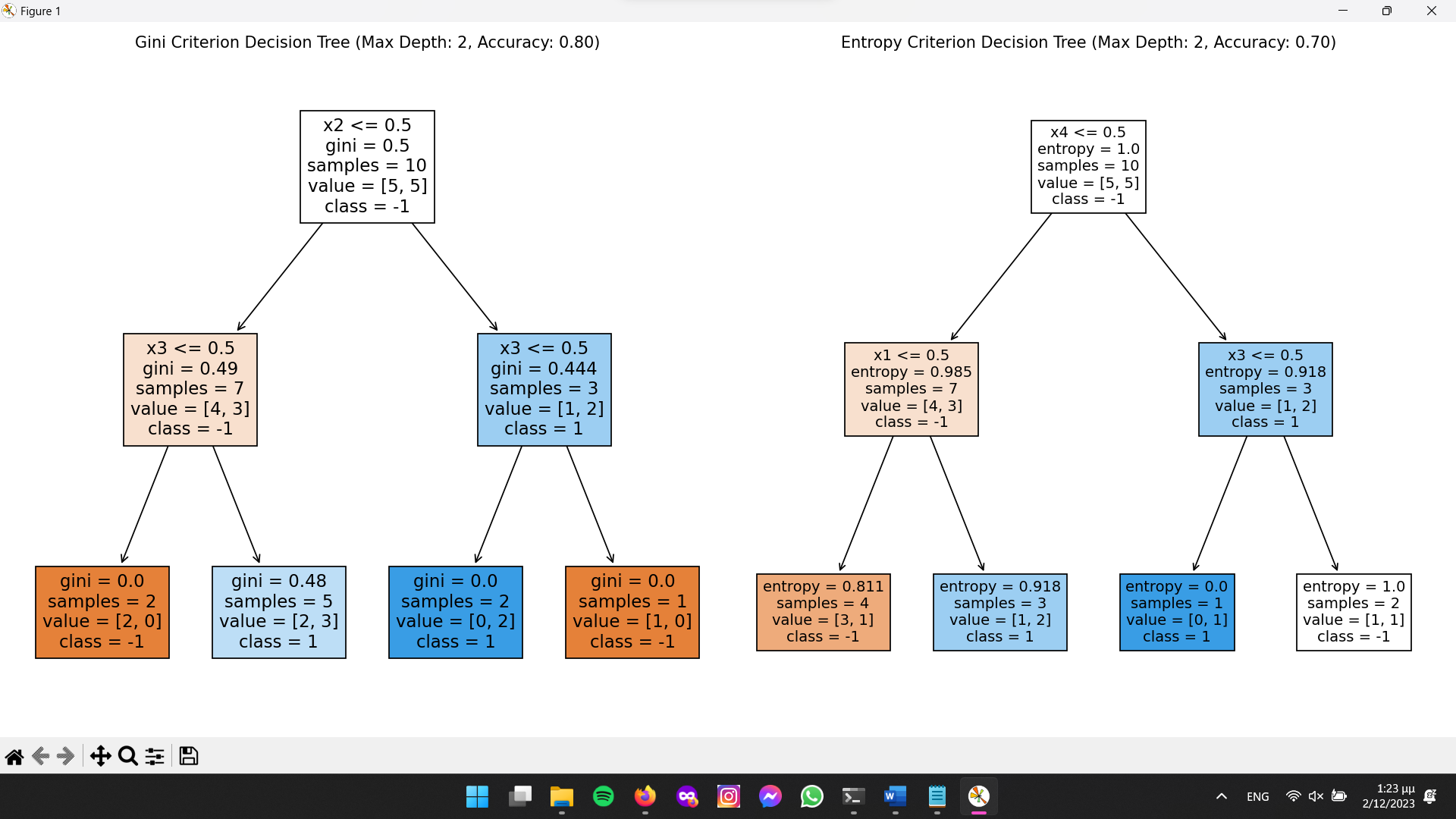


Ερώτημα (β)

**Υποερώτημα (1)**:[[10]](#footnote-10)



**Υποερώτημα (2)**:[[11]](#footnote-11)



**Υποερώτημα (3)**:

Υπάρχουν διάφορες τεχνικές για να βελτιστοποιήσουμε την αποδοτικότητα. Μεταξύ αυτών, είναι να παραλληλοποιήσουμε τις αποφάσεις σε διάφορα βήματα του δέντρου, ώστε να γίνονται ταυτόχρονα, ενώ μπορούμε να δημιουργήσουμε ένα πιο αποδοτικό dataset, π.χ. σε μορφή binary search tree. Μπορούμε να σταματήσουμε την εκτέλεση ενός βήματος, εάν προσθέτοντας ένα επιπλέον επίπεδο δεν βελτιώνει την. Τέλος, μπορούμε να ελέγχουμε συστηματικά τα βάθη των δέντρων χρησιμοποιώντας αναζήτηση πλέγματος και να διασταυρώνουμε για να ελέγξουμε την απόδοση τους σε κάθε βάθος. Τελικά, έχουμε[[12]](#footnote-12):

A screenshot of a computer

Description automatically generated

1. Κάθε ερώτημα έχει το αντίστοιχο αρχείο του: “Ex\_1\_1a.py”, “Ex\_1\_1b.py”, “Ex\_1\_1c.py”, “Ex\_1\_1d.py”, “Ex\_1\_1e.py” [↑](#footnote-ref-1)
2. “Ex\_1\_2\_plot.py” [↑](#footnote-ref-2)
3. “Ex\_1\_3b\_eq.py” [↑](#footnote-ref-3)
4. “Ex\_1\_3c\_eq.py” [↑](#footnote-ref-4)
5. “Ex\_1\_3d\_plot\_a.py, Ex\_1\_3d\_plot\_b.py, Ex\_1\_3d\_plot\_c.py” [↑](#footnote-ref-5)
6. “Ex\_1\_4a.py” [↑](#footnote-ref-6)
7. “Ex\_1\_4b.py” [↑](#footnote-ref-7)
8. “Ex\_1\_6a\_1.py” [↑](#footnote-ref-8)
9. “Ex\_1\_6a\_2.py” [↑](#footnote-ref-9)
10. “Ex\_1\_6b\_1.py” [↑](#footnote-ref-10)
11. “Ex\_1\_6b\_2.py” [↑](#footnote-ref-11)
12. “Ex\_1\_6b\_3.py” [↑](#footnote-ref-12)