



Τεχνητή Νοημοσύνη και Μηχανική Μάθηση

Κωνσταντίνος Καραμανής

The University of Texas at Austin & Archimedes/Athena RC

constantine@utexas.edu

<https://caramanis.github.io/>





Ας θυμηθούμε τα
προηγούμενα...



Πρόβλεψη Διαβήτη

Ο Διαβήτης είναι η 8η κυριότερη αιτία θανάτου στις ΗΠΑ. Ένα από τα βασικότερα προβλήματα είναι η σωστή διάγνωση, όσο και η πρόβλεψη, και η πρόληψη.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

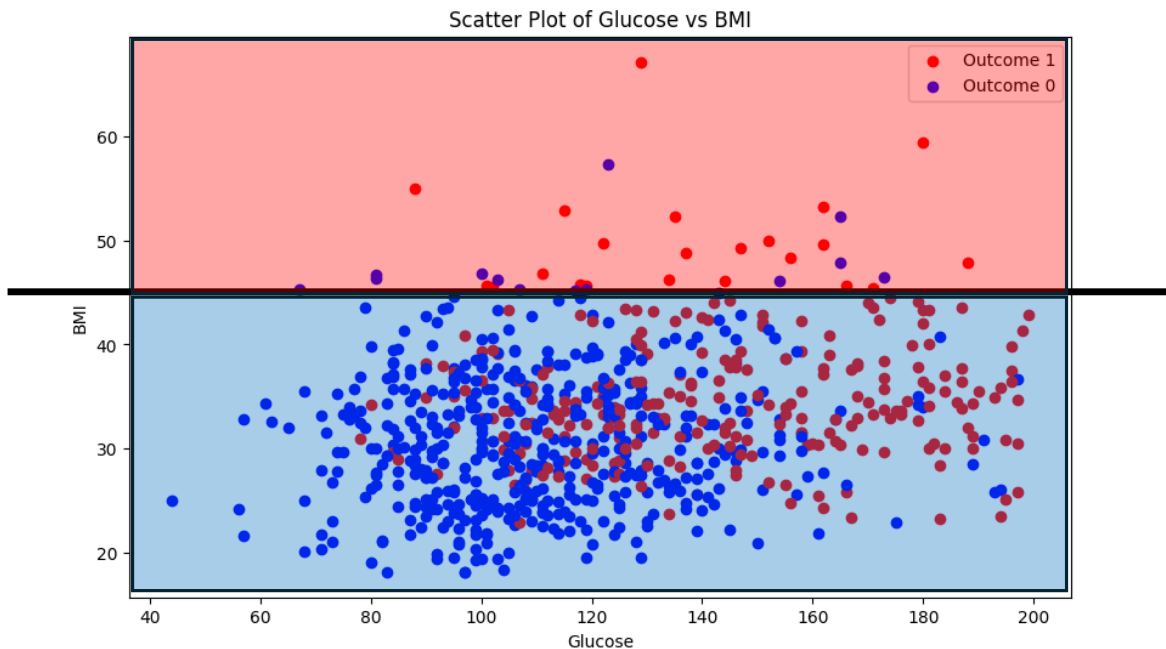
Πρόβλεψη Διαβήτη

Ο Διαβήτης είναι η 8η κυριότερη αιτία θανάτου στις ΗΠΑ. Ένα από τα βασικότερα προβλήματα είναι η σωστή διάγνωση, όσο και η πρόβλεψη, και η πρόληψη.

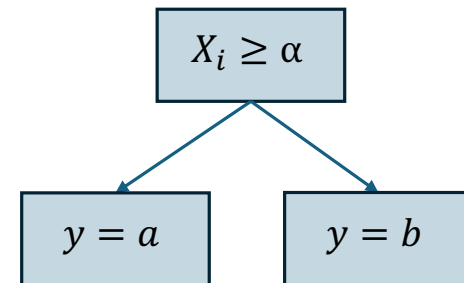
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

X_1 X_2 y

Πρώτος Αλγόριθμος: Δέντρο Απόφασης



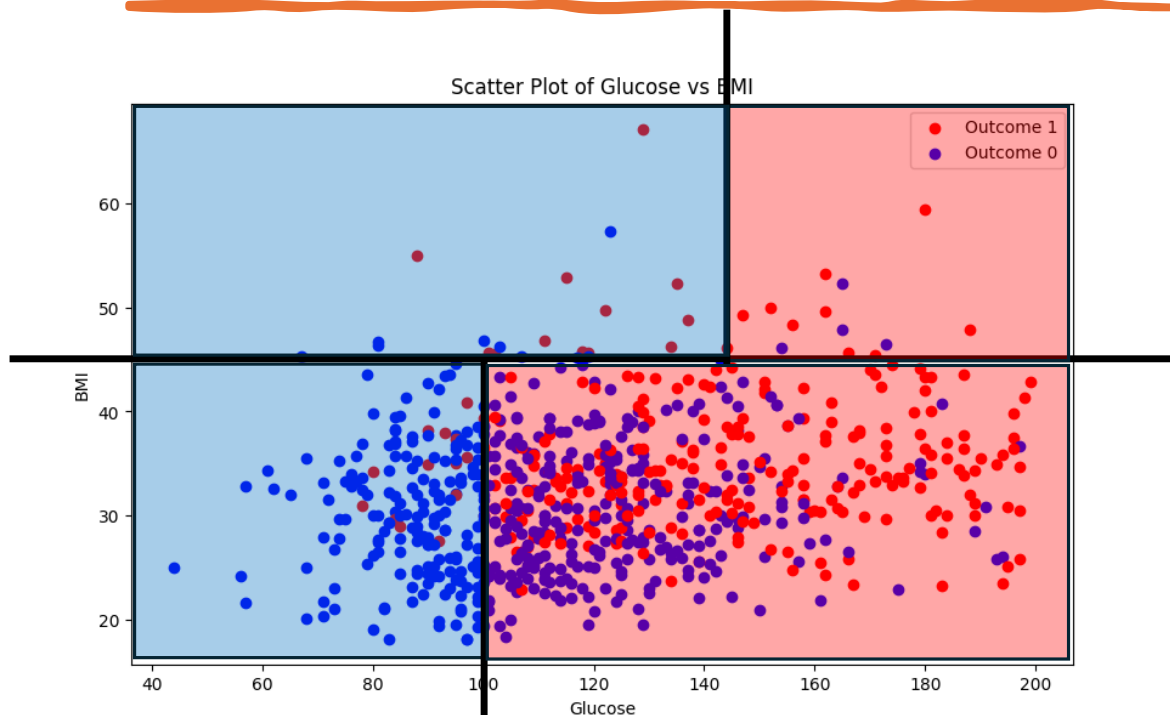
Δέντρο απόφασης με βάθος = 1



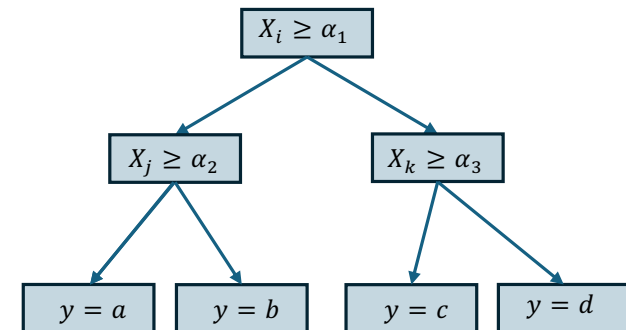
Έχουμε 4 παραμέτρους:

- (1) $i = 2$
- (2) $\alpha = 45$
- (3) $a = \text{μπλέ}$
- (4) $b = \text{κόκκινο}$

Πρώτος Αλγόριθμος: Δέντρο Απόφασης



Δέντρο απόφασης με βάθος = 2



Έχουμε 10 παραμέτρους:

(1-3) $i = 2, j = 1, k = 1$

(4-6) $\alpha_1 = 45, \alpha_2 = 100, \alpha_3 = 145,$

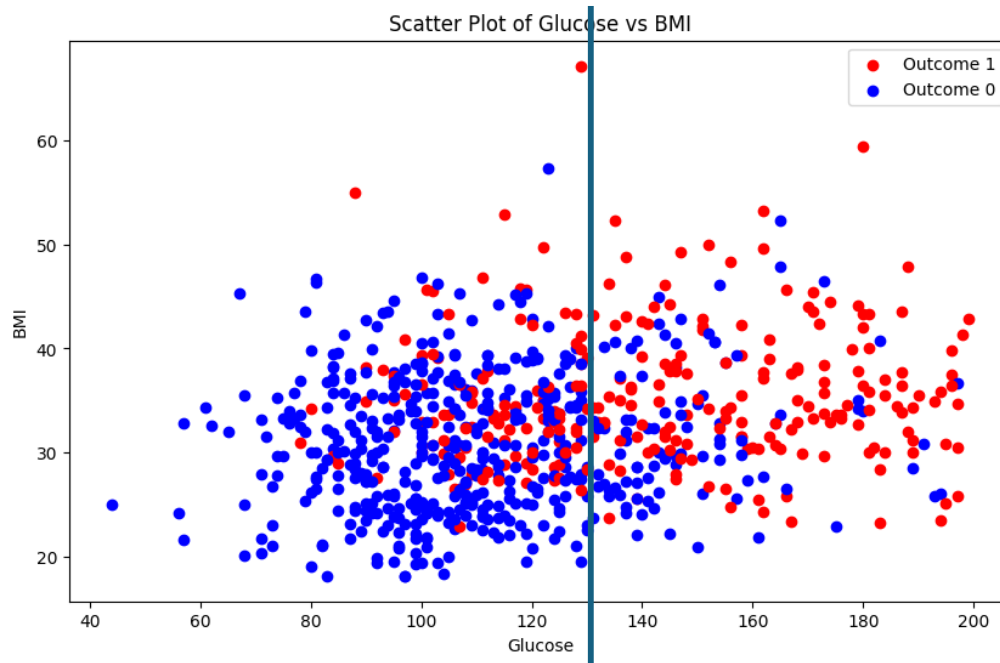
(7) $a = \text{μπλέ}$

(8) $b = \text{κόκκινο}$

(9) $c = \text{μπλέ}$

(10) $d = \text{κόκκινο}$

Το «καλύτερο» δέντρο βάθους 1



Το καλύτερο δέντρο είναι αυτό που ελαχιστοποιεί τα λάθη!

Εκπαίδευση Αλγορίθμου

1. Επιλέγουμε την οικογένεια αλγορίθμων: παράδειγμα – δέντρα απόφασης βάθους 3

2. Βρίσκουμε τις παραμέτρους που ελαχιστοποιούν τα λάθη στα δεδομένα μας

Εκπαίδευση Αλγορίθμου

1. Επιλέγουμε την οικογένεια αλγορίθμων:
παράδειγμα – δέντρα απόφασης βάθους 3

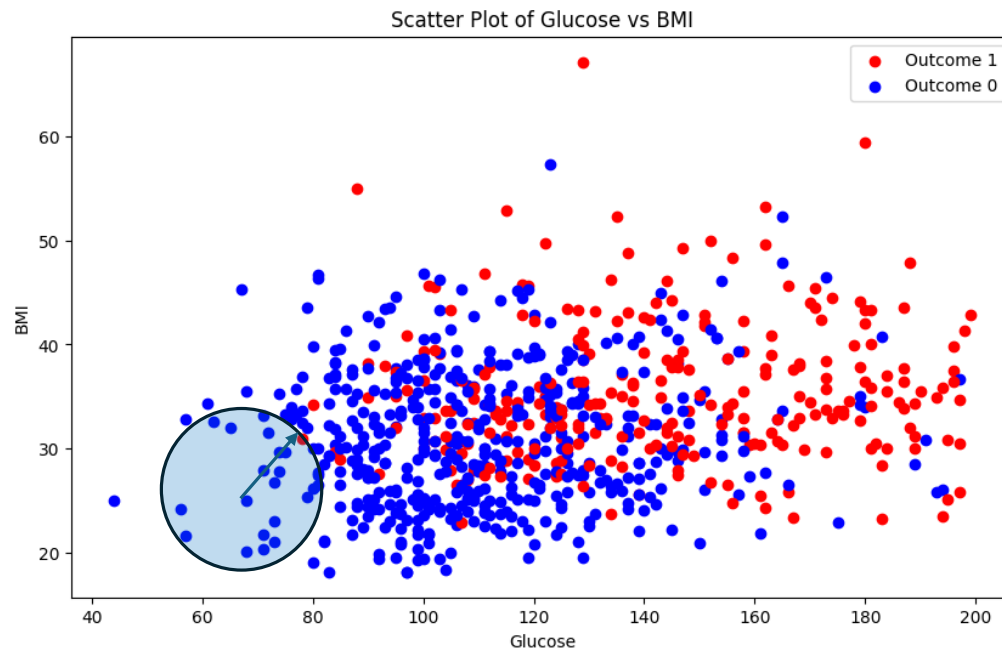
```
model = DecisionTreeClassifier(max_depth = 3)
```

2. Βρίσκουμε τις παραμέτρους που
ελαχιστοποιούν τα λάθη στα δεδομένα μας

```
model.fit(X,y)
```

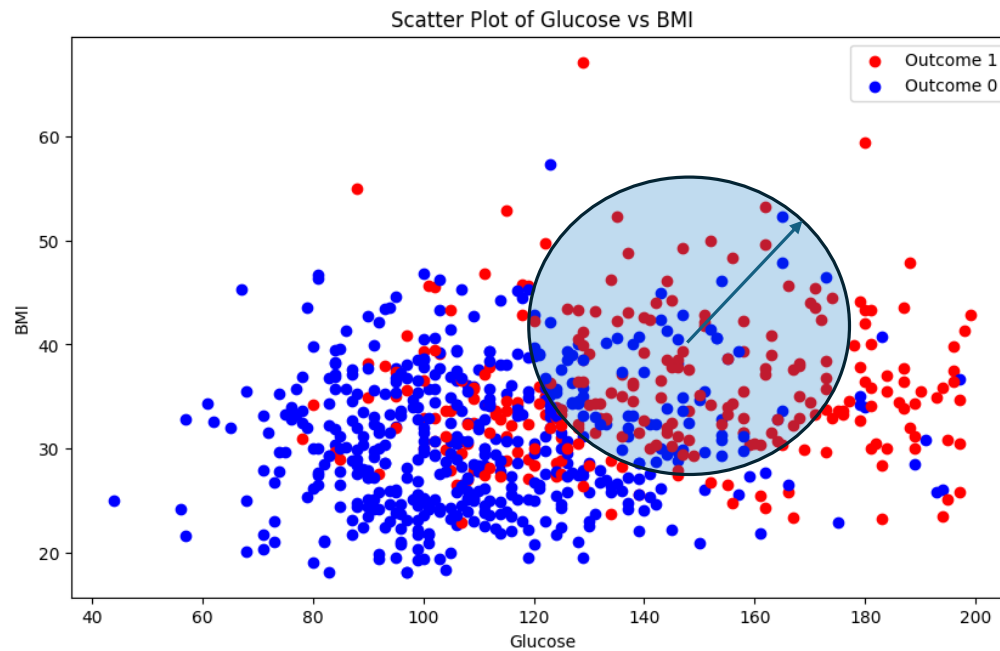
```
model.predict(x)
```

Δεύτερο Παράδειγμα



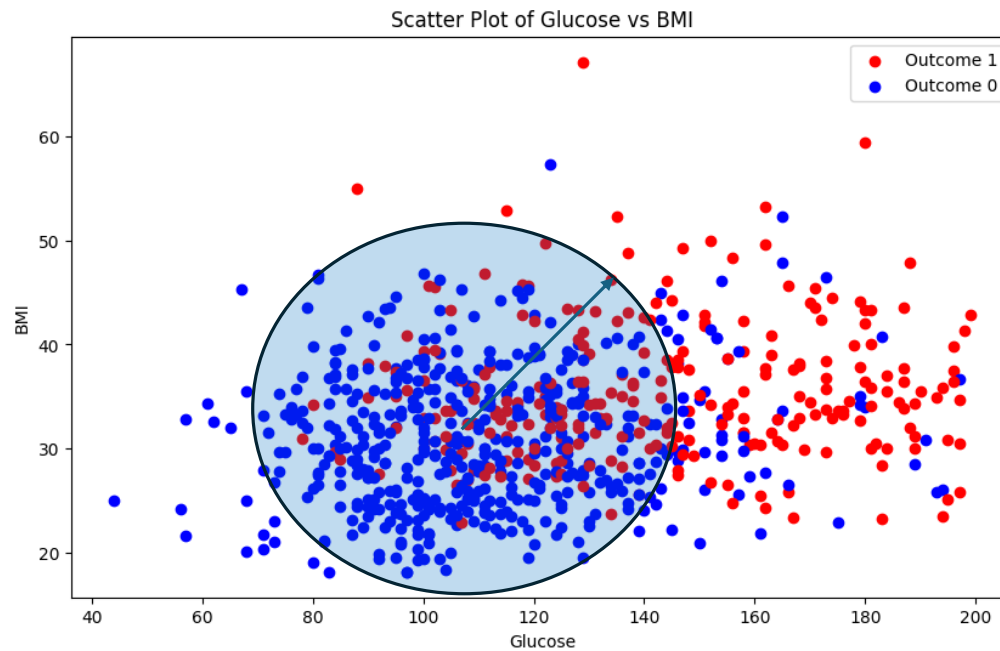
Κέντρο \mathbf{c} με ακτίνα \mathbf{p} :
εντός κύκλου μπλέ,
εκτός κόκκινο

Δεύτερο Παράδειγμα



Κέντρο \mathbf{c} με ακτίνα \mathbf{p} :
εντός κύκλου μπλέ,
εκτός κόκκινο

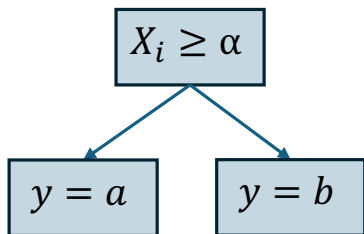
Δεύτερο Παράδειγμα



Κέντρο c με ακτίνα p :
εντός κύκλου μπλέ,
εκτός κόκκινο

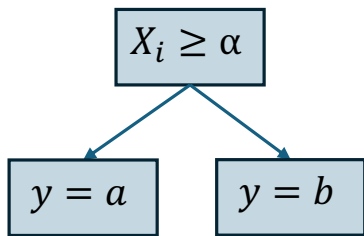
Τρίτο Παράδειγμα

Δέντρο απόφασης με βάθος = 1

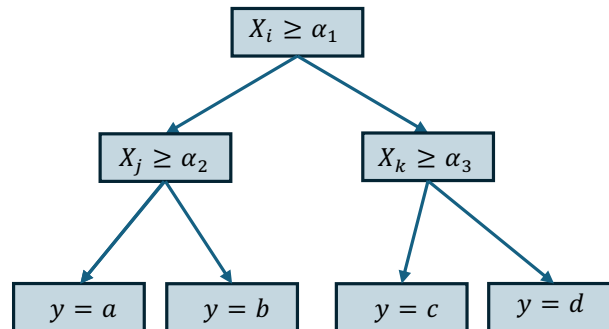


Τρίτο Παράδειγμα

Δέντρο απόφασης με βάθος = 1

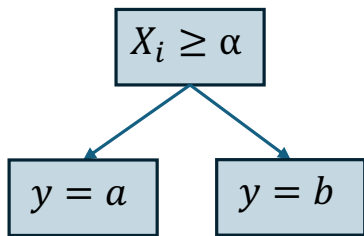


Δέντρο απόφασης με βάθος = 2

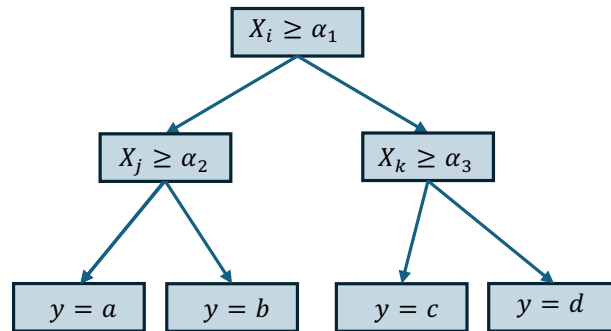


Τρίτο Παράδειγμα

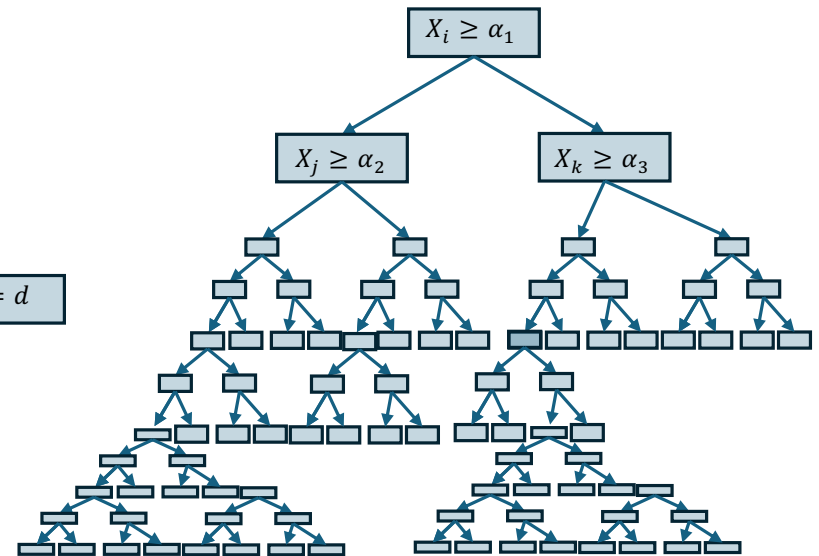
Δέντρο απόφασης με βάθος = 1



Δέντρο απόφασης με βάθος = 2

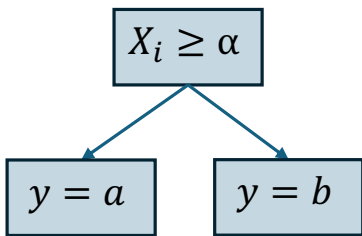


Δέντρο απόφασης με βάθος = 10+

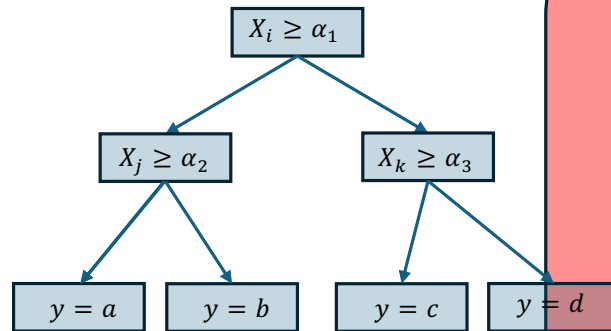


Τρίτο Παράδειγμα

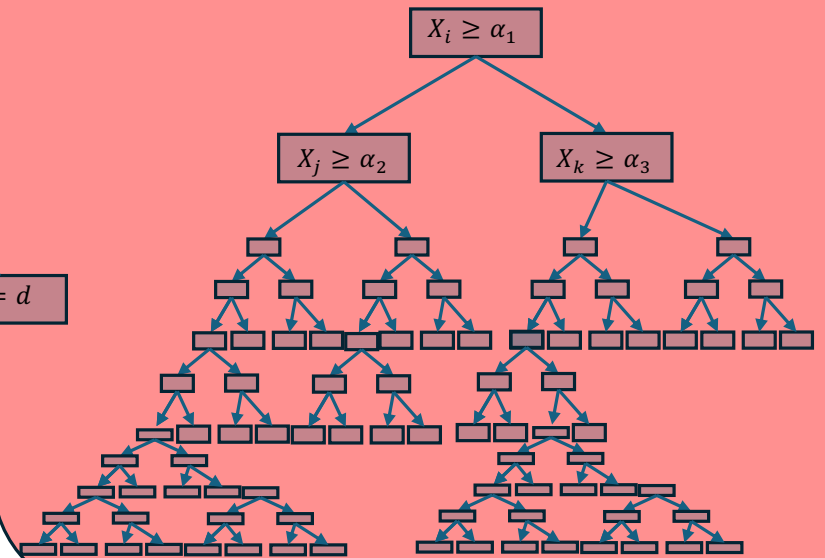
Δέντρο απόφασης με βάθος = 1



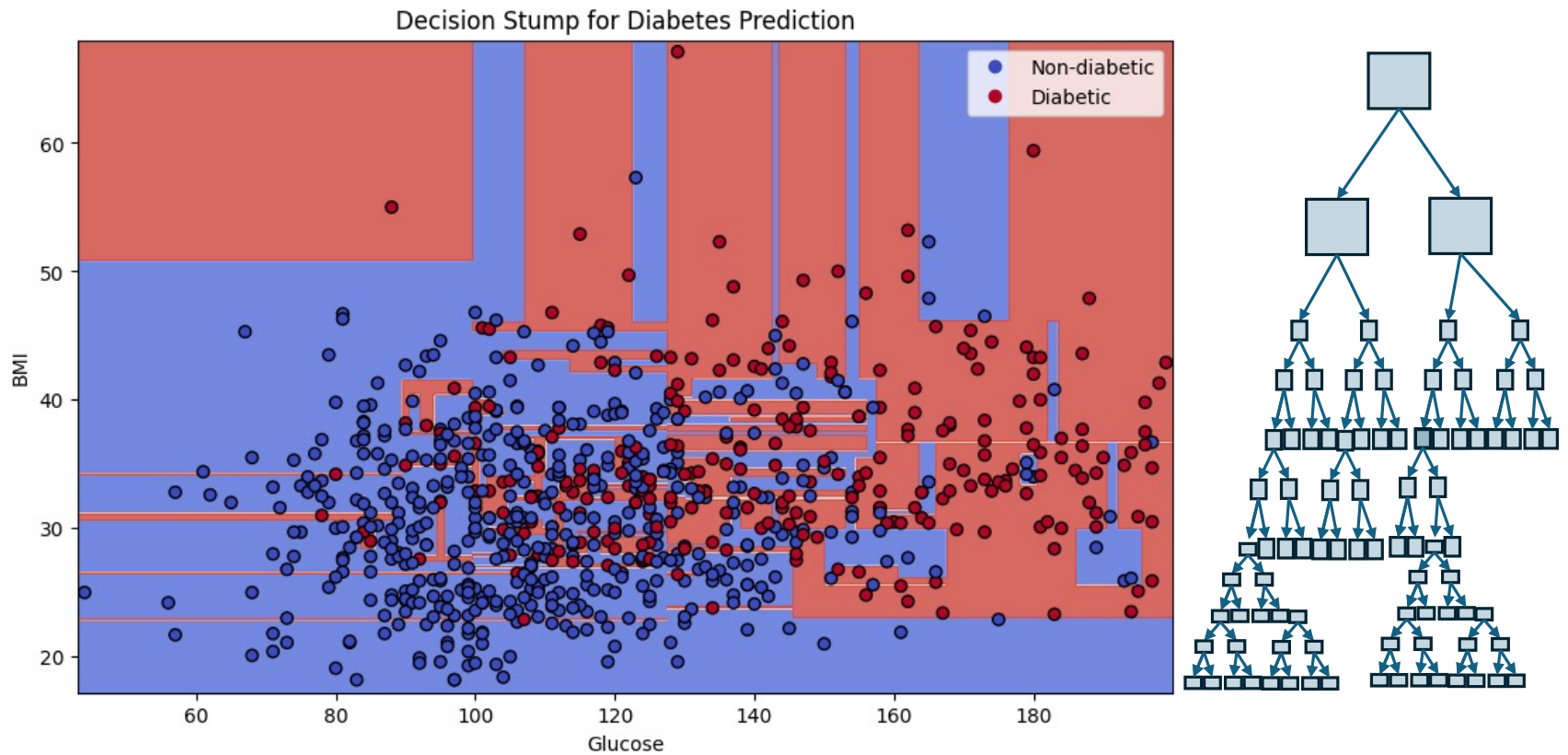
Δέντρο απόφασης με βάθος = 2



Δέντρο απόφασης με βάθος = 10+

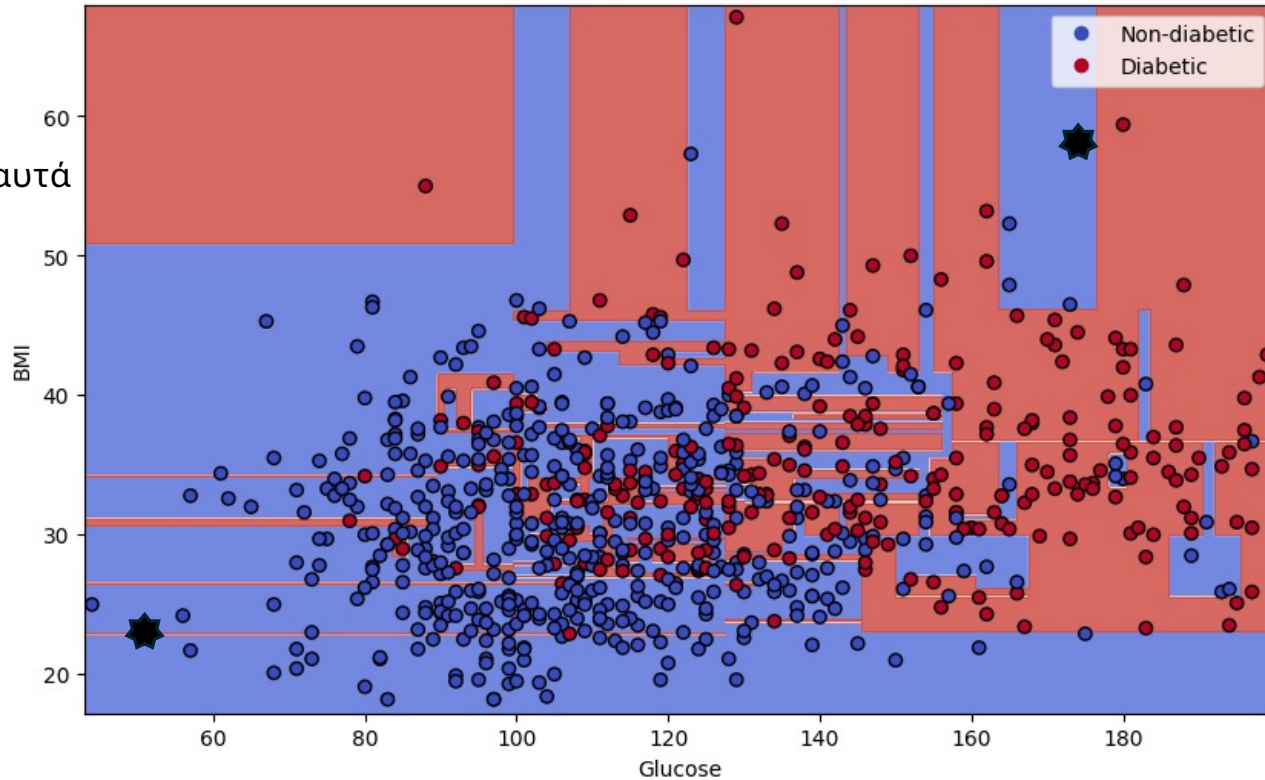


Τρίτο Παράδειγμα

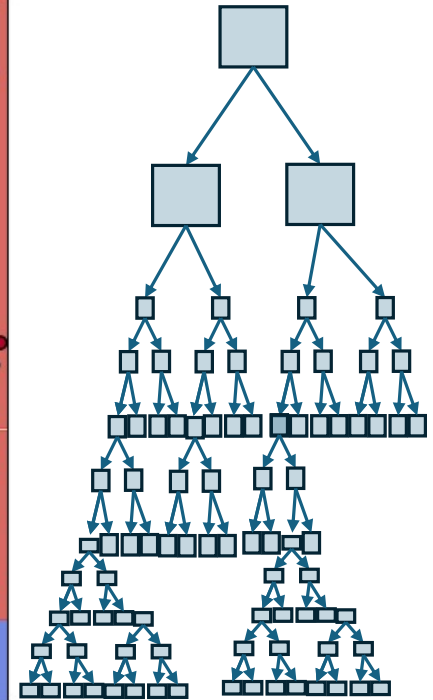


Τρίτο Παράδειγμα: Γιατί δεν μας αρέσει;

Decision Stump for Diabetes Prediction

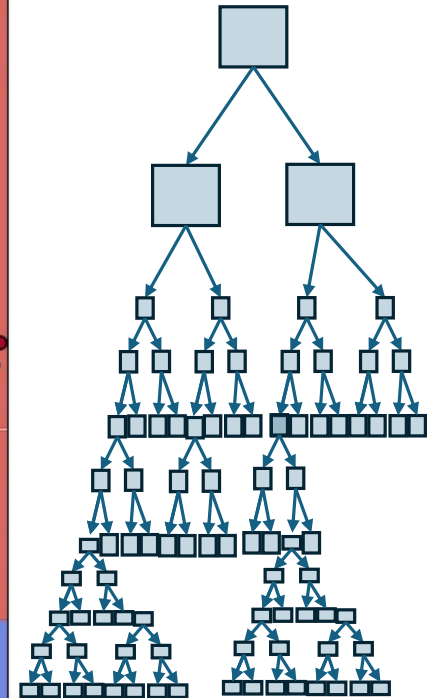
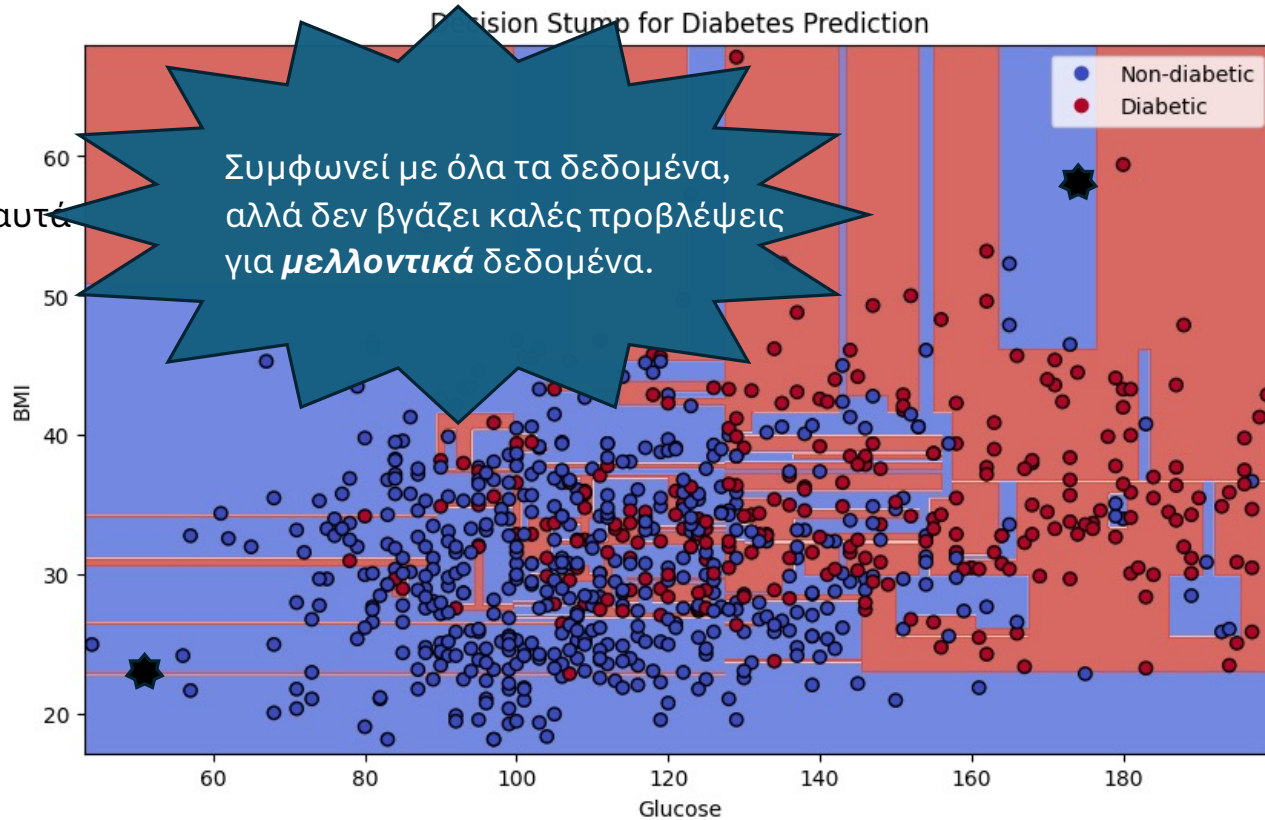


Τι λέει αυτός ο
αλγόριθμος για αυτά
τα δύο σημεία;

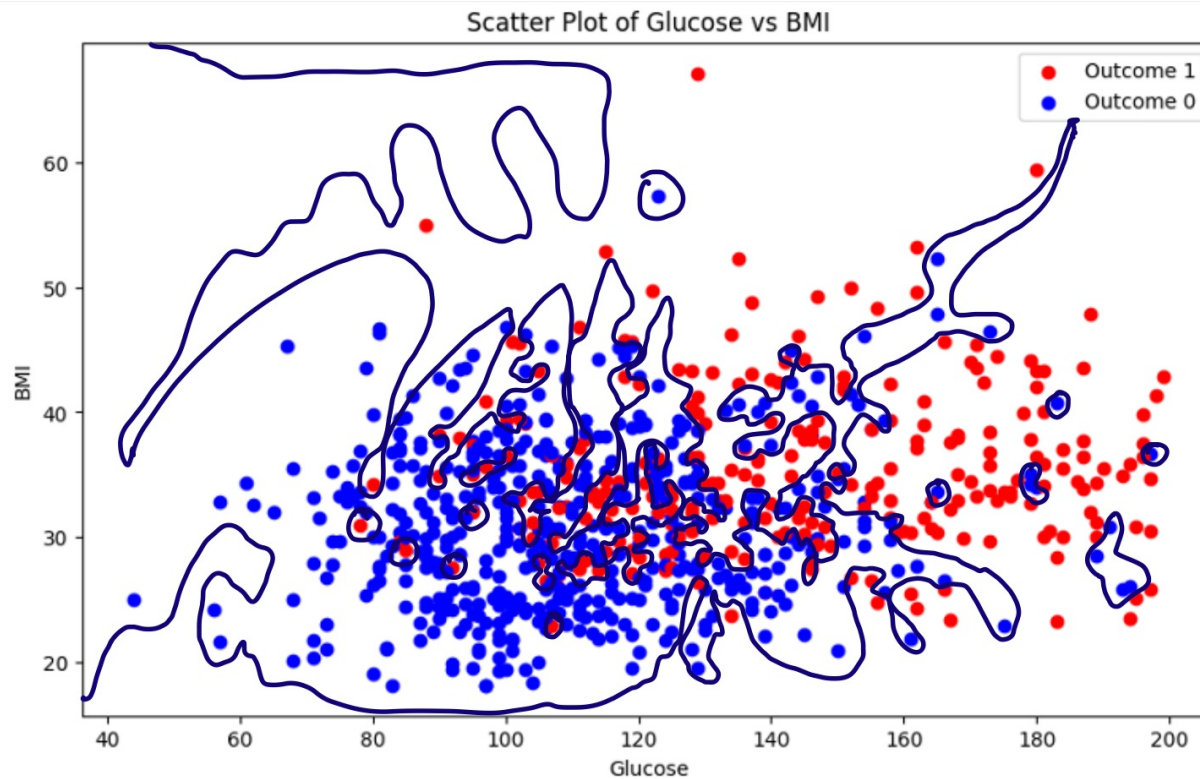


Τρίτο Παράδειγμα: Γιατί δεν μας αρέσει;

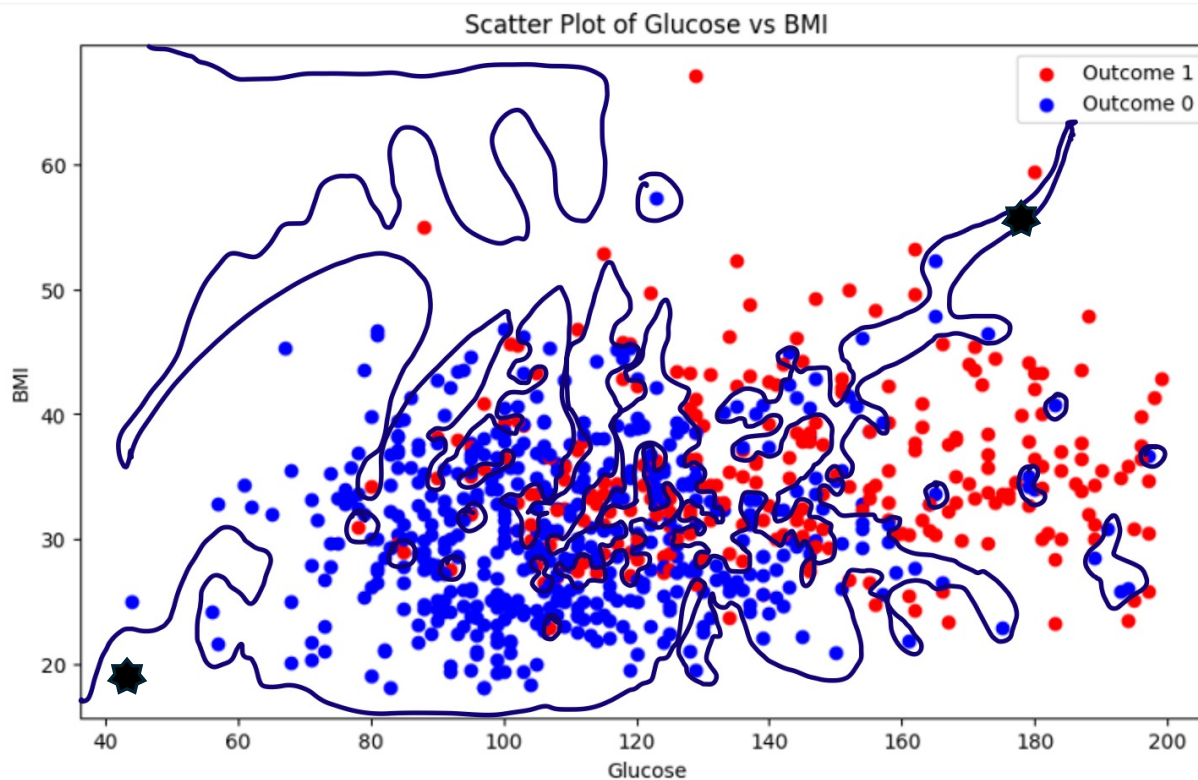
Τι λέει αυτός ο
αλγόριθμος για αυτά
τα δύο σημεία;



Τέταρτο Παράδειγμα

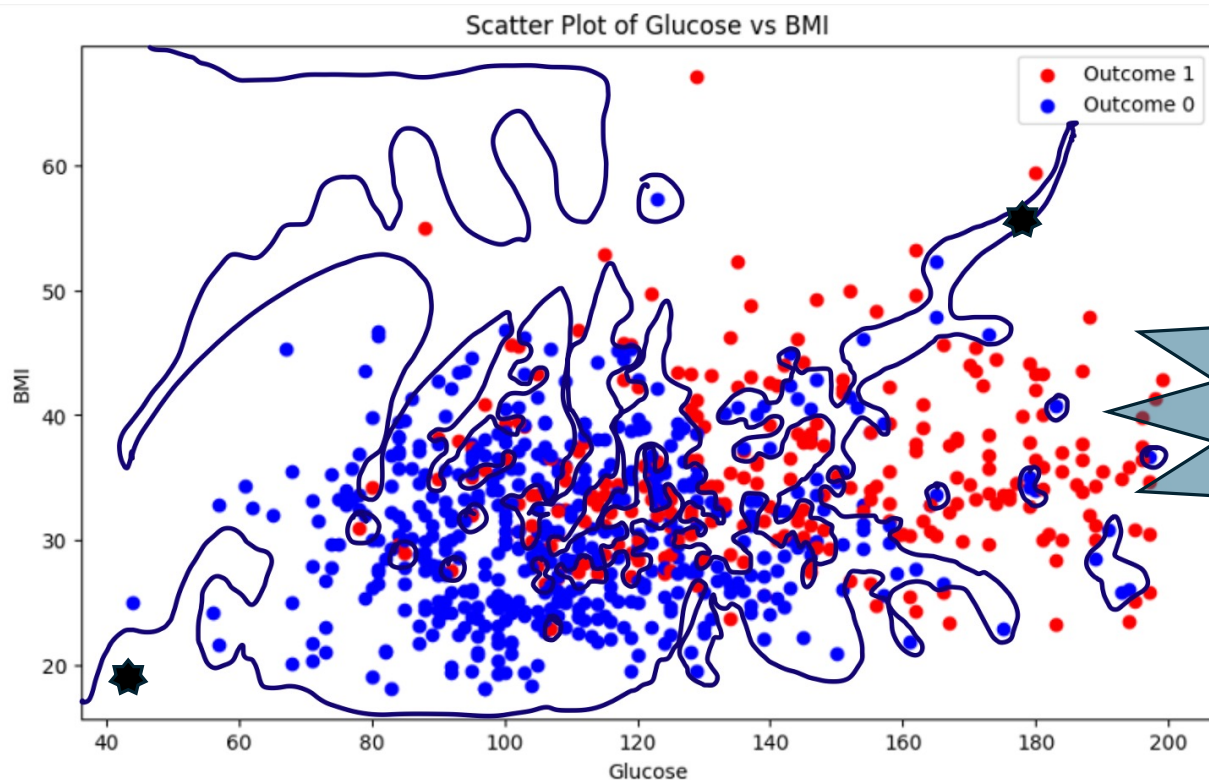


Τέταρτο Παράδειγμα: Γιατί δεν μας αρέσει;



★ Τι λέει αυτός ο αλγόριθμος για αυτά τα δύο σημεία;

Τέταρτο Παράδειγμα: Γιατί δεν μας αρέσει;



★ Τι λέει αυτός ο αλγόριθμος για αυτά τα δύο σημεία;

Συμφωνεί με όλα τα δεδομένα, αλλά δεν βγάζει καλές προβλέψεις για **μελλοντικά** δεδομένα.

Υπερμοντελοποίηση - Overfitting

- Ένας αλγόριθμος **υπερμοντελοποιεί** εάν πετυχαίνει πολύ καλύτερη ακρίβεια στα δεδομένα με τα οποία εκπαιδεύτηκε (**training data**), συγκριτικά με την ακρίβεια που πετυχαίνει σε μελλοντικά δεδομένα (**testing data**).
- Πως μπορούμε να καταλάβουμε εάν ο αλγόριθμός μας υπερμοντελοποιεί; Δεν έχουμε (ακόμα) τα μελλοντικά δεδομένα!

Υπερμοντελοποίηση - Overfitting

- Δεν έχουμε μελλοντικά δεδομένα, αλλά μπορούμε να θυσιάσουμε ένα ποσοστό των δεδομένων μας, για να προσομοιώσουμε αυτό που θέλουμε:

$$\text{Data} = (x_1, y_1), \dots, (x_n, y_n)$$

$$\text{Split: } (x_1, y_1), \dots, (x_n, y_n) = (x_1, y_1), \dots, (x_m, y_m) \cup (x_{m+1}, y_{m+1}), \dots, (x_n, y_n)$$



Όλα τα δεδομένα



Δεδομένα για εκπαίδευση
(training data)



Δεδομένα για εκτίμηση ακρίβειας
(testing data)

Υπερμοντελοποίηση - Overfitting

- Δεν έχουμε μελλοντικά δεδομένα, αλλά μπορούμε να θυσιάσουμε ένα ποσοστό των δεδομένων μας, για να προσομοιώσουμε αυτό που θέλουμε:

$$\text{Data} = (x_1, y_1), \dots, (x_n, y_n)$$

$$\text{Split: } (x_1, y_1), \dots, (x_n, y_n) = (x_1, y_1), \dots, (x_m, y_m) \cup (x_{m+1}, y_{m+1}), \dots, (x_n, y_n)$$



Όλα τα δεδομένα



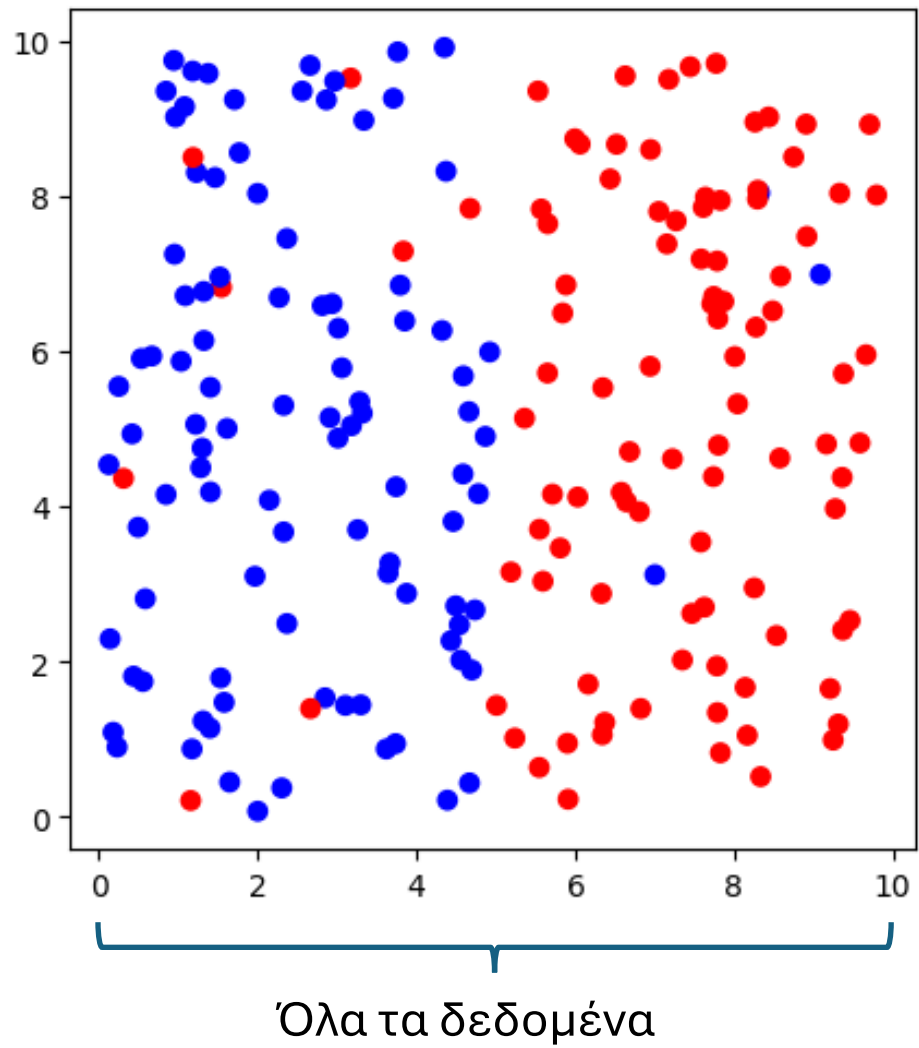
Δεδομένα για εκπαίδευση
(training data)



Δεδομένα για εκτίμηση ακρίβειας
(testing data)

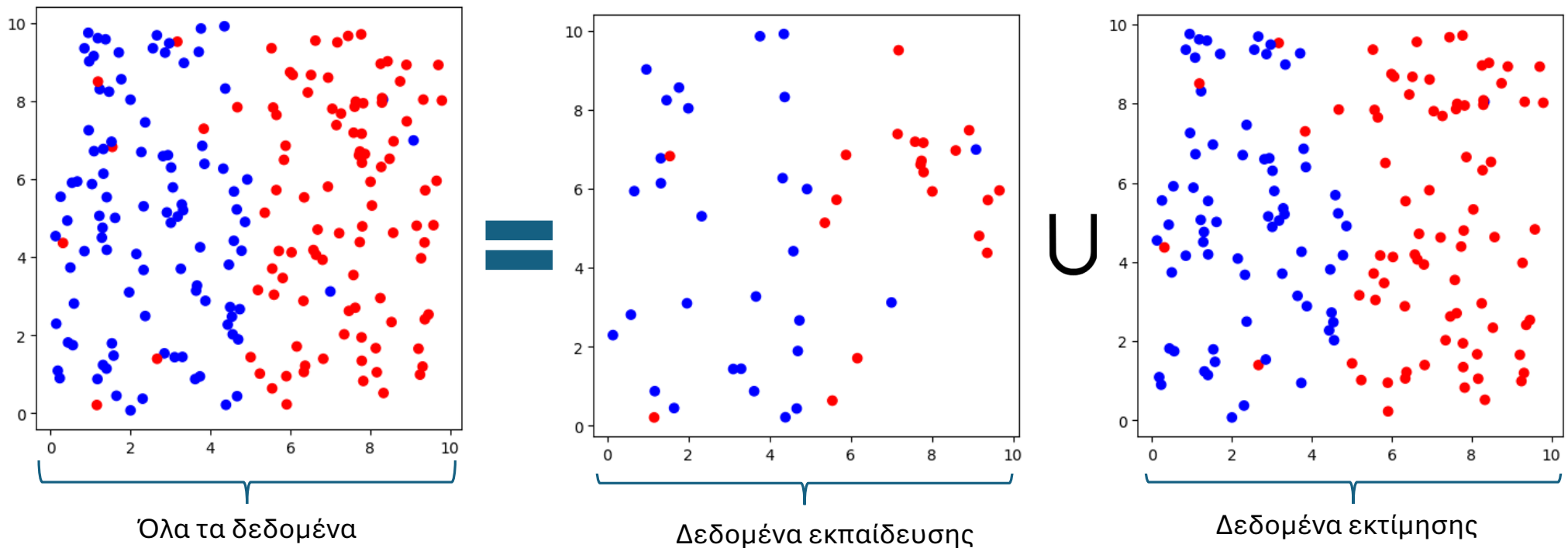
$$(X, y) = (X_{train}, y_{train}) \cup (X_{test}, y_{test})$$

Υπερμοντελοποίηση – Παράδειγμα

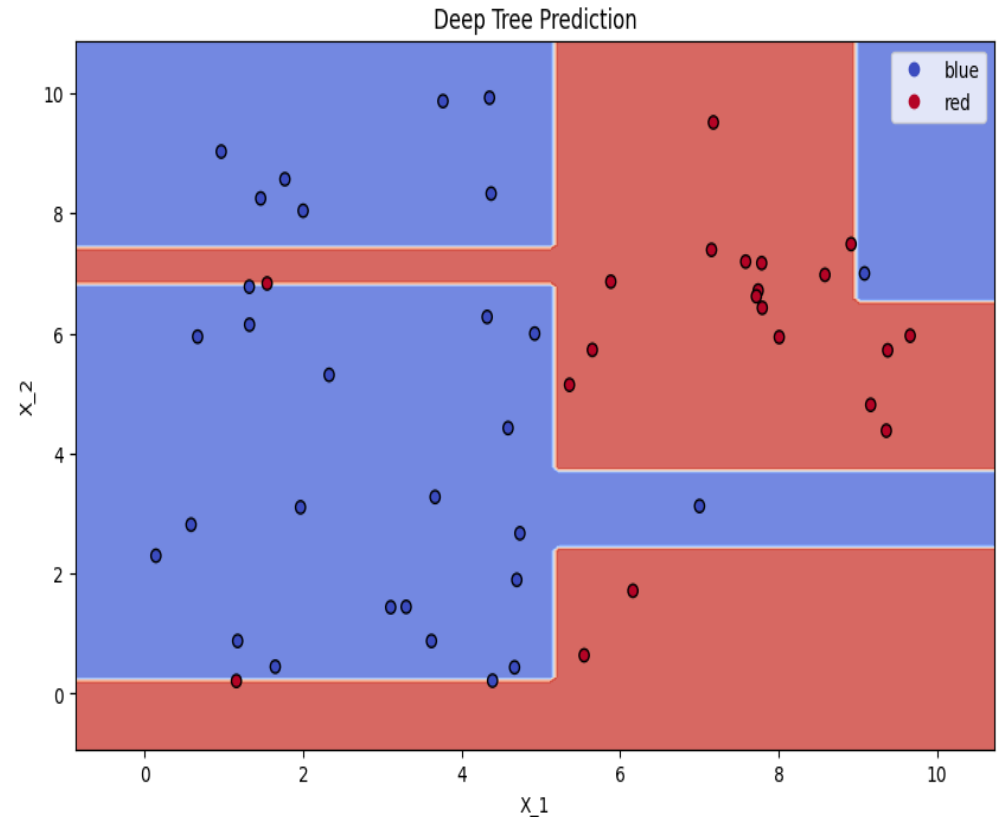
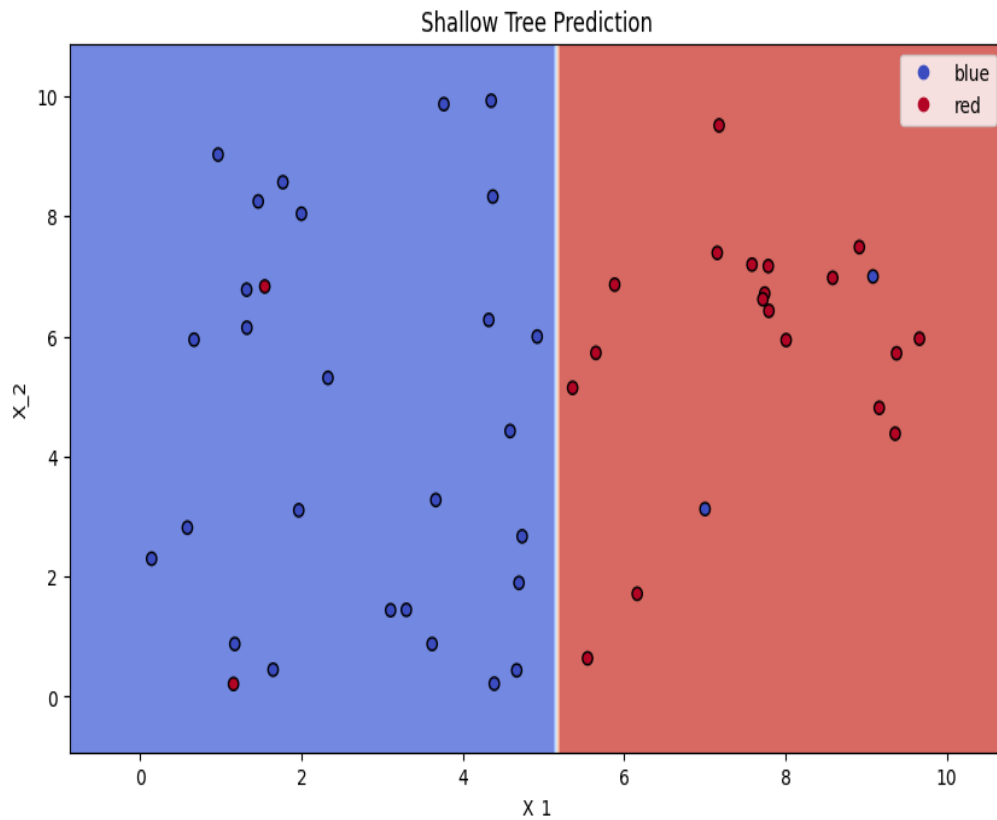


Υπερμοντελοποίηση – Παράδειγμα

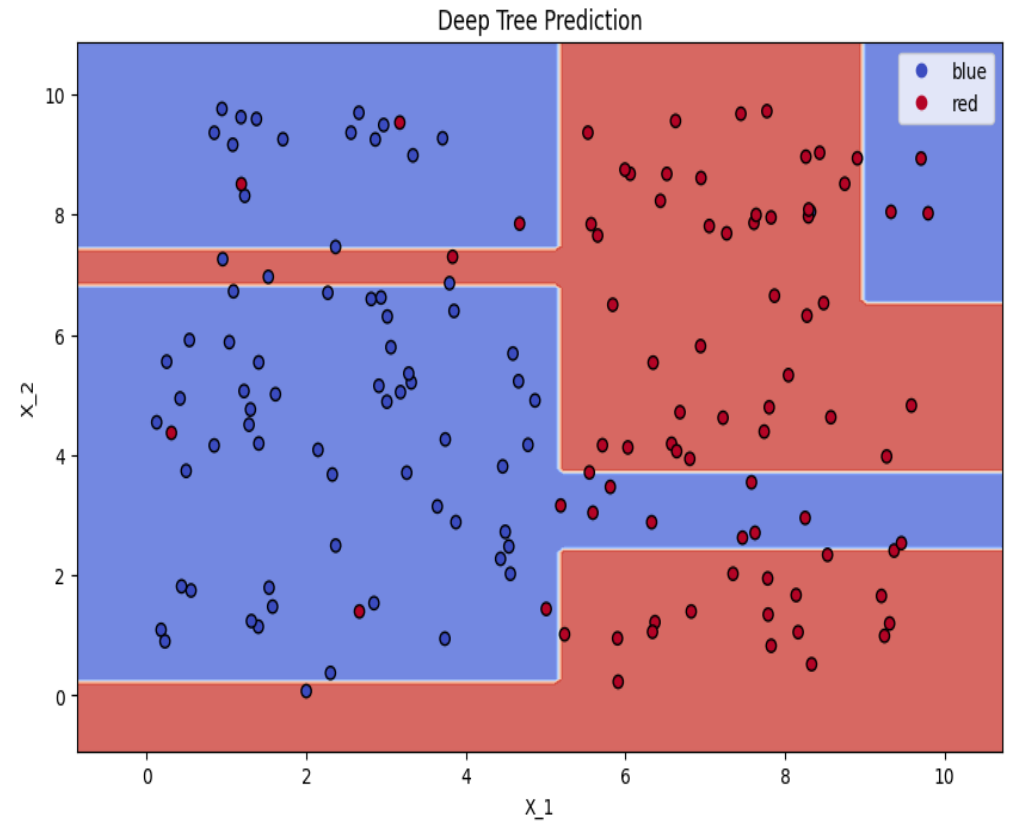
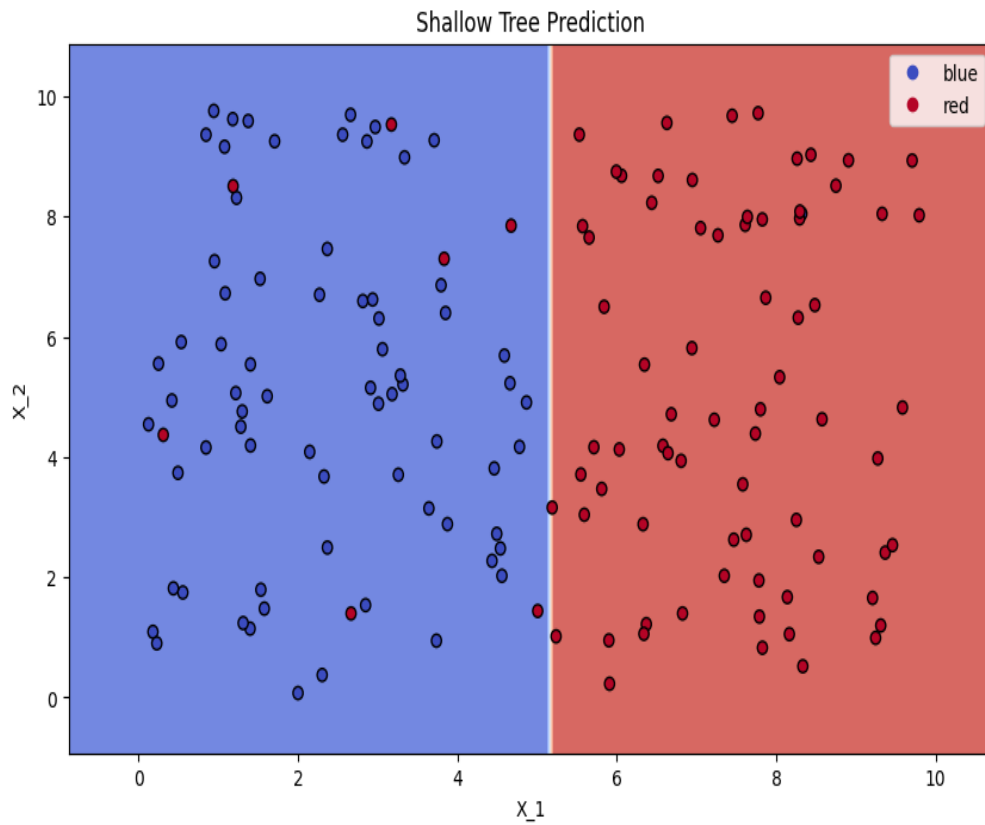
Χωρίζουμε σε δεδομένα εκπαίδευσης και δεδομένα εκτίμησης: training data + testing data



Βαθύ και ρηχό δέντρο: δεδομένα εκπαίδευσης



Βαθύ και ρηχό δέντρο: δεδομένα εκτίμησης



Υπερμοντελοποίηση

Ενας αλγόριθμος *υπερμοντελοποιεί* εάν πετυχαίνει πολύ καλύτερη ακρίβεια στα δεδομένα εκπαίδευσης (**training data**), συγκριτικά με την ακρίβεια που πετυχαίνει στα δεδομένα εκτίμησης (**testing data**).

Πάντα χωρίζουμε τα δεδομένα μας σε δεδομένα εκπαίδευσης και δεδομένα εκτίμησης: η ακρίβεια του αλγορίθμου δεν είναι η ακρίβεια στα δεδομένα εκπαίδευσης, αλλά η ακρίβεια στα δεδομένα εκτίμησης

Ενας αλγόριθμος *υπερμοντελοποιεί* εάν πετυχαίνει πολύ καλύτερη ακρίβεια στα

Εάν υπερμοντελοποιούμε, πρέπει να περιορίσουμε την πολυπλοκότητα του αλγόριθμου: π.χ., δέντρο με μικρότερο βάθος. Υπάρχουν και άλλοι τρόποι να περιοριστεί η πολυπλοκότητα, που θα συναντήσουμε στην πορεία.

Ο βασικός σκοπός αυτής της διάλεξης είναι να καταλάβουμε τι θα πεί **υπερμοντελοποίηση**, και πως καταλαβαίνουμε εάν ο αλγόριθμος μας υπερμοντελοποιεί.

Υπερμοντελοποίηση