

Boosting

Birol Yüceođlu

Migros T.A.Ş.

www.veridefteri.com

biroly@migros.com.tr

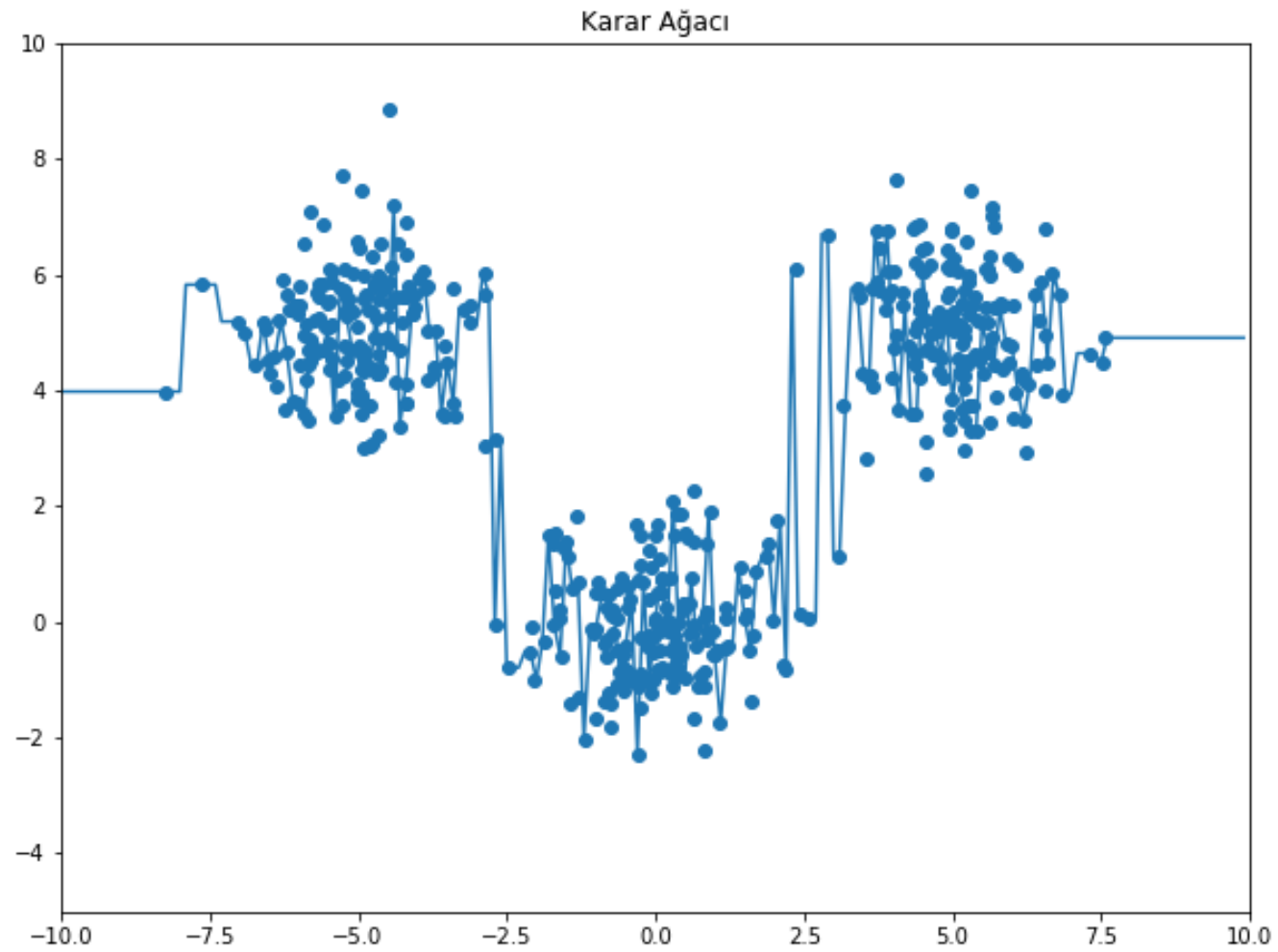
İçerik

- ▶ Karar ağaçları
- ▶ Bagging
- ▶ Boosting
 - ▶ Ana fikir
 - ▶ Boosting vs. Bagging
- ▶ LightGBM
- ▶ Scikit-learn
 - ▶ AdaBoost
 - ▶ Calibration
 - ▶ Gradient Boosting

Karar Ağaçları

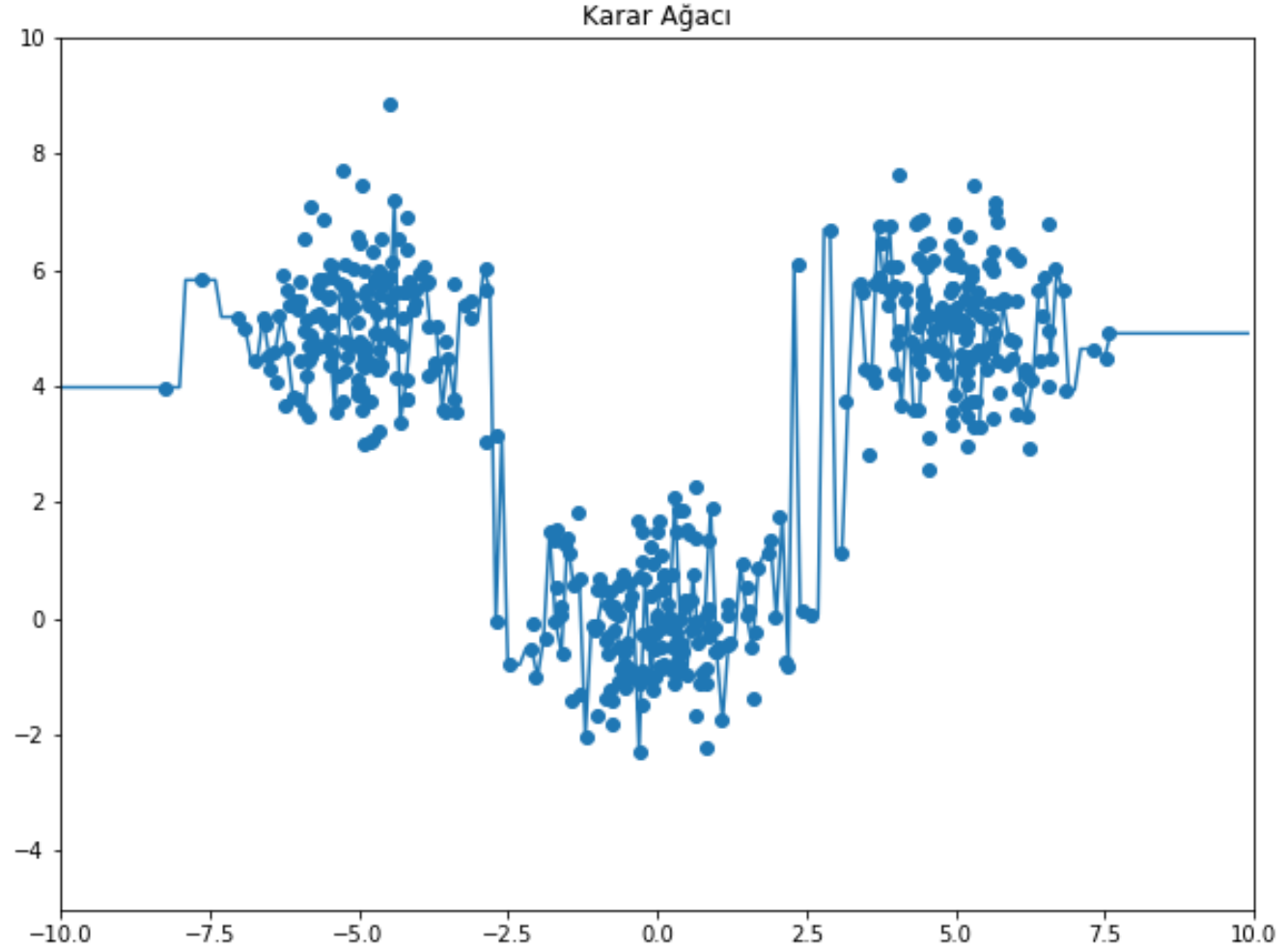
- ▶ Özniteliklerin değerlerine göre yapılan kırılımlarla birbirine benzeyen örnekleri ayırtırmaya dayanır.
- ▶ Kırılımlarla elde edilen veri kümeleri kendi içlerinde **benzer** olmalıdır.
- ▶ Yapılan ayırtırmalarla problem için kurallar oluşturulur.
- ▶ Uygulaması ve anlaması kolaydır (ağacın kendisi ve özniteliklerin önemi).
- ▶ Olası problemler:
 - ▶ Düşük yanlılık (bias), yüksek varyans (variance)
 - ▶ Parametre eniyilemesi
 - ▶ Derinlik, yapraklardaki örnek sayısı
 - ▶ Veriye ve alınan kararlara aşırı duyarlı

Karar Ağaçları



Karar Ağaçları

- ▶ Regresyon problemi
- ▶ Gözlemlerin x değerlerine göre y değerlerini tahmin edeceğiz.
- ▶ Aşırı uyum



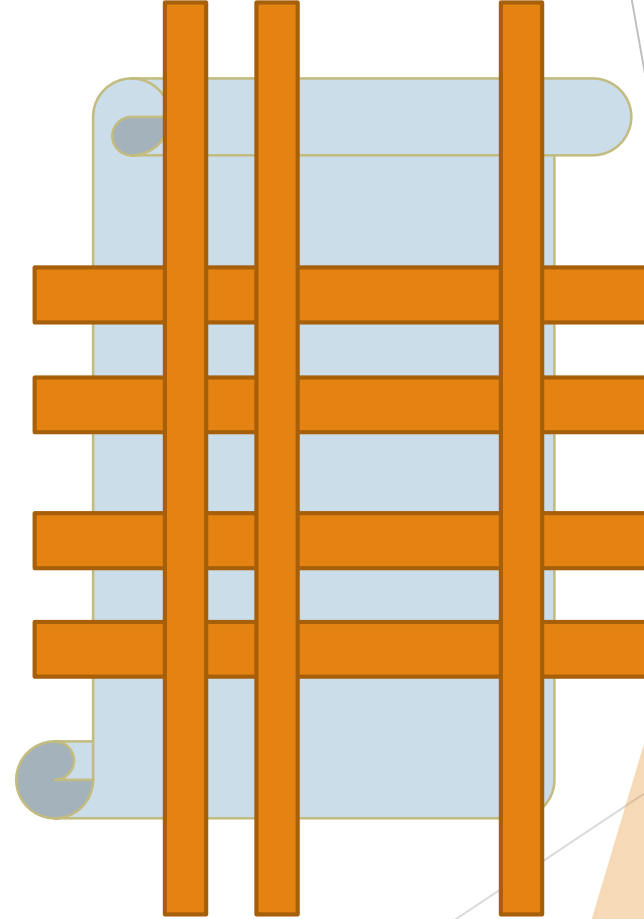
Bagging

- ▶ Hedefimiz varyansı azaltmak.
- ▶ **Birbirinden bağımsız** olarak 0.51 olasılıkla doğru tahmin yapan modellere sahip olalım. (Weak learner)
- ▶ Bu modellerden 1 tane kullanırsak %51 ihtimalle doğru tahmin yaparız.
- ▶ Bu modellerden 5 tane kullanırsak %52 ihtimalle doğru tahmin yaparız.
- ▶ Bu modellerden 100 tane kullanırsak %58 ihtimalle doğru tahmin yaparız.
- ▶ Bu modellerden 1000 tane kullanırsak %74 ihtimalle doğru tahmin yaparız.

- ▶ **Birbirinden bağımsız modeller!!!!**

Bagging

- ▶ Birbirinden bağımsız modeller nasıl oluşturulur?
- ▶ Bağımsız yerine korelasyonu düşük modeller
 - ▶ Gözlemler
 - ▶ Öznitelikler
- ▶ Gözlemler
 - ▶ Bootstrap ile yeni örneklemeler oluşturmak (Bagging)
- ▶ Öznitelikler
 - ▶ Rassal özniteliklerle modelleme (Rassal orman)



Boosting

- ▶ Yavaş öğrenmeye dayalı bir yöntem
- ▶ Zayıf öğrenen (weak learner) modellerle bir topluluk öğrenmesi gerçekleştirilir.
- ▶ Sıralı/iteratif bir yöntem
- ▶ Ana fikir: Hatadan öğrenmek
 - ▶ AdaBoost: hatalı tahmin edilen gözlemlerin ağırlıklandırılarak hatayı düzeltmeye çalışır.
 - ▶ Gradient Boosting: iteratif olarak hatayı tahmin eder
 - ▶ Stochastic Gradient Boosting: her adımda veri kümesinden örneklem seçerek ilerler.

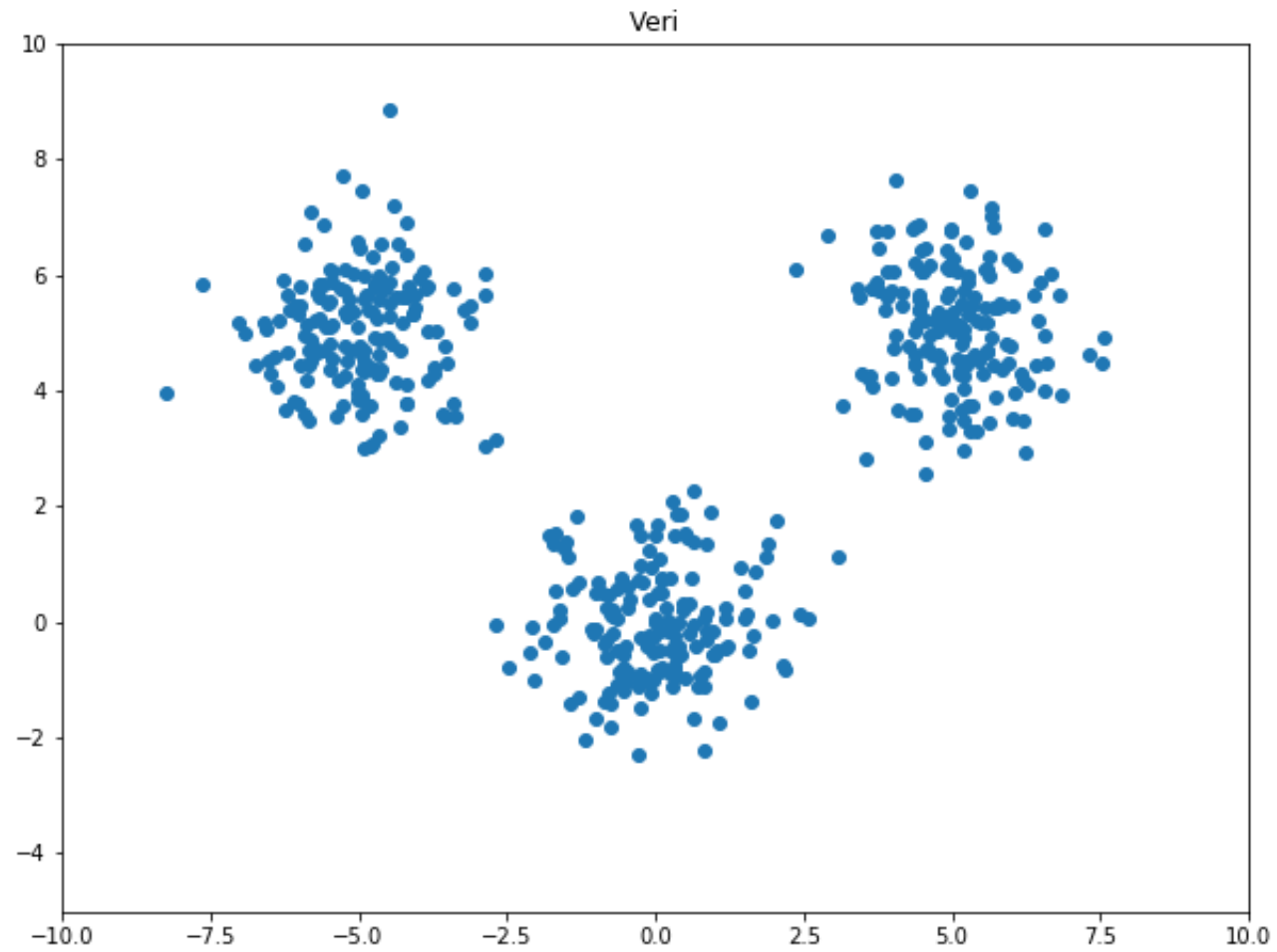
Boosting vs. Bagging

Boosting	Bagging
İteratif	Paralel
Ağırlıklı ortalama	Ortalama (voting)
Değiştirilmiş veri kümesi	Rassal örneklem / öznitelikler
Bütün veri kümesi	Rassal örneklem (korrelasyon azaltma)
Aşırı uyum olabilir	Aşırı uyuma dayanıklı
Yavaş öğrenme	Rassal öğrenme
Bias (yanlılık) azaltma	Variance (varyans) azaltma

Gradient Boosting

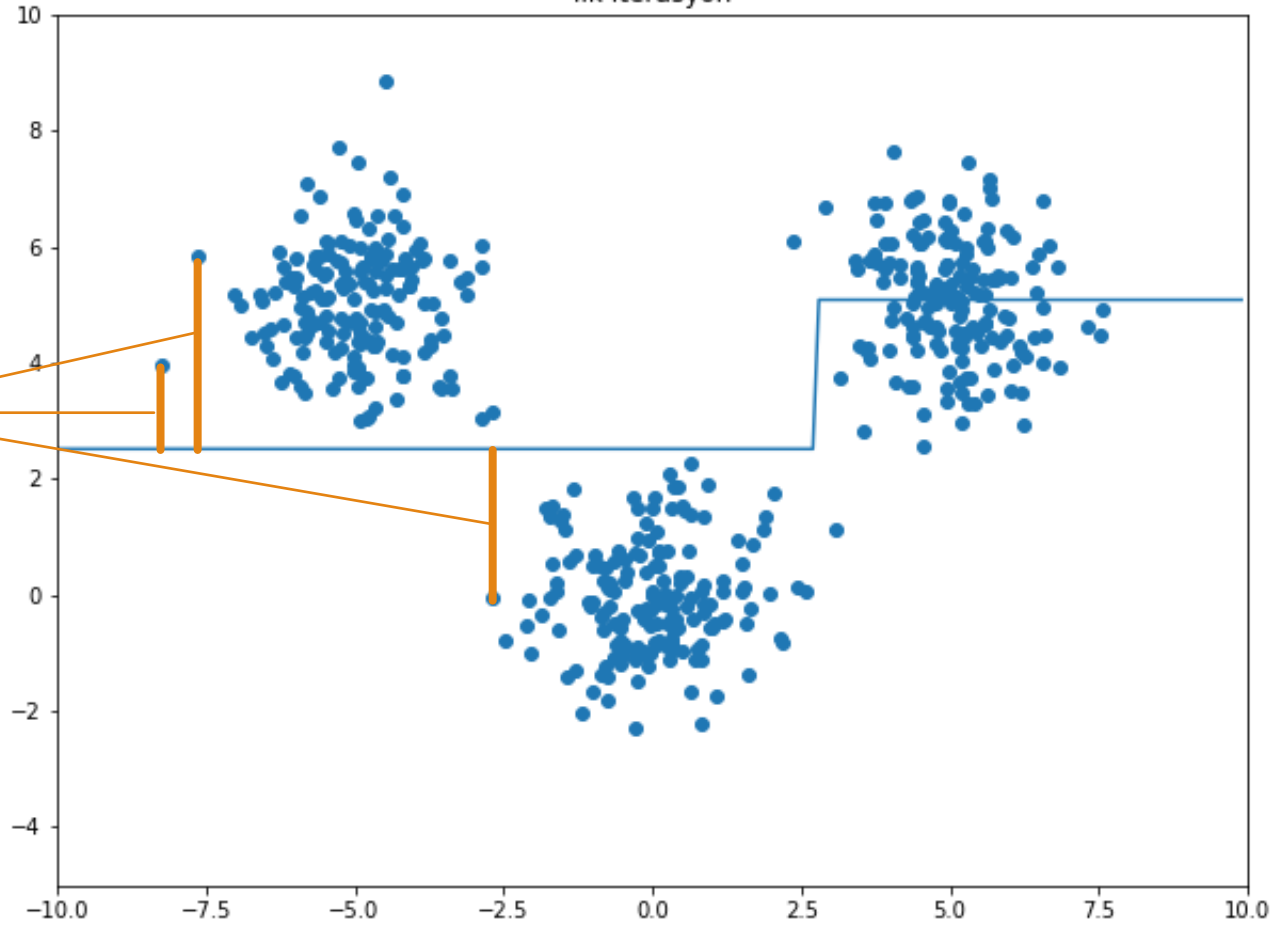
- ▶ Gradient Boosting yönteminde her iterasyonda bir önceki iterasyondaki hata tahmin edilmeye çalışılır.
- ▶ Hatadan öğrenmek gradyanın tersi yönde ilerlemeye denk gelir.
- ▶ Örneklem ve öznitelikler rassal olarak seçilerek de ilerlenebilir (Stochastic Gradient Boosting).

Gradient Boosting



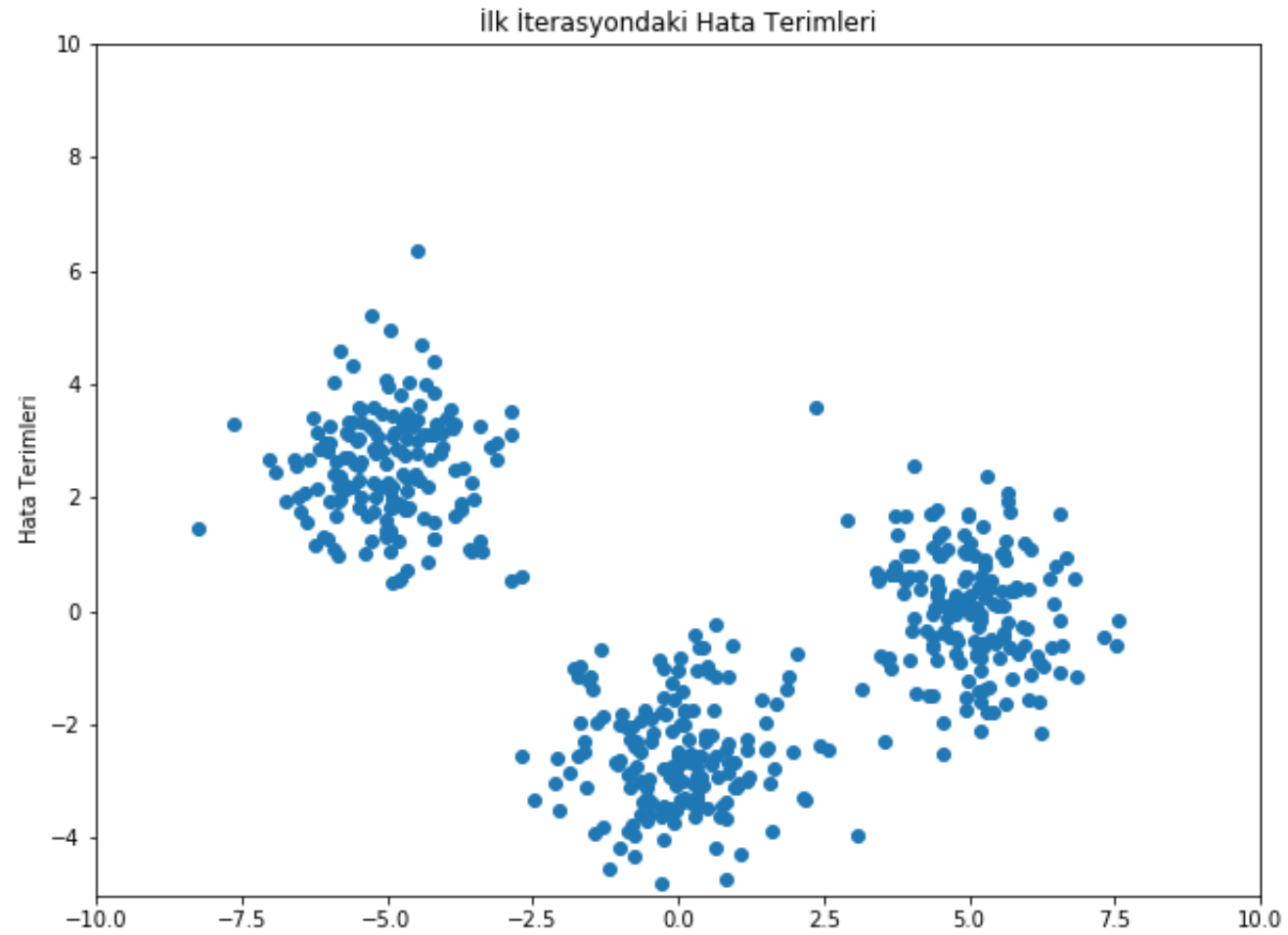
Gradient Boosting

ilk iterasyon



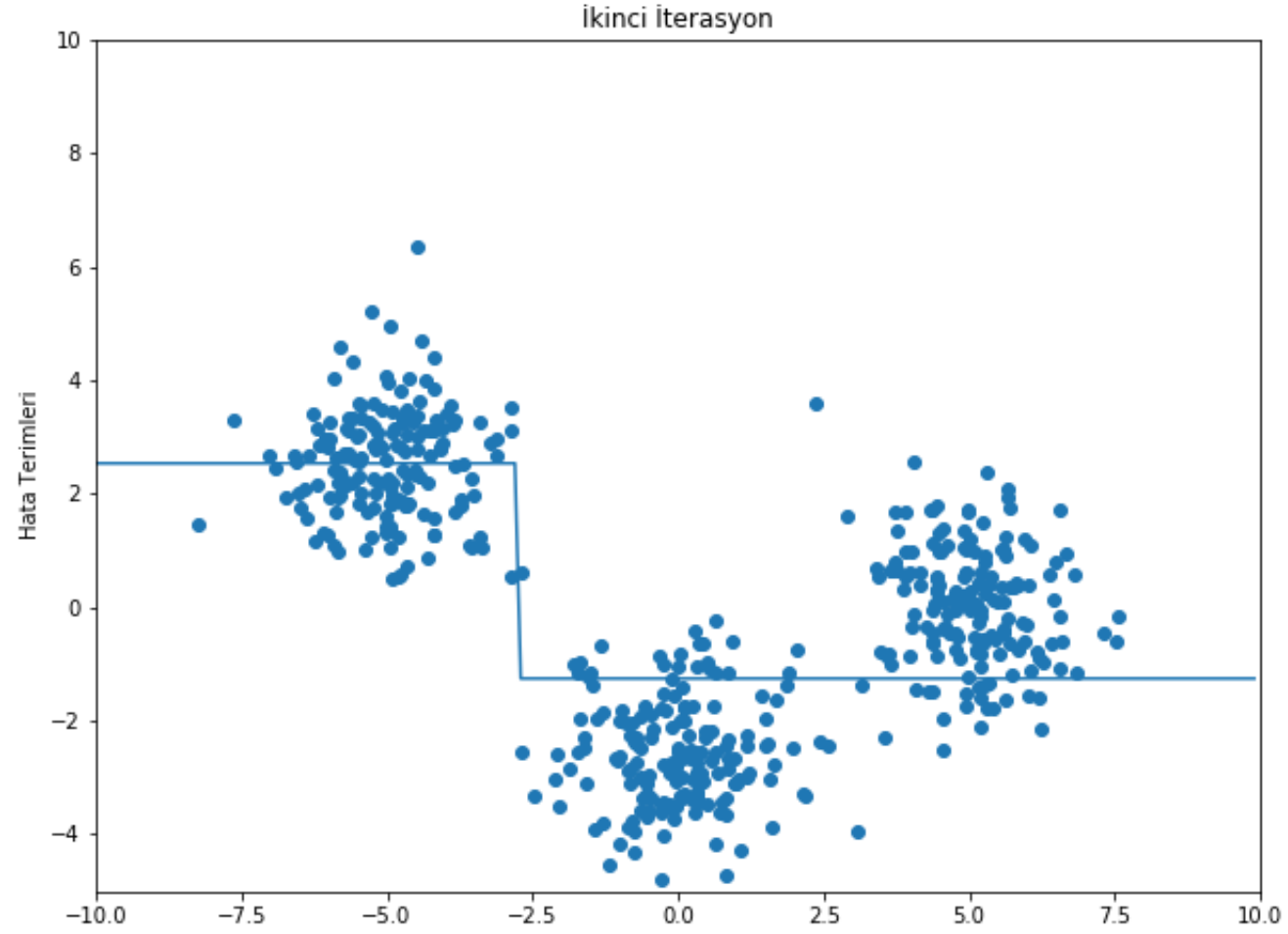
Hata terimleri

Gradient Boosting

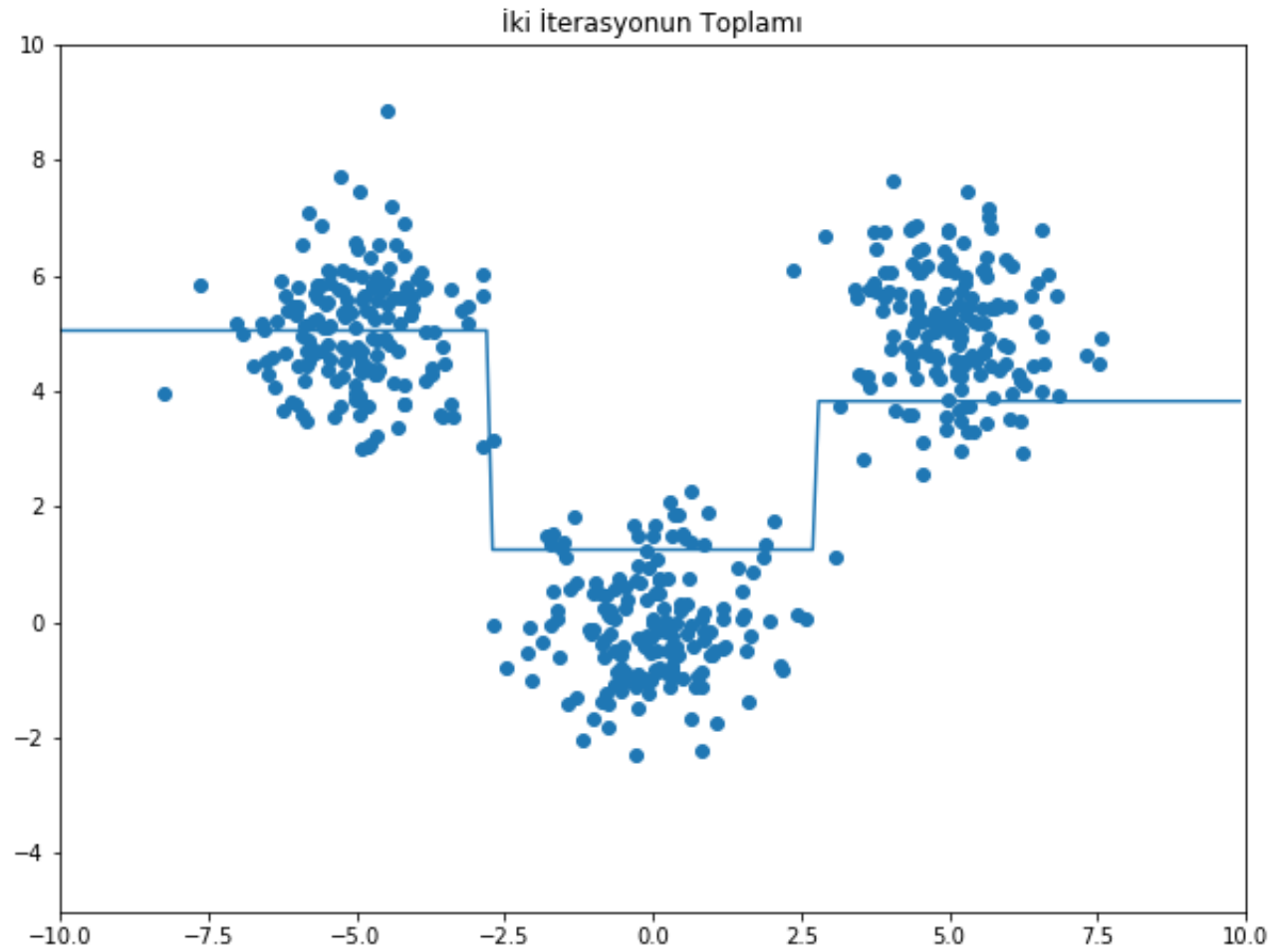


Gradient Boosting

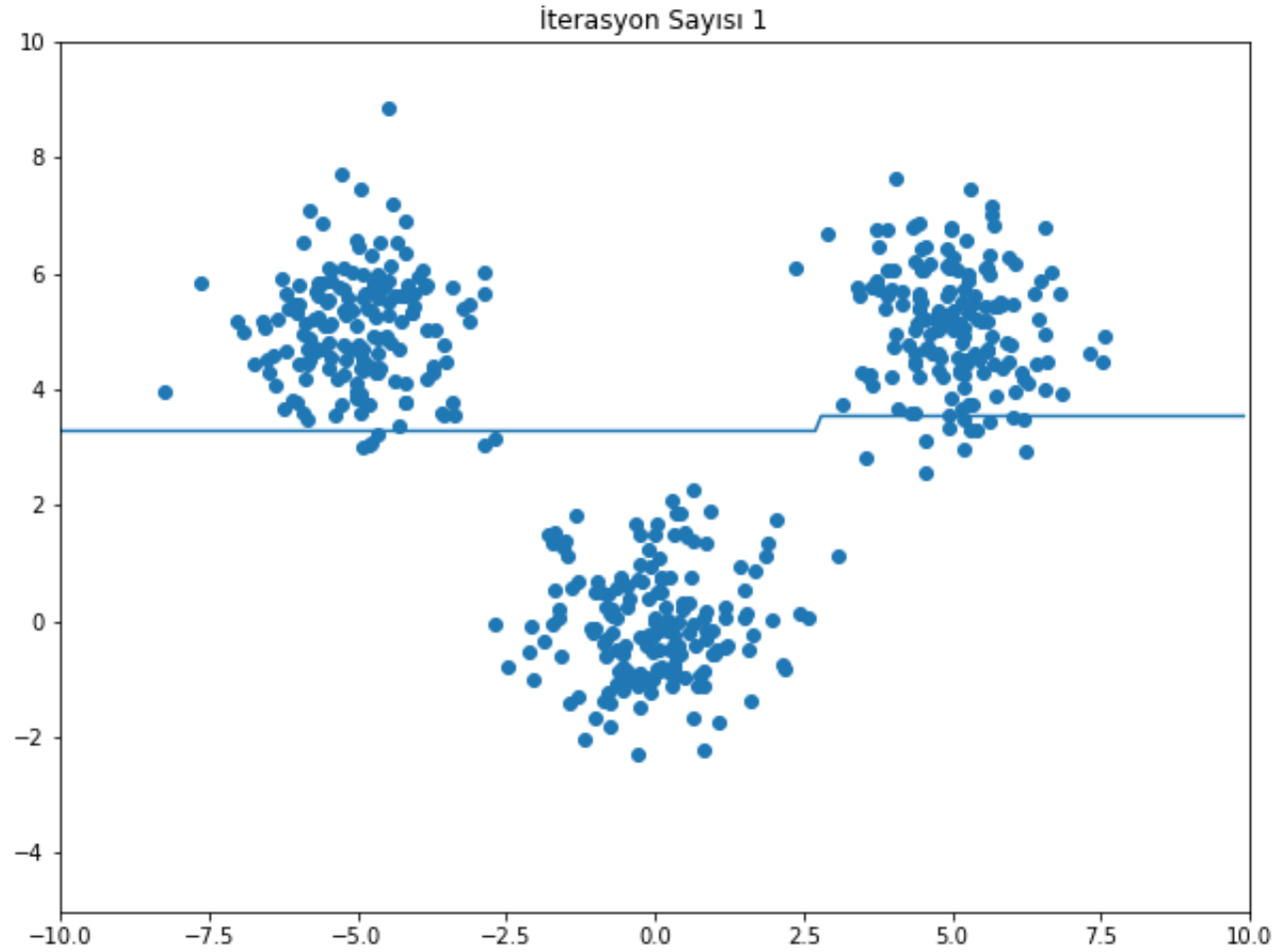
Hatanın tahmini



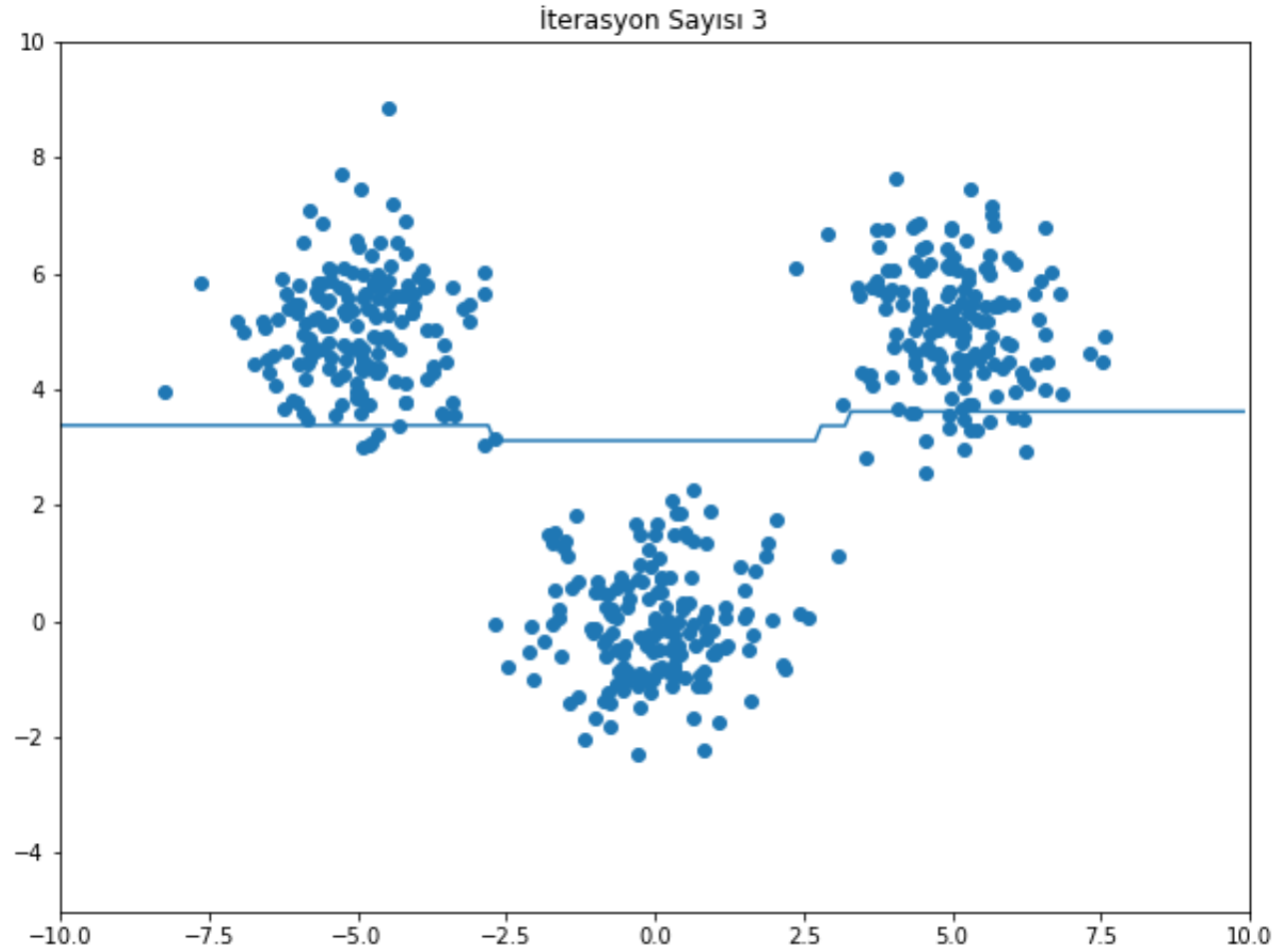
Gradient Boosting



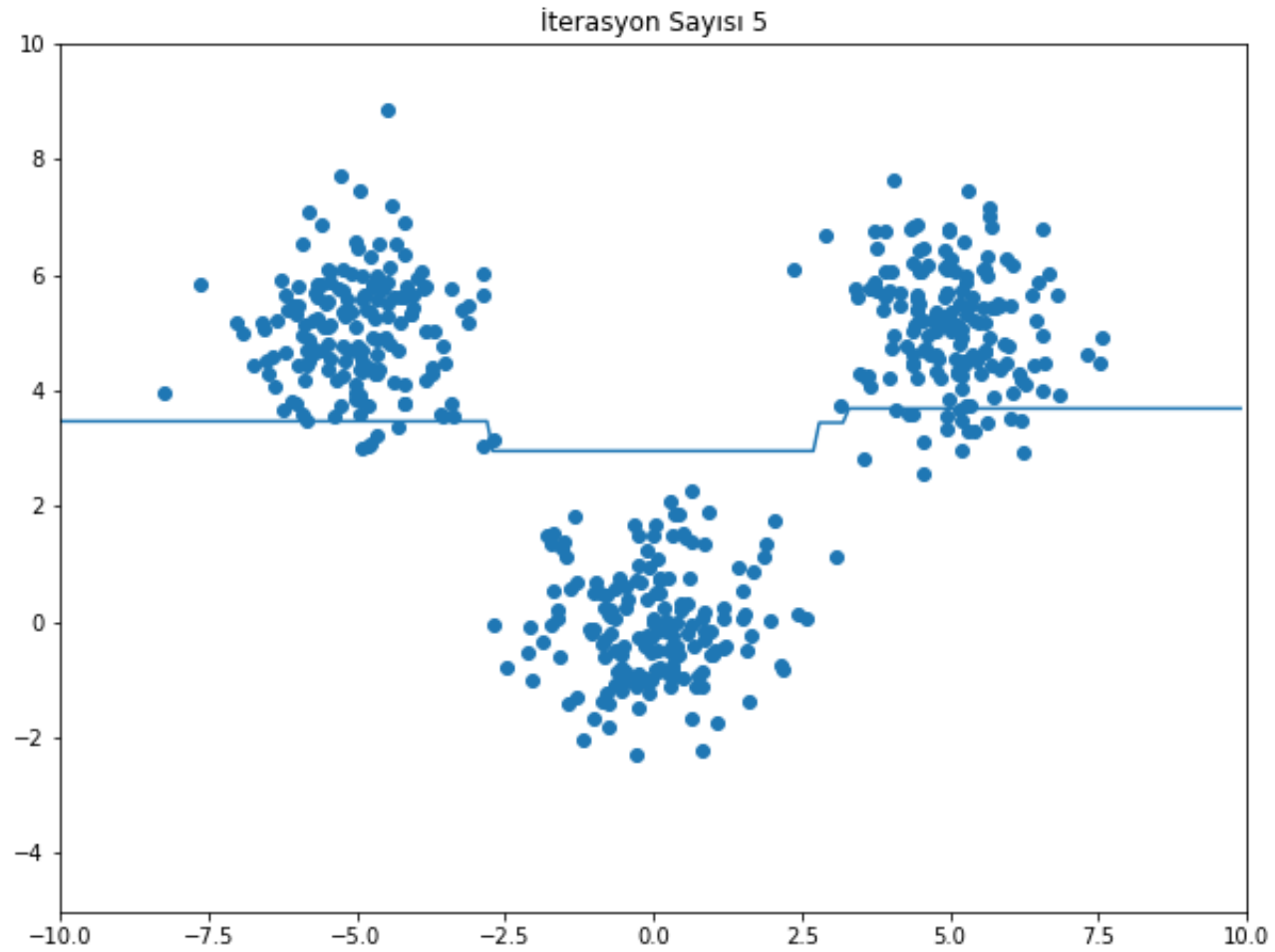
Gradient Boosting (Scikit-learn)



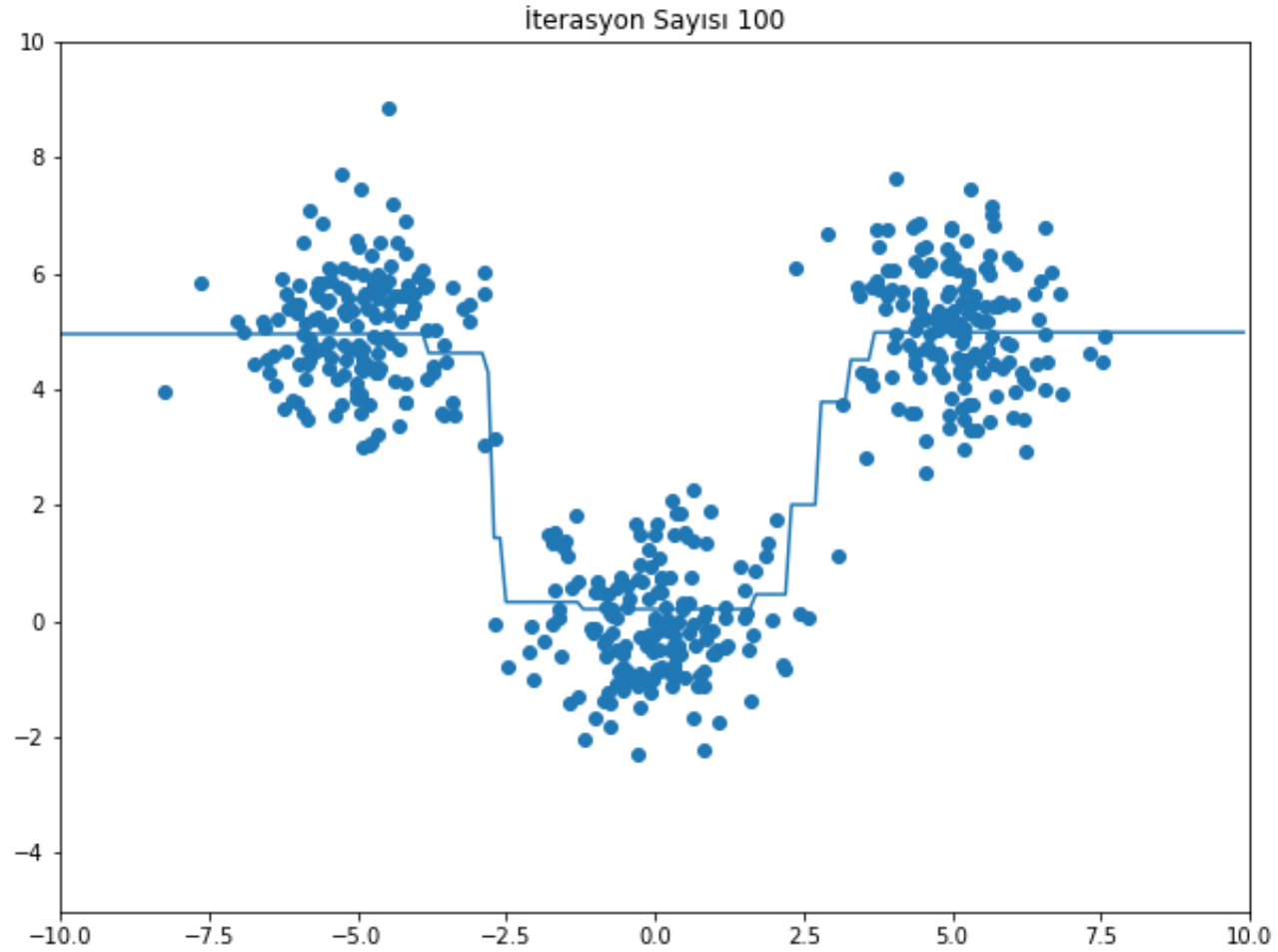
Gradient Boosting (Scikit-learn)



Gradient Boosting (Scikit-learn)

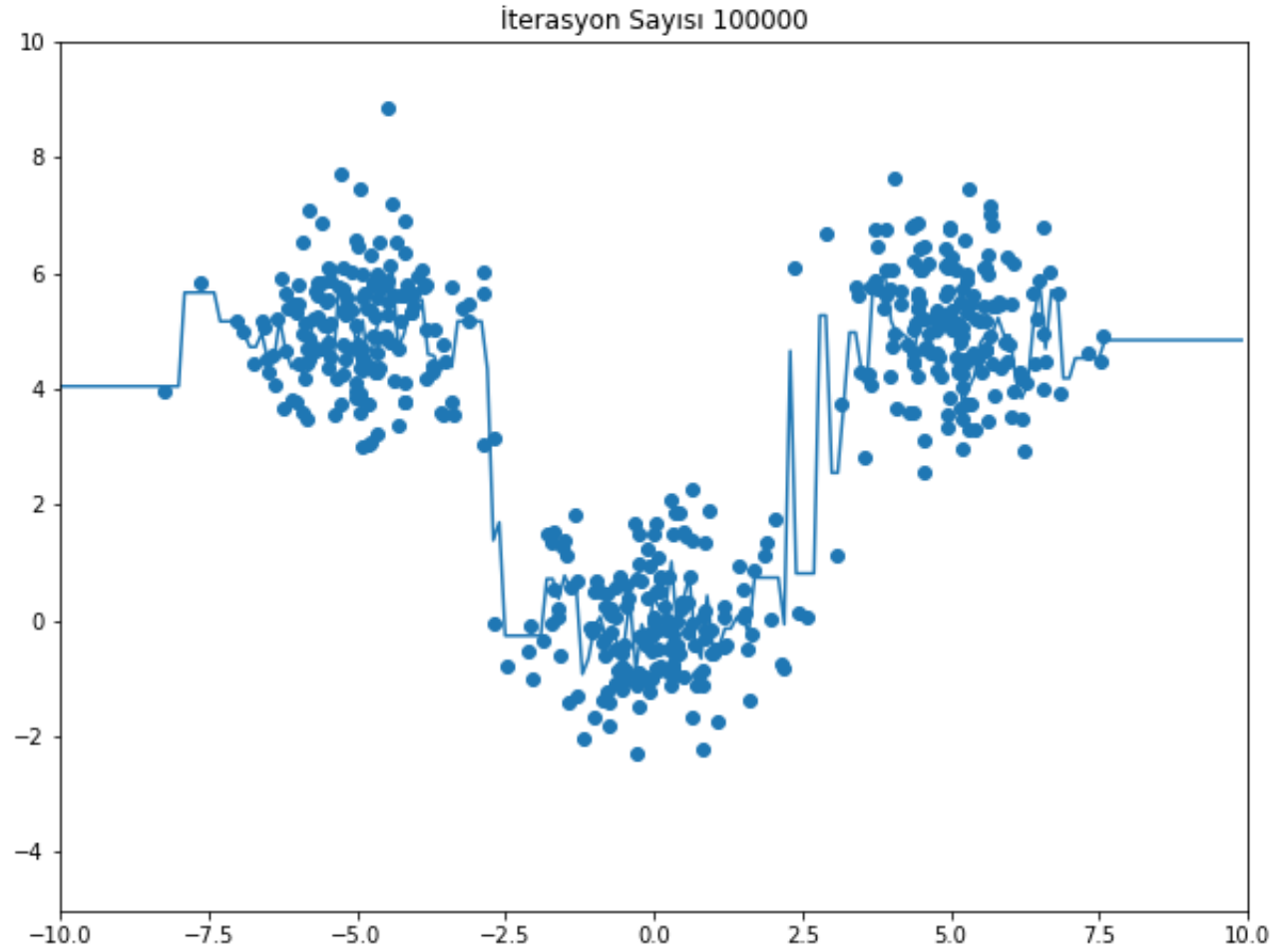


Gradient Boosting (Scikit-learn)



Gradient Boosting (Scikit-learn)

Aşırı uyuma

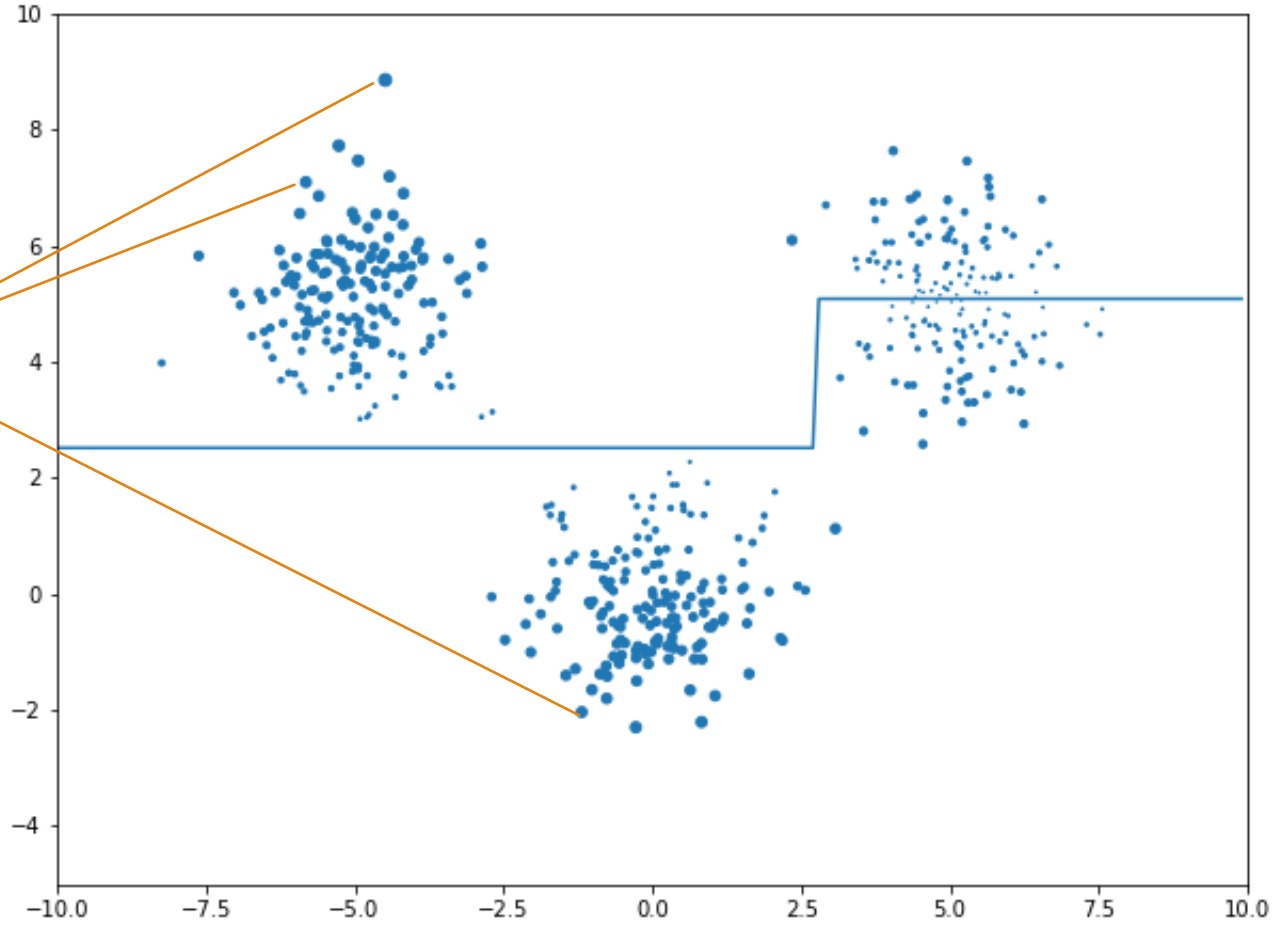


AdaBoost

- ▶ Zayıf öğrenen modellerle ilk tahmin
- ▶ Hataya göre veri kümesindeki gözlemlerin ağırlıklandırılması
- ▶ Ağırlıklandırma ya da ağırlıklı bootstrap
- ▶ Ağırlıklara göre tahmin
- ▶ Ağırlıkların güncellenmesi
- ▶ Son modelin oluşturulması (ağırlıklı toplam)

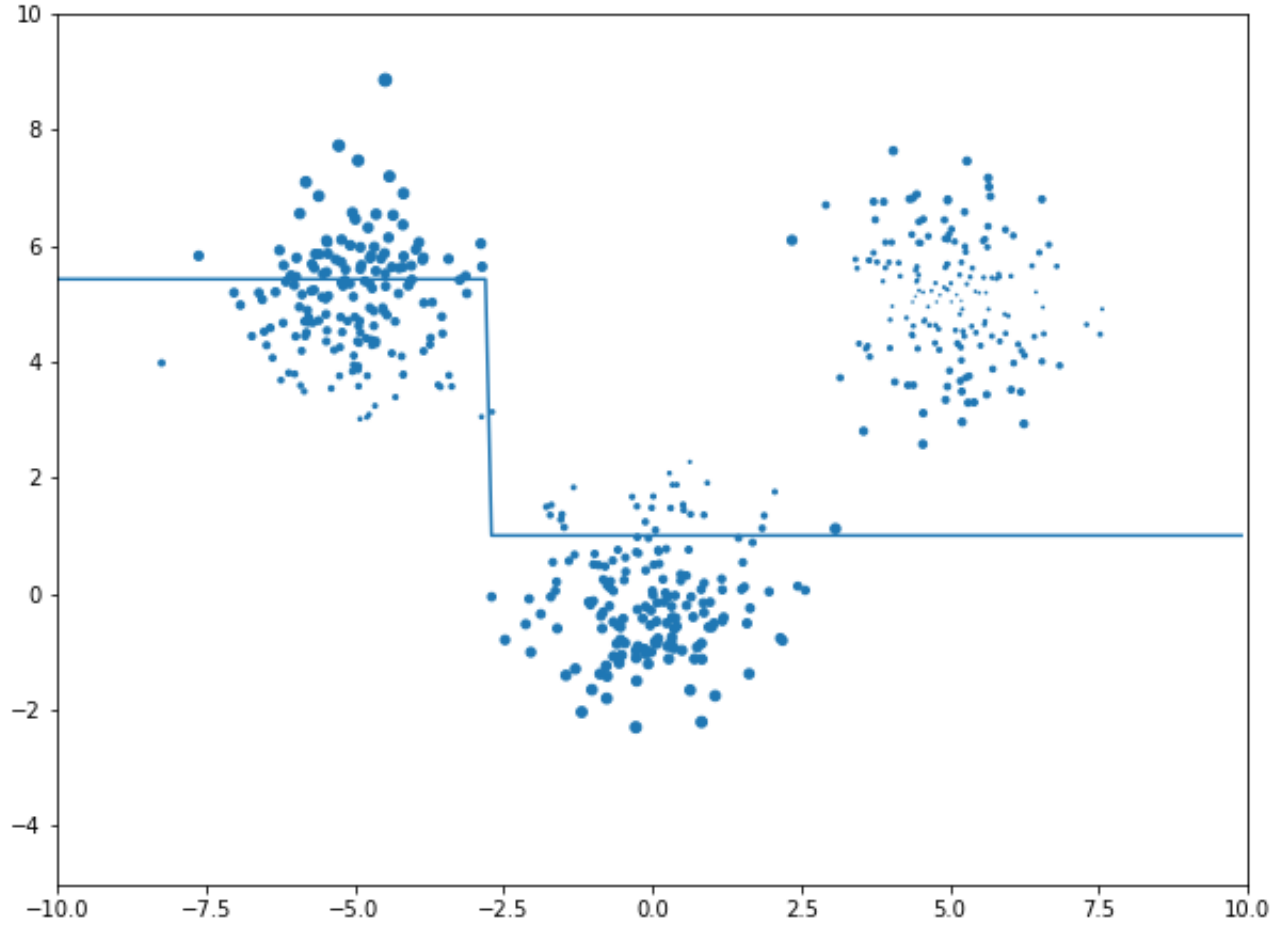
AdaBoost (Örnek)

Hatalı
gözlemlerin
ağırlıklandırılması

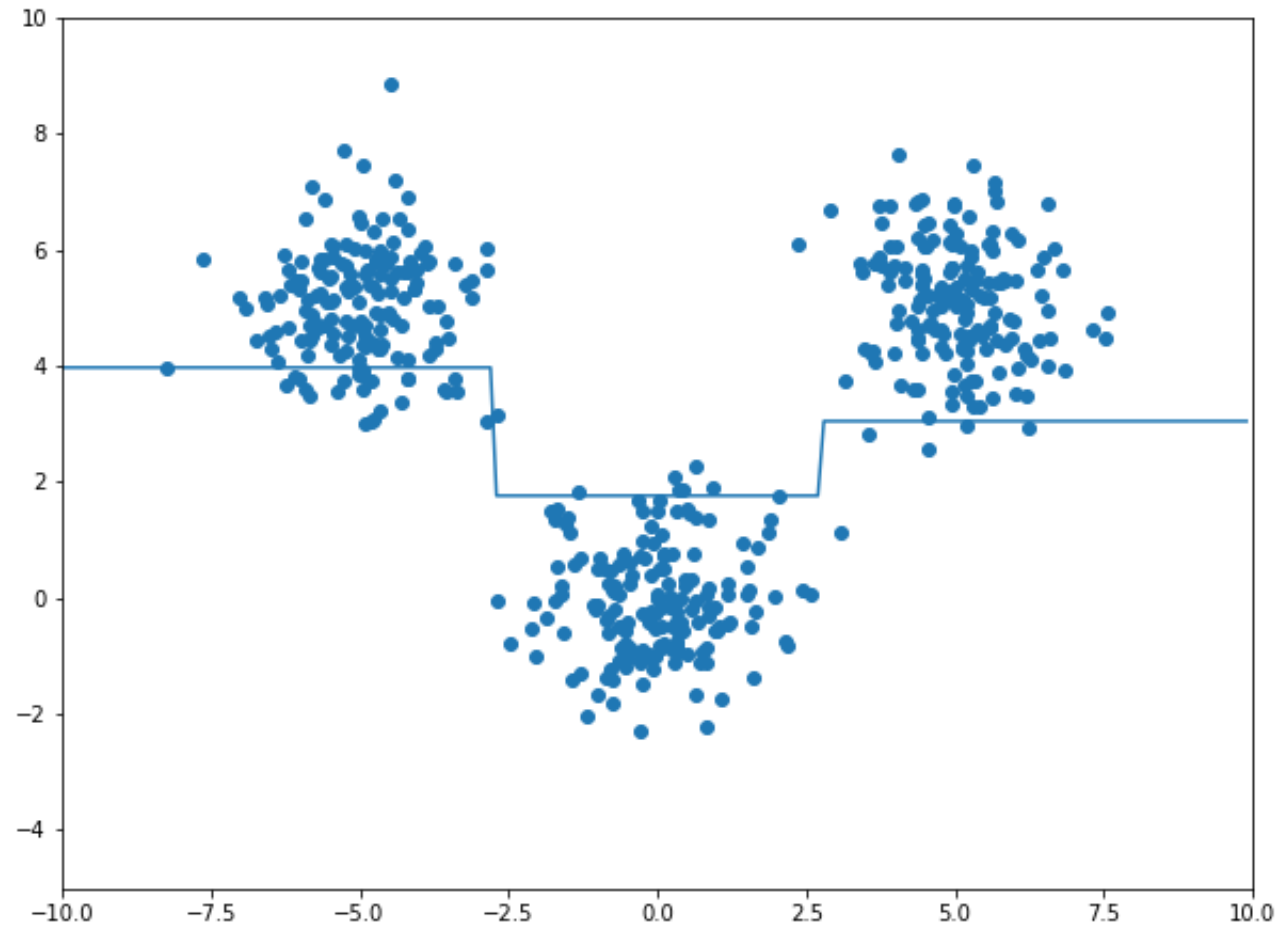


AdaBoost (Örnek)

Ağırlıklı
gözlemlere
göre tahmin



AdaBoost (Örnek)



Paketler

- ▶ Gradient Boosting
 - ▶ **LightGBM** (R, Python)
 - ▶ **scikit-learn** (Python)
 - ▶ **xgboost** (R, Python)
 - ▶ **Gbm** (R)
- ▶ AdaBoost
 - ▶ **scikit-learn** (Python)