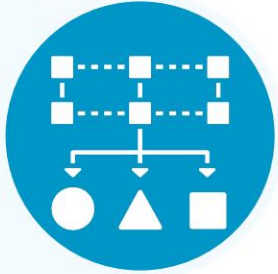


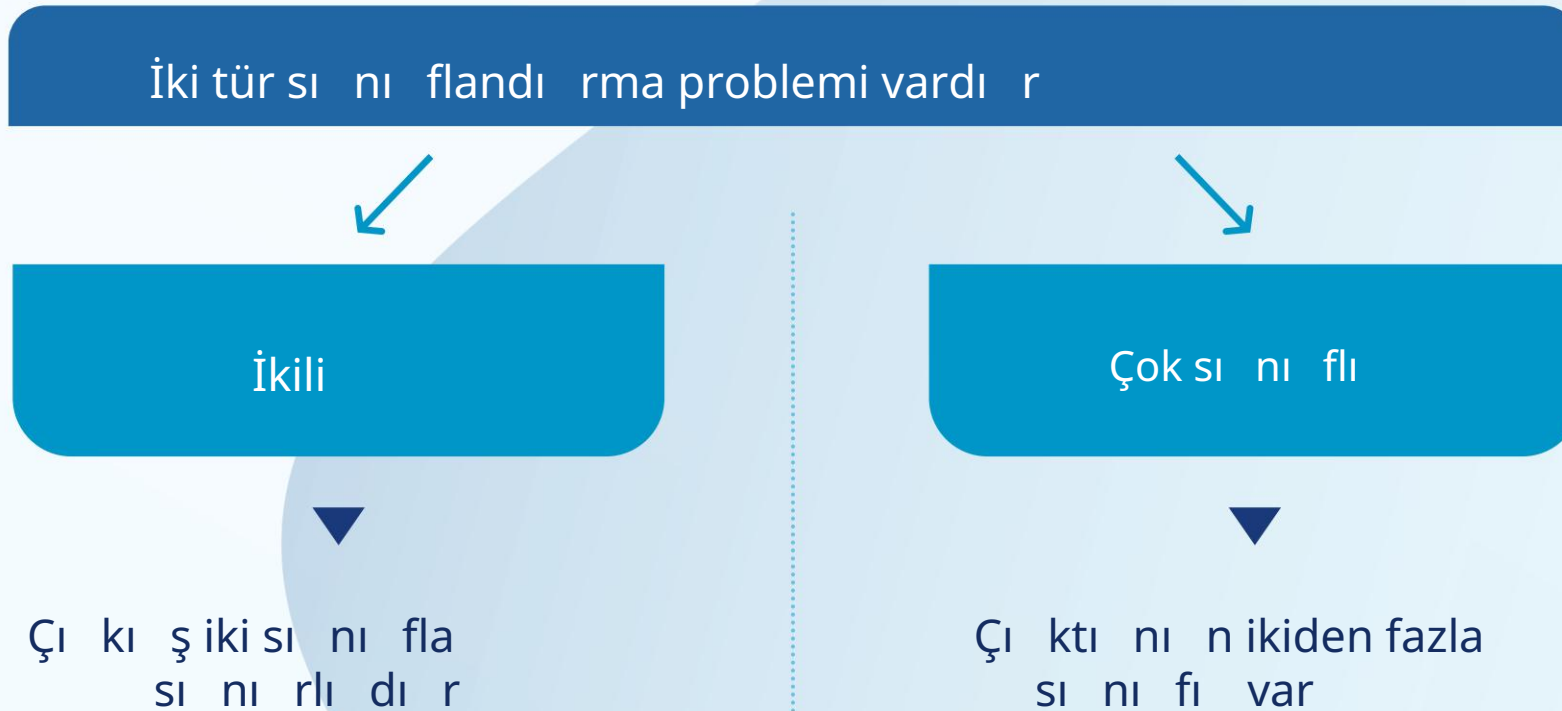
# Denetimli öğrenme tekniği: sınıflandırma



## Sınıflandırma nedir?

- Sınıflandırma, belirli bir veri kümesini sınıflara ayırma işlemidir. Önceden tanımlanmış sınıflar etiketlerimiz veya temel gerçeklerimiz olarak hareket eder.
- Model, etiketlerini tahmin etmek için bir nesnenin özelliklerini kullanır. Örneğin, spam olmayan e-postalardan spam filtrelemek veya meyve türlerini renklerine, ağırlıklarına ve boyutlarına göre sınıflandırmak.

# Sınıflandırma ne tür sorunları çözer?



# Sınıflandırma problemlerini çözmek için: lojistik regresyon

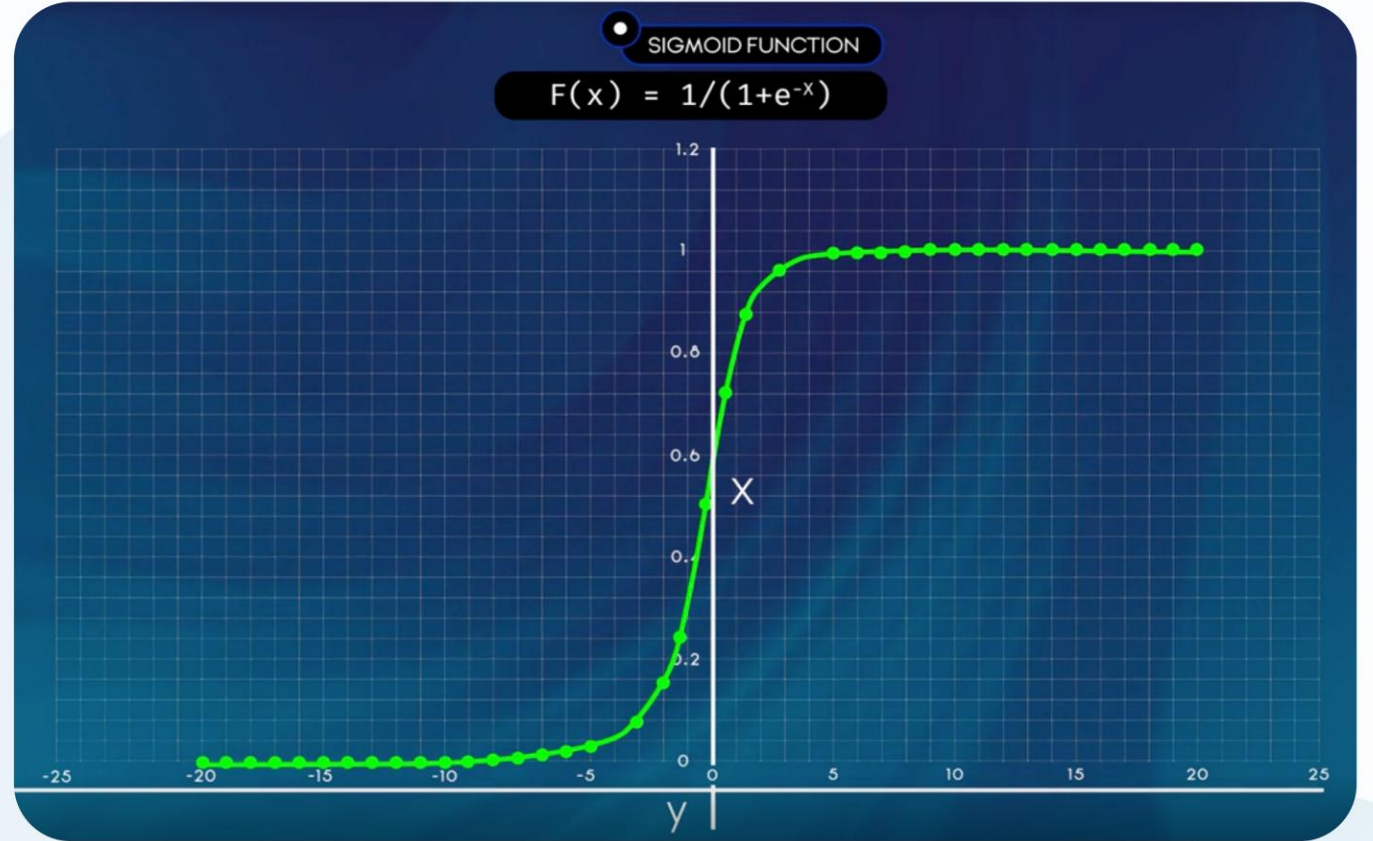


## Lojistik regresyon nedir?

Lojistik regresyon doğrusal bir regresyondur ancak sınıflandırma problemleri içindir. Doğrusal regresyonun aksine, lojistik regresyonun girdi ve çıktı değişkenleri arasında doğrusal bir ilişkiye ihtiyacı yoktur.

# Lojistik regresyon bir lojistik fonksiyon kullanır: sigmoid fonksiyon

Sigmoid fonksiyonu herhangi bir gerçek girişi alır ve sıfırı ile bir arasında bir değer verir.



# Lojistik regresyon sınıflandırıcısının performansı nasıl ölçülebilir?

- Sınıflandırmaya modelimizden (sınıflandırıcı) tahmin edilen sonuçları aldıktan sonra, sonuçlar gerçek etiketle (temel gerçek) karşılaştırılır.
- Daha sonra performansı model karışıklık matrisi kullanılarak değerlendiriliyor

CONFUSION MATRIX

		PREDICTED CLASS	
		NEGATIVE	POSITIVE
ACTUAL CLASS	NEGATIVE	TN	FP
	POSITIVE	FN	TP

# Model performansı nı ölçmek için karışık matrisini uygulama

- Gerçek pozitifler (TP) - pozitif ve temel gerçek olarak tahmin edilen sonuçlar da pozitif.
- Yanlış pozitifler (FP) - pozitif olarak tahmin edilen ancak aslında negatif olan örnekler.
- Gerçek negatifler (TN) - şu şekilde tahmin edilen örnekler olumsuzluklar ve onların temel gerçekleri de olumsuzdu.
- Yanlış negatifler (FN) - olumsuz olarak tahmin edilen ancak temel gerçekleri olumlu olan örnekler.



# Olası İşbirliği Alanları



## Kesinlik

Toplam numune miktarı ndan bir sonucun ne kadar doğru tahmin edilebileceğini gösterir



## Kesinlik

Olumlu örneklerin ne kadar doğru tahmin edildiğini ve bunlardan kaçının olumlu olduğunu gösterir



## Geri Çağırma (Hassasiyet)

Sınıflandırıcı nın kaç tane pozitif örneği yanlış tahmin ettiğini gösterir



## F1 puanı (F ölçüsü)

Kesinlik ve geri çağırma arasındaki dengeyi gösterir

Amaç, gerçek olumluları ve gerçek olumsuzları en üst düzeye çıkarmaktır; Yanlış pozitifleri ve negatifleri en aza indirin



# Değerlendirme metrikleri

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

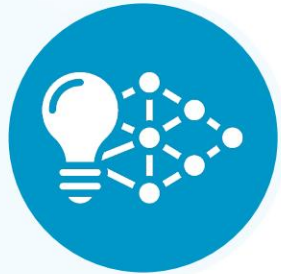
$$Precision = \frac{(TP)}{(TP+FP)}$$

$$Recall = \frac{(TP)}{(TP+FN)}$$

$$F1 \text{ Score} = \frac{2 * Precision * Recall}{Precision + Recall}$$



# Destek vektör makinesi (SVM)



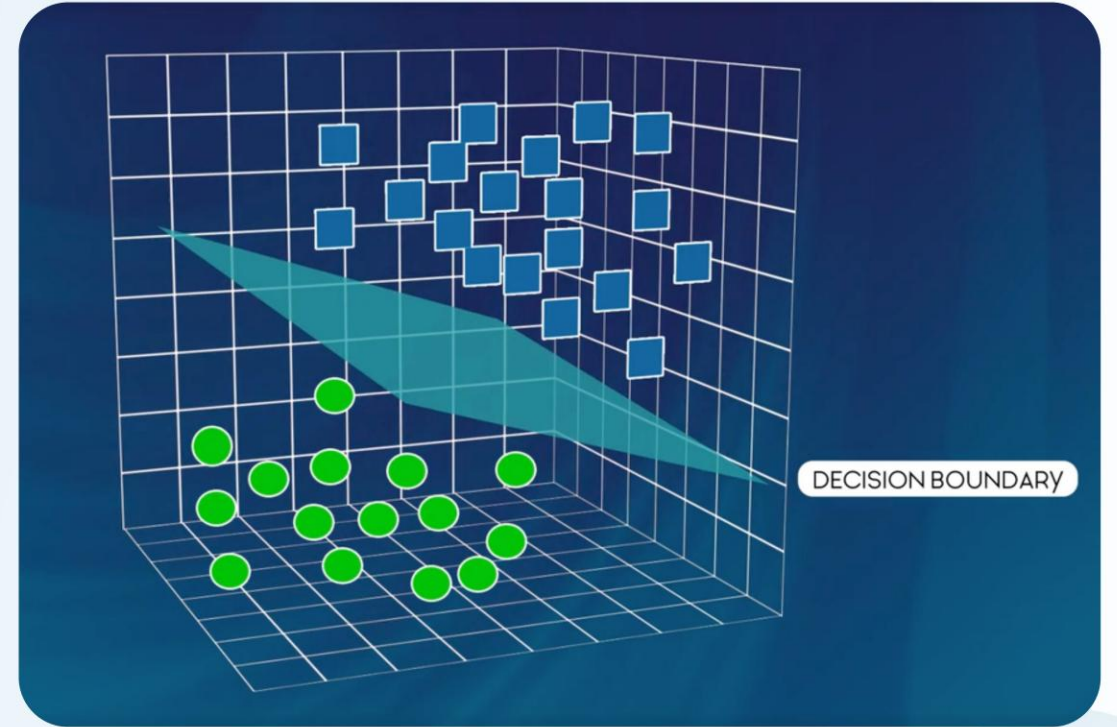
## Destek vektör makinesi (SVM) nedir?

- Destek vektör makinesi (SVM), denetlenen bir ML'dir sınıflandırma ve regresyon problemlerini çözmek için kullanılabilecek bir tekniktir. Ancak çoğunlukla sınıflandırma amacıyla kullanılır.
- Bu algorithma her özellik ve veri noktası uzayda işaretlenir. Daha sonra SVM modeli, farklı veri örneklerini belirli sınıflara ayırmak için sınırlar bulur.

# Pratik bir örnek: iki sınıfı farklı laştıran 2 boyutlu bir düzlem bulma

Diyelim ki iki sınıfı farklı hayvanlardan oluşan bir veri kümemiz var: kuşlar ve balıklar

- Yalnızca üç özellik vardı: vücut ağırlığı, vücut uzunluğu ve günlük yiyecek tüketimi
- 3 boyutlu bir ızgara çiziyoruz ve tüm bu noktaları işaretliyoruz
- Bir SVM modeli bir çözüm bulmaya çalışacaktı. 2 sınıfı farklı laştıran 2 boyutlu düzlem



# Üçten fazla özellik varsa, bir hiper uzaya sahip oluruz.

Hiper-uzay, 4D, 5D gibi 3'ten daha yüksek boyutlara sahip bir alandır ve 3'ten daha büyük bir boyuttaki ayrımcı çizgiye hiper-düzlem denir .

- Hiper düzlemler doğrusal ise SVM çağrılır  
Doğrusal Çekirdek SVM
- Doğrusal olmayan hiper düzlemler için Polinom Çekirdek veya diğer gelişmiş SVM'ler kullanılır

