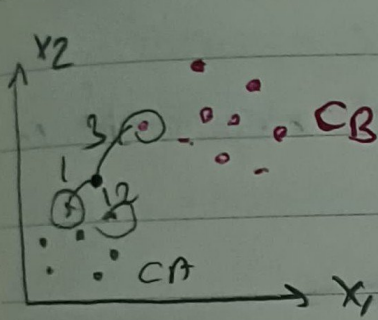# K-En Yakın Komşu (KNN) (K-Nearest Neighbors)

- Sınıflandırma ve Regresyon problemleri için kullanılır.
- Bir veri noktasının sınıfını tahmin etmez veya bir değeri öngörmek için en yakın komşuların etrafın daki veri noktalarının etiketlerini veya değerlerini kullanır.
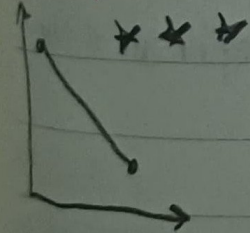

X_1 ve X_2 feature

etizet: CA, CB

K: 3 : Hyper parametre hangi takımına ait olduğunu bulmayan çalışmaz için seçeceğimiz nokta sayısı (3 en yakın komşu A'dan 2 en yakın komşu B'den 1 tae en yakın komşu var. O zaman belirsiz noktay A kategoride derler.
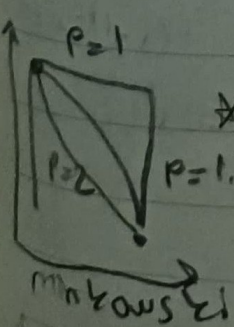
Mesefe (Distance) (Hyperparameter) → Bizler tarafından seçilmesi gereken → Düzüz kayıtlı veri setlerinde ve kategorik değişkenlerle çalışırkenki

Euclidean (Kuş uçuşu)   manhottan (Dikdörtgen)

P=1
P=2
P=1.3   P parametresine göre karar verir.

* Öklid ve manhatta mesefelerinin ı genelleştirin

minkowski

1	0	0	1	1

1	1	0	1	1

Hamming

* kategorik değişkenler arasındaki farklılıkları ölçer. 2 vektör arasındaki öğe sayısını bulur. İki metin arasındaki uzaklığı farklılıklar hesaplar. (değerler kategoriridir.

Heperparametre Optimizasyonu (K)

Elbow methodu ile seçim gerçekleşir.
* fıklı k değerını deneyerek modelin performansını değerladirı.
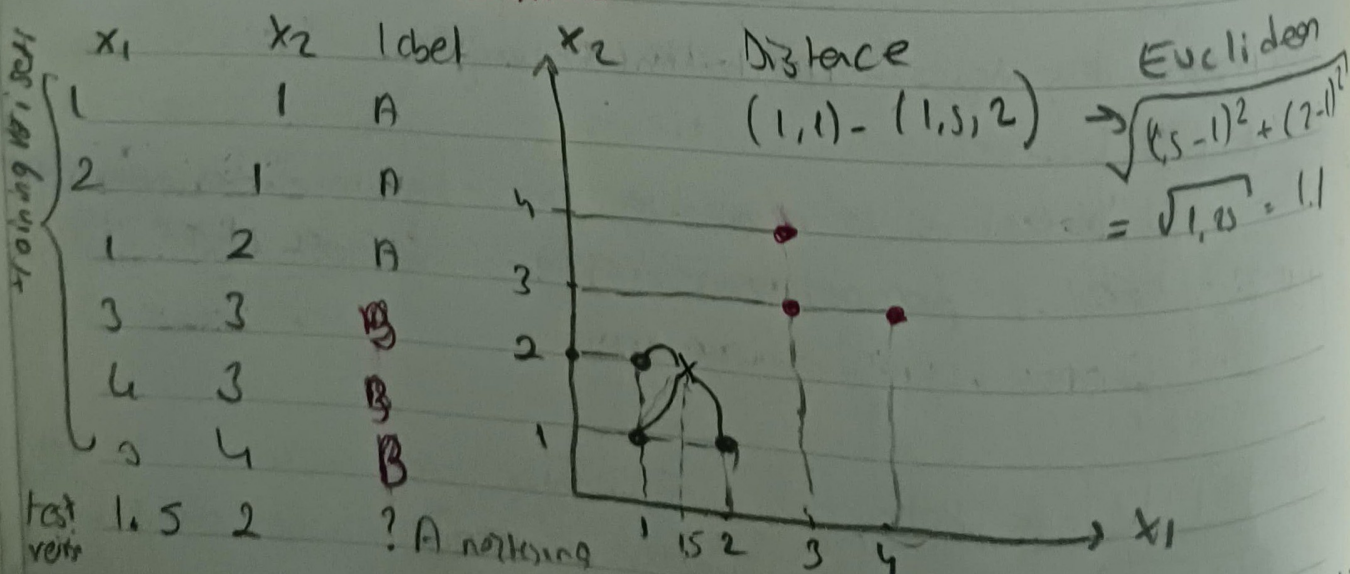

performans (total)

1 2 3 4 5 6 7 → K
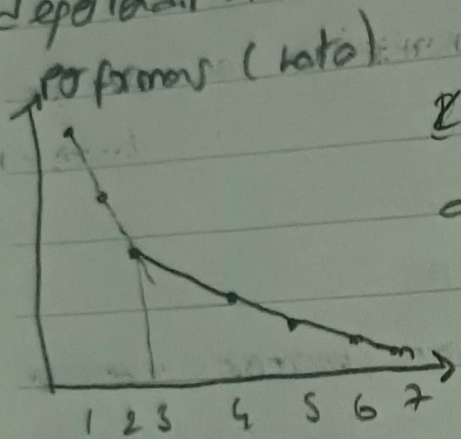
$K = 3$ olduğunda ani bir eğim
değişimi var 3 nozlerinden sonra
herhangi bir K değeri benim
K problemime ya.

* Söpliz - ticaret ve finansta bullanldbilir KNN.

Avantaj	Dezavata
• Basit ve doğruluk	• Boyutluluk problemi
• Parametre ayarı kolaylığı	• veriye duyarlı (outlinelda etkili)
• örnek bazlı öğrenme	• yüksek hesaplama maliyeti

→ ML dezi train pez söz brusu değil veriye
algoritması doğrultusnda karar veriyor.

KNN Durum Çalışması

X₁	X₂	label
1	1	A
2	1	A
1	2	A
3	3	B
4	3	B
3	4	B

Distance

$(1,1) - (1.5,2)$

Euclidean

$$\sqrt{(1.5-1)^2 + (2-1)^2}$$
$$= \sqrt{1.25} = 1.1$$

test	1.5	2	?A noktasına

aittir.

$K = 3$ için

X₂

4
3
2
1

1.5 2 3 4 → X₁

K yı çift seçersez sıkıntı obbir.
Tez sayıda seçmeye çalış

Karar Ağaçları

Karar ağaçları, ağaç yapısı şeklinde bir dizi karar kurarak veri kümesini bölerek hedef değişkenin tahmin edilmesini sağlar.

Root node → kök düğüm

internal node — internal node → iç düğümler

leaf node " " " — i → yaprak düğüm
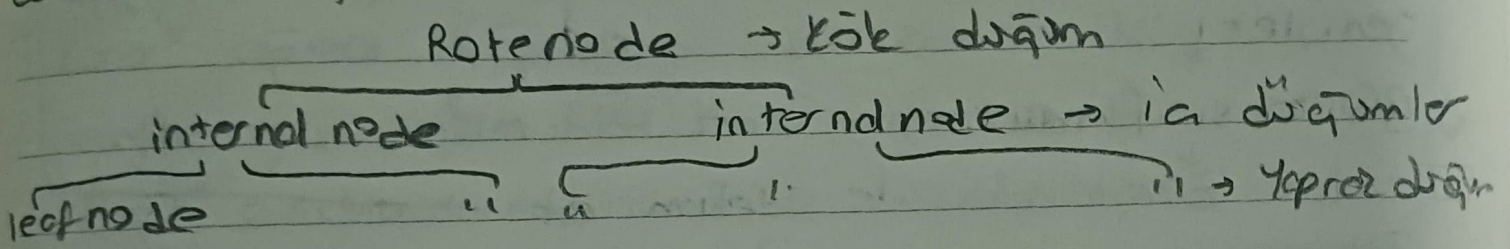
Karar Ağaçları Türleri

- ID3 (Iterative Dichotomiser 3)
→ alt bölünmeleri değerlendirmek için entropi ve bilgi kazancını kullanır.

- C4.5
→ ID3'ün daha sonrazi versiyonudur

- CART (sınıflandırma ve regresyon ağaçları)
→ ideal bölünme için Gini kullanılır. Gini başarsızlığı rastgele seçilen bir özelliğin ne sıklıkla yanlış sınıflandırıldığını ölçer.

	ID3	C4.5	CART
Veri tipi	Kategorik	Sürekli ve kate.	Sürek ve kate.
Veri boyutu	orta ve küçük	Orta ve büyük	Büyük
Hız	Düşük	Orta	Yüksek
Boosting	Yok	%2	Var (Rastgele ormanlar, Gradient bootin)
Kayıp veri	Hassas	Orta Hassas	Yüksek (kayıp verinin üstesinden gelir)
Algoritma	Entropi, Bilgi kazancı	bölme Bilgisi kazanç oranı	Gini Impurity