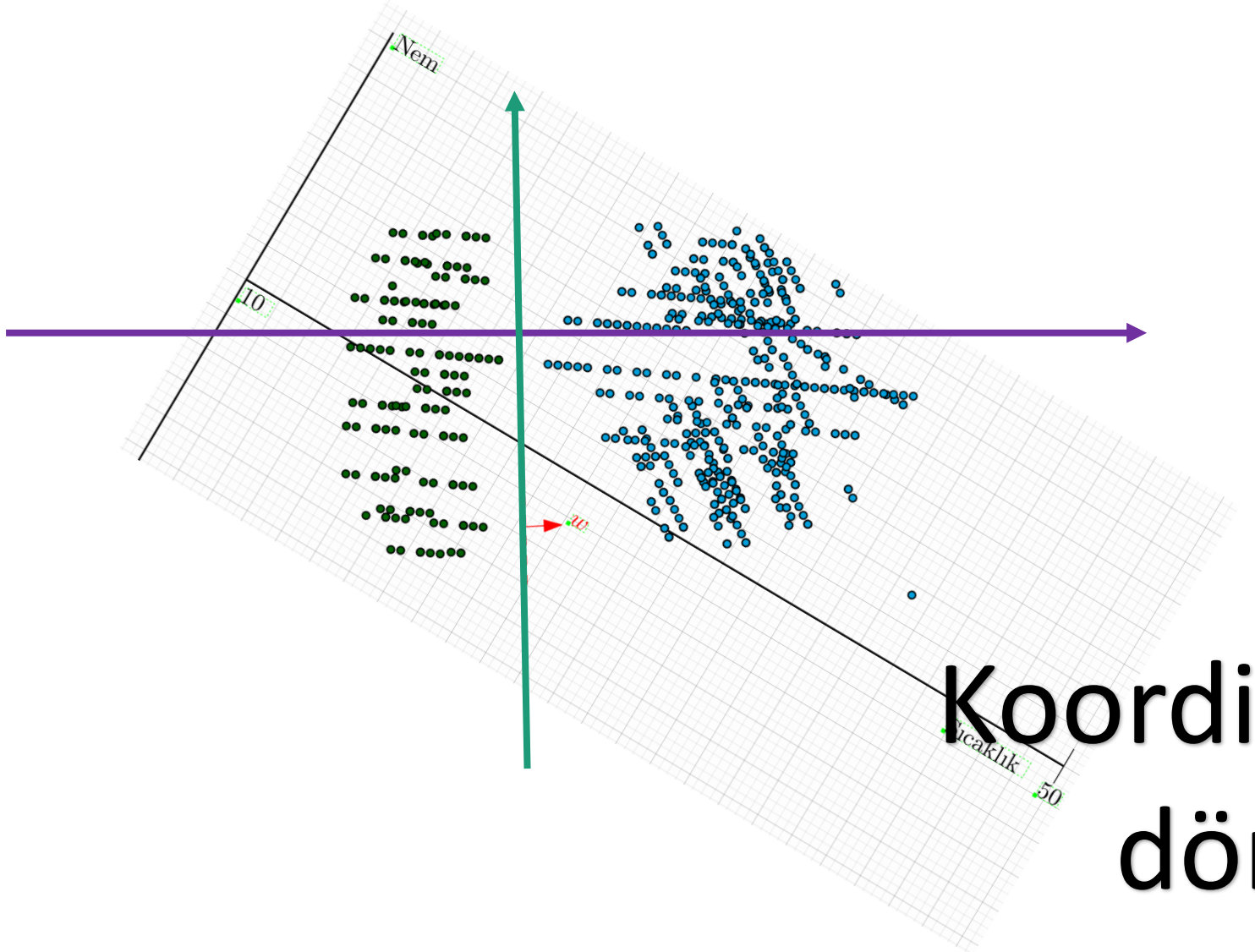
BİL 475 Örüntü Tanıma

Hafta-3:

K-En Yakın Komşu Algoritması

Temel Olasılık

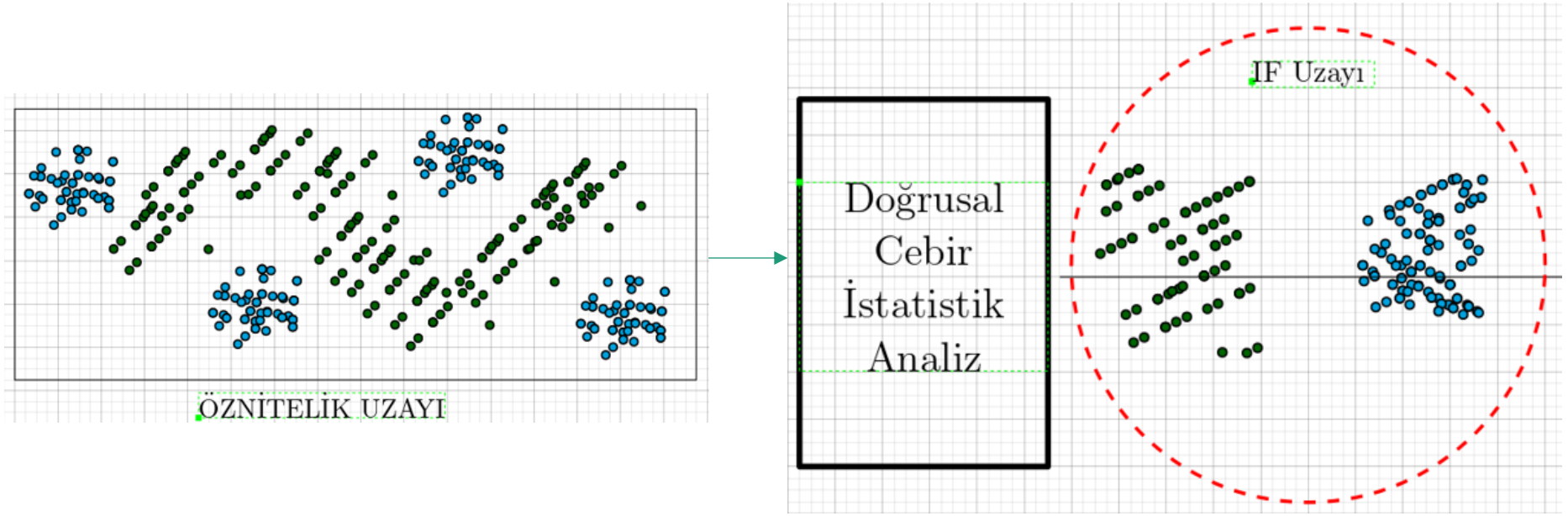
Öznitelik Kavramı



Koordinat sistemi dönüşümü

Öznitelik Kavramı

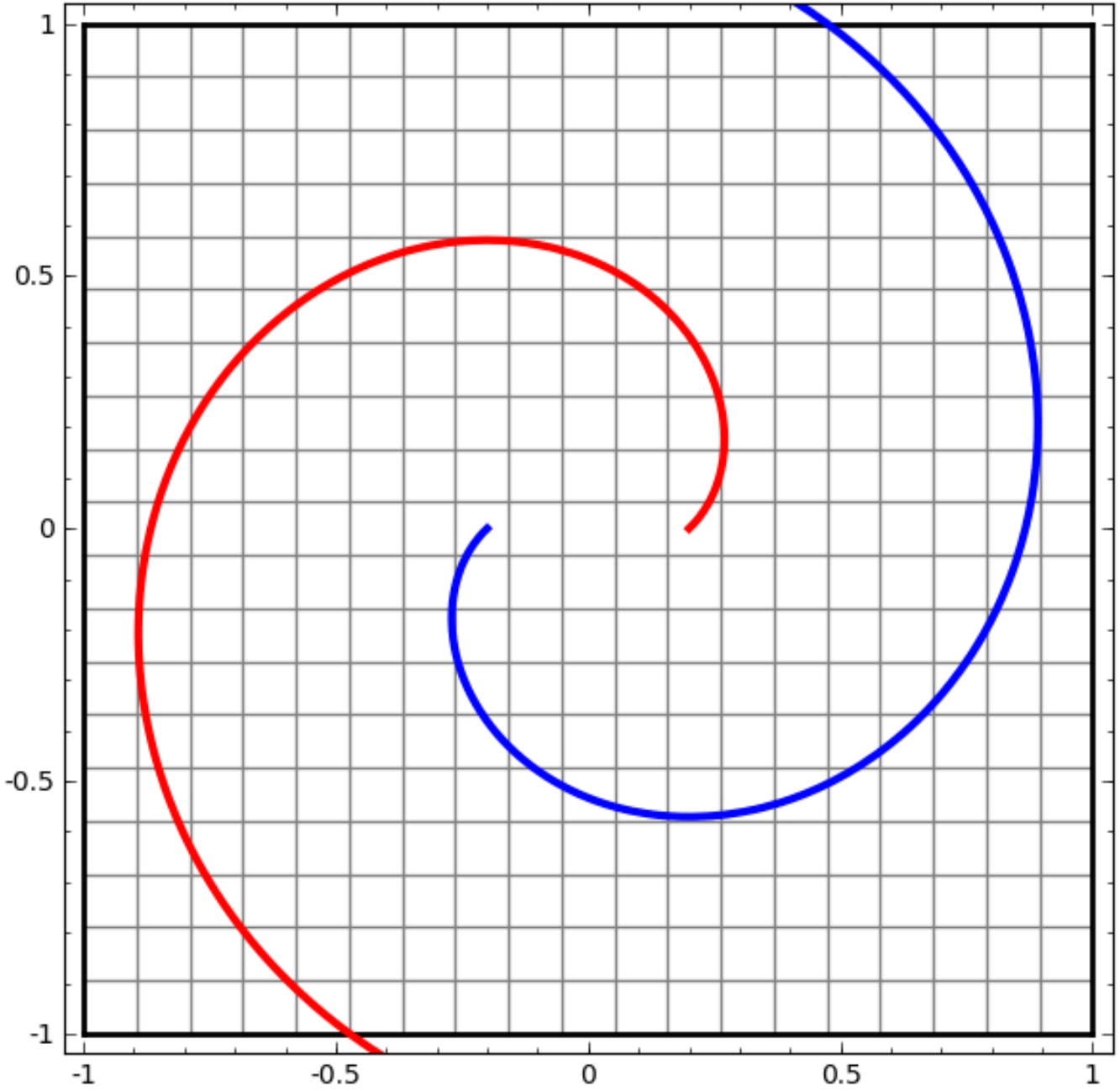
Öznitelik Kavramı



Öznitelik Kavramı

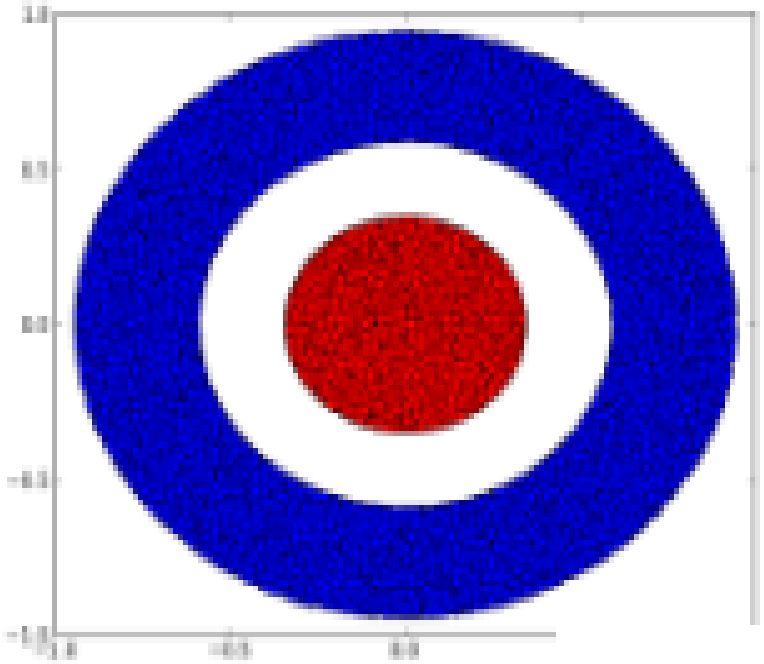
Makine Öğrenmesinin Amacı

Doğrusal ayrıştırılabilir bir uzaya koordinat dönüşümü yapmak



http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

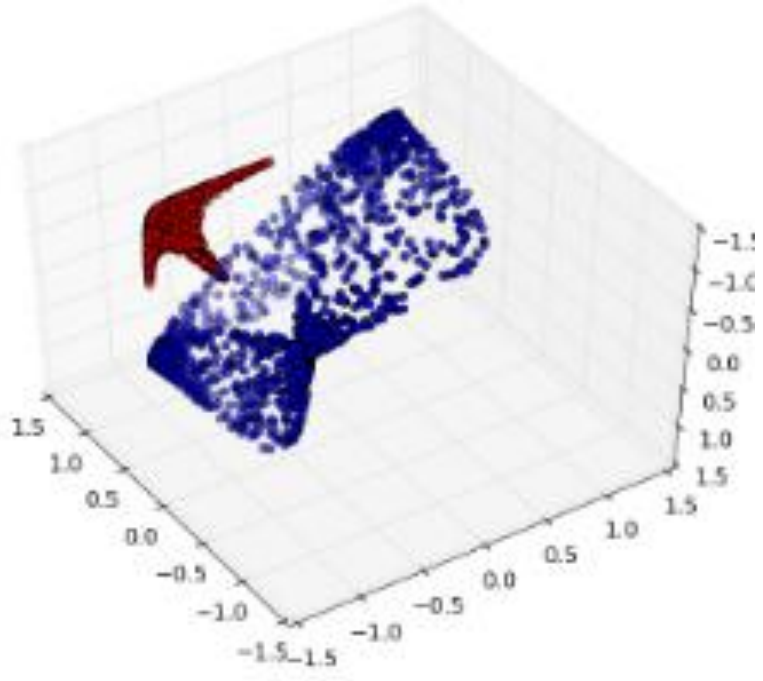
Öznitelik Kavramı



Makine Öğrenmesinin Amacı

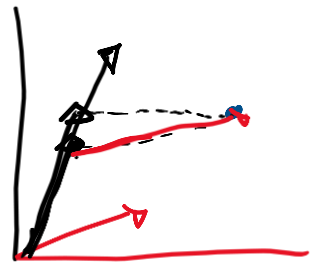
Doğrusal ayrıştırılabilir bir uzaya

koordinat dönüşümü yapmak

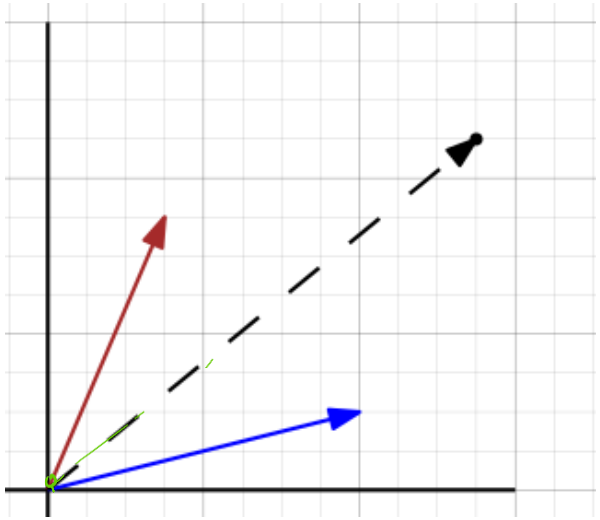


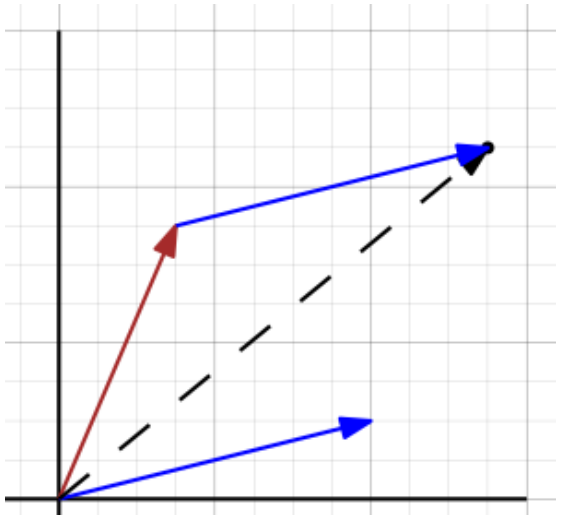
http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

Sayılar, Vektörler ve Matrisler



Fiziksel Olgulardan Matematiğe Geçiş





Baz vektörleri ile diğer vektörleri ifade edebilme (Taylor , Fourier)

$$\begin{bmatrix}3\\2\\2\end{bmatrix}x+\begin{bmatrix}7\\3\\2\end{bmatrix}y+\begin{bmatrix}3\\2\\2\end{bmatrix}z=\begin{bmatrix}11\\1\\-5\end{bmatrix}$$

Sayılar, Vektörler ve Matrisler

Vektör İşlemleri

Toplama

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_1+y_1 \\ x_1+y_2 \\ \vdots \\ x_N+y_N \end{bmatrix}$$

Transpoze

$$\bar{x}^T = [x_1, x_2, \dots x_N]$$

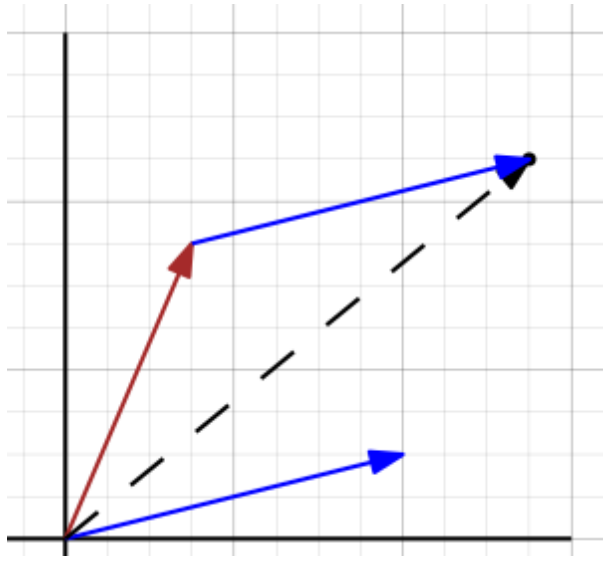
İç çarpım (Bra-ket)

$$\langle \bar{x}, \bar{y} \rangle = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}$$

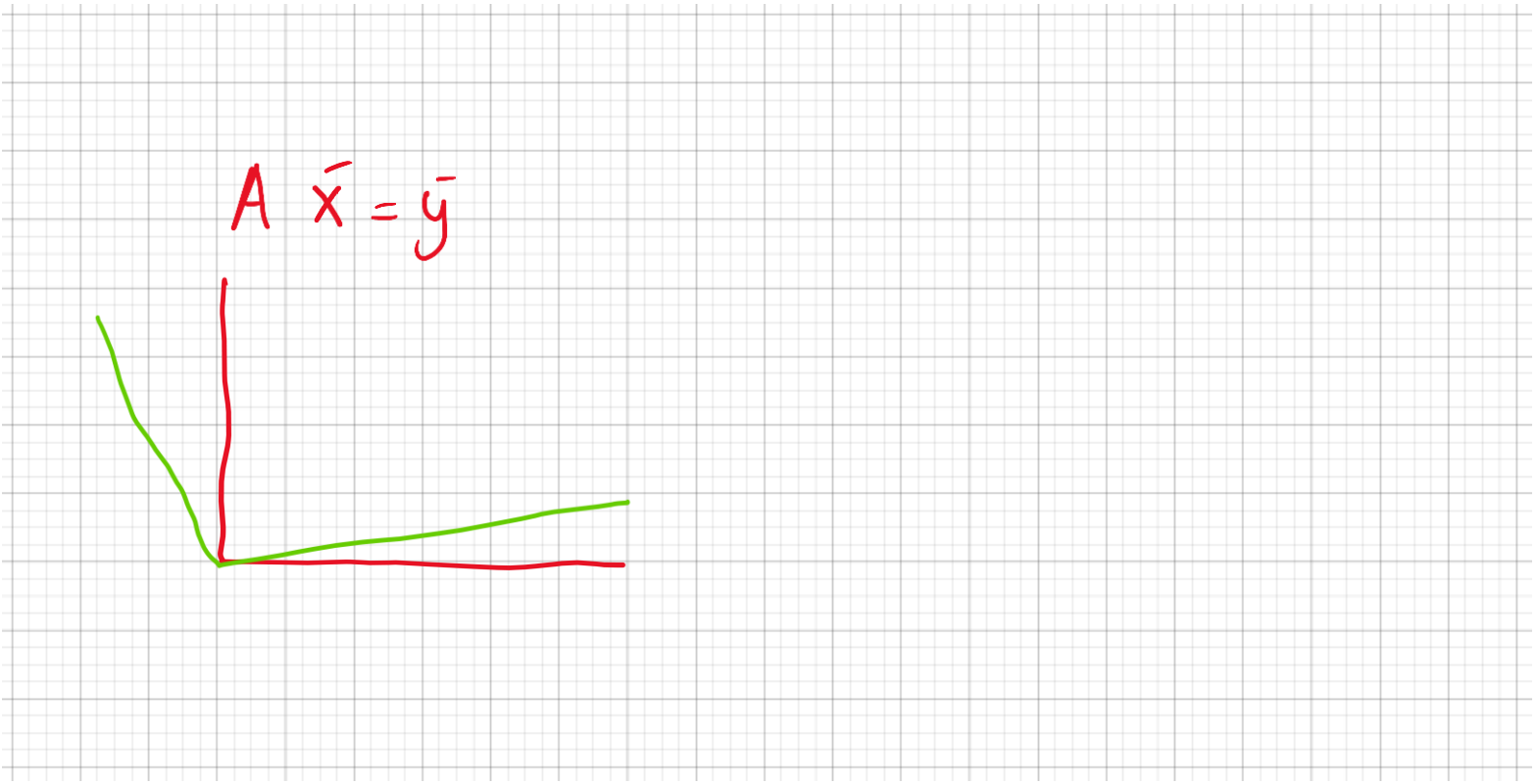
$$\bar{x} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \quad \bar{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \Rightarrow 1.2 + 4.3 = 14$$

Sayılar, Vektörler ve Matrisler

Vektör İşlemleri



Doğrusal Dönüşüm
(Linear Transformation)



Sayılar, Vektörler ve Matrisler

Vektör İşlemleri

```
clear all, close all; clc

v1 = [1,2,3]';
v2 = [2,2,3]';
v3 = [1,1,1]';

V = [v1,v2,v3]

rank(V)
```

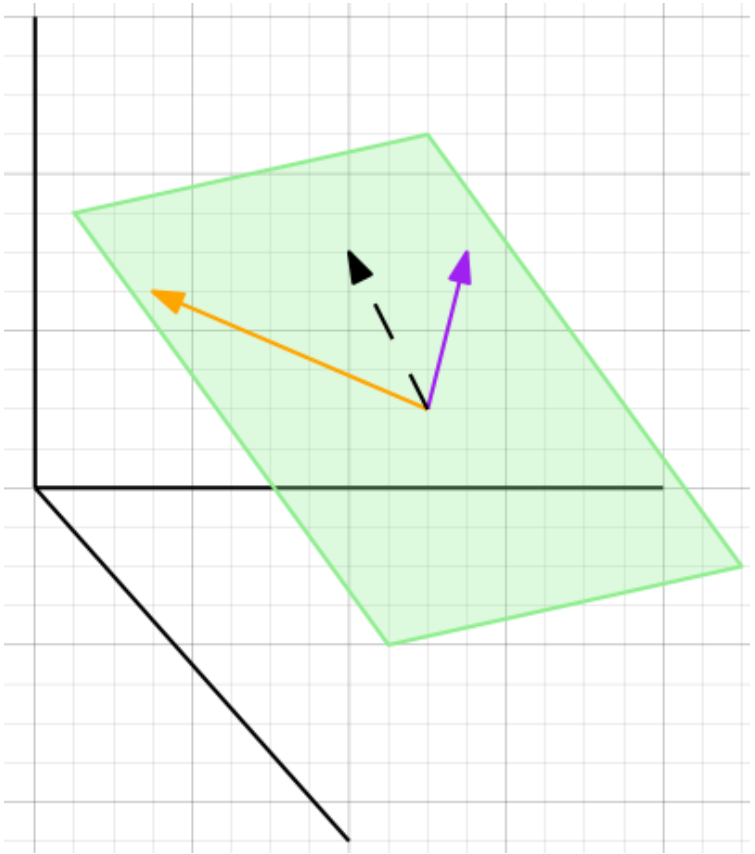
```
V =

     1     2     1
     2     2     1
     3     3     1


ans =

     3
```

Uzayın Germe: İstanbul'da Yemek Kültürü



Sayılar, Vektörler ve Matrisler

Vektör İşlemleri

```
clear all, close all; clc

v1 = [1,2,3]';
v2 = [2,2,3]';
v3 = randn(1)*v1 + randn(1)*v2;

V = [v1,v2,v3]

rank(V)
```

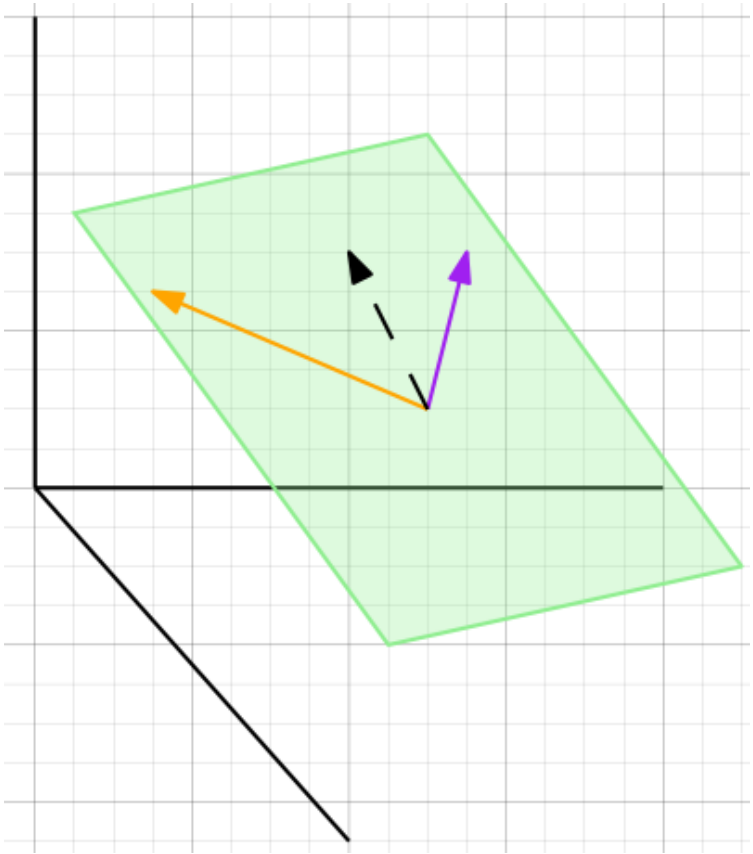
```
V =

    1.0000    2.0000   -3.7624
    2.0000    2.0000   -4.0244
    3.0000    3.0000   -6.0366


ans =

     2
```

Uzayın Germe: İstanbul'da Yemek Kültürü

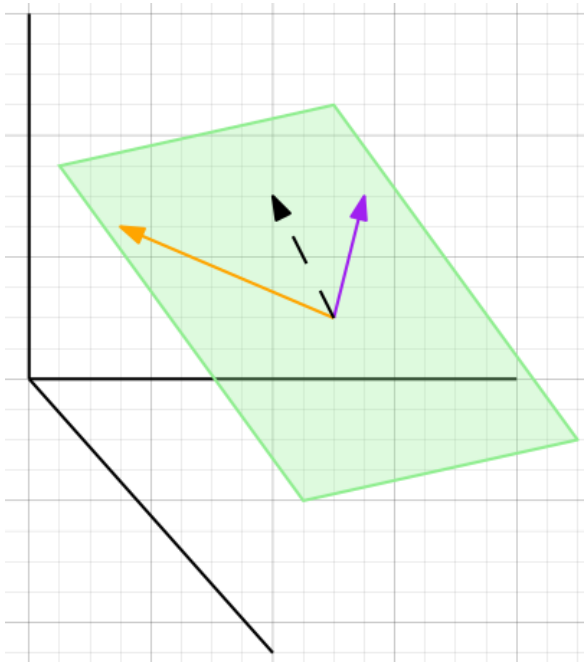


Sayılar, Vektörler ve Matrisler

Vektör İşlemleri

Uzayın Germe ve Doğrusal Bağımsızlık

Baz vektörleri,
gerdikleri uzay içinde
herhangi bir noktayı oluşturabilirler.

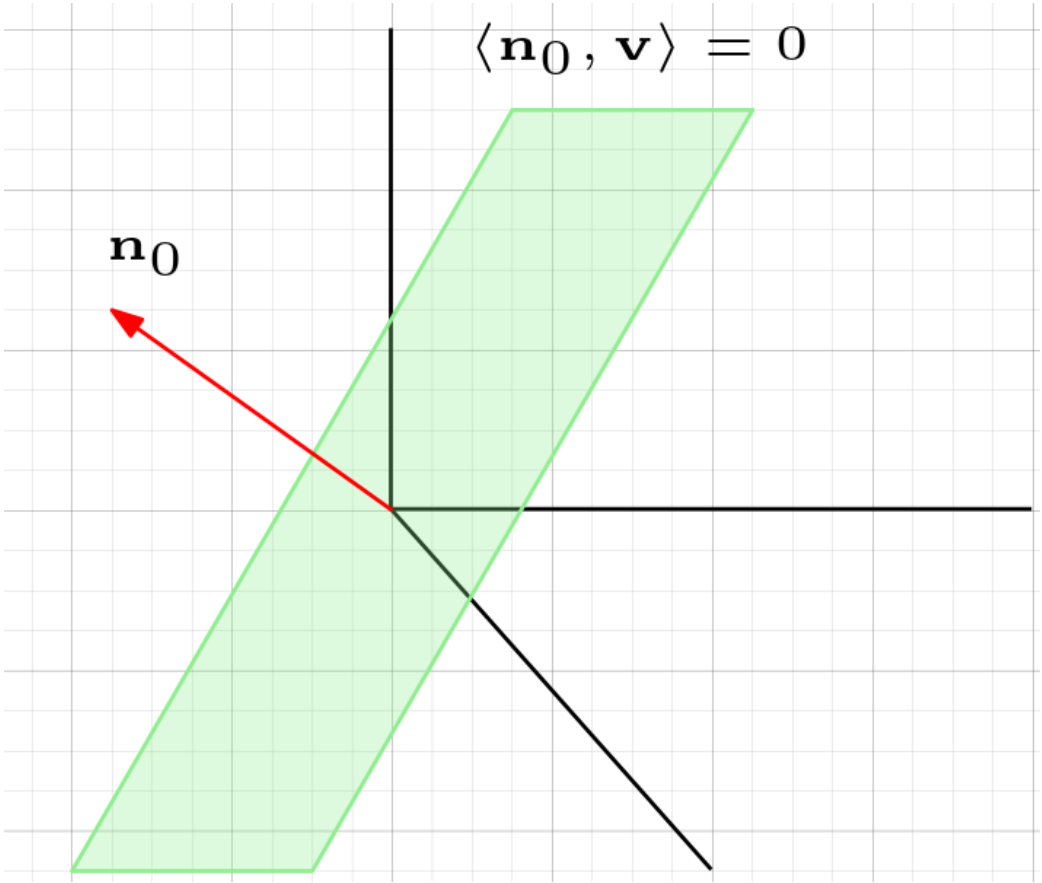


Bu kural N boyutlu uzaylar için geçerlidir. !!

Göremesek bile sayılar bize yol gösterici
Örnek: Norm

Sayılar, Vektörler ve Matrisler

- Düzlem Denklemi (İpucu-İki vektör dikse iç çarpım 0'dır):



$\bar{n}_0 = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$ $\bar{x}_0 = \begin{bmatrix} 8 \\ 1 \end{bmatrix}$ $\bar{x}_1 \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ $\bar{x}_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

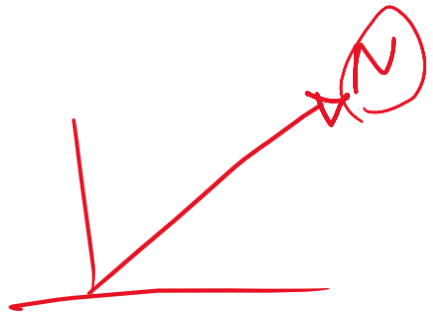
$\langle \bar{n}_0, \bar{x}_0 \rangle = -8 + 4 = -4$

$\langle \bar{n}_0, \bar{x}_1 \rangle = -4 + 8 = 4$

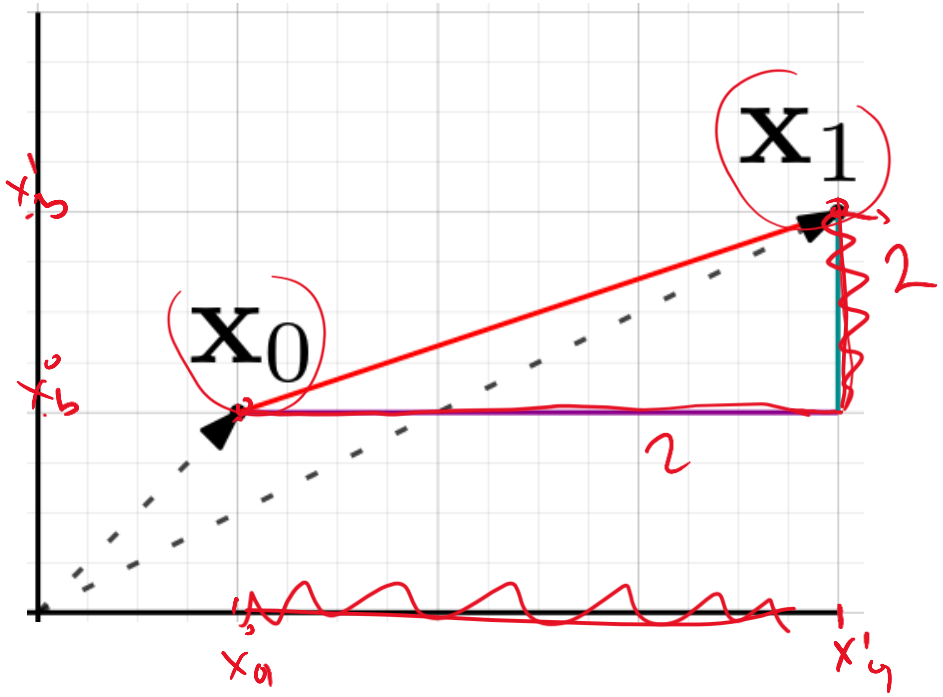
$\langle \bar{n}_0, \bar{x}_2 \rangle = -3 + 4 = 2$

Uzayı İkiye Böler !!!

Sayılar, Vektörler ve Matrisler



NORM KAVRAMI



$$\bar{x}_1 = \begin{bmatrix} x'_0 \\ x'_1 \end{bmatrix}$$

$$\sqrt{\left|x'_0 - x^0_{01}\right|^2 + \left|x'_1 - x^0_1\right|^2}$$

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

$3^2 + 4^2 = 5^2$

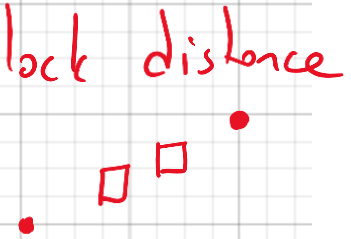
$$|\bar{x}| = \sqrt{\sum_{i=1}^{N} |x_i|^2}$$

Sayılar, Vektörler ve Matrisler

NORM KAVRAMI

$\mathbf{x} \in \Re^{N \times 1}$ olsun

ℓ_1 -norm $= |\mathbf{x}|_1 \longrightarrow \sum_{i=1}^{N} |\mathbf{x}(i)|$



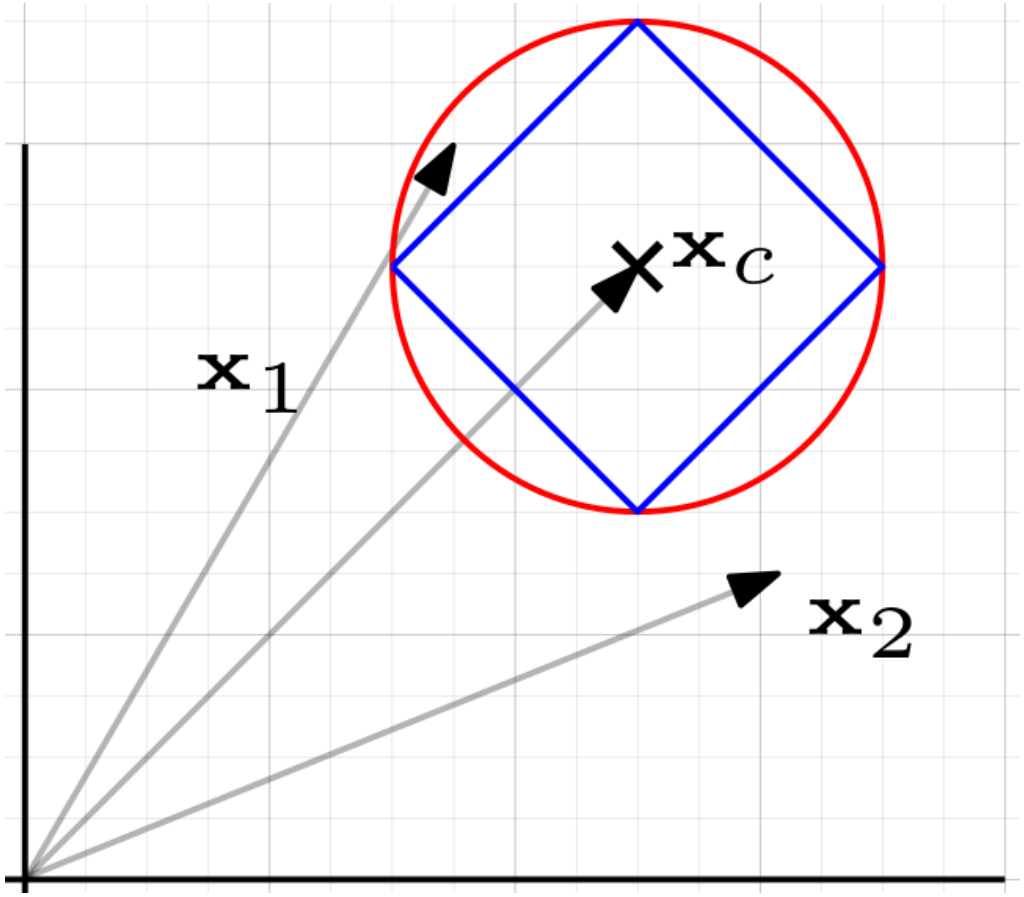
ℓ_2 -norm $= |\mathbf{x}|_2 \longrightarrow \sqrt[2]{\sum_{i=1}^{N} |\mathbf{x}(i)|^2}$

ℓ_p -norm $= |\mathbf{x}|_p \longrightarrow (\sum_{i=1}^{N} |\mathbf{x}(i)|^p)^{\frac{1}{p}}$

ℓ_∞ -norm $= |\mathbf{x}|_\infty \longrightarrow max\{|\mathbf{x}(i)| : i \in 1, 2, 3, .., N\}$

Sayılar, Vektörler ve Matrisler

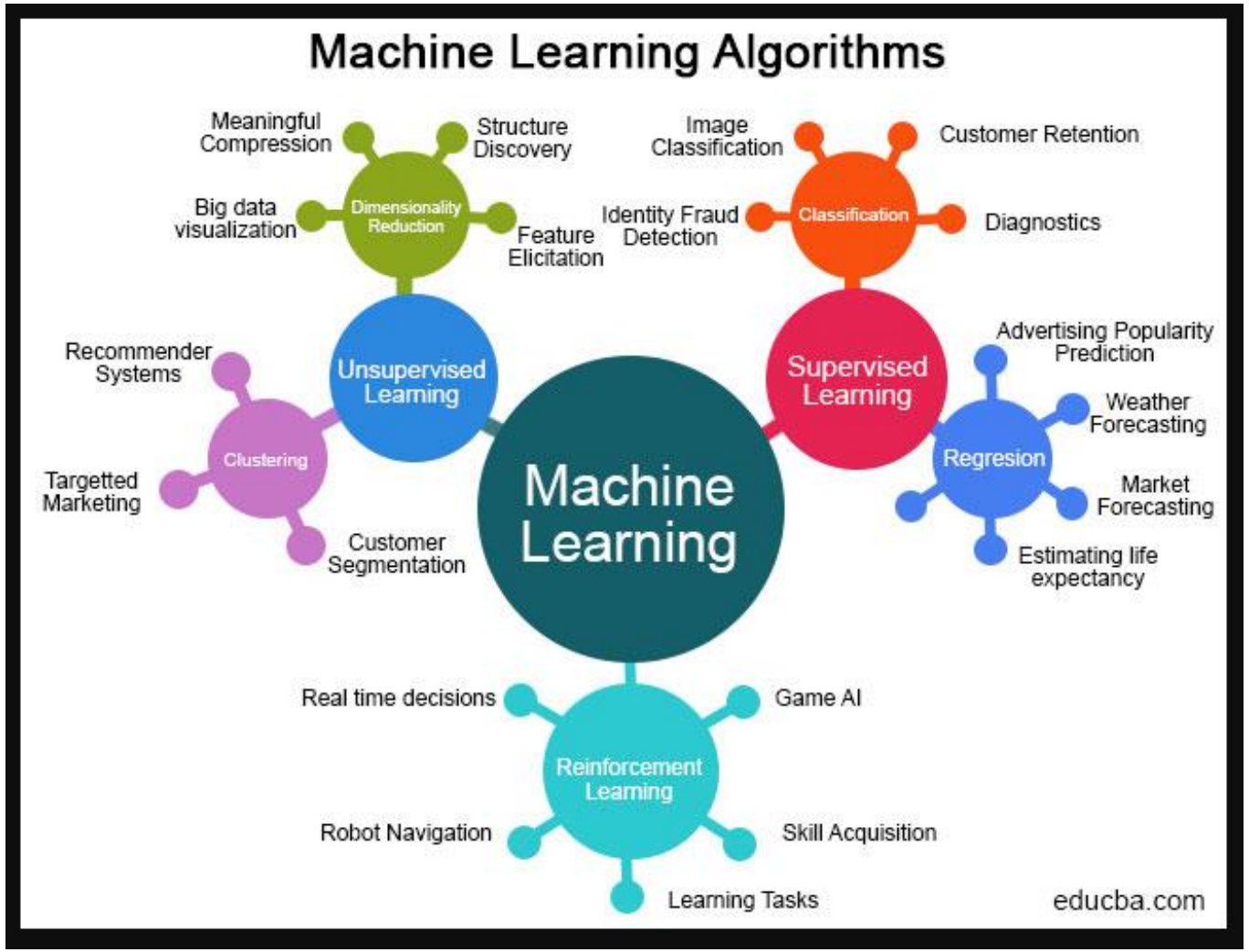
NORM KAVRAMI



$$\boldsymbol{x} \in \Re^{2\times 1} \quad |\boldsymbol{x} - \boldsymbol{x_c}|_p = 1$$

Verinin etrafını sarma

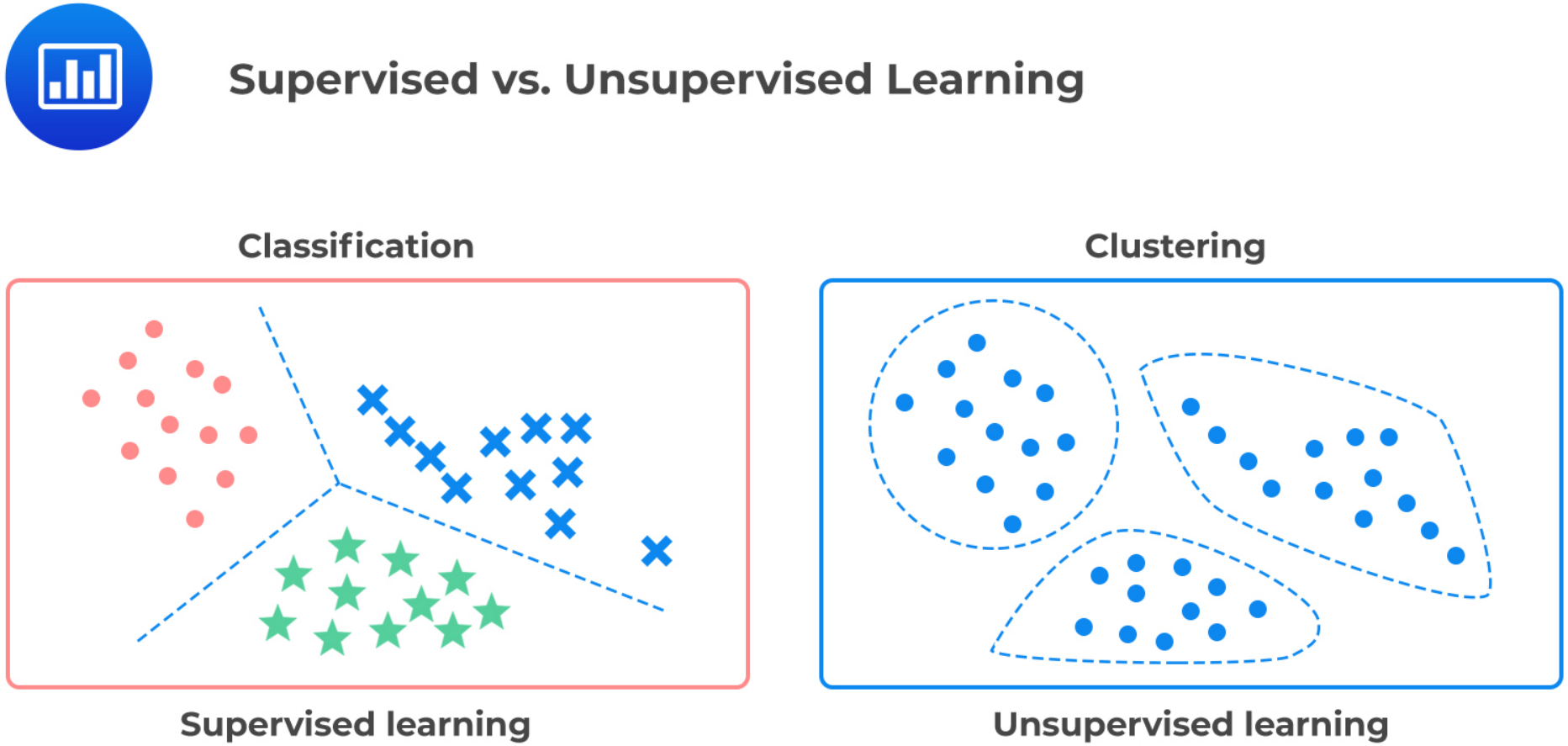
Denetimli Öğrenme



Denetimli Öğrenme

- Denetimli Öğrenme
- Veri
- Veri seti
- Eğitim verisi
- Test verisi

Denetimli Öğrenme (Tersi ??)



https://analystprep.com/study-notes/cfa-level-2/quantitative-method/supervised-machine-learning-unsupervised-machine-learning-deep-learning/

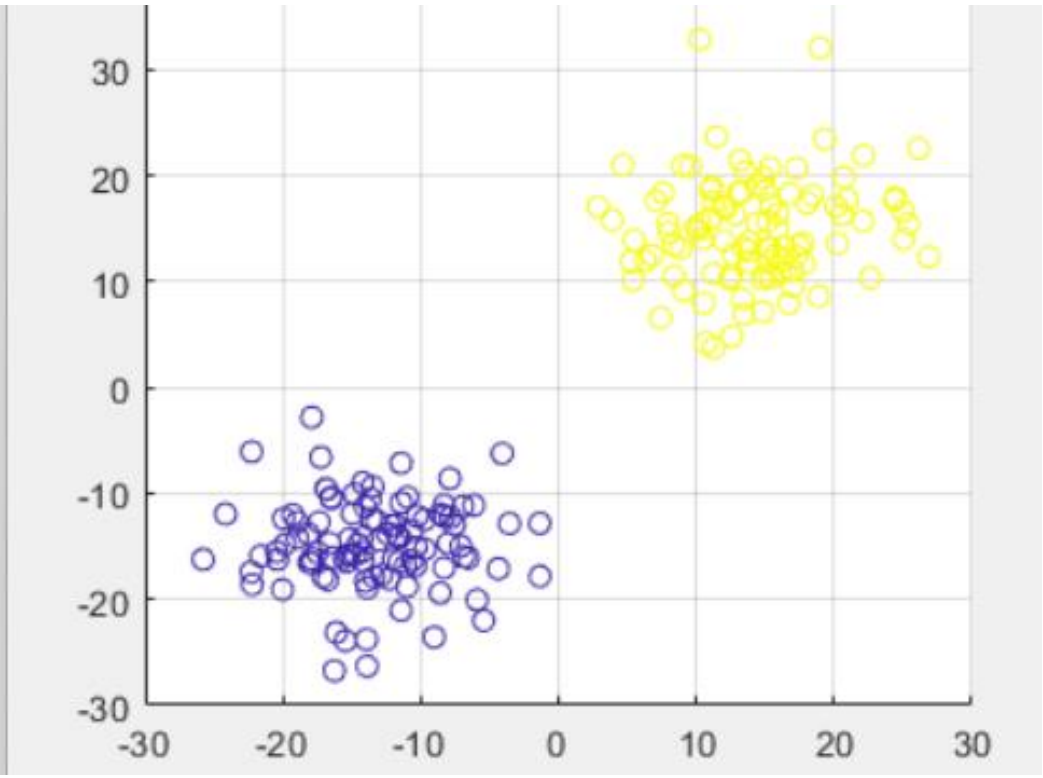
```
clear all, close all; clc


x = -15 + 5*randn(100,2);
y =  15 + 5*randn(100,2);

z = [x ; y];

c = [ones(1,100) , 2*ones(1,100)];

figure
scatter(z(:,1) , z(:,2) ,[] ,c),grid on
```

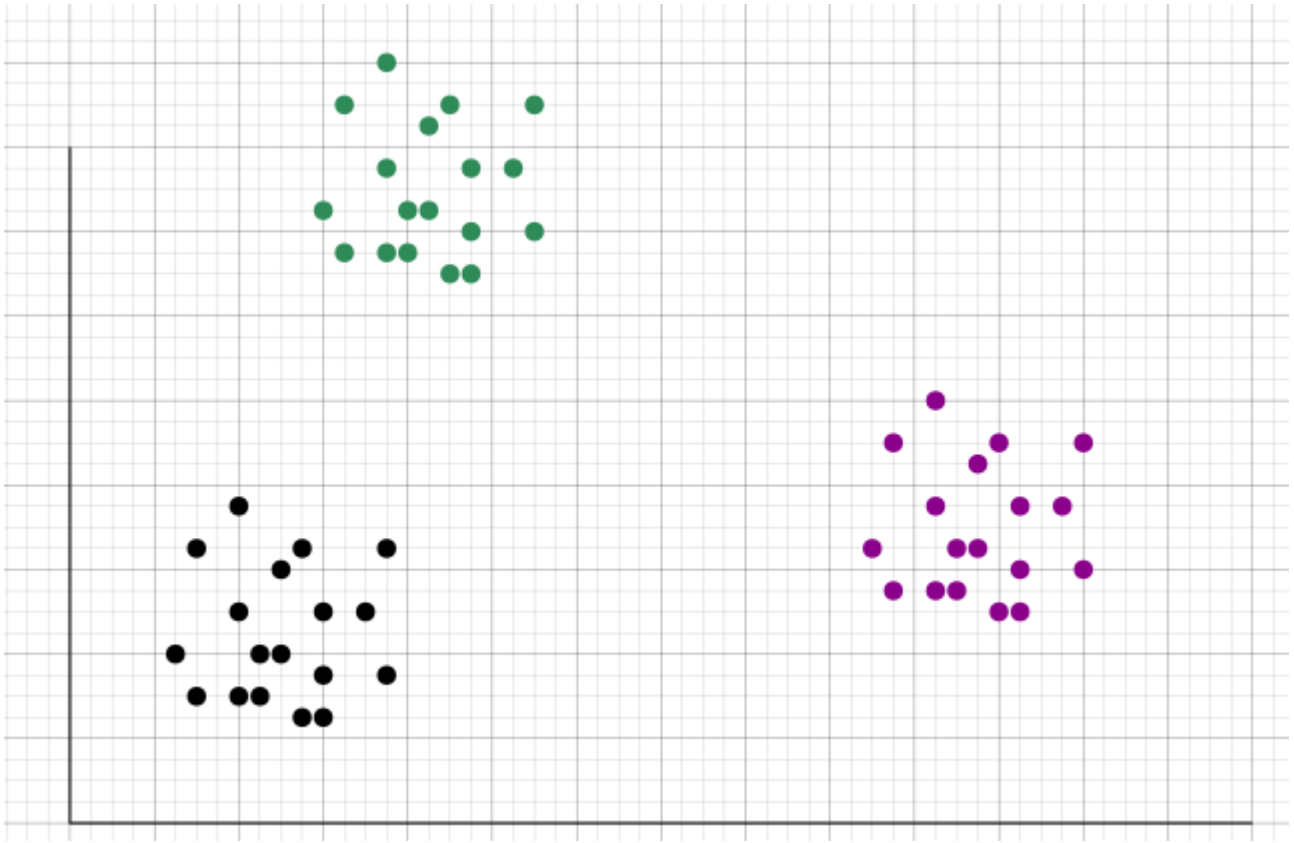


z

200x2 double

	1	2
1	-8.2906	-17.0246
2	-19.9422	-12.3607
3	-5.9103	-20.0348
4	-16.8722	-9.5549
5	-22.2587	-6.0756
6	-18.0934	-16.5188
7	-10.3275	-15.0435
8	-9.7204	-12.4668
9	-14.1989	-8.9837
10	-13.5630	-12.3899
11	-11.8355	-13.0148
12	-22.2952	-17.4141
13	-17.9085	-16.1575
14	-24.1507	-11.9331
15	-17.2455	-6.5857
16	-10.2536	-12.1580

K-En Yakın Komşu Sınıflandırma (k-NN)



K-En Yakın Komşu Sınıflandırma (k-NN)

```
X = [1,2        L = [1        v = [6,5];
     4,4             2
     7,2             3
     8,6             3
     11,5];          4];
```

$Metrik: \ell_2 - Norm$

K-En Yakın Komşu Sınıflandırma (k-NN)

```
X = [1,2
     4,4
     7,2
     8,6
     11,5];

v = [6,5];
```

```
D =
```

-5	-3
-2	-1
1	-3
2	1
5	0

25	9
4	1
1	9
4	1
25	0

34
5
10
5
25

5.8310
2.2361
3.1623
2.2361
5.0000

```
L = [1
     2
     3
     3
     4];
```

K-En Yakın Komşu Sınıflandırma (k-NN)

$$Metrik: \ell_1 - Norm\ Olsaydı$$

```
D =

   -5    -3
   -2    -1
    1    -3
    2     1
    5     0

    5     3
    2     1
    1     3
    2     1
    5     0

    8
    3
    4
    3
    5

L = [1
     2
     3
     3
     4];
```

K-En Yakın Komşu Sınıflandırma (k-NN)

- Euclidian Distance (ℓ_2)
- City Block (Manathan) (ℓ_1)
- Minkowski Dist.
- Chebyshev (ℓ_∞)

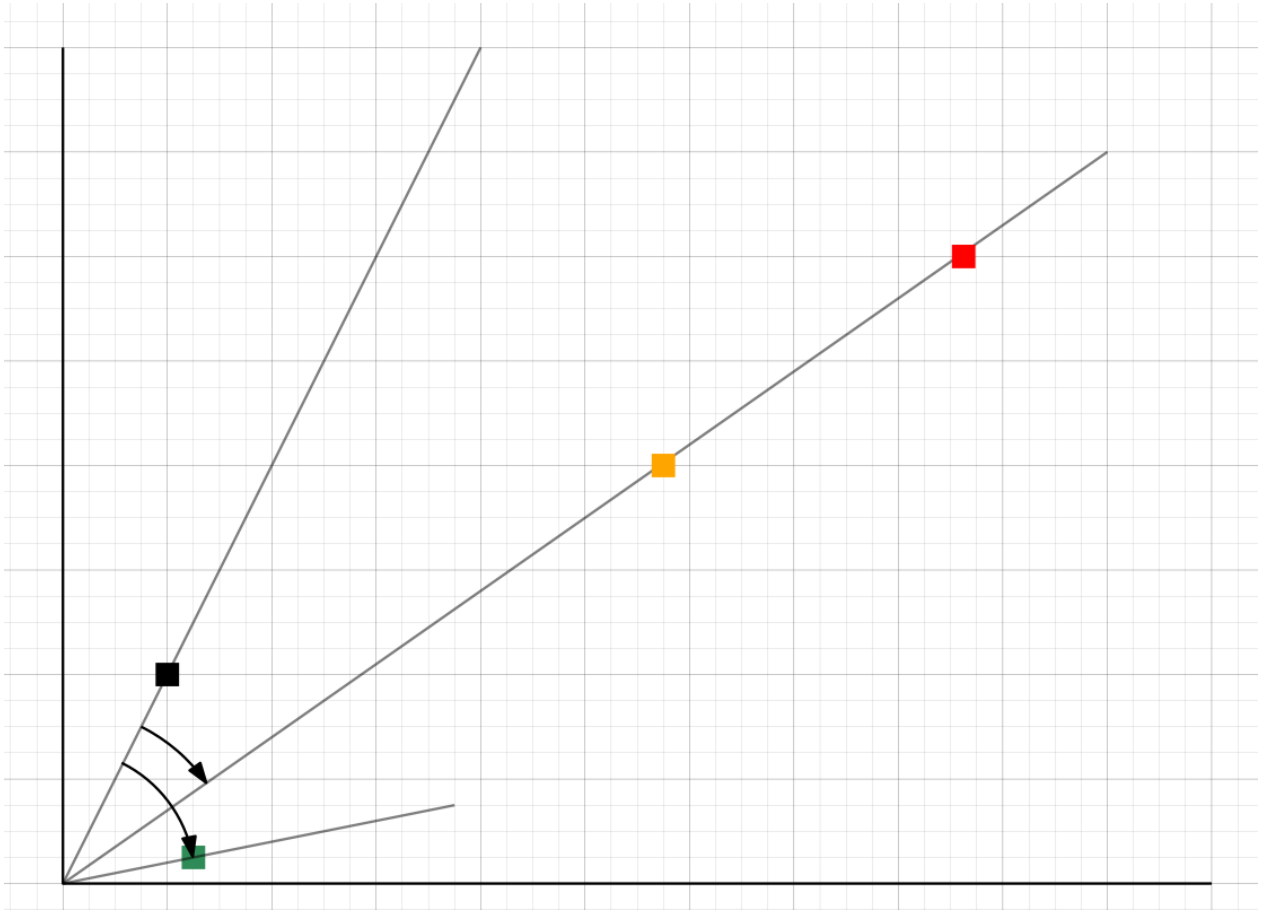
$$\left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}$$

K-En Yakın Komşu Sınıflandırma (k-NN)

- Euclidian Distance (ℓ_2)
- City Block (Manathan) (ℓ_1)
- Minkowski Dist.
- Chebyshev (ℓ_∞)
- Cosine Similarity

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}$$

K-En Yakın Komşu Sınıflandırma (k-NN)



$$\frac{\langle x_1, x_2 \rangle}{|x_1||x_2|}$$

$$\mathbf{u}.\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$$

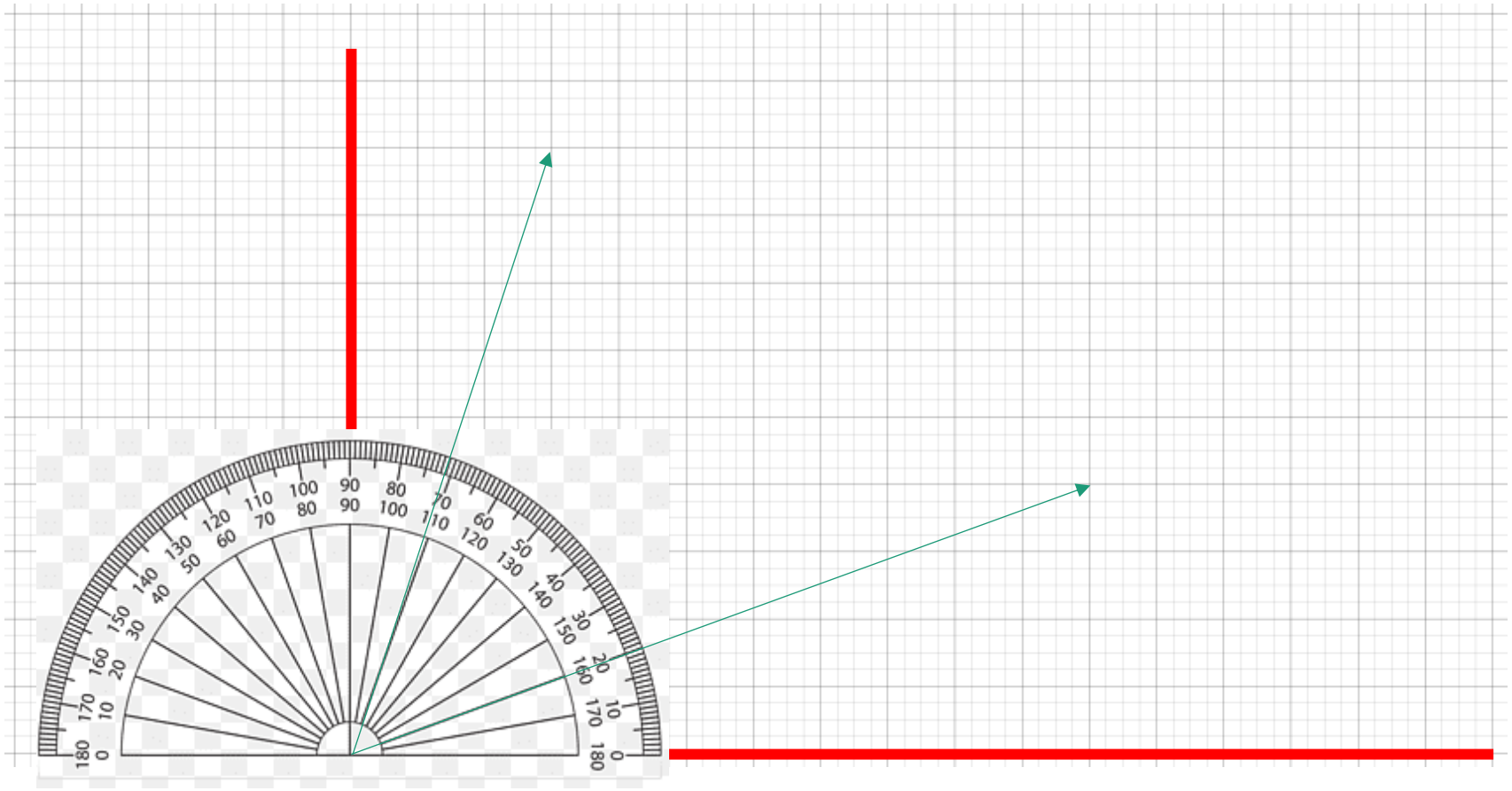
$$\mathbf{u}.\mathbf{v} = ||\mathbf{u}||||\mathbf{v}|| \cos\theta$$

K-En Yakın Komşu Sınıflandırma (k-NN)

$$\mathbf{u}.\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$$

$$\mathbf{u}.\mathbf{v} = ||\mathbf{u}||||\mathbf{v}|| \cos\theta$$

$$\frac{\langle x_1, x_2 \rangle}{|x_1||x_2|}$$



K-En Yakın Komşu Sınıflandırma (k-NN)

Algoritma adımları

- Etiketli bir veri seti
- Test örneği
- T-V her örnekle mesafeler hesaplanacak
- Sort (en düşük mesafe , indis)
- K tane en düşük örnek arasında baskın olan sınıfı

- PYTHON

Parameters::

n_neighbors : *int, default=5*
Number of neighbors to use by default for `kneighbors` queries.

weights : *{'uniform', 'distance'} or callable, default='uniform'*
Weight function used in prediction. Possible values:

- 'uniform' : uniform weights. All points in each neighborhood are weighted equally.
- 'distance' : weight points by the inverse of their distance. in this case, closer neighbors of a query point will have a greater influence than neighbors which are further away.
- [callable] : a user-defined function which accepts an array of distances, and returns an array of the same shape containing the weights.

algorithm : *{'auto', 'ball_tree', 'kd_tree', 'brute'}, default='auto'*
Algorithm used to compute the nearest neighbors:

- 'ball_tree' will use `BallTree`
- 'kd_tree' will use `KDTree`
- 'brute' will use a brute-force search.
- 'auto' will attempt to decide the most appropriate algorithm based on the values passed to `fit` method.

Note: fitting on sparse input will override the setting of this parameter, using brute force.

leaf_size : *int, default=30*
Leaf size passed to BallTree or KDTree. This can affect the speed of the construction and query, as well as the memory required to store the tree. The optimal value depends on the nature of the problem.

p : *int, default=2*
Power parameter for the Minkowski metric. When p = 1, this is equivalent to using manhattan_distance (l1), and euclidean_distance (l2) for p = 2. For arbitrary p, minkowski_distance (l_p) is used.

metric : *str or callable, default='minkowski'*

Metric to use for distance computation. Default is "minkowski", which results in the standard Euclidean distance when p = 2. See the documentation of scipy.spatial.distance and the metrics listed in `distance_metrics` for valid metric values.

If metric is "precomputed", X is assumed to be a distance matrix and must be square during fit. X may be a sparse graph, in which case only "nonzero" elements may be considered neighbors.

If metric is a callable function, it takes two arrays representing 1D vectors as inputs and must return one value indicating the distance between those vectors. This works for Scipy's metrics, but is less efficient than passing the metric name as a string.

sklearn.metrics.pairwise.distance_metrics

`sklearn.metrics.pairwise.distance_metrics()` [source]

Valid metrics for pairwise_distances.

This function simply returns the valid pairwise distance metrics. It exists to allow for a description of the mapping for each of the valid strings.

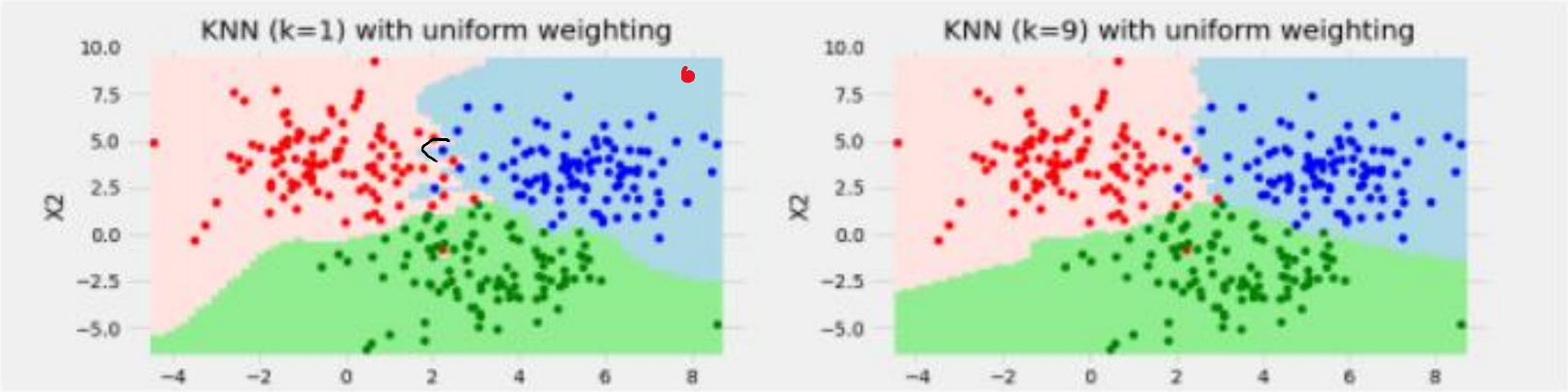
The valid distance metrics, and the function they map to, are:

metric	Function
'cityblock'	metrics.pairwise.manhattan_distances
'cosine'	metrics.pairwise.cosine_distances
'euclidean'	metrics.pairwise.euclidean_distances
'haversine'	metrics.pairwise.haversine_distances
'l1'	metrics.pairwise.manhattan_distances
'l2'	metrics.pairwise.euclidean_distances
'manhattan'	metrics.pairwise.manhattan_distances
'nan_euclidean'	metrics.pairwise.nan_euclidean_distances

K-NN algoritması

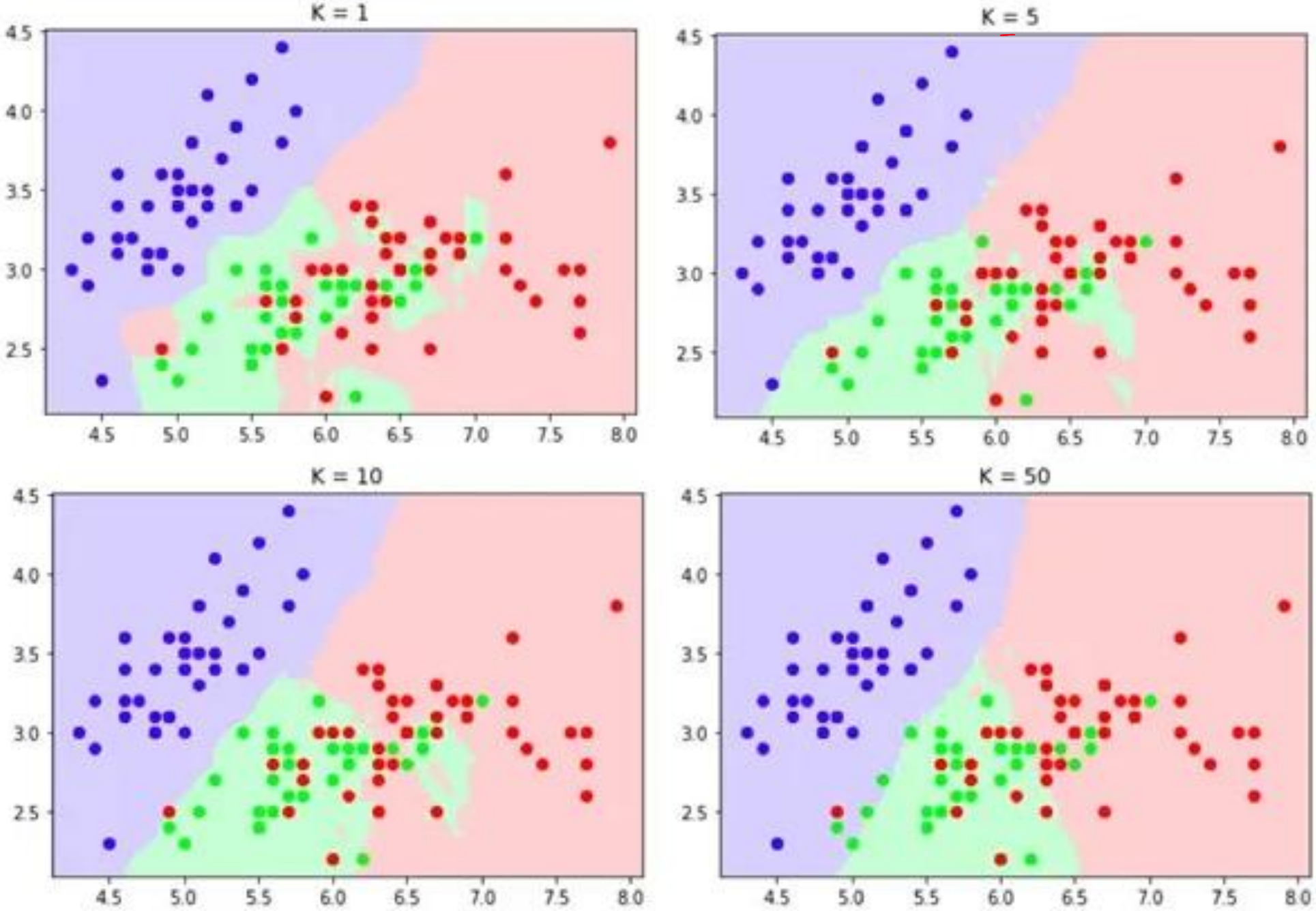
- Avantajları
  - Gerçekleme kolaylığı
  - Herhangi bir ön kabule ihtiyacı yoktur.
  - Eğitim yok
  - Yeni örnekler geldiğinde hızlı adaptasyon sağlar
  - Hem sınıflandırma hem de regresyon için kullanılır.
  - Birkaç parametre (k ve norm)
  - Doğrusal olmayan veriler sınıflandırılabilir
- Dezavantajları
  - Yavaş bir algoritmadır (büyük veri)
  - Homojen öznitelikler olması gerekir
  - Aykırı örneklere takılabilir.
  - K sayısının tespiti
  - RAM ihtiyacı

K-NN algoritması



K-NN (k=1) vs. K-NN (k=9) Classifiers (all images are generated by the author)

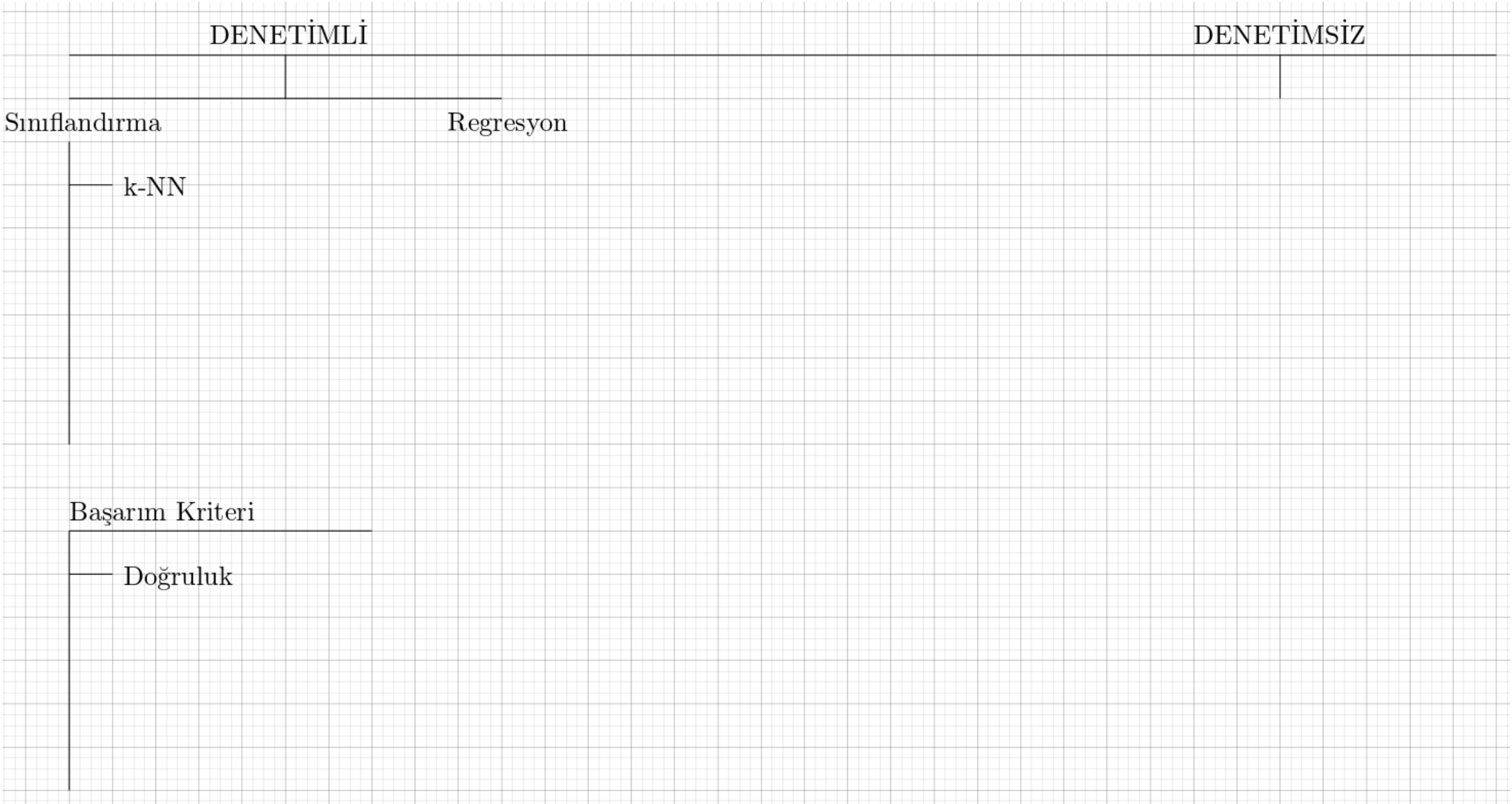
https://towardsdatascience.com/k-nearest-neighbors-k-nn-explained-8959f97a8632



https://kdagiit.medium.com/k-nearest-neighbor-knn-algorithm-9a0eefe1f148

K-NN algoritması

- Doğruluk Hesaplaması


DENETİMLİ
DENETİMSİZ
Sınıflandırma
Regresyon
k-NN
Başarım Kriteri
Doğruluk

K-NN algoritması

- İleri Konular (Büyük Veri)

PAPER • OPEN ACCESS

Analysis of KNN Algorithm with Mapreduce Technique on Big Data

Tatikonda Bhavana[1], J. Padmavathy[1], R. Sethuraman[2] and J.K. Jeevitha[3]

IOP Conference Series: Materials Science and Engineering, Volume 590, International Conference on Frontiers in Materials and Smart System Technologies 10 April 2019, Tamil Nadu, India

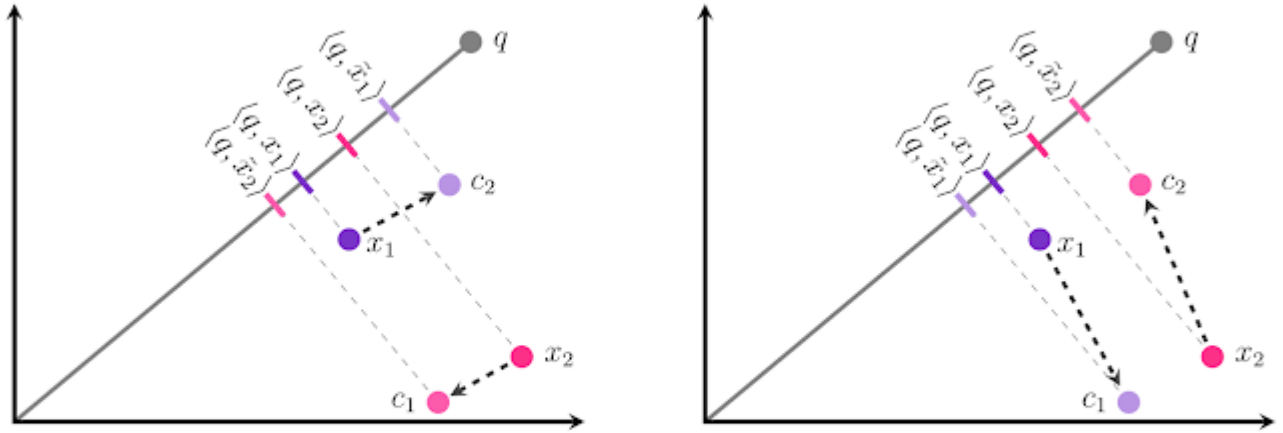
Citation Tatikonda Bhavana *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **590** 012028

DOI 10.1088/1757-899X/590/1/012028

Article PDF

K-NN algoritması

- İleri Konular (Büyük Veri)



The goal is to quantize each x_i to $\tilde{x}_i = c_1$ or $\tilde{x}_i = c_2$. Traditional quantization (left) results in the incorrect ordering of x_1 and x_2 for this query. Even though our approach (right) chooses centers farther away from the data points, this in fact leads to lower inner product error and higher accuracy.

https://blog.research.google/2020/07/announcing-scann-efficient-vector.html

K-NN algoritması – Yarıçap Tabanlı

Bayes Karar Teorisi

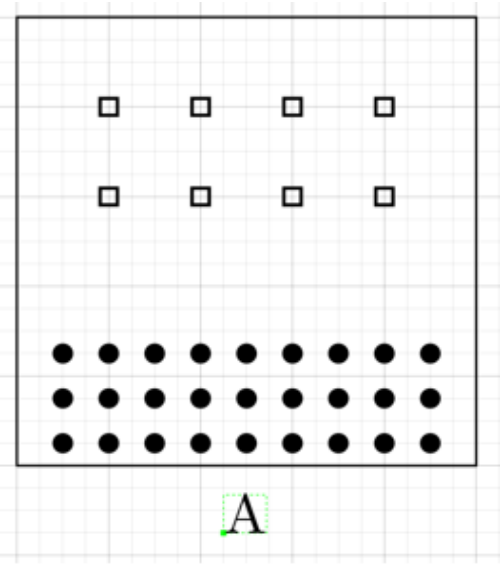
Olasılık 101

- Olasılık nedir?
  - ❑Bir şeyin olmasına ait matematiksel yüzdesi (wiki)
  - ❑Popülasyonu betimleyen sayısal bilgiler

- Yazı tura
- Zar atma
- Okula varma süresi
- 5 günlük hava raporunun

sonunda meteoroloji tahmini

Olasılık 101



Kare ve Daire

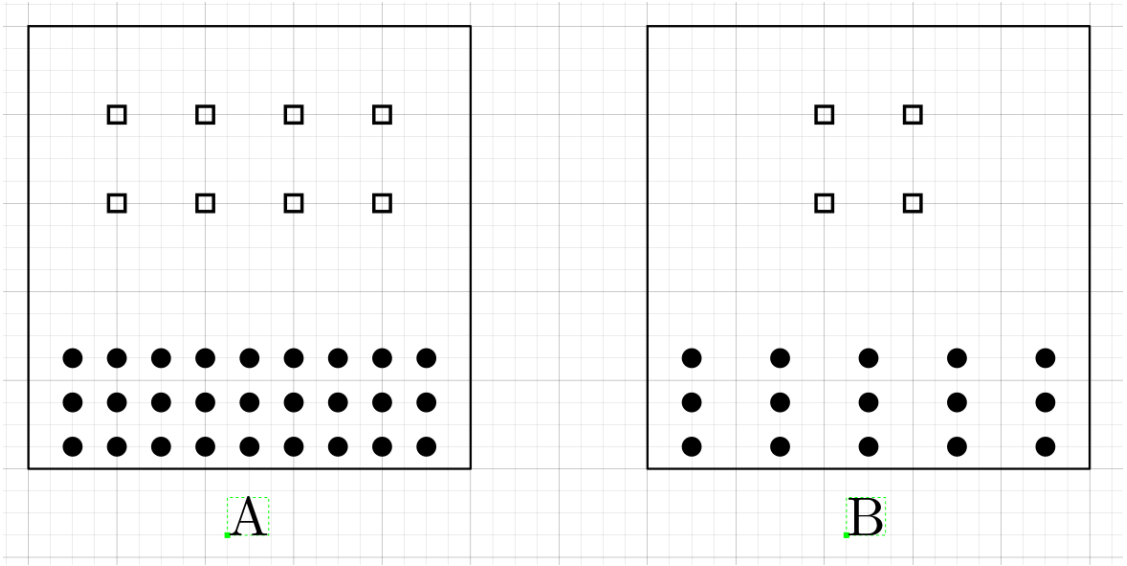
S: 27 D , 8 K

Axioms of Probability:

- Axiom 1: For any event A, $P(A) \geq 0$.
- Axiom 2: Probability of the sample space S is $P(S) = 1$.
- Axiom 3: If $A_1, A_2, A_3, \cdots$ are disjoint events, then $P(A_1 \cup A_2 \cup A_3 \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$

https://www.probabilitycourse.com/chapter1

Olasılık 101 – Bayes Teoremi

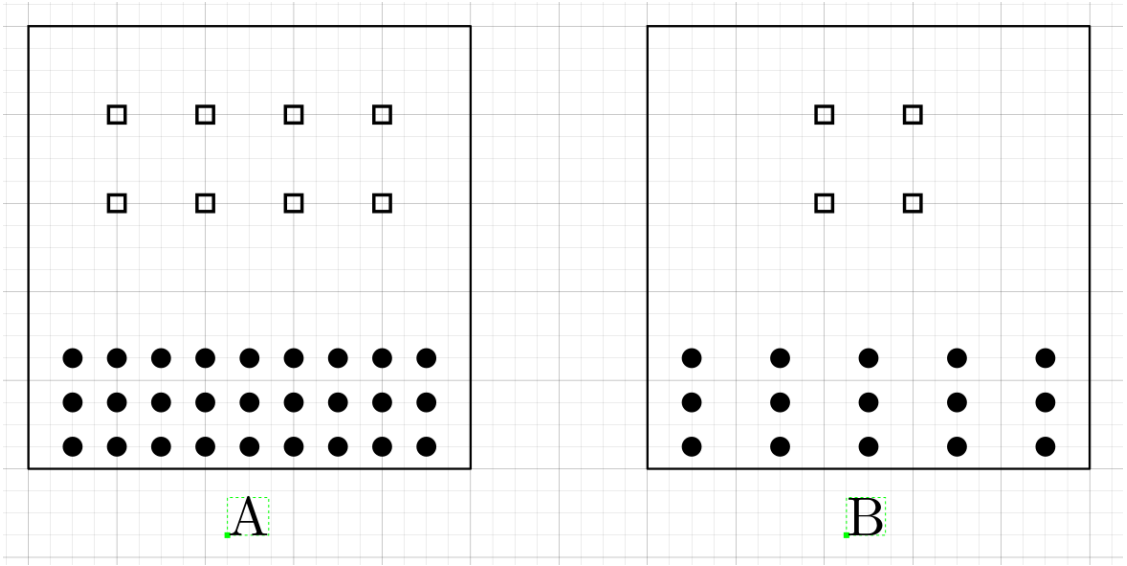


Kare ve Daire

A: 27 D , 8 K

B: 15 D , 4 K

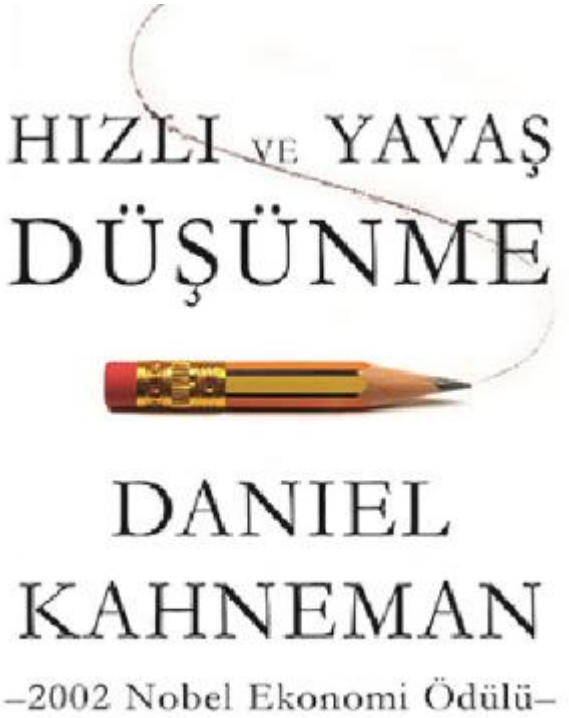
Olasılık 101 – Bayes Teoremi



Kare ve Daire

A: 27 D , 8 K

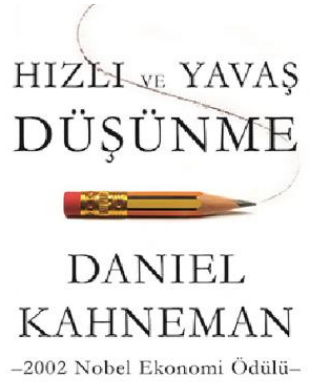
B: 15 D , 4 K



If A and B are two events in a sample space S, then the **conditional probability of** A **given** B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) > 0.$$

Olasılık 101 – Bayes Teoremi



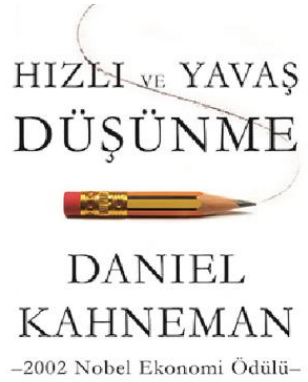
Kim Bu? Çiftçi yada Kütüphaneci

- Kendisi içine kapanık ve duygusal biriydi.
- Sosyal çevresi pek yok, keni ağır işleriyle ilgilenmeyi sever.
- Oldukça entelektüel ve derinlikli fikir sahibi.

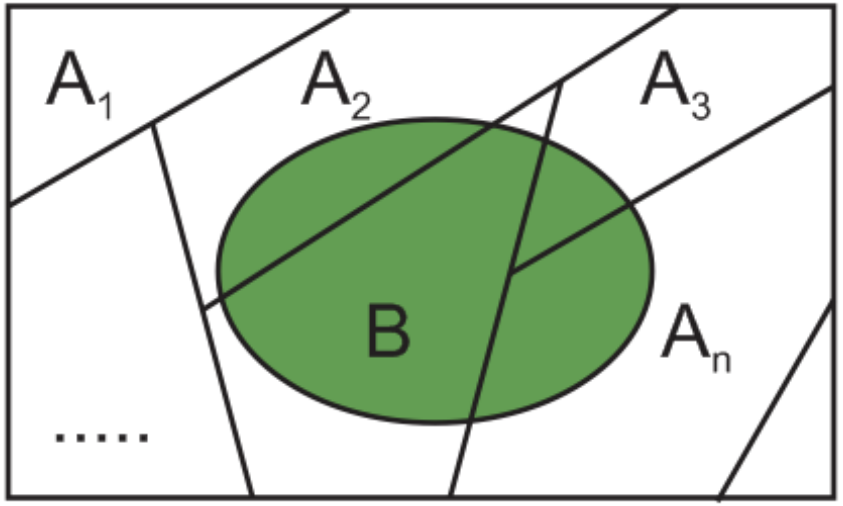
Olasılık 101 – Bayes Teoremi

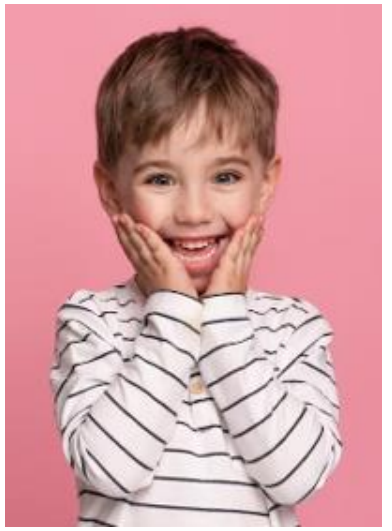
Adobe Stock | #221094422

https://gokmenarslan.com/mutlu-cocuk-denilince-ne-anlamaliyiz/



Olasılık 101 – Bayes Teoremi





Olasılık 101 – Bayes Teoremi

- Bir toplumda kanser vakaları %0.1 olsun (0.001)
- Bir test cihazı hasta (C) iken %98 (+), (C’) iken %95 (-) hassasiyetle hastaları tespit edebiliyor.
- Eğer cihaz bir kişiye (+) demiş ise bu kişinin kanser olma (C) ihtimali nedir?

Olasılık 101 – Bayes Teoremi

- Bir toplumda kanser vakaları %0.1 olsun (0.001)
- Bir test cihazı hasta (C) iken %98 (+), (C’) iken %95 (-) hassasiyetle hastaları tespit edebiliyor.
- Eğer cihaz bir kişiye (+) demiş ise bu kişinin kanser olMAma (C’) ihtimali nedir?

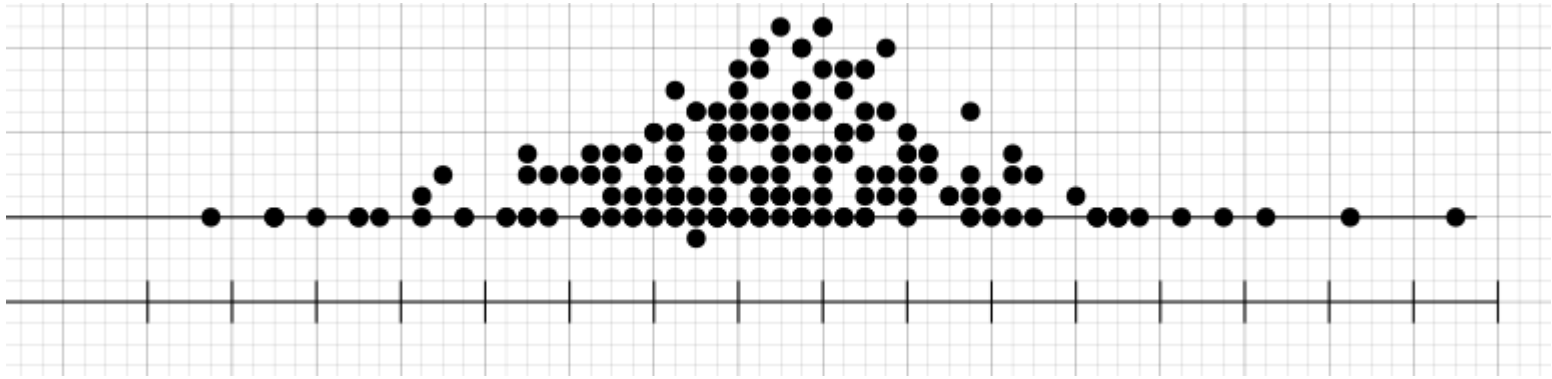
Olasılık 101 – Rassal Değişken ve İstatistikler

Olasılık 101 – Rassal Değişken ve İstatistikler

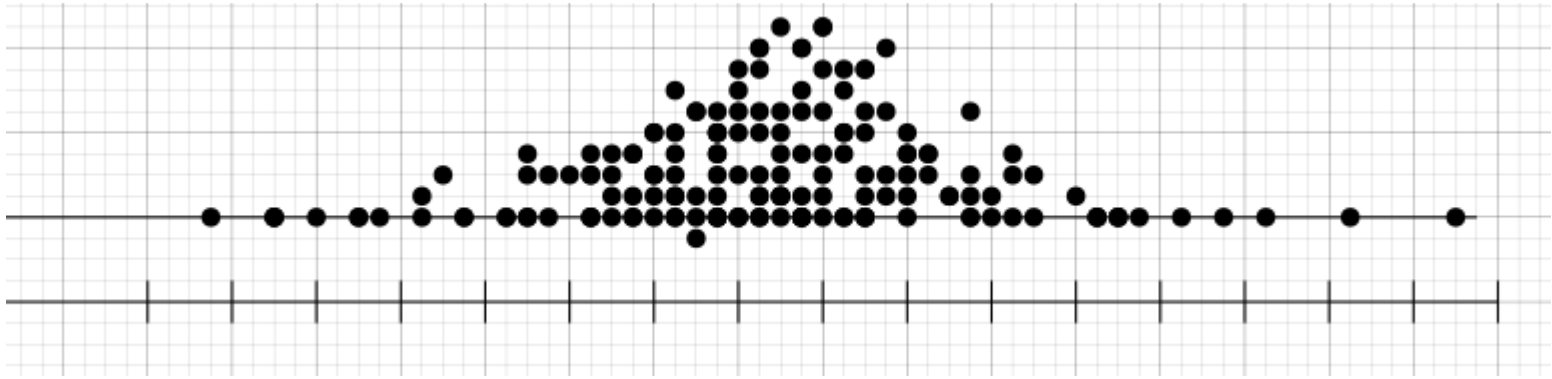
Olasılık 101 – Rassal Değişken ve İstatistikler

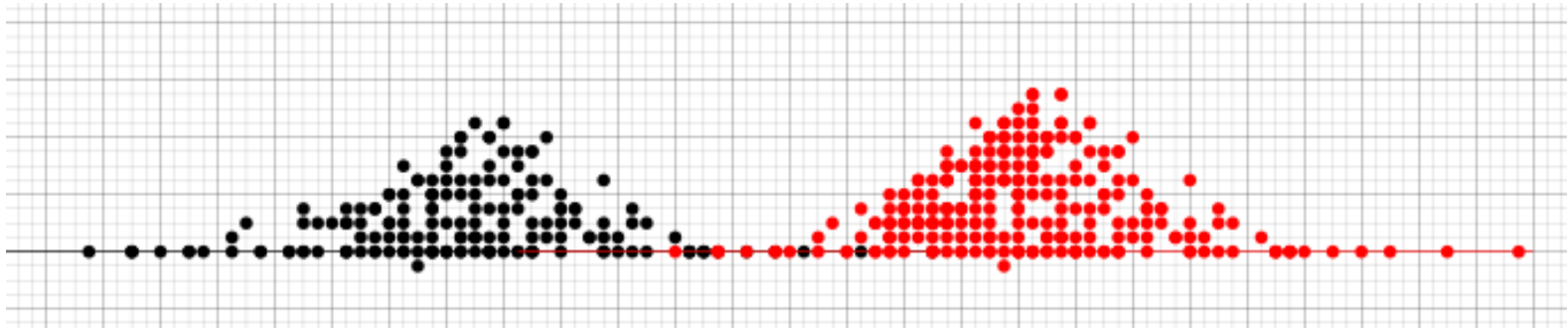
- Elimizde p(x) yoksa

Olasılık 101 – Rassal Değişken ve İstatistikler



Olasılık 101 – Rassal Değişken ve İstatistikler





Olasılık 101 – Rassal Değişken ve İstatistikler

- Denizden palamut çıkma olasılığı P(A) = 0.6
- Hamsi çıkma olasılığı P(B) = 0.4
- P(10 cm | B) = 0.5 , P(10 cm | A) = 0.2
- Eğer 10 cm bir balık geldiyse H mi P mi?

Olasılık 101 – Rassal Değişken ve İstatistikler

$$P(\text{class} \mid \text{data}) \qquad P(\text{data} \mid \text{class}) \times P(\text{class})$$

Olasılık 101 – Rassal Değişken ve İstatistikler

- Merhaba,
- Bugün gerçekleştirilen bir çekilişte tam 3M TL kazandınız.
- TC kimlik ve IBAN adresinizi xxxx adresine gönderdiğiniz takdirde size büyük ödül iletilecektir.

Uskumru (C1)		Palamut (C2)	
Uzunluk (cm)	Adet	Uzunluk (cm)	Adet
05-10	5	15-20	5
10-15	15	20-25	10
15-20	20	25-30	20
20-25	15	30-35	30
25-30	5	35-40	10
30-35	0	40-45	5
35-40	0	45-50	0

1. Sınıf olasılıklarını bul
2. Sınıf içi olasılıkları bul
3. Formülde yerine koy

$$P(C1|u = 22) = ?$$