

# IKT484 Makine Öğrenmesi

## 2. İlk kNN Modeli

Öğretim Görevlisi:

**Bora GÜNGÖREN**

[bora.gungoren@atilim.edu.tr](mailto:bora.gungoren@atilim.edu.tr)

# K-Nearest Neighbor (kNN)

- kNN, sınıflandırma (ve regresyon) için kullanılan çok basit bir makine öğrenme tekniğidir.
  - Başlangıçta bir sınıf kümesinin tanımlanmasını gerektirdiği için denetlenen bir öğrenme tekniğidir.
  - Başlangıçtaki sınıf kümesi, etiketlerin sınıflar olduğu bir dizi etiketli veri noktası olarak sunulur. Bu, gerekli eğitim kümesidir.
  - Daha sonra sunulan her yeni veri noktası eğitim kümesiyle karşılaştırılır ve etiketlenir.
  - Eğitim kümesine yeni veri noktaları ekleyebilir veya eklemeyebilirsiniz. Bu, uygulamınıza bağlıdır.

# K-Nearest Neighbor (kNN)

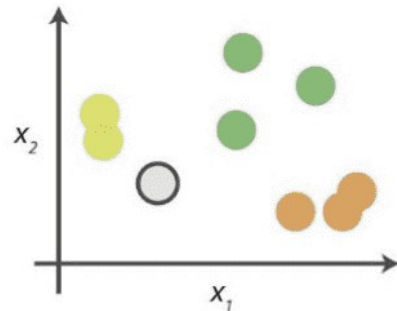
- Yöntem basit.
- Her yeni veri noktası için
  - Eğitim veri setindeki veri noktaları ile yeni veri noktası arasındaki mesafeleri hesaplıyoruz.
    - “Mesafe” kavramını nasıl tanımladığımız çok önemlidir.
  - En yakın komşuları seçiyoruz, böylece mahallede zaten var olan sınıflandırmayı gösteren bir örneğimiz var.

# K-Nearest Neighbor (kNN)

- Yöntem basit.
  - Her yeni veri noktası için
    - Sınıflandırma atamalarını sayıyor ve çoğunluğu buluyoruz ve yeni veri noktasını bu çoğunluğa atıyoruz. ("Mahalledeki çoğunluğa aitsiniz")
      - Sayım sürecine genellikle oy denir.
      - Çoğunluk kuralı uygulamanıza göre değişebilir.
  - Bu atamanın bir tahmin olduğunu unutmayın.
    - Dolayısıyla KNN, sınıflandırma sonucunu tahmin etmek için bir tür regresyon kullanır.

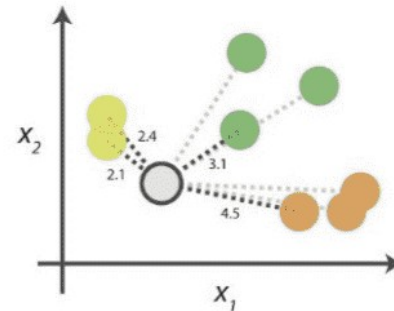
# K-Nearest Neighbor (kNN)

## 0. Look at the data











Say you want to classify the grey point into a class. Here, there are three potential classes - lime green, green and orange.

## 1. Calculate distances









Start by calculating the distances between the grey point and all other points.

## 2. Find neighbours

Point Distance			
		2.1	→ 1st NN
		2.4	→ 2nd NN
		3.1	→ 3rd NN
		4.5	→ 4th NN

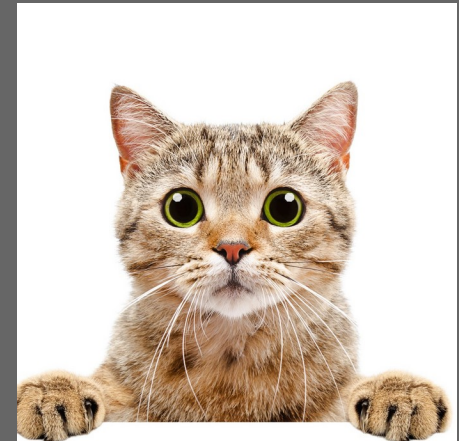
Next, find the nearest neighbours by ranking points by increasing distance. The nearest neighbours (NNs) of the grey point are the ones closest in dataspace.

## 3. Vote on labels

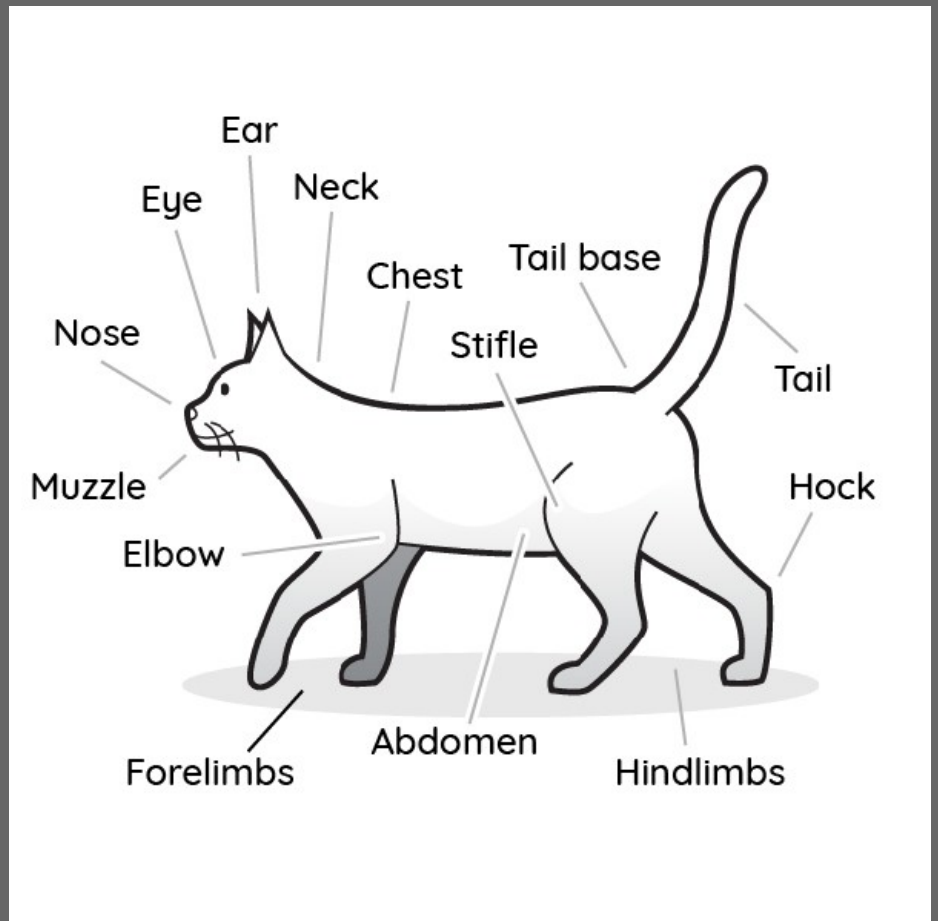
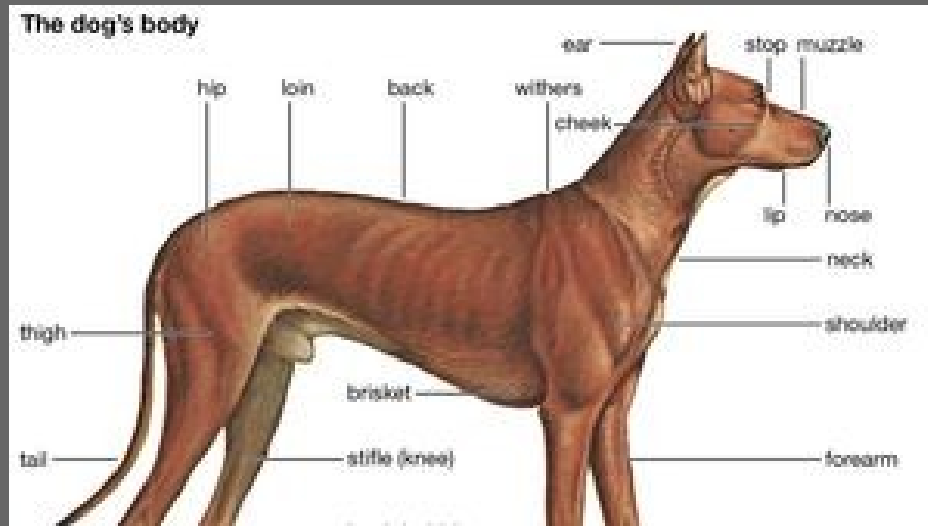
Class	# of votes	
	2	→ Class  wins the vote! Point  is therefore predicted to be of class  .
	1	
	1	

Vote on the predicted class labels based on the classes of the k nearest neighbours. Here, the labels were predicted based on the k=3 nearest neighbours.

# K-Nearest Neighbor (kNN)



# K-Nearest Neighbor (kNN)





# K-Nearest Neighbor (kNN)

- Dikkatli seçilmiş özellikler ve iyi tasarlanmış bir mesafe metriği olmadan KNN'nin saf bir uygulaması genellikle büyük bir yanlış sınıflandırma oranına sahiptir.
- Yanlış sınıflandırma oranı, olumlu ve olumsuz tahminler arasında ayırım yapmadan yanlış tahminlerin bir kısmını söyleyen bir performans metriğidir.
- Öte yandan, veri kümesi dengesiz olduğunda (prevalans çok yüksek veya çok düşük olduğunda) yanlış sınıflandırma oranı çok yanıltıcı bir metrik olabilir.



# K-Nearest Neighbor (kNN)

- Karışıklık Matrisini (Confusion Matrix) hatırlayalım.
  - Doğruluk: Genel olarak, sınıflandırıcı ne sıklıkla doğrudur?
  - Yanlış sınıflandırma oranı: Genel olarak, ne sıklıkla yanlış?
  - Prevalans: EVET koşulu gerçekten örneğimizde ne sıklıkla meydana gelir?
- Çok dengesiz bir eğitim veri seti ile başladığımızda ne olur?
- Eğitim veri setinin gerçeği yeterince iyi temsil ettiğinden nasıl emin olabiliriz?

# K-Nearest Neighbor (kNN)

- KNN öğretimde çok popülerdir, çünkü denetimli öğrenmenin hem güçlü hem de zayıf yönlerini kolayca gösterir.
  - Bu nedenle hemen hemen tüm yöntemler performans metriklerini KNN ile karşılaştırır.
  - Bazı yöntemler KNN'den daha kötü performans gösterirse, uygun değildir.

# K-Nearest Neighbor (kNN)

- Hız ve bellek nasıl?
  - En yakın komşuları tanımlamak için kolay bir mesafe metriği ve hızlı bir arama yöntemi göz önüne alındığında, KNN nispeten hızlı çalışır.
  - Bellek kullanımını optimize etmek için gelişmiş veri yapılarınız yoksa KNN çok fazla bellek kullanma eğilimindedir.
  - Bir tür veri seti için, bu ikisinin gelmesi zordur, bu nedenle KNN çok kötü bir seçimdir.
- Sadece bir PC kullanarak veri kümeleri ile oynadığınızda, bunu yapmanız gerekmedikçe büyük veri kümelerinde KNN kullanmayın.

# K-Nearest Neighbor (kNN)

- Önden Öğrenen ve Geç Öğrenen
  - Erken Öğrenen algoritmalar öğrenmek için daha fazla zaman harcar ve daha az zaman harcarlar.
  - Geç Öğrenen Algoritmalar öğrenmek ve tahmin etmek için daha az zaman harcarlar.
- KNN hangisidir? Neden?

# K-Nearest Neighbor (kNN)

- Boyutsallık laneti
  - KNN, çok sayıda özellikten daha az sayıda özellik ile daha iyi performans gösterir.
  - Aşırı öğrenmeyi önlemek için, boyut sayısını artırdıkça gerekli verilerin katlanarak büyümesi gerekecektir.
- Özellik seçimi KNN için önemlidir.

# K-Nearest Neighbor (kNN)

- Bu algoritma için  $k$  değerini nasıl seçeriz?
  - Bu zor bir seçim.
  - Genellikle  $k = 2$ 'den bir üst sınırdan deneriz ve farklı  $k$  değerleri için performans metriklerini karşılaştırırız.
- Neden  $k = 1$  olmasın? Bu herhangi bir pratik uygulamada mantıklı olur mu?

# K-Nearest Neighbor (kNN)

- KNN için verileri nasıl işleriz?
  - Örneğimiz, birisinin çikolata muz milkshake'i sevip sevmediğini tahmin etmekle ilgilidir.
- İşte eğitim veri seti.
  - Bazı değişkenler kategoriktir.
  - KNN, sayısal verilerle mesafe ve ortak mesafe metrikleri çalışmasını gerektirir.
    - Şanslıyız çünkü çoğu kategori ikili.
    - Likert benzeri ölçeklerle de çalışabiliriz.

No	Age	Sex	Occupation	Likes Bananas	Likes Chocolate	Lactose Intolerant	Likes Chocolate Banana Milkshake
1	35	Male	Engineer	Yes	Yes	No	Yes
2	36	Male	Teacher	Yes	No	No	No
3	35	Male	Engineer	No	No	No	No
4	39	Male	Medical Doctor	No	Yes	Yes	No
5	37	Male	Truck Driver	Yes	Yes	Yes	Yes
6	40	Male	Truck Driver	Yes	Yes	No	Yes
7	24	Male	Teacher	Yes	No	No	Yes
8	58	Male	Chef	Yes	Yes	No	Yes
9	18	Male	Cashier	No	Yes	No	Yes
10	19	Male	Student	Yes	Yes	No	Yes
11	50	Male	Engineer	Yes	Yes	No	Yes
12	43	Male	Teacher	Yes	No	Yes	No
13	59	Male	Retired	Yes	No	No	No
14	44	Male	Waiter	Yes	Yes	Yes	Yes
15	18	Male	Student	Yes	Yes	Yes	No
16	58	Male	Lawyer	Yes	No	Yes	No
17	33	Female	Medical Doctor	Yes	Yes	No	Yes
18	65	Female	Retired	No	Yes	No	Yes
19	24	Female	Student	No	Yes	Yes	Yes
20	61	Female	Retired	Yes	Yes	No	Yes
21	23	Female	Student	Yes	Yes	Yes	Yes
22	59	Female	Medical Doctor	No	No	Yes	No
23	56	Female	Teacher	Yes	Yes	Yes	No
24	26	Female	Student	No	Yes	No	Yes
25	34	Female	Teacher	No	Yes	No	Yes
26	48	Female	Teacher	Yes	No	No	Yes
27	49	Female	Medical Doctor	No	Yes	No	Yes
28	18	Female	Student	Yes	Yes	Yes	Yes
29	44	Female	Engineer	Yes	Yes	No	Yes
30	32	Female	Medical Doctor	Yes	Yes	Yes	No
31	20	Female	Student	Yes	Yes	Yes	Yes
32	50	Female	Nurse	No	Yes	Yes	No
33	38	Female	Lawyer	Yes	Yes	No	Yes
34	58	Female	Retired	No	Yes	Yes	No



# K-Nearest Neighbor (kNN)

No	Age	Sex	Occupation	Likes Bananas	Likes Chocolate	Lactose Intolerant	Likes Chocolate Banana Milkshake	Binary Sex	Binary Banana	Binary Chocolate	Binary Lactose	Binary Milkshake
1	35	Male	Engineer	Yes	Yes	No	Yes	0	1	1	0	1
2	36	Male	Teacher	Yes	No	No	No	0	1	0	0	0
3	35	Male	Engineer	No	No	No	No	0	0	0	0	0
4	39	Male	Medical Doctor	No	Yes	Yes	No	0	0	1	1	0
5	37	Male	Truck Driver	Yes	Yes	Yes	Yes	0	1	1	1	1
6	40	Male	Truck Driver	Yes	Yes	No	Yes	0	1	1	0	1
7	24	Male	Teacher	Yes	No	No	Yes	0	1	0	0	1
8	58	Male	Chef	Yes	Yes	No	Yes	0	1	1	0	1
9	18	Male	Cashier	No	Yes	No	Yes	0	0	1	0	1
10	19	Male	Student	Yes	Yes	No	Yes	0	1	1	0	1
11	50	Male	Engineer	Yes	Yes	No	Yes	0	1	1	0	1
12	43	Male	Teacher	Yes	No	Yes	No	0	1	0	1	0
13	59	Male	Retired	Yes	No	No	No	0	1	0	0	0
14	44	Male	Waiter	Yes	Yes	Yes	Yes	0	1	1	1	1
15	18	Male	Student	Yes	Yes	Yes	No	0	1	1	1	0
16	58	Male	Lawyer	Yes	No	Yes	No	0	1	0	1	0
17	33	Female	Medical Doctor	Yes	Yes	No	Yes	1	1	1	0	1
18	65	Female	Retired	No	Yes	No	Yes	1	0	1	0	1
19	24	Female	Student	No	Yes	Yes	Yes	1	0	1	1	1
20	61	Female	Retired	Yes	Yes	No	Yes	1	1	1	0	1
21	23	Female	Student	Yes	Yes	Yes	Yes	1	1	1	1	1
22	59	Female	Medical Doctor	No	No	Yes	No	1	0	0	1	0
23	56	Female	Teacher	Yes	Yes	Yes	No	1	1	1	1	0
24	26	Female	Student	No	Yes	No	Yes	1	0	1	0	1
25	34	Female	Teacher	No	Yes	No	Yes	1	0	1	0	1
26	48	Female	Teacher	Yes	No	No	Yes	1	1	0	0	1
27	49	Female	Medical Doctor	No	Yes	No	Yes	1	0	1	0	1
28	18	Female	Student	Yes	Yes	Yes	Yes	1	1	1	1	1
29	44	Female	Engineer	Yes	Yes	No	Yes	1	1	1	0	1
30	32	Female	Medical Doctor	Yes	Yes	Yes	No	1	1	1	1	0
31	20	Female	Student	Yes	Yes	Yes	Yes	1	1	1	1	1
32	50	Female	Nurse	No	Yes	Yes	No	1	0	1	1	0
33	38	Female	Lawyer	Yes	Yes	No	Yes	1	1	1	0	1
34	58	Female	Retired	No	Yes	Yes	No	1	0	1	1	0

# K-Nearest Neighbor (kNN)

- Böylece aşağıdaki değişkenleri uzaktan metrikte girdi olarak kullanabiliriz:
  - Yaş (sayısal)
  - Cinsiyet (ikili)
  - Muz sevmek (ikili)
  - Çikolata sevmek (ikili)
  - Laktoz intolerant (ikili)

# K-Nearest Neighbor (kNN)

- Komşuları seçmek için mesafeleri kullanabiliriz.
  - Manhattan mesafesi
  - Öklid mesafesi.
- “Komşularımız” mesafede kullanacağımız ölçütlere göre “bize en yakın olanlar”
  - Dolayısı ile buradaki ölçütler nedeni ile “bize en yakın” olanlar nasıl hesaplanacak? (Buna geleceğiz)

# K-Nearest Neighbor (kNN)

- Örneklem temsil gücü?
  - 4 ikili değişken ve bir adet sayısal değişken.
  - $2^4 = 16$  varyasyon (ikili değişkenler)
  - 34 örnekten gelen veri.
  - Çok iyi bir durum değil.
- Belki daha az değişken ile denemek gerekir.
  - Daha belirgin bir etkisi olan değişkenler bulunabilir.

# K-Nearest Neighbor (kNN)

- Sadece eğitim verilerimizi anlamak için,
  - İkili olanlar da dahil olmak üzere bu değişkenler için basit korelasyon katsayılarını hesaplayacağız.
- Normalde bunu ikili veriler için yapmamalıyız.

	Age	Binary Sex	Binary Banana	Binary Chocolate	Binary Lactose	Binary Milkshake
Age	1,000	0,094	-0,088	-0,214	-0,071	-0,286
Binary Sex		1,000	-0,274	0,311	0,126	0,167
Binary Banana			1,000	-0,087	-0,019	0,147
Binary Chocolate				1,000	0,074	<b>0,461</b>
Binary Lactose					1,000	<b>-0,459</b>
Binary Milkshake						1,000

# K-Nearest Neighbor (kNN)

- Sonuç?
  - Kötü bir eğitim veri seti değil.
    - İki özellik (çikolata sevmek ve laktoz intoleransı) diğerlerinden çok daha güçlü etkiye sahip.
    - İki özelliğe dayalı sınıflandırıcı ile hızlı biçimde karar alabilir miyiz?
  - Acaip iyi de değil.

	Age	Binary Sex	Binary Banana	Binary Chocolate	Binary Lactose	Binary Milkshake
Age	1,000	0,094	-0,088	-0,214	-0,071	-0,286
Binary Sex		1,000	-0,274	0,311	0,126	0,167
Binary Banana			1,000	-0,087	-0,019	0,147
Binary Chocolate				1,000	0,074	<b>0,461</b>
Binary Lactose					1,000	<b>-0,459</b>
Binary Milkshake						1,000

# K-Nearest Neighbor (kNN)

- Tek kişilik deneme örnekleme:
  - Yaş: 42, erkek, muz seviyor, çikolatayı seviyor, laktoz intolerör değil.
- Tahminde Adımlar:
  - 1. Adım: mesafeleri hesapla.
  - 2. Adım: En yakın k komşuyu ara ( $K = 2,3,4,5$ )
  - 3. Adım: Oylama yoluyla tahmin gerçekleştir.



# K-Nearest Neighbor (kNN)

- Ayrıca ölçeklendirme ile ilgili bir sorunumuz var.
- “Doğrusal ölçeklendirme” sonrası yaş aralıkları.

	Age	Binary Sex	Binary Banana	Binary Chocolate	Binary Lactose	
Test Subject	42	0	1	1	0	
No	Age Distance	Sex Distance	Banana Distance	Chocolage Distance	Lactose Distance	Total Distance
1	0.149	0.000	0.000	0.000	0.000	0.149
2	0.128	0.000	0.000	1.000	0.000	1.128
3	0.149	0.000	1.000	1.000	0.000	2.149
4	0.064	0.000	1.000	0.000	1.000	2.064
5	0.106	0.000	0.000	0.000	1.000	1.106
6	0.043	0.000	0.000	0.000	0.000	0.043
7	0.383	0.000	0.000	1.000	0.000	1.383
8	0.340	0.000	0.000	0.000	0.000	0.340
9	0.511	0.000	1.000	0.000	0.000	1.511
10	0.489	0.000	0.000	0.000	0.000	0.489
11	0.170	0.000	0.000	0.000	0.000	0.170
12	0.021	0.000	0.000	1.000	1.000	2.021
13	0.362	0.000	0.000	1.000	0.000	1.362
14	0.043	0.000	0.000	0.000	1.000	1.043
15	0.511	0.000	0.000	0.000	1.000	1.511
16	0.340	0.000	0.000	1.000	1.000	2.340
17	0.191	1.000	0.000	0.000	0.000	1.191
18	0.489	1.000	1.000	0.000	0.000	2.489
19	0.383	1.000	1.000	0.000	1.000	3.383
20	0.404	1.000	0.000	0.000	0.000	1.404
21	0.404	1.000	0.000	0.000	1.000	2.404
22	0.362	1.000	1.000	1.000	1.000	4.362
23	0.298	1.000	0.000	0.000	1.000	2.298
24	0.340	1.000	1.000	0.000	0.000	2.340
25	0.170	1.000	1.000	0.000	0.000	2.170
26	0.128	1.000	0.000	1.000	0.000	2.128
27	0.149	1.000	1.000	0.000	0.000	2.149
28	0.511	1.000	0.000	0.000	1.000	2.511
29	0.043	1.000	0.000	0.000	0.000	1.043
30	0.213	1.000	0.000	0.000	1.000	2.213
31	0.468	1.000	0.000	0.000	1.000	2.468
32	0.170	1.000	1.000	0.000	1.000	3.170
33	0.085	1.000	0.000	0.000	0.000	1.085
34	0.340	1.000	1.000	0.000	1.000	3.340

# K-Nearest Neighbor (kNN)

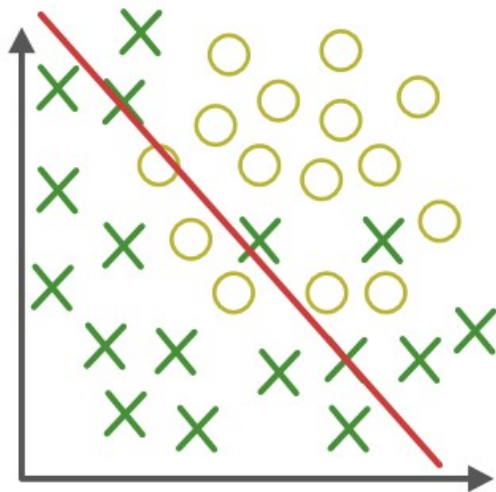
No	Age Distance	Sex Distance	Banana Distance	Chocolage Distance	Lactose Distance	Total Distance	Binary Milkshake
6	0,043	0,000	0,000	0,000	0,000	0,043	1
1	0,149	0,000	0,000	0,000	0,000	0,149	1
11	0,170	0,000	0,000	0,000	0,000	0,170	0
8	0,340	0,000	0,000	0,000	0,000	0,340	1
10	0,489	0,000	0,000	0,000	0,000	0,489	1
14	0,043	0,000	0,000	0,000	1,000	1,043	0
29	0,043	1,000	0,000	0,000	0,000	1,043	0
33	0,085	1,000	0,000	0,000	0,000	1,085	0
5	0,106	0,000	0,000	0,000	1,000	1,106	1
2	0,128	0,000	0,000	1,000	0,000	1,128	1
17	0,191	1,000	0,000	0,000	0,000	1,191	1
13	0,362	0,000	0,000	1,000	0,000	1,362	0
7	0,383	0,000	0,000	1,000	0,000	1,383	1
20	0,404	1,000	0,000	0,000	0,000	1,404	0
9	0,511	0,000	1,000	0,000	0,000	1,511	0
15	0,511	0,000	0,000	0,000	1,000	1,511	0
12	0,021	0,000	0,000	1,000	1,000	2,021	1
4	0,064	0,000	1,000	0,000	1,000	2,064	1
26	0,128	1,000	0,000	1,000	0,000	2,128	1
3	0,149	0,000	1,000	1,000	0,000	2,149	1
27	0,149	1,000	1,000	0,000	0,000	2,149	0
25	0,170	1,000	1,000	0,000	0,000	2,170	1
30	0,213	1,000	0,000	0,000	1,000	2,213	1
23	0,298	1,000	0,000	0,000	1,000	2,298	1
16	0,340	0,000	0,000	1,000	1,000	2,340	0
24	0,340	1,000	1,000	0,000	0,000	2,340	1
21	0,404	1,000	0,000	0,000	1,000	2,404	0
31	0,468	1,000	0,000	0,000	1,000	2,468	1
18	0,489	1,000	1,000	0,000	0,000	2,489	1
28	0,511	1,000	0,000	0,000	1,000	2,511	1
32	0,170	1,000	1,000	0,000	1,000	3,170	1
34	0,340	1,000	1,000	0,000	1,000	3,340	1
19	0,383	1,000	1,000	0,000	1,000	3,383	1
22	0,362	1,000	1,000	1,000	1,000	4,362	0

# K-Nearest Neighbor (kNN)

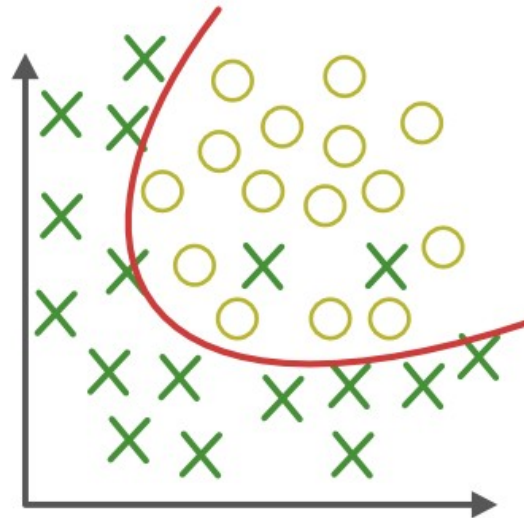
k	Çoğunluk	Sonuç
2	1 (%100)	Seviyor
3	1 (%67)	Seviyor
4	1 (%75)	Seviyor
5	1 (%80)	Seviyor
6	1 (%67)	Seviyor

- Gerçekten güzel bir sonuç, ama tüm örneklem için kesin bir gösterge midir?
- Sonuçlarımızı nasıl değerlendirmeliyiz?

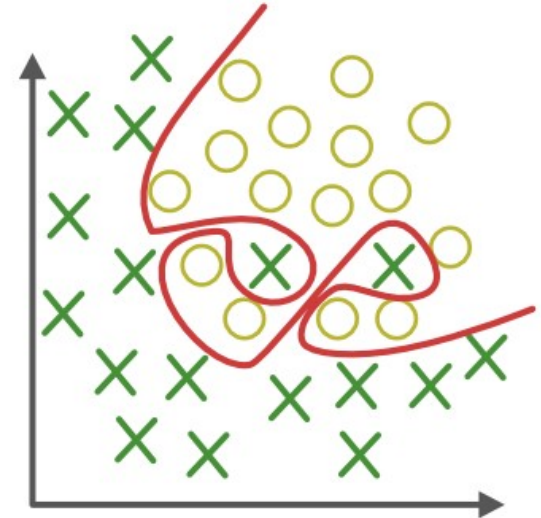
# K-Nearest Neighbor (kNN)




**Under-fitting**  
(too simple to  
explain the variance)



**Appropriate-fitting**

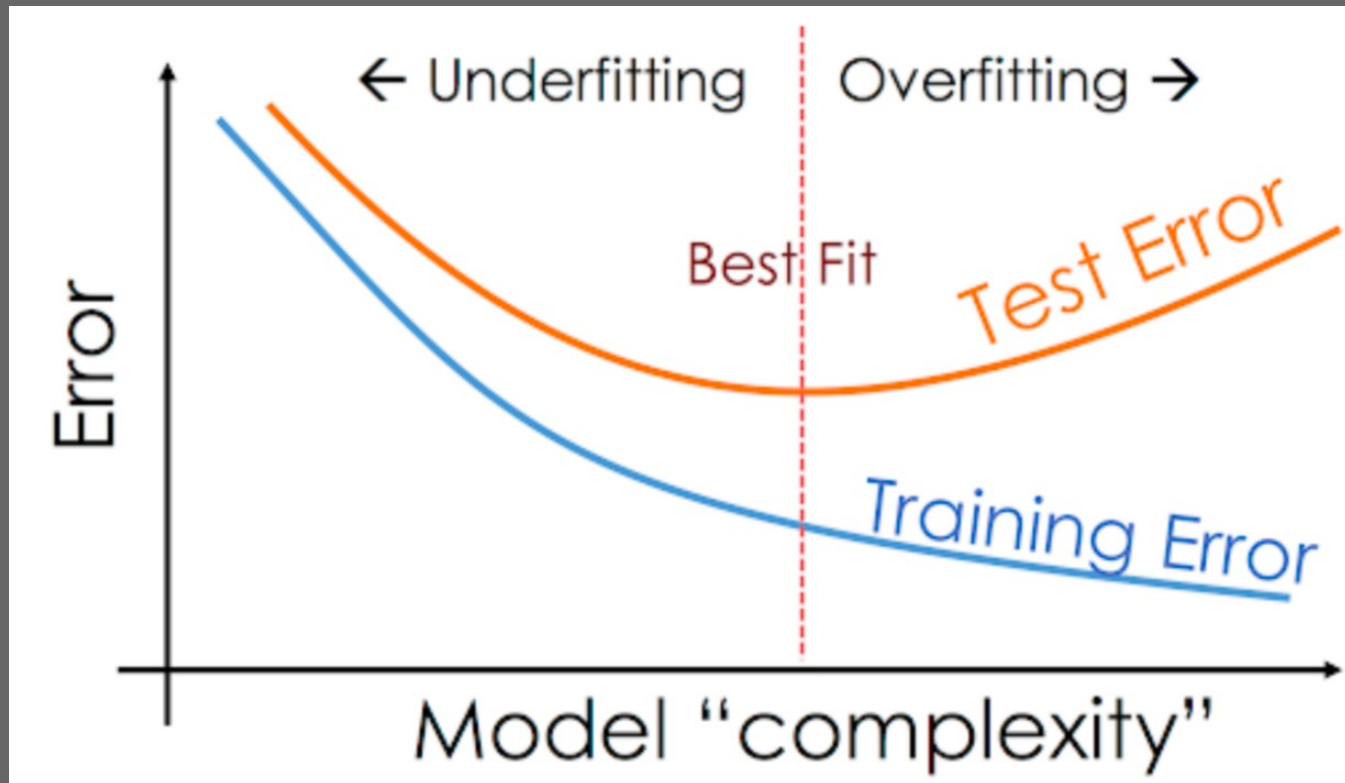


**Over-fitting**  
(forcefitting--too  
good to be true) 

# K-Nearest Neighbor (kNN)

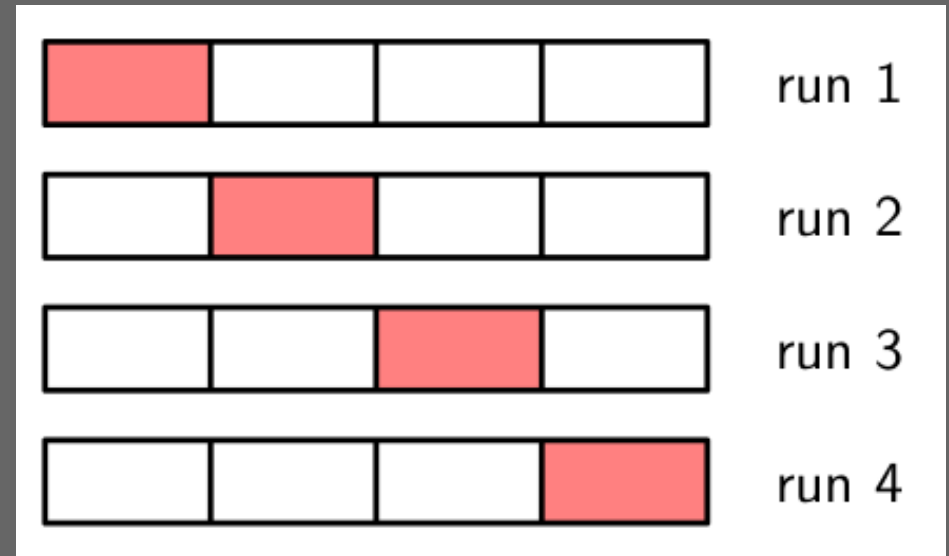
- Aşırı takma, istatistiksel bir model tam olarak eğitim verilerine uyduğunda gerçekleşir.
  - Model gürültüyü ezberlediğinde ve eğitim setine çok yakından uyduğunda, model “aşırı şişirme” olur ve yeni verilere iyi genelleme yapamaz.
  - Bir model yeni veriler için iyi genelleme yapamazsa, amaçlandığı sınıflandırma veya tahmin görevlerini yerine getiremez.
- Düşük hata oranları ve yüksek varyans, aşırı uymanın iyi göstergeleridir.
- Eğitim verilerinin düşük bir hata oranı varsa ve test verilerinin yüksek bir hata oranı varsa, aşırı uymaya işaret eder.

# K-Nearest Neighbor (kNN)



# K-Nearest Neighbor (kNN)

- Modellerimiz gerçek tahminler yapmak için kullanılmadan önce kaçınabilmemiz için aşırı öğrenmeyi tespit edebilmeliyiz (yani sınıflandırmalar).
- S-kat-çapraz validasyonu hatırlayın.
- Her çalışma için performans metriklerini hesaplayın.
- Bu metriklerin varyansını gözlemleyin.





# K-Nearest Neighbor (kNN)

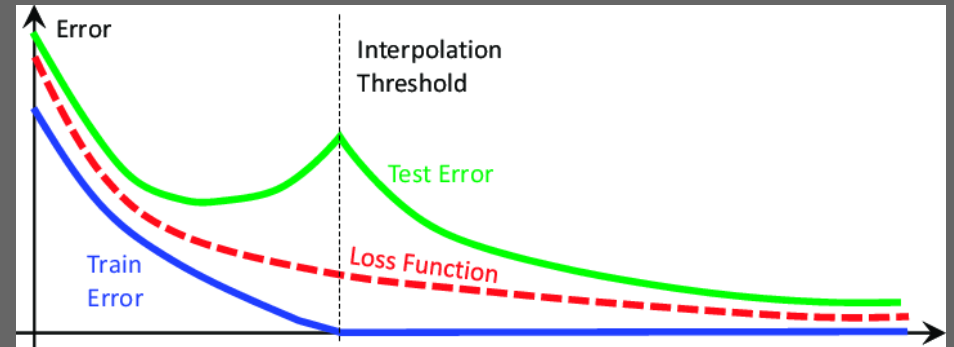
- Aşırı öğrenmeden kaçınabilir miyiz? Kesin çözüm yoktur, ancak bazı teknikler vardır.
  - Erken durdurma, model modeldeki gürültüyü öğrenmeye başlamadan önce eğitimi durdurur. Bu yaklaşım, eğitim sürecini çok erken durdurma riskiyle karşı karşıya kalır ve bunun tersine eksik öğrenme sorununa yol açabilir.
  - Veri büyütme, daha kararlı bir model yaratmak için gürültü içeren veri ekler. Teorik olarak eklenen gürültü, mevcut gürültüye aşırı uyum sağlamayı zorlaştıracaktır. Ancak, bu yaklaşımı yönetmenin güvenilir bir yolu yoktur.

# K-Nearest Neighbor (kNN)

- Aşırı öğrenmeden kaçınabilir miyiz? Kesin çözüm yoktur, ancak bazı teknikler vardır.
  - Özellik seçimi, eğitim verileri içindeki en önemli olanları tanımlama ve daha sonra alakasız veya gereksiz olanları ortadan kaldırma sürecidir. Bu genellikle boyutsallık azaltma ile karıştırılır, ancak farklıdır.
  - Düzenli hale getirme (regularization), daha sonra modeldeki varyans miktarını sınırlayan daha büyük katsayılara sahip giriş parametrelerine bir ceza uygular. Bu yöntemler verilerdeki gürültüyü tanımlamayı ve azaltmayı amaçlamaktadır. Ayrıca hiperparametre optimizasyon yöntemleri altında sınıflandırılırlar.

# K-Nearest Neighbor (kNN)

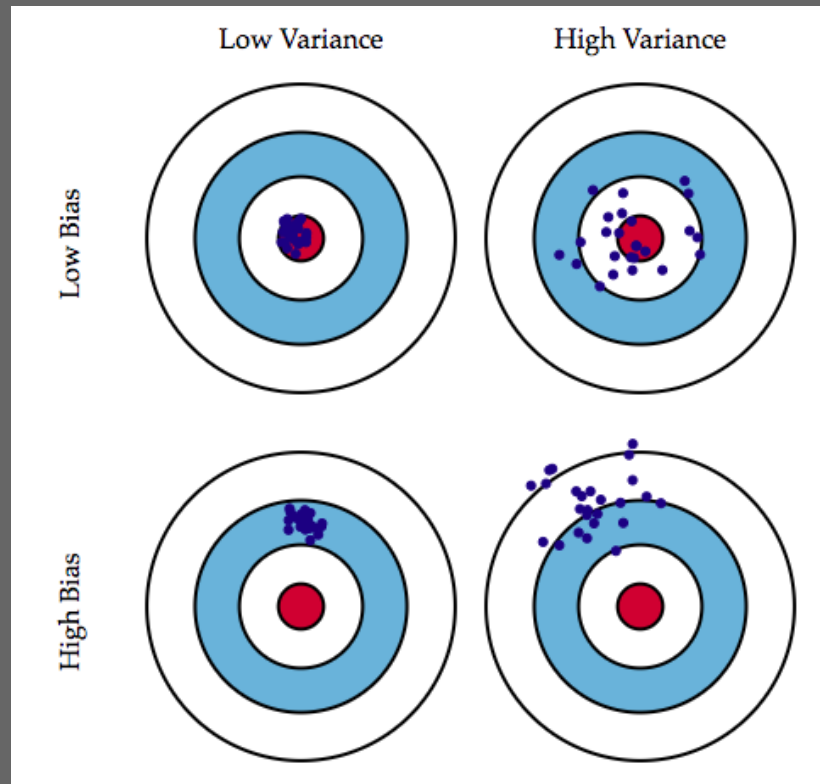
- Aşırı öğrenmeden kaçınabilir miyiz?
- Hiper parametre optimizasyonları çok başarılı bir şekilde kullanabiliyorsanız, sunulan gibi çift yönlü başarılı bir eğri ile sonuçlanabilirsiniz.
- Çoğu derin öğrenme modelleri bunu vaat ediyor. Elbette sundukları şey değişir.



# K-Nearest Neighbor (kNN)

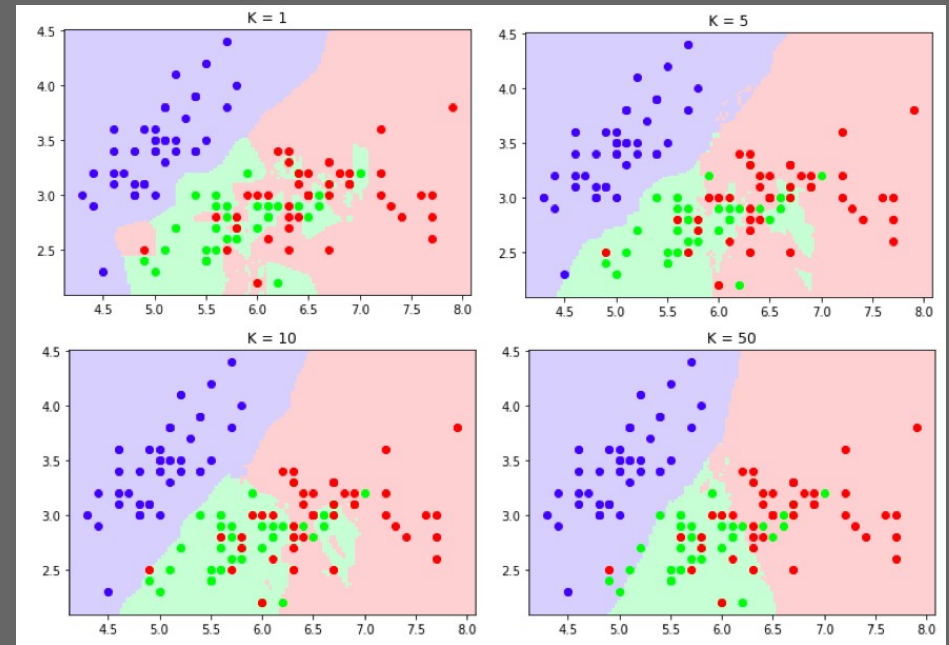
- Önyargı-varyans değiş tokuşu (dengesi)
  - Önyargı, öğrenme algoritmasındaki yanlış varsayımlardan kaynaklanan bir tür hatadır. Yüksek önyargı, bir algoritmanın seçilen girdiler ve hedef çıktılar arasındaki ilgili ilişkileri kaçırmasına neden olabilir.
  - Varyans, eğitim setindeki küçük dalgalanmalara duyarlılıktan kaynaklanan bir hatadır. Yüksek varyans, bir algoritmanın gerçek sonuçlar yerine eğitim verilerindeki rastgele gürültüyü modellemesine neden olabilir.

# K-Nearest Neighbor (kNN)



# K-Nearest Neighbor (kNN)

- KNN ile ilgili olarak, girişe bakılmaksızın sonucu K'ye göre tanımlayan bir “karar sınırı” vardır.
  - Ne olursa olsun neden?
  - Bu "iyi bir K?"
- Çok güzel bir açıklama ile karar sınırını gösteren örnek kod:
  - <https://medium.com/30-Days-of-machine-learning/day-3-k-nearest-neighbors-and-variance-tadeoff-75f84d515bdb>



# Sorular?

CONTACT:

[bora.gungoren@atilim.edu.tr](mailto:bora.gungoren@atilim.edu.tr)

License: Creative Commons Attribution Non-Commercial Share Alike 4.0 International (CC BY-NC-SA 4.0)