



# *Yapay Zeka Performans Değerlendirme Analizi*

# Değerlendirme

- Kesinlik
- Hassasiyet ve geri çağırma
- Kare hatası
- Olasılık
- Arka olasılık
- Maliyet / Fayda
- Marj
- Entropi
- K-L sapma
- Vesaire.

# Yetenek geliştirme

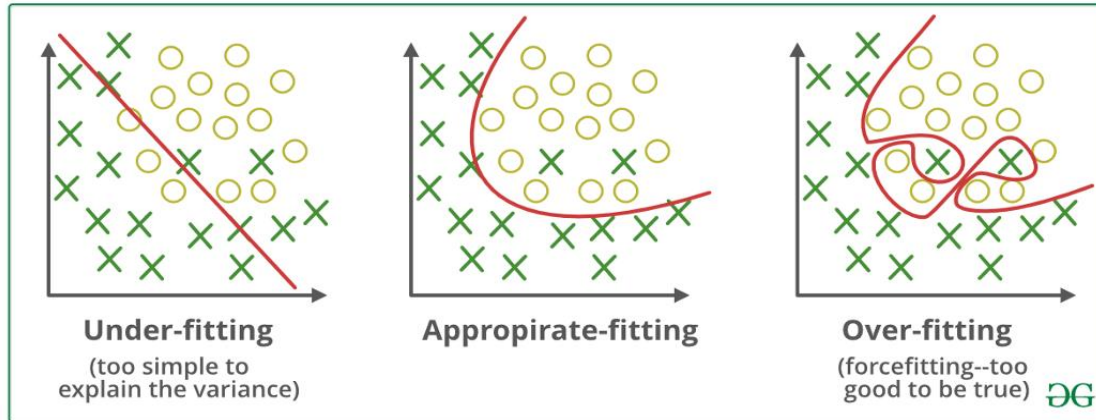
- Hangi algoritma sizin probleminizi akıllandırmada kullanılır? Bir metot var mı? Yok. Ne var? Deneyim, yetenek...
- Eğitim ve test verisi nedir?
- Model nedir?
- Görselleştirme?
- Yorum, öngörude bulunmak, karar vermek
- Büyük veriden model çıkarılmaz. Eğitim ve test veri kümesi.
- Eğitim veri kümesi: model oluşturulur.
- Test veri kümesi: Model test edilir, doğrulanır.

# Değerlendirme - Evaluation

- Accuracy
  - Precision and recall
  - Squared error
  - Likelihood
  - Posterior probability
  - Cost / Utility
  - Margin
  - Entropy
  - K-L divergence
  - Etc.
- Kesinlik
  - Hassasiyet ve geri çağırma
  - Kare hatası
  - Olasılık
  - Arka olasılık
  - Maliyet / Fayda
  - Marj
  - Entropi
  - K-L sapma
  - Vesaire.

# Genelleme

- Eğitim verilerinden bir sınıflandırıcı, tahmin edici ya da model geliştirilir.
- Ancak eğitim verileri üzerinden model geliştirmek iyi performans göstermek yeterli **DEĞİLDİR!**
- Genelleme yapabilmek için test verilerindeki görünmeyen performansı değerlendirmek de önemlidir.
- Eğitim verilerindeki hassasiyet çok düşükse, çok fazla varyansa (**ağırlık ortalama ile her bir veri arasındaki sapma**) sahip ise, bu aşırı uyuma neden olabilir: eğitim verilerinde çok iyi, ancak test verilerinde zayıf yüksek varyans (Sapma) var demektir. Varyans, ortalamadan standart sapma olarak da bilinen farkların karelerinin ortalamasıdır. Veri yığını içindeki verilerin birbirlerinden ne kadar farklı olduklarının göstergesidir.



# Genelleme Hatası

- Genelleme hatasının bileşenleri:
  - **Önyargı – Eşik seviyesi: Akıllı modellerdeki bazı parametrelerin göz ardı edilmesi, ihmal edilmesidir.** Tüm eğitim setlerindeki ortalama model, gerçek modelden ne kadar farklıdır? Model tarafından yapılan yanlış varsayımlardan ya da basitleştirmelerden kaynaklanan hatadır.
  - **Varyans:** Farklı eğitim setlerinden tahmin edilen modelleri oluşturan verilerin birbirinden ne kadar farklı olduğunun göstergesidir.
- Eksik önyargı: model, ilgili tüm sınıf özelliklerini temsil etmek için çok “basit”
  - Yüksek önyargı ve düşük varyans
  - Yüksek eğitim hatası ve yüksek test hatası
- Aşırı uyum: Tüm parametrelerin belirlenmesi. Veri yığınının tamamen temsil edel model oluşturulması. Model çok "karmaşık" ve verilerdeki alakasız özelliklere (gürültü) uyuyor
  - Düşük önyargı ve yüksek varyans
  - Düşük eğitim hatası ve yüksek test hatası

# Hassasiyet

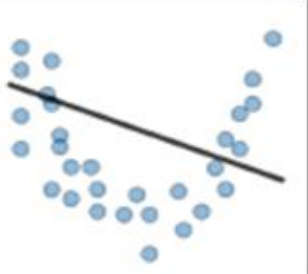

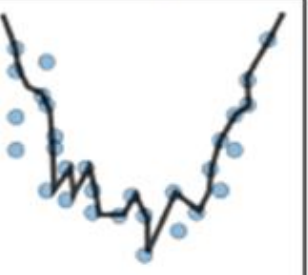
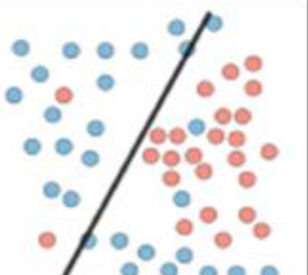
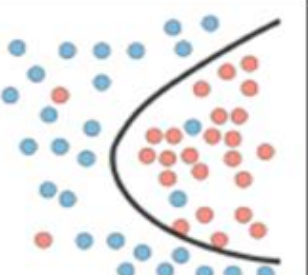
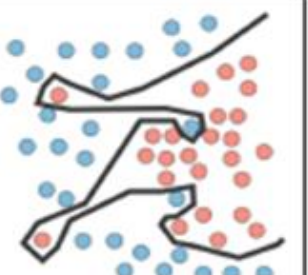
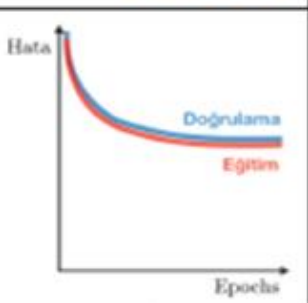
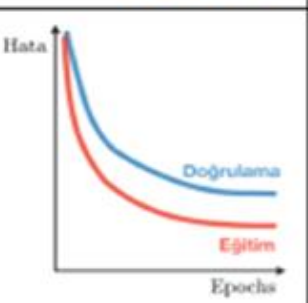
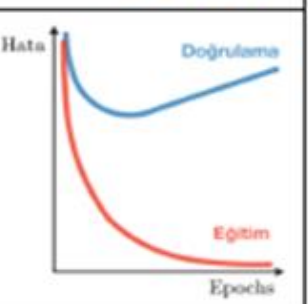
- Modeli oluřturan parametrelerden herhangi birindeki küçük bir deęişim çok büyük bir sapmaya ya da kaotik davranıřa neden olmasdır.

# Uyum, Varyans, Eğitim hatası

Underfitting: Yetersiz uyum

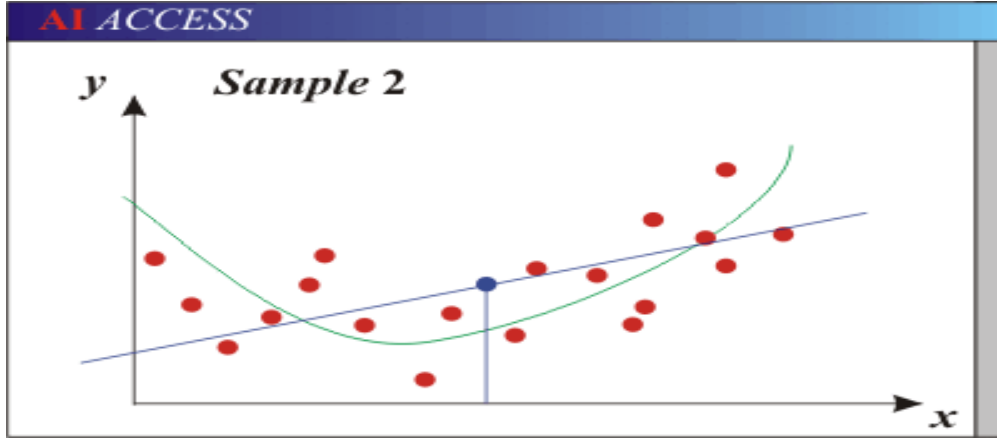
Just right: Eğitim hatasından biraz daha düşük test hatası

- Önyargı
- Varyans
- Aşırı uyum
- Uyumsuzluk
- Eğitim hatası

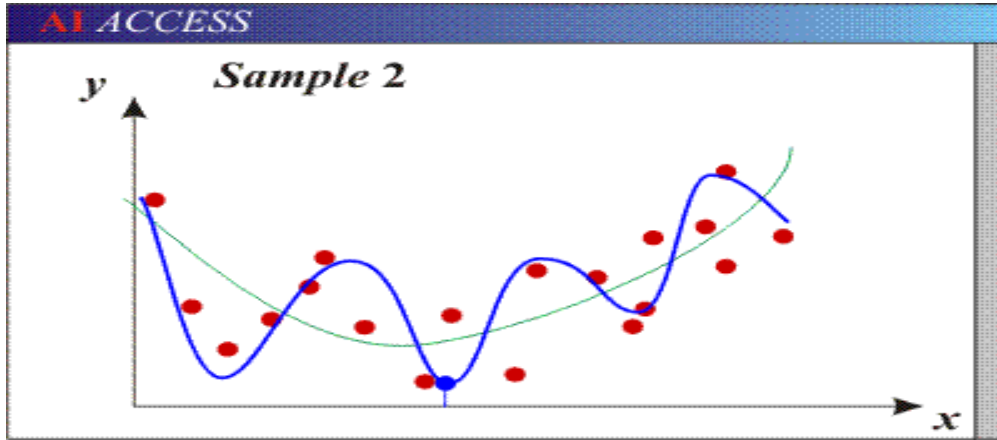
	Underfitting	Just right	Overfitting
Belirtiler	- Yüksek eğitim hatası - Test hatasına yakın eğitim hatası - Yüksek önyargı	- Eğitim hatasından biraz daha düşük <b>test</b> hatası	- Çok düşük eğitim hatası - Eğitim hatası test hatasının çok altında - Yüksek varyans
Regresyon			
Sınıflandırma			
Derin öğrenme			



# Önyargı-Varyans Arasındaki ilişki



- Ön yargıdan dolayı çok az parametrelili modeller, büyük bir sapma (yeterli esneklik) nedeniyle hatalıdır.



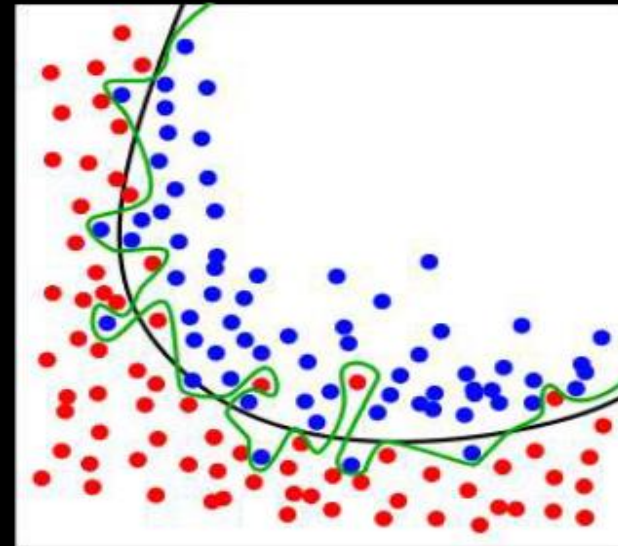
- Ön yargıdan çok fazla parametreye sahip modeller, büyük bir varyans, aşırı uyum (örneğe karşı çok fazla hassasiyet) nedeniyle hatalıdır.

# Varyans Nedir?

- **Verilerdeki eğilim, eşik seviyesi, sapma** bize verilerde tutarsızlık olduğunu söyler. Tutarsızlık, birbirini dışlamayan çeşitli nedenlerle ortaya çıkabilir.
- Örneğin, Amazon gibi bir teknoloji devi, işe alım sürecini hızlandırmak için 100 özgeçmiş verecekleri bir arama motoru inşa ediyor, ilk beşi seçecek ve onları işe alacak. Şirket, yazılımın cinsiyetten bağımsız sonuçlar üretmediğini fark ettiğinde, bu önyargıyı ortadan kaldırmak için ayarlamalar yapmak zorunda kalmıştır.

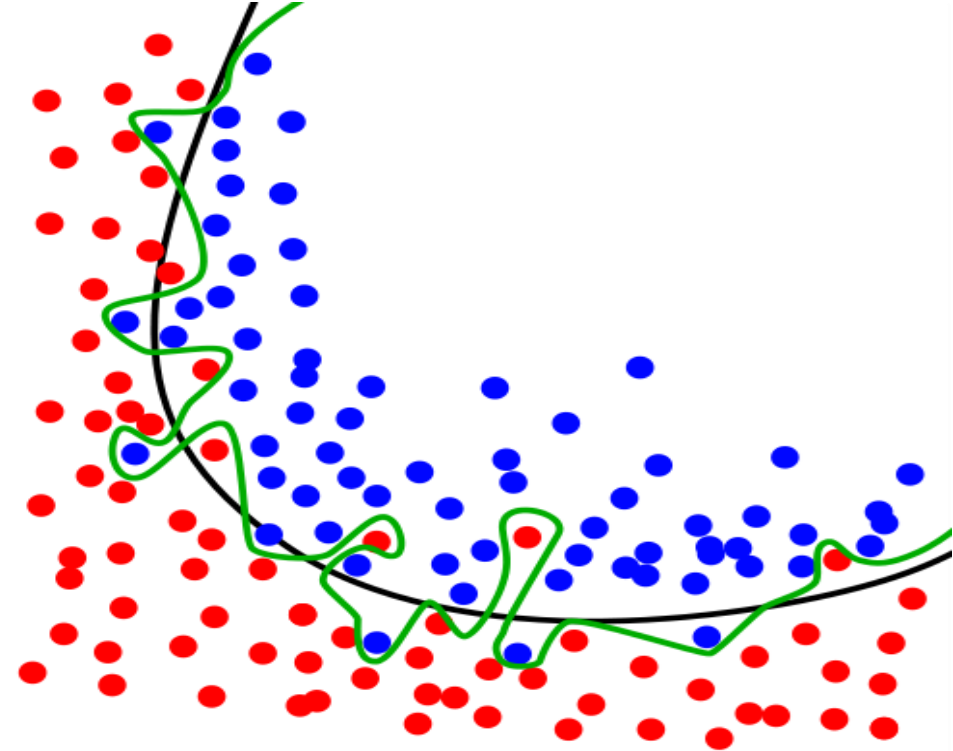
## Overfitting

- **Regularization**
  - Introduce additional information such as a lambda parameter in the cost function (to update the theta parameters in the gradient descent algorithm)
  - Dropout: prevent complex adaptations on training data by dropping out units (both hidden and visible)
- **Test new datasets**

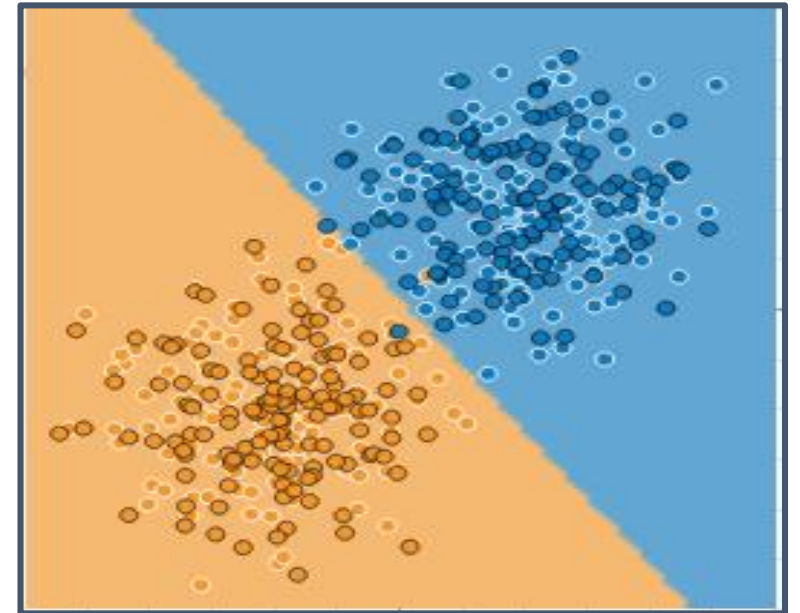
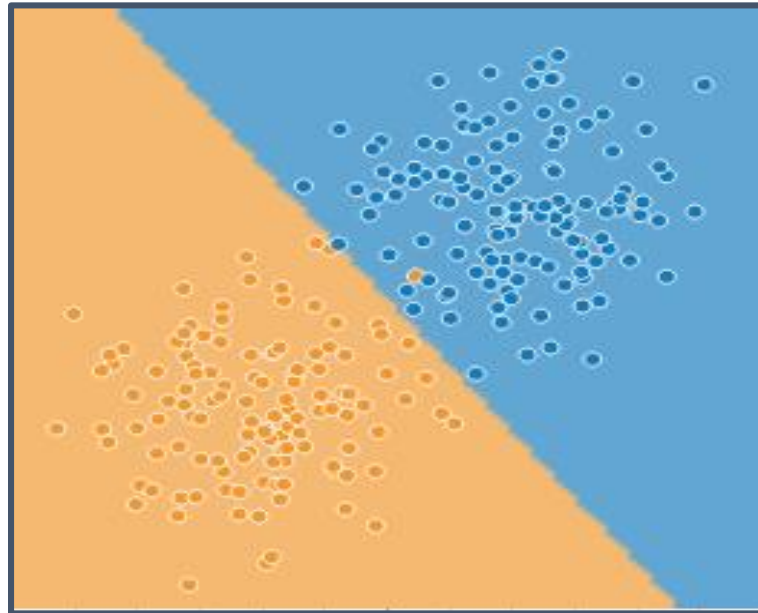
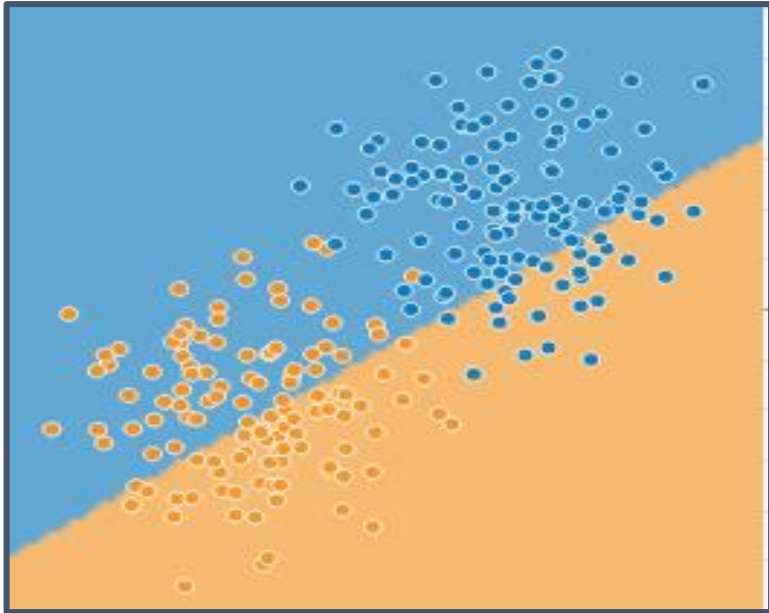


# Aşırı ve yetersiz uyum, genelleme, düzenlileştirme

- Çok sayıda parametreye sahip modeller, eğitim verilerine kolaylıkla uyarlanabilir.
- Genelleme: ML modelinin kalitesi yeni, görülmemiş örnekler üzerinde ölçülür.
- Düzenlileştirme: Daha fazla veri eklemek dışında aşırı uyum basitliğini, seyrekliği, bırakmayı, erken durdurmayı \*) önleyen herhangi bir yöntem\*



# Model – learning/training – inference



$$\hat{y} = f(\mathbf{x}; \theta)$$

- parameters  $\theta$  and hyperparameters

# Remember...

- Hiçbir sınıflandırıcı doğası gereği diğerlerinden daha iyi değildir: genellemek için varsayımlarda bulunmanız gerekir.
- Three kinds of error
  - Doğuştan: kaçınılmaz (Inherent: unavoidable)
  - Önyargı: aşırı basitleştirmeler nedeniyle (Bias: due to over-simplifications)
  - Varyans: Sınırlı verilerden parametrelerin mükemmel şekilde tahmin edilememesi nedeniyle



Sadece varsayımlarla genelleme yapabilirsiniz.

# Unutmayın...

- Hiçbir sınıflandırıcı doğası gereği diğerlerinden daha iyi değildir: genellemek için varsayımlarda bulunmanız gerekir.
- Three kinds of error
  - Doğuştan: kaçınılmaz (Inherent: unavoidable)
  - Önyargı: aşırı basitleştirmeler nedeniyle (Bias: due to over-simplifications)
  - Varyans: Sınırlı verilerden parametrelerin mükemmel şekilde tahmin edilememesi nedeniyle
  - Aşırı Uyum

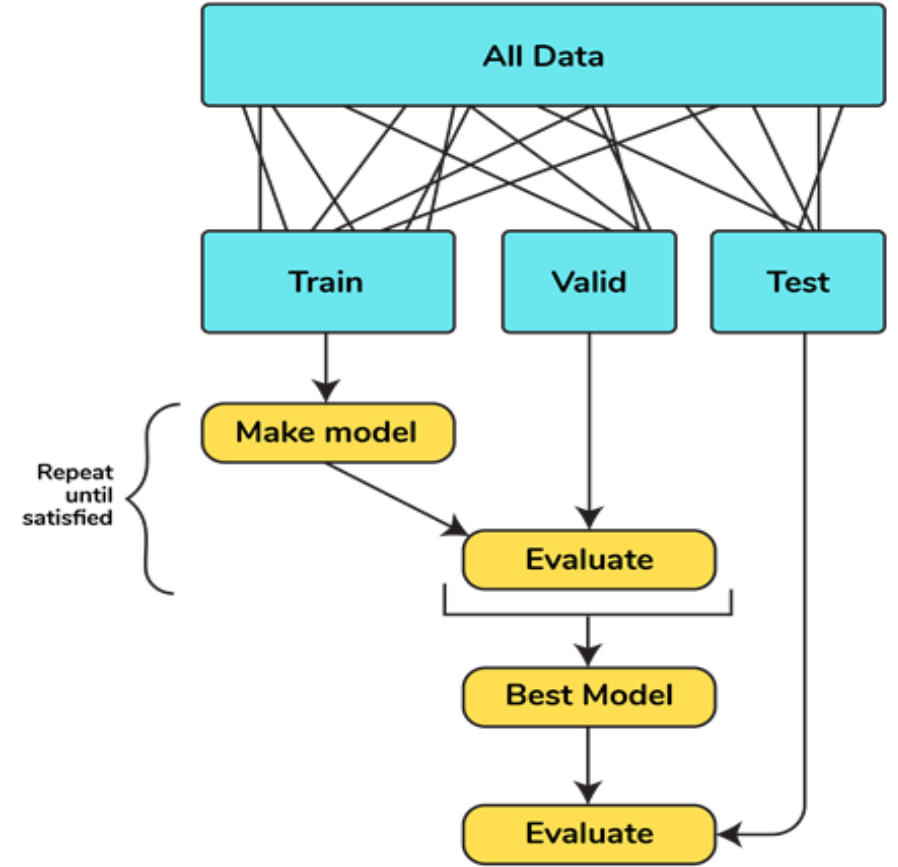
# Varyans nasıl azaltılır?

- Daha basit bir sınıflandırıcı seçilir
- Parametreler düzenli hale getirilir.
- Daha fazla eğitim verisi alınır.
- Daha basit sınıflandırıcı veya düzenlileştirme, yanlılığı artırabilir ve daha fazla hataya yol açabilir.



# Çapraz doğrulama

- Çapraz doğrulama, tüm verileri üç bölüme ayırma yöntemidir: eğitim, test ve doğrulama verileri. Veriler k alt kümeye bölünür ve model bu veri kümelerinin k-1'i üzerinde eğitim almıştır.
- Son alt küme test için tutulur. Bu, alt kümelerin her biri için yapılır. Bu, k kat çapraz doğrulamadır. Son olarak, nihai puanı üretmek için tüm k-katlarından elde edilen puanların ortalaması alınır.



# Parametrik ve Parametrik Olmayan Modeller

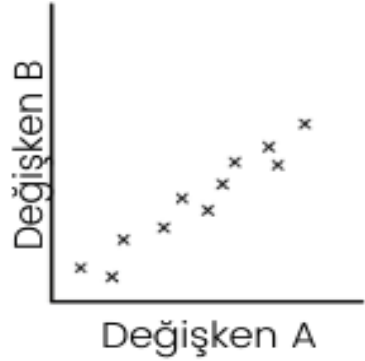
- Etkin Faktörler (Agents)
- Parametrik modeller sınırlı parametrelere sahip olacak ve yeni verileri tahmin etmek için sadece modelin parametresini bilmeniz yeterli.
- Parametrik olmayan modellerin bir dizi parametre alma konusunda hiçbir sınırı yoktur, bu da daha fazla esnekliğe ve yeni verileri tahmin etmeye olanak tanır. Verilerin durumunu ve model parametrelerini bilmeniz gerekir.

# Korelasyon ve Kovaryans

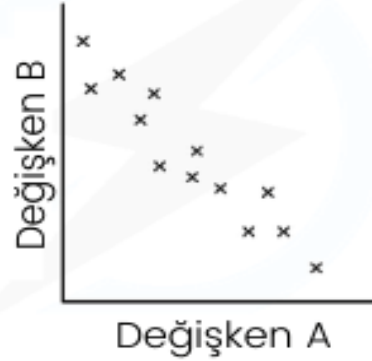
- Korelasyon, iki deęişken arasındaki nicel ilişkiyi ölçmek ve tahmin etmek için kullanılır. Korelasyon, iki deęişkenin ne kadar güçlü bir şekilde ilişkili olduğunu ölçer. Örnekler, gelir ve gider, talep ve arz vb.
- Kovaryans, iki deęişken arasındaki korelasyonu ölçmenin basit bir yoludur. Kovaryansla ilgili sorun, normalleştirme olmadan karşılaştırmanın zor olmasıdır.

# Korelasyon Analizi Nedir ?

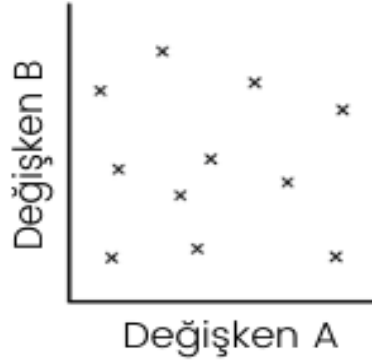
Pozitif Korelasyon



Negatif Korelasyon



Korelasyon yok



iki nümerik nitelik arasında bir bağlantının kuvvetini tanımlar. Pearson korelasyon katsayısına bakılır. Korelasyon (Pearson) katsayısı, X bağımsız değişken veya neden değişkenidir. Y ise bağımlı değişken veya sonuç değişkenidir. r değeri her zaman -1 ile +1 arasında bir değer alır. Bu değerlere yakınsaması bağımsız değişken ile bu değişkene bağımlı olan değişken arasında kuvvetli bir ilişki olduğunu ifade eder.

# Korelasyon Analiz Formulu

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}}$$

Burada:

- $r$  = Pearson korelasyon katsayısı,
- $n$  = gözlem sayısı,
- $X$  ve  $Y$  = sırasıyla ilk ve ikinci değişkenin değerleri,
- $\sum$  = toplamı ifade eder.

$x$ ,: bağımsız değişken, dizidir.  $y$ : bağımlı değişken, dizi.

- $r$ , değeri 0 ise korelasyon yoktur.
- $R$  değeri -1 ya da +1 değerine eşit ya da yaklaşırsa güçlü ilişki (korelasyon) vardır.
- -0.5, +0.5 değeri etrafında ise zayıf korelasyon vardır.

# Hipotezler

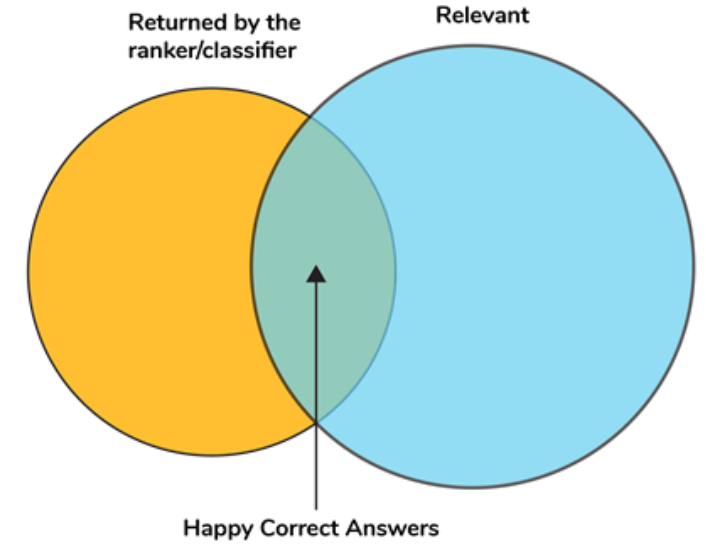
- Olasılıksal olarak belirlenen değerin doğrulanmasıdır.
- Hipotezler, eğitim verilerinde olmayan örnekleri doğru bir şekilde sınıflandırmak için genelleştirme yapar.
- Basitçe eğitim örneklerini ezberlemek, genellemeyen tutarlı bir hipotezdir.
- Basit bir hipotez bulmak, genellemenin sağlanmasına yardımcı olur.

# What is P-value?

- P değerleri, bir hipotez testi hakkında karar vermek için kullanılır. P değeri, sıfır hipotezini reddedebileceğiniz minimum anlamlı düzeydir. P değeri ne kadar düşükse, boş hipotezi reddetme olasılığınız o kadar yüksek olur.

# Kesinlik ve Hatırlama

- Kesinlik ve hatırlama, makine öğrenimi uygulamasının gücünü izlemenin yollarıdır. Ancak çoğu zaman aynı anda kullanılırlar.
- Kesinlik, "Sınıflandırıcının tahmin ettiği öğelerden kaç tanesi gerçekten alakalı?" sorusuna yanıt verir.
- Hatırlama, "Gerçekten alakalı olan tüm öğelerden kaç tanesi sınıflandırıcı tarafından bulunur?"
- Genel olarak kesinliğin anlamı, kesin ve doğru olma gerçeğidir. Aynı şey makine öğrenimi modelimizde de geçerli olacak. Modelinizin alakalı olması için tahmin etmesi gereken bir dizi öğeniz varsa. Kaç öğe gerçekten alakalı?
- Yandaki şekil, kesinlik ve hatırlama sağlayan Venn şemasını göstermektedir.
- Matematiksel olarak kesinlik ve hatırlama şu şekilde tanımlanabilir:
  - kesinlik = doğru cevap/ toplam öğe
  - hatırlama = doğru cevap/ toplam alakalı (relevant ) cevap





# Öneri Sistemleri

- Öneri sistemleri için gerekli veriler, bir film izledikten veya bir şarkı dinledikten sonra açık kullanıcı derecelendirmelerinden, gizli arama motoru sorgularından ve satın alma geçmişlerinden veya kullanıcılar/öğelerin kendileri hakkındaki diğer bilgilerden kaynaklanır.
- Kullanıcı farkında olmadan, ekranda ürüne odaklanma, şık seçme, yorumlara bakma

## Bir veri setinin Normalliği nasıl kontrol edilir?

- Görsel olarak, grafikleri kullanabiliriz. Normallik kontrollerinden birkaçı aşağıdaki gibidir:
  - Shapiro-Wilk Test
  - Anderson-Darling Test
  - Martinez-Iglewicz Test
  - Kolmogorov-Smirnov Test
  - D'Agostino Skewness Test

## Difference Between Sigmoid and Softmax functions?

- Karar verme aşamasında,
- İkili sınıflandırma için sigmoid işlevi kullanılır. Olasılıklar toplamı 1 olmalıdır.
- Oysa çoklu sınıflandırma için Softmax işlevi kullanılır. Olasılıklar toplamı 1 olacaktır.

# Aykırı Değerler Nasıl İşlenir?

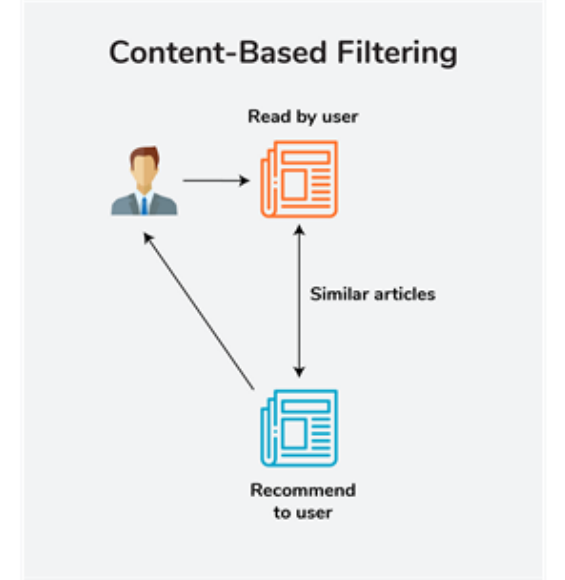
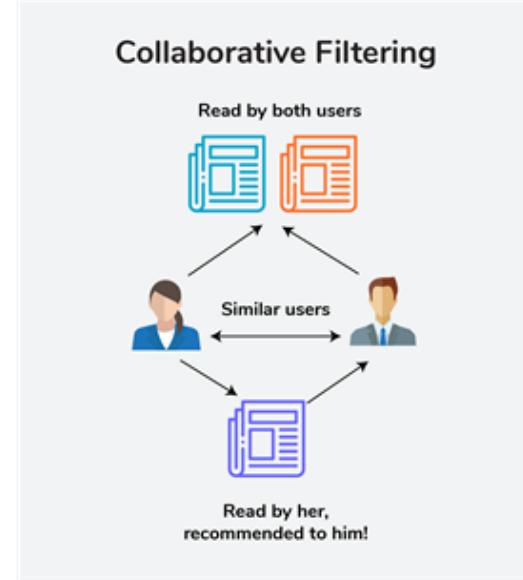
- Aykırı Değer, veri kümesindeki diğer gözlemlerden çok uzakta olan veri kümesindeki bir gözlemdir. Aykırı değerleri keşfetmek için kullanılan araçlar şunlardır:
  - Box plot
  - Z-score
  - Scatter plot, etc.
- Tipik olarak, aykırı değerleri ele almak için dört basit strateji izlememiz gerekir:
  - Onları bırakabiliriz.
  - Bunları aykırı değer olarak işaretleyebilir ve bir özellik olarak ekleyebiliriz.
  - Aynı şekilde, aykırı değerın etkisini azaltmak için özelliği dönüştürebiliriz.
  - Manipülasyon amaçlıdır, alarm üretilir.

# Kayıp Fonksiyonu ve Maliyet Fonksiyonları nedir? Aralarındaki Anahtar Farkı Açıklayın?

- Kaybı hesaplarırken yalnızca tek bir veri noktasını dikkate alıyoruz, o zaman kayıp fonksiyonu terimini kullanıyoruz.
- Oysa birden fazla veri için hata toplamını hesaplarırken maliyet işlevini kullanırız. Önemli bir fark yok.
- Başka bir deyişle, kayıp işlevi, tek bir kayıt için gerçek ve tahmin edilen değerler arasındaki farkı yakalarken, maliyet işlevleri tüm eğitim veri kümesi için farkı toplar.
- En sık kullanılan kayıp fonksiyonları, Ortalama kare hatası kaybıdır.
- **Ortalama Kare Hata (MSE):** Basit bir deyişle, modelimizin gerçek değerlere karşı değerleri nasıl tahmin ettiğini söyleyebiliriz.
- $MSE = \frac{1}{n} \sum (predicted\ value - actual\ value)^2$
- Hinge loss: It is used to train the machine learning classifier, which is
- $L(y) = \max(0, 1 - yy)$
- Where  $y = -1$  or  $1$  indicating two classes and  $y$  represents the output form of the classifier. The most common cost function represents the total cost as the sum of the fixed costs and the variable costs in the equation  $y = mx + b$

# İşbirlikçi ve İçerik Tabanlı Filtreleme nedir?

- Ortak filtreleme, kişiselleştirilmiş içerik önerileri için kanıtlanmış bir tekniktir. İşbirlikçi filtreleme, bireysel kullanıcının ilgi alanlarını birçok kullanıcının tercihleriyle eşleştirerek yeni içeriği tahmin eden bir tür öneri sistemidir.
- İçerik tabanlı öneri sistemleri sadece kullanıcının tercihlerine odaklanmaktadır. Kullanıcının önceki tercihlerine göre benzer içeriklerden kullanıcıya yeni öneriler yapılır.

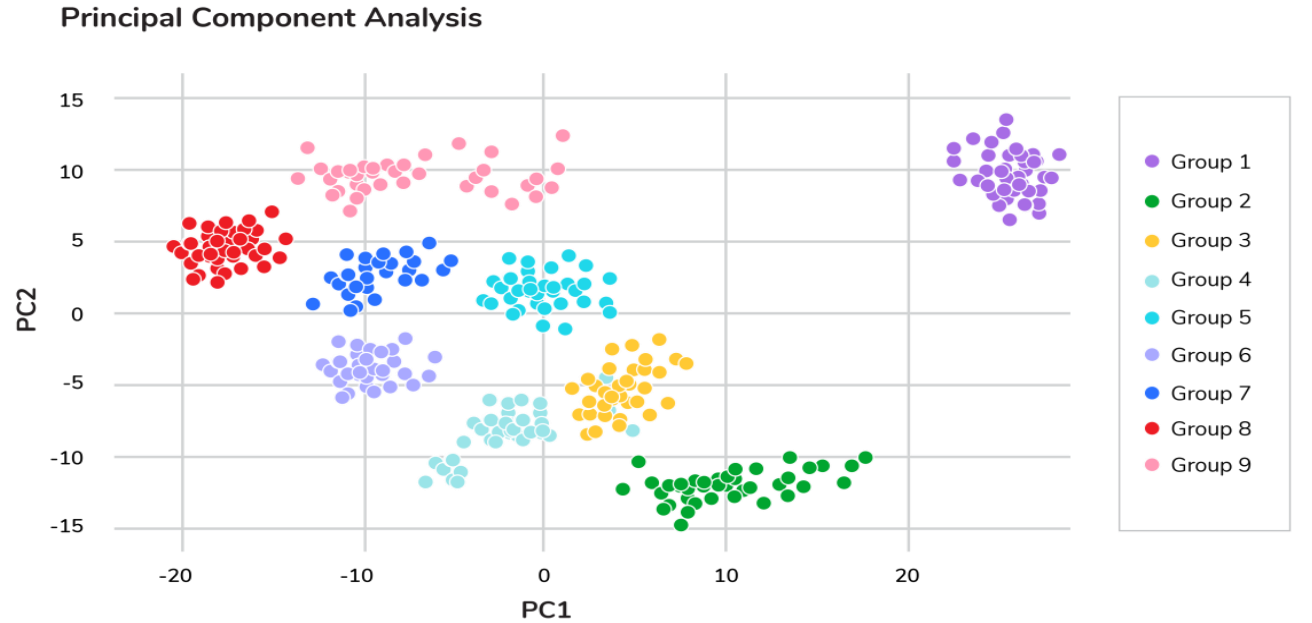


## Aşırı Uyum ve Eksik Uyumla Nasıl Başa Çıkılır?

- Aşırı uyum, modelin eğitim verilerine çok iyi uyduğu anlamına gelir, bu durumda, verileri yeniden örneklememiz ve k-kat çapraz doğrulama gibi teknikleri kullanarak model doğruluğunu tahmin etmemiz gerekir.
- Uygun olmayan durum için verilerden kalıpları anlayamıyor veya yakalayamıyoruz, bu durumda algoritmaları değiştirmemiz veya modele daha fazla veri noktası beslememiz gerekiyor.

# What is Principal Component Analysis (PCA)? When do you use it?

- Temel bileşen analizi (PCA) en yaygın olarak boyut küçültme için kullanılır.
- Bu durumda, PCA her bir değişkendeki (veya tablodaki sütundaki) varyasyonu ölçer. Küçük bir varyasyon varsa, aşağıdaki şekilde gösterildiği gibi değişkeni dışarı atar.
- Böylece veri kümesinin görselleştirilmesini kolaylaştırır. PCA finans, sinirbilim ve farmakolojide kullanılır.
- Özellikle özellikler arasında doğrusal korelasyonlar olduğunda, bir ön işleme adımı olarak çok kullanışlıdır..





# What is F1 score? How would you use it?

- İkili sınıflandırmanın istatistiksel analizinde, F puanı veya F ölçüsü, bir testin doğruluğunun bir ölçüsüdür. Kesinliğin, doğru olarak tanımlanmayanlar da dahil olmak üzere tüm pozitif sonuçların sayısına bölünmesiyle elde edilen gerçek pozitif sonuçların sayısına bölüdüğü ve testin kesinliği ve geri çağırılmasıyla hesaplanır ve geri çağırma, gerçek pozitif sonuçların sayısının bölünen gerçek pozitif sonuçların sayısıdır. pozitif olarak tanımlanması gereken tüm örneklerin sayısı. Kesinlik, pozitif tahmin değeri olarak da bilinir ve hatırlama, tanısal ikili sınıflandırmada duyarlılık olarak da bilinir.
- Doğrudan F1 skoruna geçmeden önce bu tabloya bir göz atalım.
- İkili sınıflandırmada, F1 puanını modelin doğruluğunun bir ölçüsü olarak kabul ederiz. F1 puanı, kesinlik ve hatırlama puanlarının ağırlıklı ortalamasıdır.
- $F1 = 2TP / (2TP + FP + FN)$
- F1 için 0 ile 1 arasında puanlar görüyoruz, burada 0 en kötü puan ve 1 en iyi puan.
- F1 puanı, genellikle bir modelin ilgili sonuçları ne kadar iyi aldığını ve modelimizin performans gösterdiğini görmek için bilgi almada kullanılır.

Prediction	Predicted Yes	Predicted No
Actual Yes	True Positive (TP)	False Negative (FN)
Actual No	False Positive (FP)	True Negative (TN)