Mekatronik Mühendisliği Uygulamalarında Yapay Zekâ

Özellik Belirleme (Feature selection and extraction)

Prof.Dr. Erhan AKDOĞAN

Özellik Seçimi

Machine learning works on a simple rule:

if you put garbage in,

you will only get garbage to come out.

By garbage here, I mean noise in data.

https://www.analyticsvidhya.com

Özellik seçimi:

- **Elimizdeki verilere ait birçok özellikten verinin kümesini, sınıfını veya değerini belirleyen özelliklerin hangilerinin olduğunu bilinmeyebilir.**

- **Elimizdeki veriye ait tüm özellik kümesinin bir alt kümesinin seçimi özellik seçimi,**

- **Bu özelliklerin birleşiminden yeni özellik elde etmeye ise özellik çıkarma denir.**

Özellik Seçim ve Çıkarımın karşılaştırılması:

Table 3: Advantages and disadvantages between feature selection and feature extraction.

Method	Advantages	Disadvantages
Selection	Preserving data characteristics for interpretability	Discriminative power Lower shorter training times Reducing overfitting
Extraction	Higher discriminating power Control overfitting when it is unsupervised	Loss of data interpretability Transformation maybe expensive

Özellik seçimini kullanmanın başlıca nedenleri:

Makine öğrenimi algoritmasının daha hızlı ilerlemesini sağlar.

Bir modelin karmaşıklığını azaltır ve yorumlanmasını kolayla

Doğru alt küme seçildiğinde bir modelin doğruluğunu artırır.

Ezberlemeyi, minimuma takılmayı azaltır, genelleştirmeyi artır

Özellik seçim yöntemleri:

1. **Bilgi kazancı (Information Gain)**

2. **Sinyal Gürültü Oranı (signal to noise ratio)**

3. **Alt Küme Seçiciler (Wrappers)**

4. **Filtre Metodları (Filter methods)**

5. **Gömülü Metodlar (Embedded methods)**

1. Bilgi Kazancı

- ***Bilgi kazancı,*** verilen bir özelliğin sınıflandırmada ne kadar etkili olduğunu gösteren ölçüt, 0 ile 1 arasında değişir.

- ***Entropi***: Bir veri kümesi içinde belirsizlik ve rasgeleliği ölçmek için kullanılır. Bu değer ne kadar büyük ise belirsizlik o kadar yüksektir.

Entropi

$$H(S) = -\sum_{t=1}^{n} p_i * log_2(p_i)$$

H: entropi
S: Kaynak
p: Olasılık

Bilgi Kazancı

$$IG(D) = H(D) - \sum_{i=1}^{n} P(D_i)H(D_i)$$

IG: Bilgi kazancı
D: veri kümesi
P: ağırlık olasılığı
H: Entropi

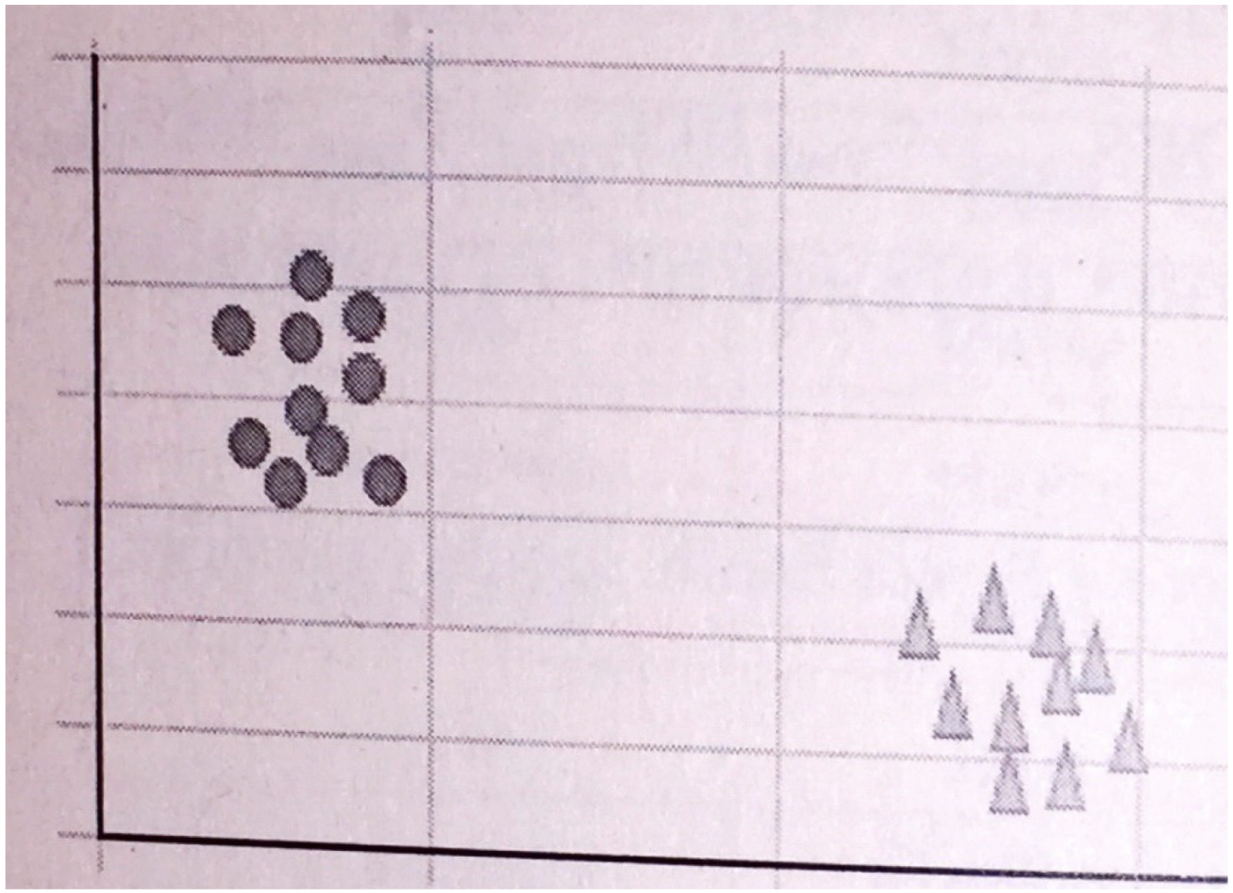
Örnek:

YAŞ	MEZUNİYET	ŞİRKET SAHİBİ	KARAR
orta	lise	E	İYİ
orta	üniv	E	FAKİR
yaşlı	lise	E	ZENGİN
genç	lise	H	İYİ
genç	üniv	E	ORTA
genç	lise	H	İYİ
yaşlı	üniv	H	İYİ
yaşlı	üniv	E	ZENGİN
yaşlı	lise	E	İYİ
orta	üniv	E	FAKİR

Hangi özelliğin bilgi kazancı daha yüksektir?

2. Sinyal Gürültü Oranı:

- **Sınıflar arası uzaklıklar fazla, sınıf içi uzaklıklar az olduğunda özellik seçiminde kullanılan bir yöntemdir.**
- **Herbir özellik için bu oran ayrı ayrı hesaplanır.**
- **Yüksek değerler yüksek ilişkiye(korelasyon) işaret eder.**



İki sınıflı örneklem kümesinin koordinat düzlemindeki görüntüsü

Sinyal gürültü oran formülü

$$S_i = \frac{m_1 - m_2}{d_1 - d_2}$$

S_i : i. özelliğin sinyal gürültü oranı

m_1 : 1.sınıfdaki özelliklerin ortalaması

d_1 : 1.sınıfdaki özelliklerin standart sapması

3. Alt Küme Seçiciler (Wrappers)

- **Herbir özellik için ayrı bir bilgi edinme yerine özellikler birlikte değerlendirilerek sınıflandırma yapılır ve özellik alt uzayları tespit edilir.**

- **Sınıflandırma başarısı yüksektir.**

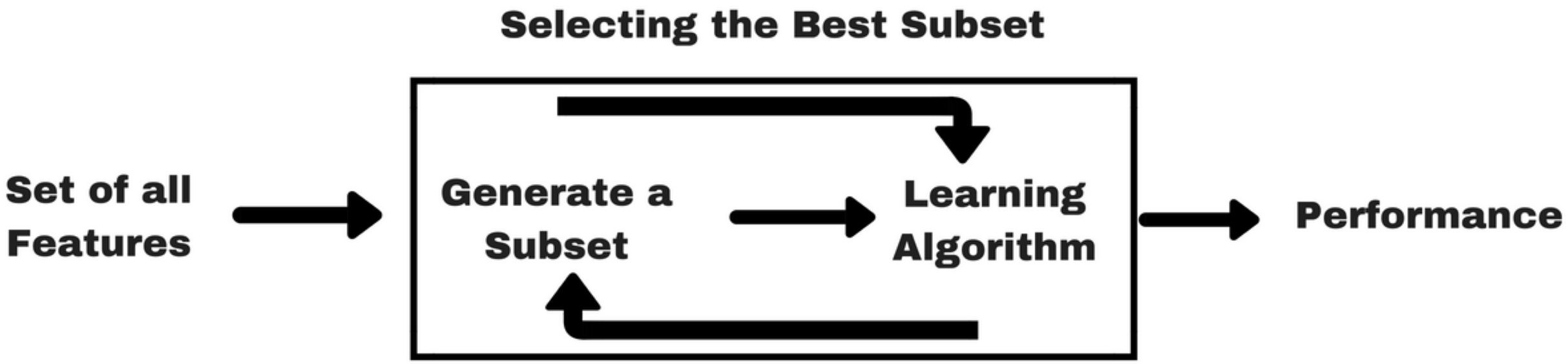
- **Hesaplama karmaşıklığı içerir. Çalışma hızları yavaştır.**

- **Özellik seçiminin her adımında sınıflandırıcıya ihtiyaç duyar.**

Alt Küme Seçiciler (devam)

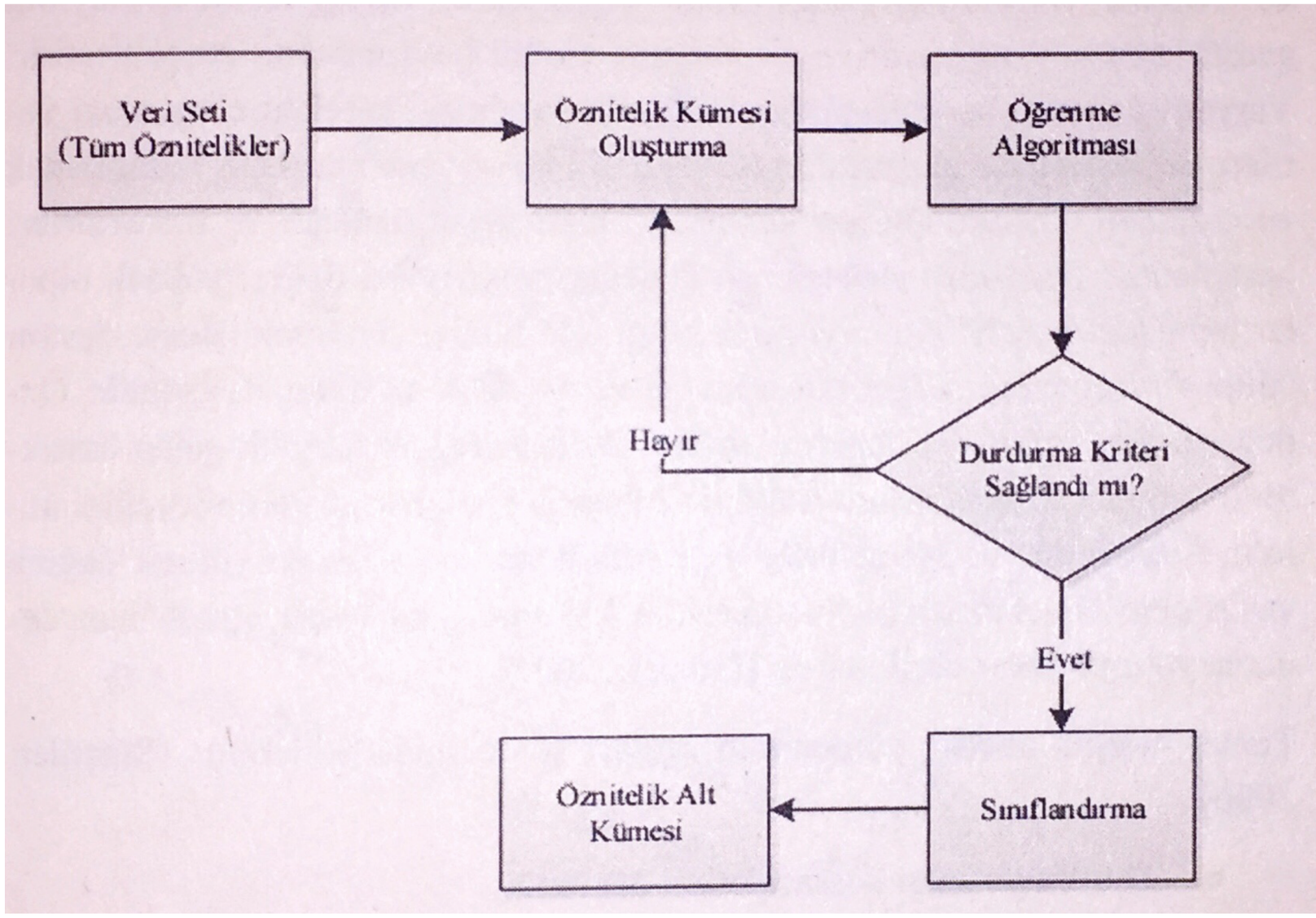
- Bu yöntemde, bir alt özellik kümesi kullanılmaya ve bunları kullanarak bir model geliştirilmeye(eğitilmeye) çalışılır.
- Önceki modelden alınan çıkarımlara dayanarak alt kümeden özellikler eklemeye veya kaldırmaya karar verilir.
- Problem aslında bir arama problemine indirgenmiştir.
- Bu yöntemler genellikle hesaplama açısından çok pahalıdır.
- Bu yöntem ile en iyi özellik çıkarımı için “Boruta” metodu kullanılabilir. Detaylı bilgi için https://www.analyticsvidhya.com/blog/2016/03/select-important-variables-boruta-package/

Alt küme seçim diyagramı:



https://www.analyticsvidhya.com

Alt küme seçim diyagramı:



Alt küme seçicilere ait akış diyagramı(Kaya, 2014)

Alt küme seçicilerde yaygın örnekler

- İleriye doğru özellik seçimi,
- Geriye doğru özellik eleme,
- Özyinelemeli (recursive) özellik eleme, vb.

İleriye doğru özellik seçimi

- İleriye doğru seçim, modelde hiçbir özelliğe sahip olmadan başladığımız yinelemeli(iteratif) bir yöntemdir.
- Her yinelemede, yeni bir değişkenin eklenmesi, modelin performansını iyileştirmeyene kadar modelimizi en iyi şekilde geliştiren özelliği eklemeye devam ediyoruz.

Geriye Doğru eleme:

- Tüm özellikler ile başlanır
- Modelin performansını artıran her yinelemede en az önemli olan özellik kaldırılır.
- Özelliklerin kaldırılmasında hiçbir gelişme gözlenmeyene kadar bu işlem devam eder.

Özyinelemeli (recursive) Özellik Elemesi:

- En iyi performans gösteren özellik alt kümesi bulunmaya çalışılır.
- Bir optimizasyon algoritmasıdır.
- Tekrarlı olarak model oluşturur ve her yinelemede en iyi veya en kötü performans özelliği elenir.
- Tüm özellikler bitene kadar bir sonraki modeli atılan özellikler ile yapılandırır.
- Daha sonra, elemelerinin sırasına göre özellikleri sıralar.

4. Filtre Metodları:

- Genellikle önişleme adımlarında kullanılır.
- Özellik seçimi herhangi makine öğrenmesi adımından bağımsızdır.
- Özellikler, çıktı değişken ile özellik korelasyonları için çeşitli istatistiksel testlerdeki puanlara dayanarak seçilir.

İstatiksel Metodlar:

- **Pearson Korelasyonu**

- **LDA:** Linear discriminant analysis

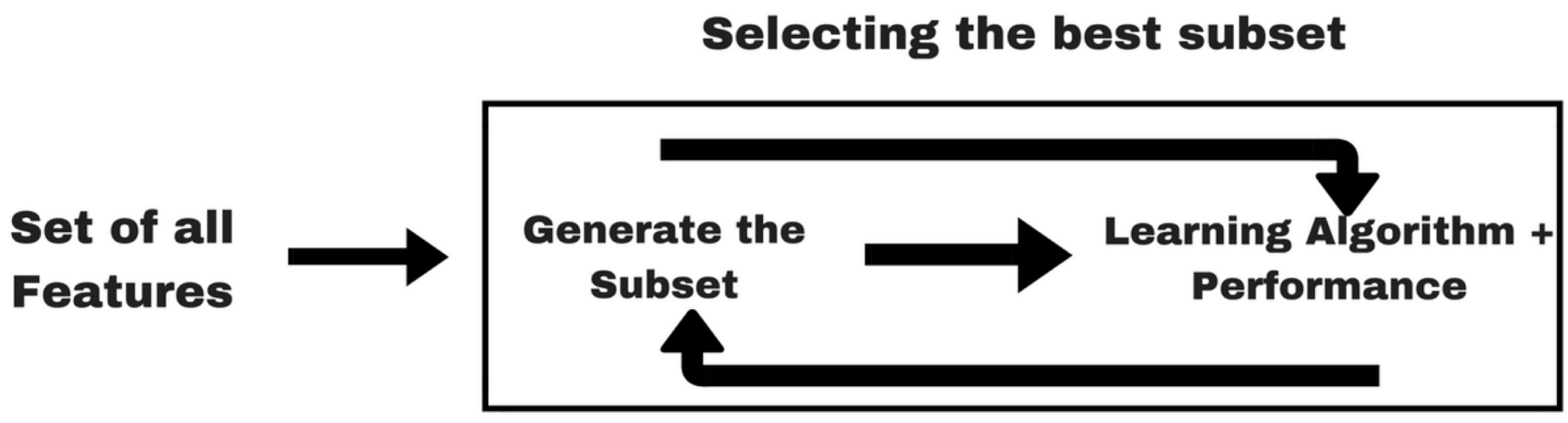
- **ANOVA:** Varyansın analizi için kullanılır.

- **Chi-Square**

5. Gömülü metodlar

Gömülü yöntemler, filtre ve wrapper yöntemlerinin niteliklerini birleştirir.

Kendi yerleşik özellik seçim yöntemlerine sahip olan algoritmalar tarafından uygulanır.



Ozellik Çıkarımı

Özellik Çıkarım Yöntemleri:

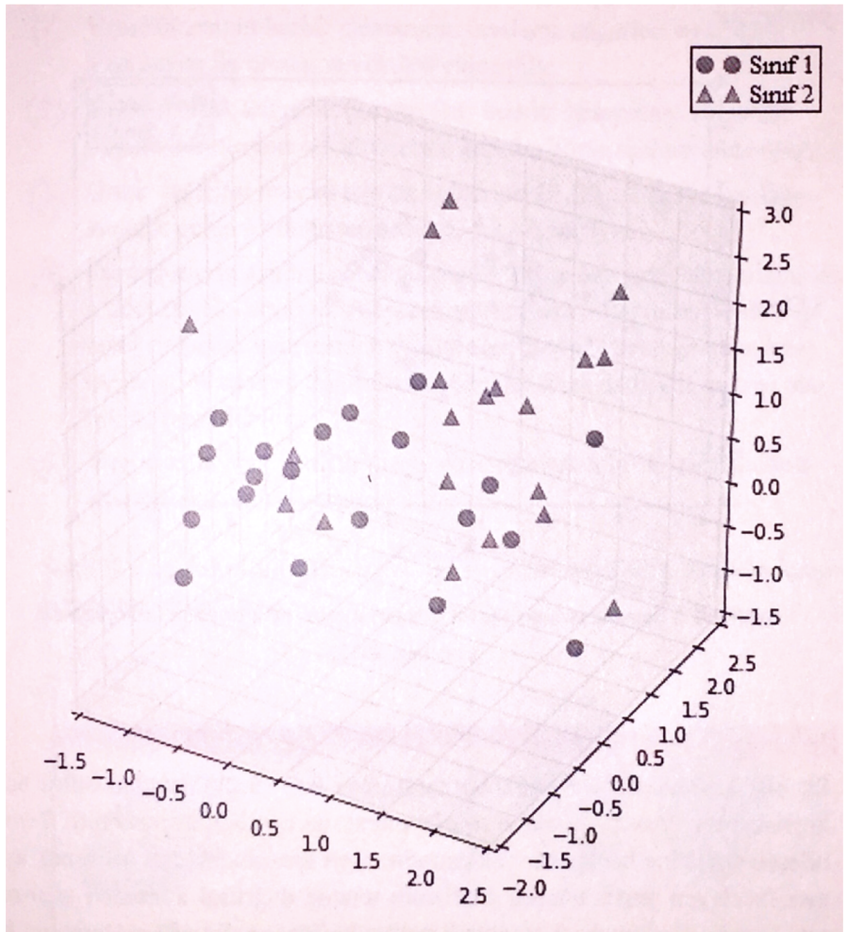
1. Temel Bileşen analizi
   (Principal Component Analysis (PCA))

2. Doğrusal Ayırt eden Analizi

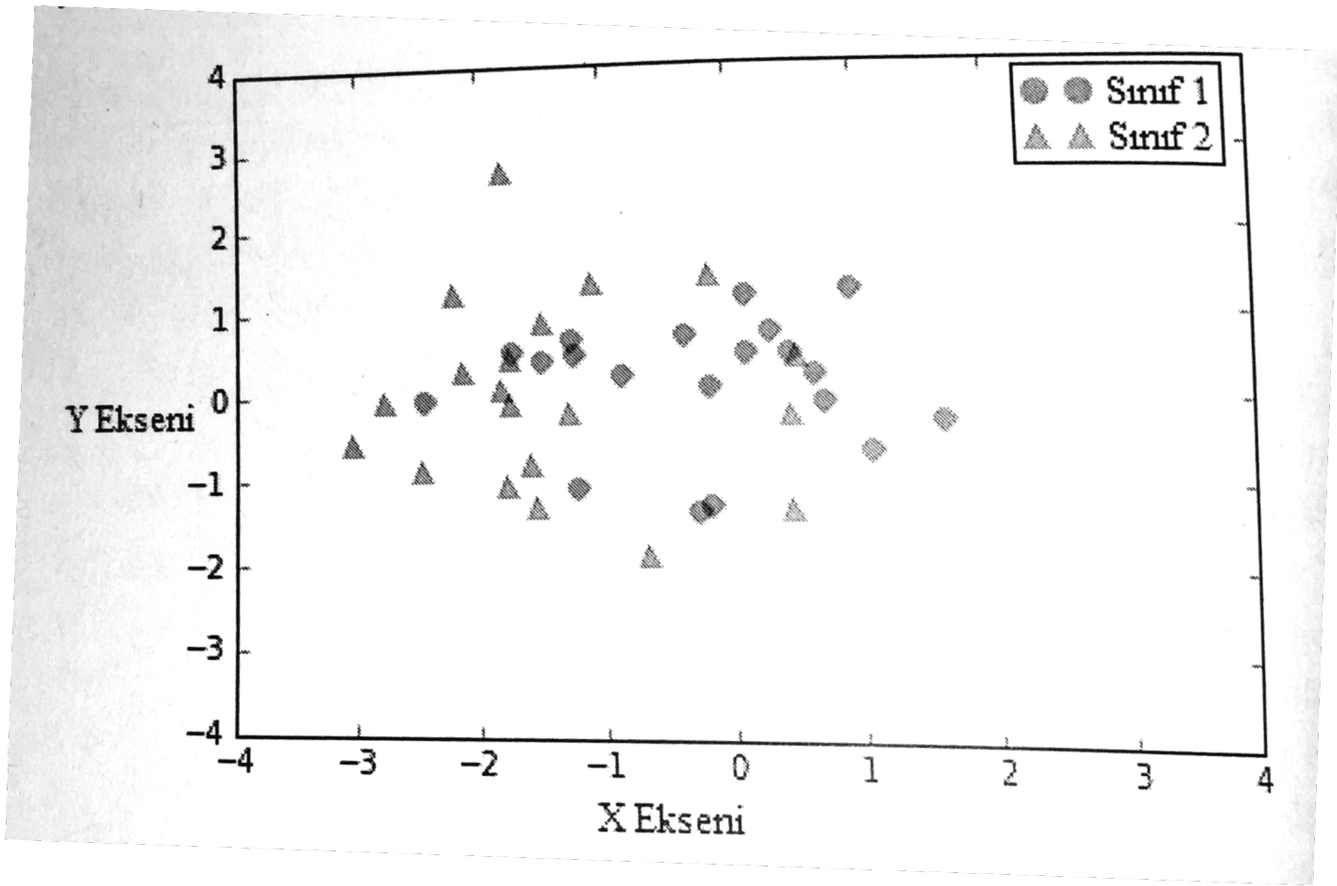
Özellik Çıkarımı Yöntemleri:

- Temel bileşen analizi birbiri ile ilişkili olan birden fazla değişkeni bulunan bir veri kümesinin boyutunu azaltmak için geliştirilmiş bir yöntemdir.
- Amaç verinin temel yapısının bulunması ve boyutunun azaltılmasıdır.
- Boyut azaltma varyans ile yapılır.
- Bunun için eldeki verilerle kovaryans matrisi hesaplanır.
- bundan sonra buna göre özdeğer (eigen value) ve özvektörler (eigen vectors) hesaplanır.
- Sayısal değerş yüksek özdeğerler ile işleme devam edilir.
- Özdeğerin sıfır olması ilgili özdeğere karşılık gelen özvektörün kendisi dışındaki özvektörlerin bileşenleri oalrak gösterilecebileceğini ifade eder.
- En yüksek değere sahip özvektörlerden hangisinin kullanılacağı deneme yanılma yolu ile belirlenir.

Örnek:



3-B Veri



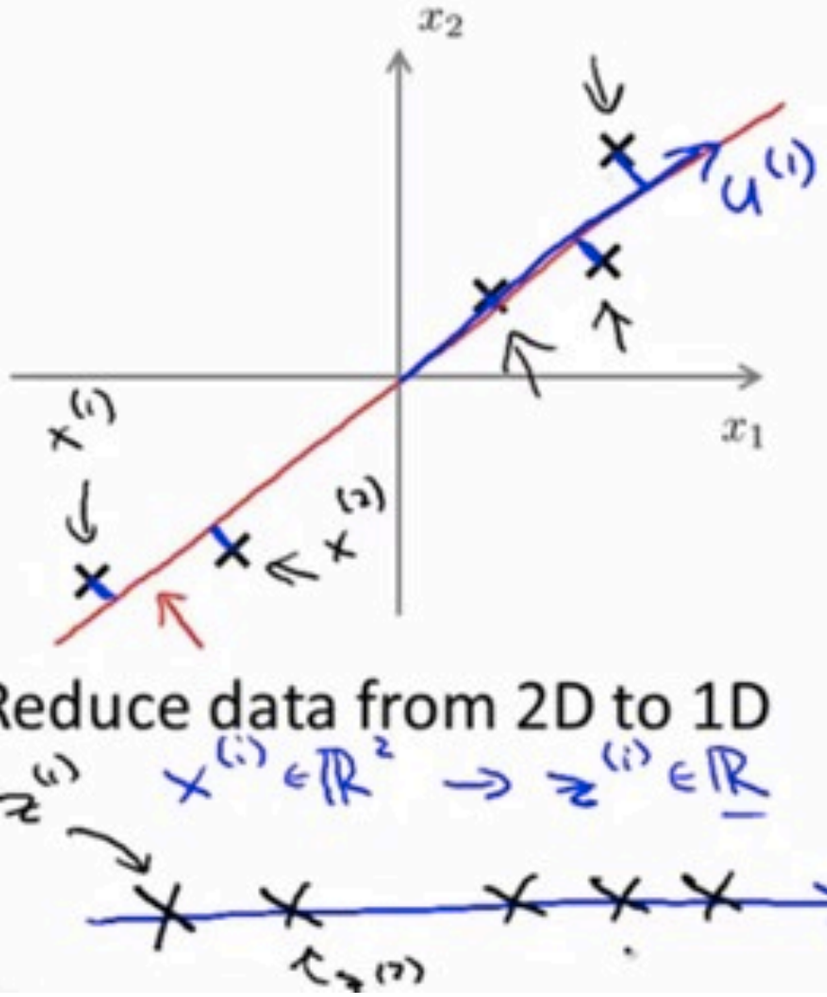
İndirgenmiş 2-B Veri

Bir başka deyişle temel bileşen analizi

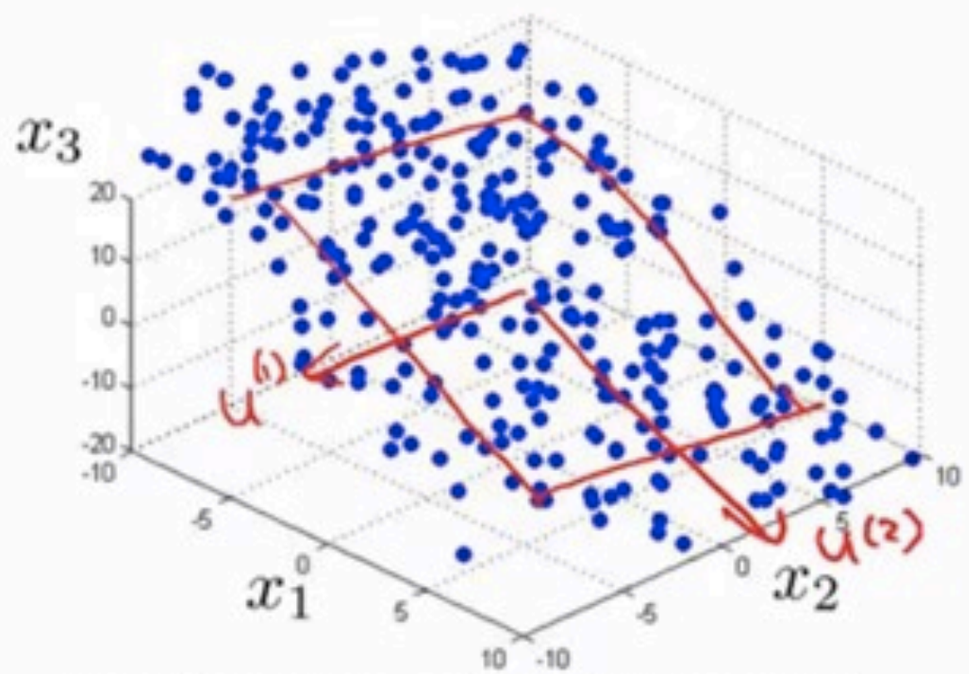
- Bir PCA analizinin ana amacı: verideki paternleri tanımlamaktır.
- PCA değişkenler arasındaki korelasyonu tespit etmeyi amaçlamaktadır.
- Değişkenler arasında güçlü bir korelasyon varsa, boyutsallığı azaltma mantıklıdır.
- Özetle, PCA'nın tamamı şu şekildedir: Yüksek boyutlu verilerde maksimum varyansın yönlerini bulmak ve bilgiyi korurken daha küçük boyutlu bir alt uzaya bunları yansıtmak.

Veri indirgeme:

Principal Component Analysis (PCA) algorithm



Reduce data from 2D to 1D



Reduce data from 3D to 2D

Kaynak: Machine Learning Lectures by Prof. Andrew NG at Stanford University

Temel Bileşen Analizi Adımları:

1. Verileri normalize et
2. Kovaryans hesabı yap
3. Özdeğer(Eigenvalue) ve Özvektörleri(Eigen Vector) Hesapla
4. Bileşenleri seç ve bir özellik vektörü haline dönüştür
5. Temel Bileşen haline getirme

2. Doğrusal Ayırteden Analizi:

- **Veri kümesinin herbir elemanın ortalama değerleri hesaplanır. Veri sayısı ile ortalama vektörü elde edilir.**

- **Veririnin ortak kovaryans matrisi hesaplanır. (Matrisin boyutu her bir veri vektörü için *d* kabul edilir ise *d* * *d* şeklinde olur.)**

- **Ortak kovaryans (değişinti) matrisi için öz vektörleri (E_1, E_2,…E_d) ve bunlara karşılık gelen öz değerleri $(\lambda_1, \lambda_2, \ldots\ldots, \lambda_d)$ hesaplanır.**

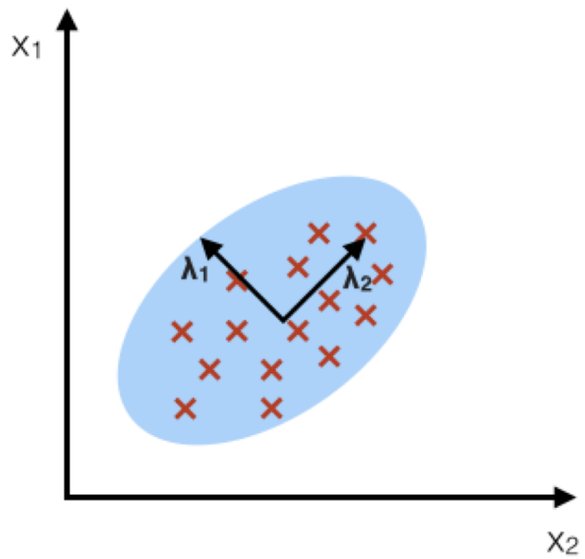
- Özdeğerlerin azalma sırasına karşılık gelen öz vektörler sıralanır.

- k adet büyük öz vektörü seçecek şekilde d*k boyutlu bir W matrisi ortak kovaryans matrisinden oluşturulur.

- Veri vektörlerini yeni alt uzaya dönüştürmek için W matrisi dönüşüm matrisi olarak kullanılır.

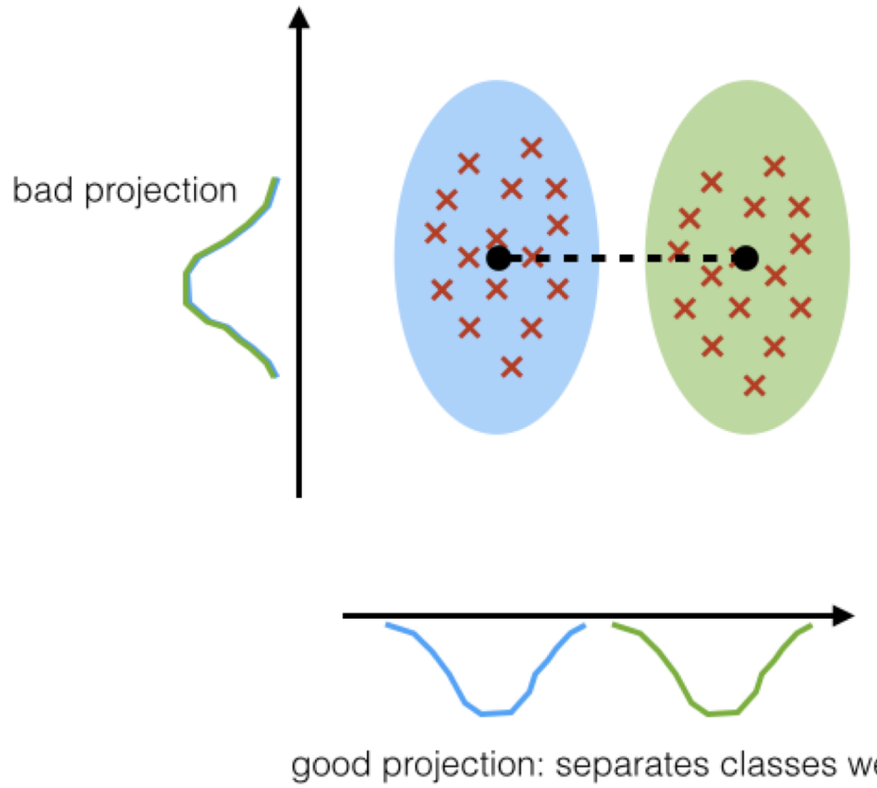
Karşılaştırma:

PCA:
component axes that maximize the variance



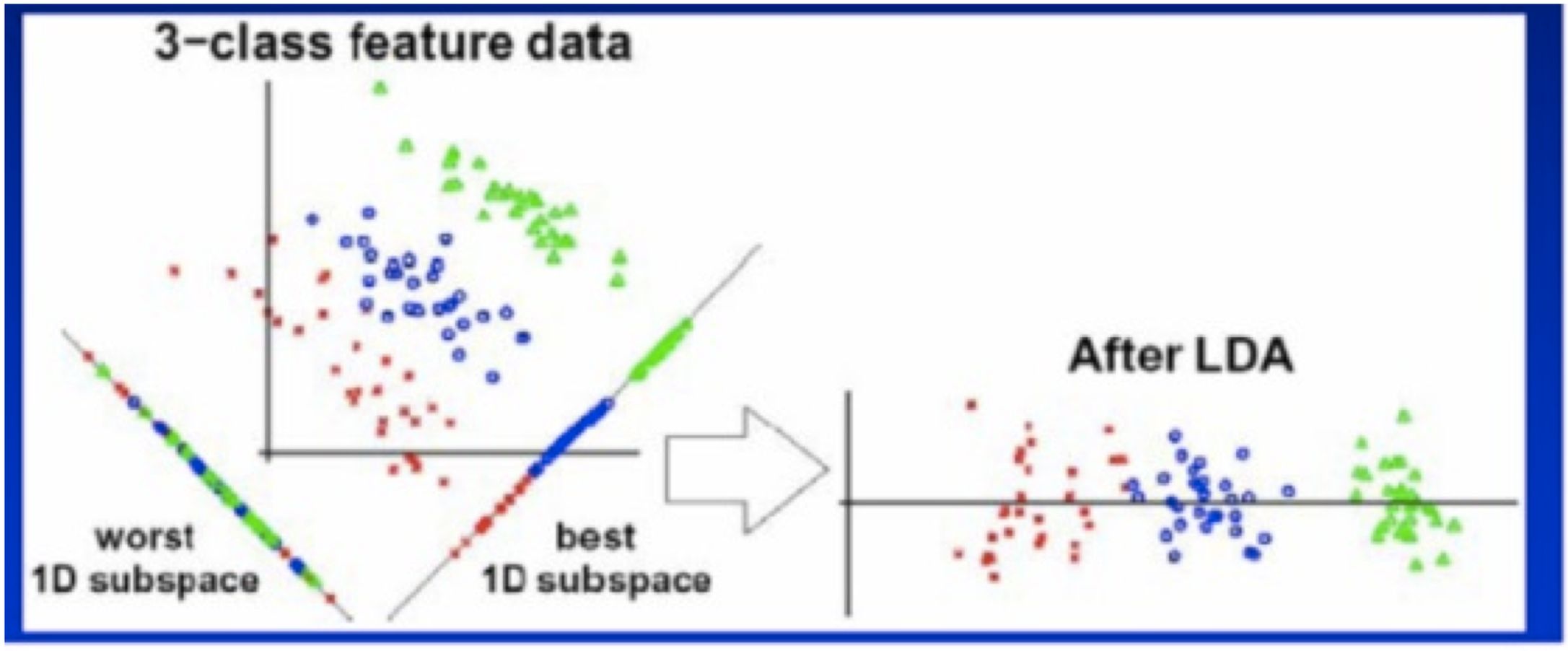
Temel Bileşen Analizi Varyansı maksimize eden bileşen ekseni

LDA:
maximizing the component axes for class-separation



Doğrusal Ayırteden Analiz
Sınıf ayrımı için bileşen eksenlerinin maksimize edilmesi

Üç sınıfa ait verilerin 1 boyuta indirgenmesi



Referans:

Makine Öğrenmesi

Papatya Yayıncılık

Dr. Metin Bilgin