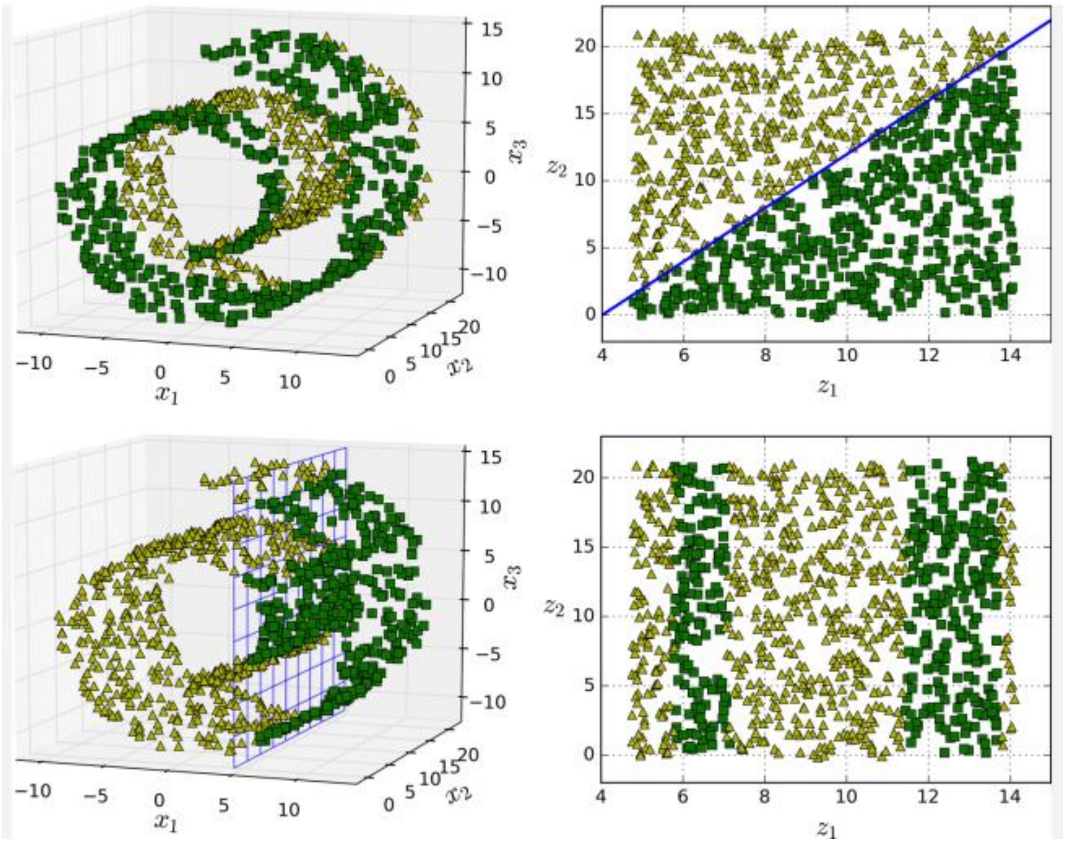
Örüntü Tanıma

Dr. Öğr. Üyesi Mehmet Zahid YILDIRIM

e-mail: m.zahidyildirim@karabuk.edu.tr

Boyut İndirgeme (Dimension Reduction)

- Boyut indirgeme, bir verinin yüksek boyutlu bir uzaydan, daha düşük boyutlu bir uzaya, anlamını kaybetmeden dönüştürülmesidir.
- Bir veri kümesinde bulunan çok sayıda öznitelik (boyut) içinden en önemli olanları seçilerek veya yeni öznitelikler çıkararak gerçekleştirilir.
- Boyut indirgeme yaklaşımları doğrusal ve doğrusal olmayan olarak ikiye ayrılır.



Boyut İndirgeme (Dimension Reduction)

Neden Önemlidir?

Hesaplama Maliyeti ve Performans İyileştirmesi:

Yüksek boyutlu veri setlerinin analizi, hesaplama maliyetini ve zamanı artırabilir. Boyut indirme, bu maliyeti azaltarak algoritmaların daha hızlı çalışmasına olanak tanır.

Görselleştirme Kolaylığı:

İki veya üç boyutlu uzayda daha rahat görselleştirme yapmak mümkündür. Boyut indirme, çok boyutlu veri setlerini daha anlamlı ve kullanıcı dostu hale getirir.

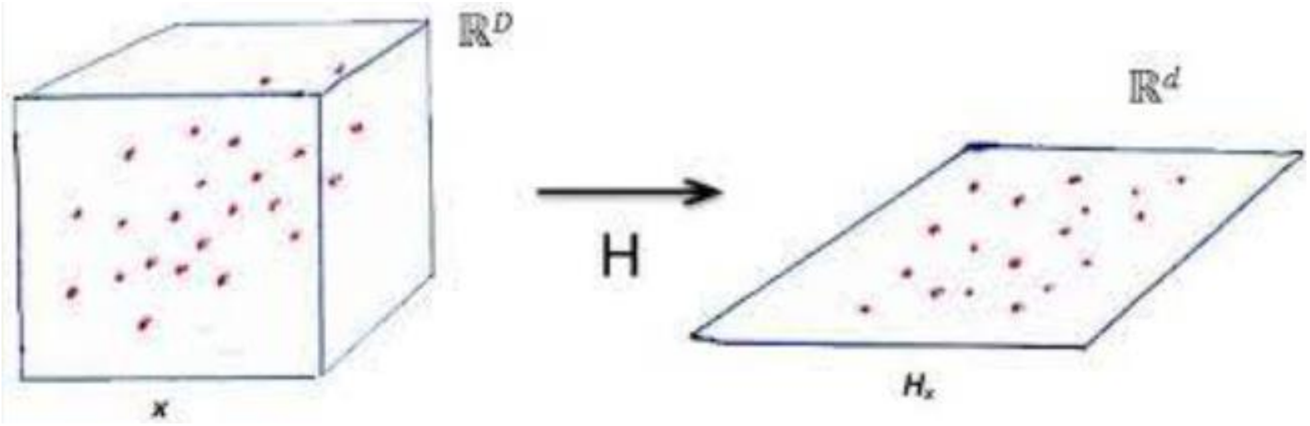
Veri Anlama ve Yorumlama Kolaylığı:

Daha az boyut, veri setinin daha anlaşılır ve yorumlanabilir olmasını sağlar. Bu, veri analistlerinin veya araştırmacıların veriyi daha iyi anlamalarına yardımcı olur.

Boyut İndirgeme (Dimension Reduction)

Boyut indirgeme, iki temel yöntem ile gerçekleştirilir.

- **Özellik Seçimi (Feature Selection)**
- **Özellik Çıkarımı (Feature Extraction)**

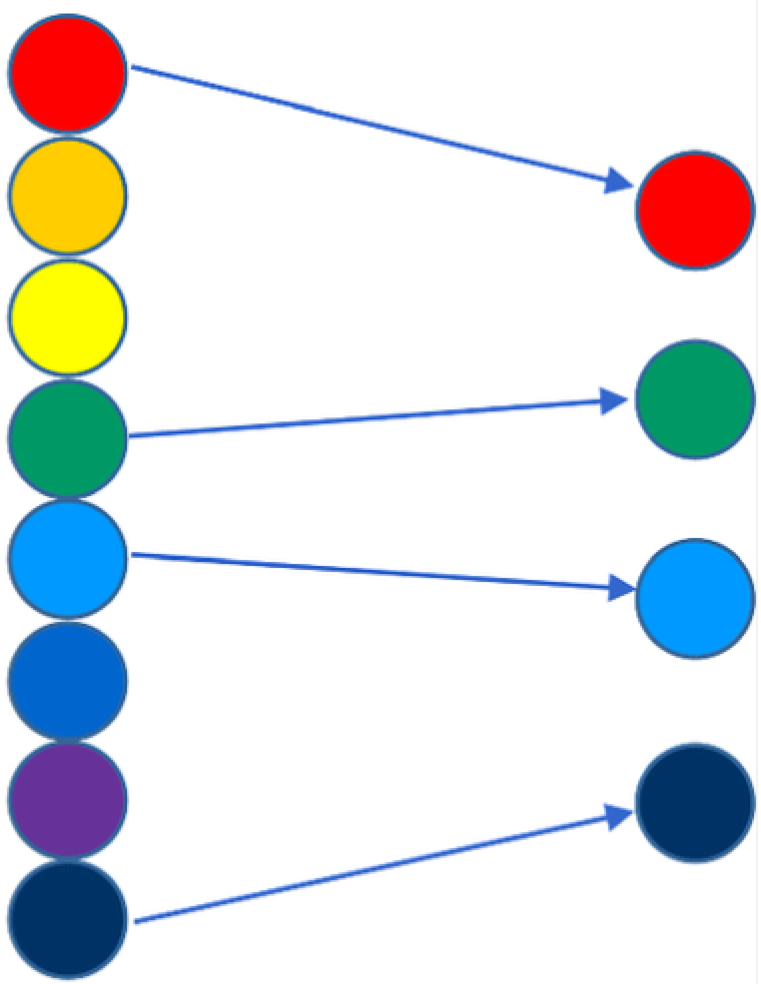


Boyut İndirgeme (Dimension Reduction)

Özellik Seçimi (Feature Selection) : Boyut küçültmenin en kolay yolu verimizi en iyi tanımlayan öznitelikleri bulup diğerlerini atmaktır. Örneğin; 30 boyuttan 5 tanesinin önemli olduğunu belirleyip kalan 25 özniteliği atmak.

Yöntemler;

- İnformasyon kazancı
- Korelasyon analizi
- Ağaç tabanlı yöntemler (Random Forest, Gradient Boosting)



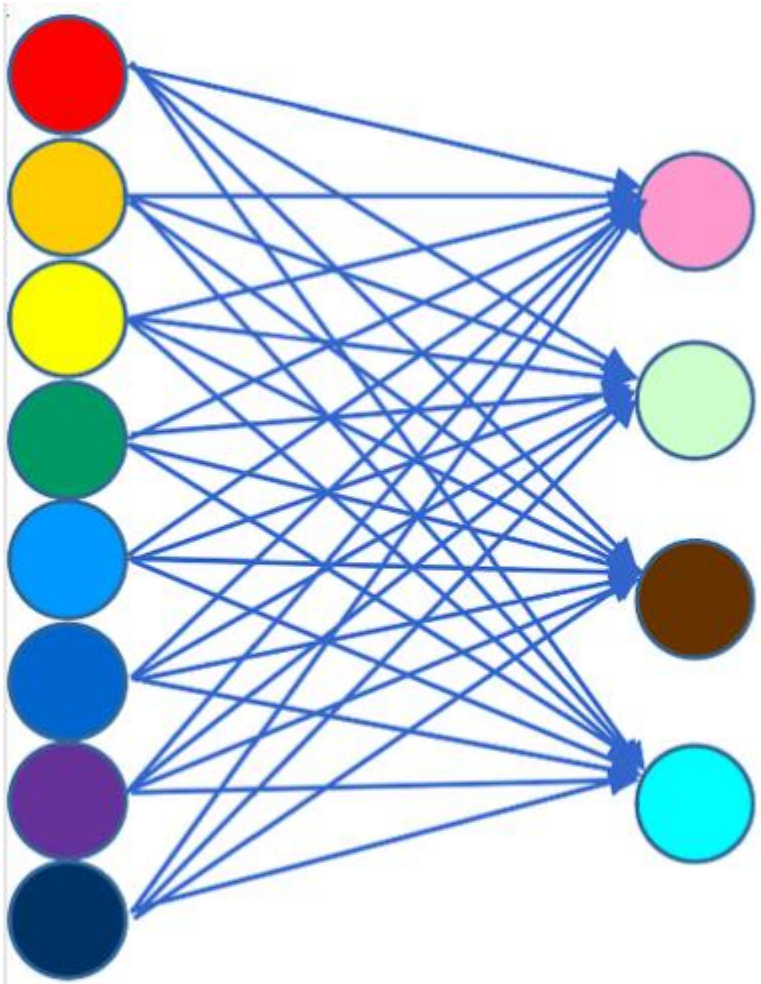
Boyut İndirgeme (Dimension Reduction)

Özellik Çıkarımı (Feature Extraction): Özellik çıkarımı, mevcut özellikleri kullanarak yeni ve daha az sayıda özellik üretme işlemidir.

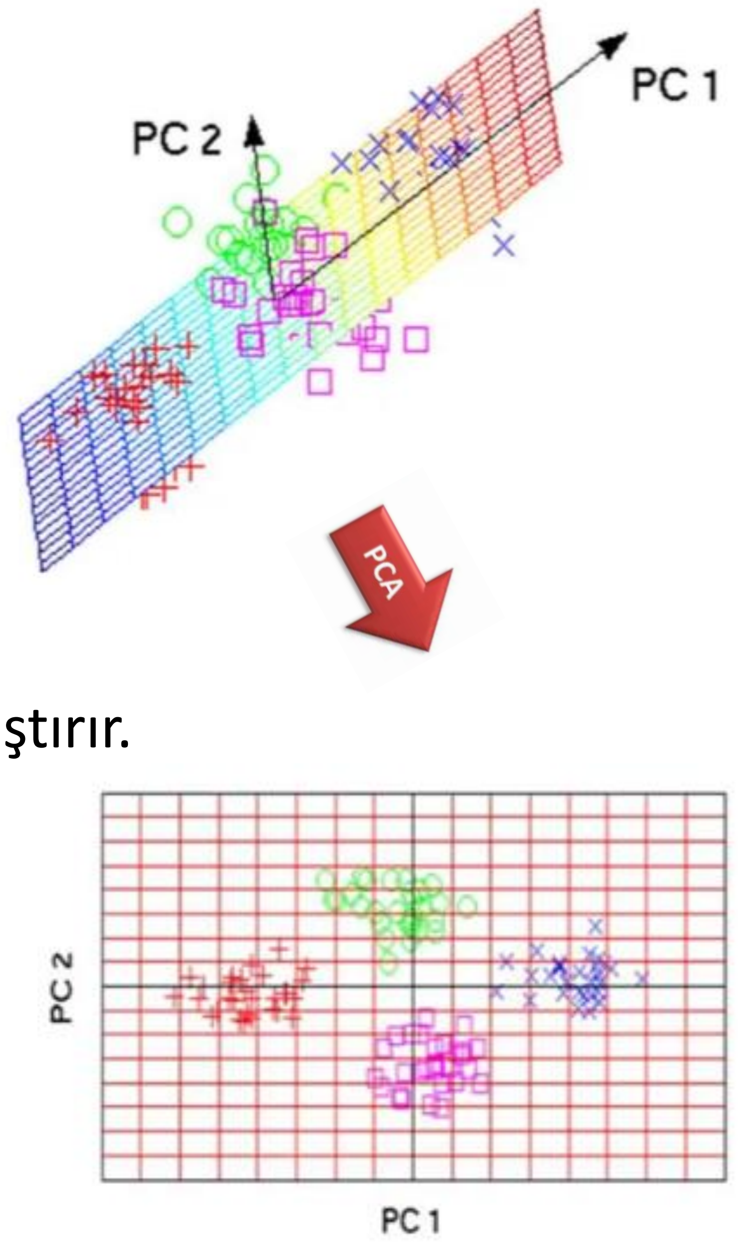
Bu, özellikle verilerin görselleştirilmesinde çok tercih edilen bir tekniktir.

Yöntemler;

- Temel Bileşen Analizi (Principal Component Analysis - PCA)
- Faktör Analizi (Factor Analysis)
- T-distributed Stochastic Neighbor Embedding (t-SNE)



Temel Bileşenler Analizi (PCA)



PCA nedir?

- Çok boyutlu verilerin boyutunu azaltan bir yöntemdir.
- Veriyi daha anlamlı ve görselleştirilebilir hale getirir.

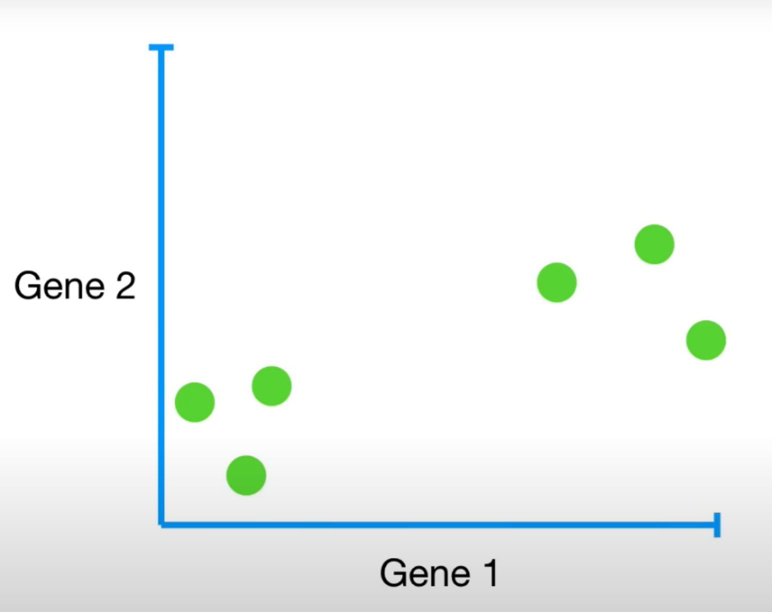
PCA neden önemlidir?

- Veri boyutunun yüksek olduğu durumlarda analizi kolaylaştırır.
- Gürültüyü azaltır ve bilgi kaybını minimize eder.
- Makine öğrenimi modellerinde performansı artırabilir.

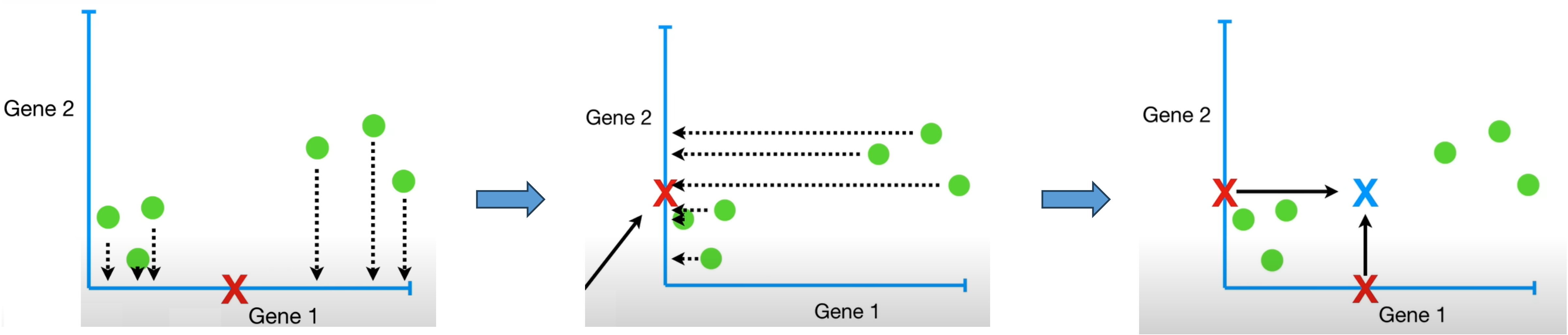
Temel Bileşenler Analizi (PCA)

PCA nasıl çalışır?

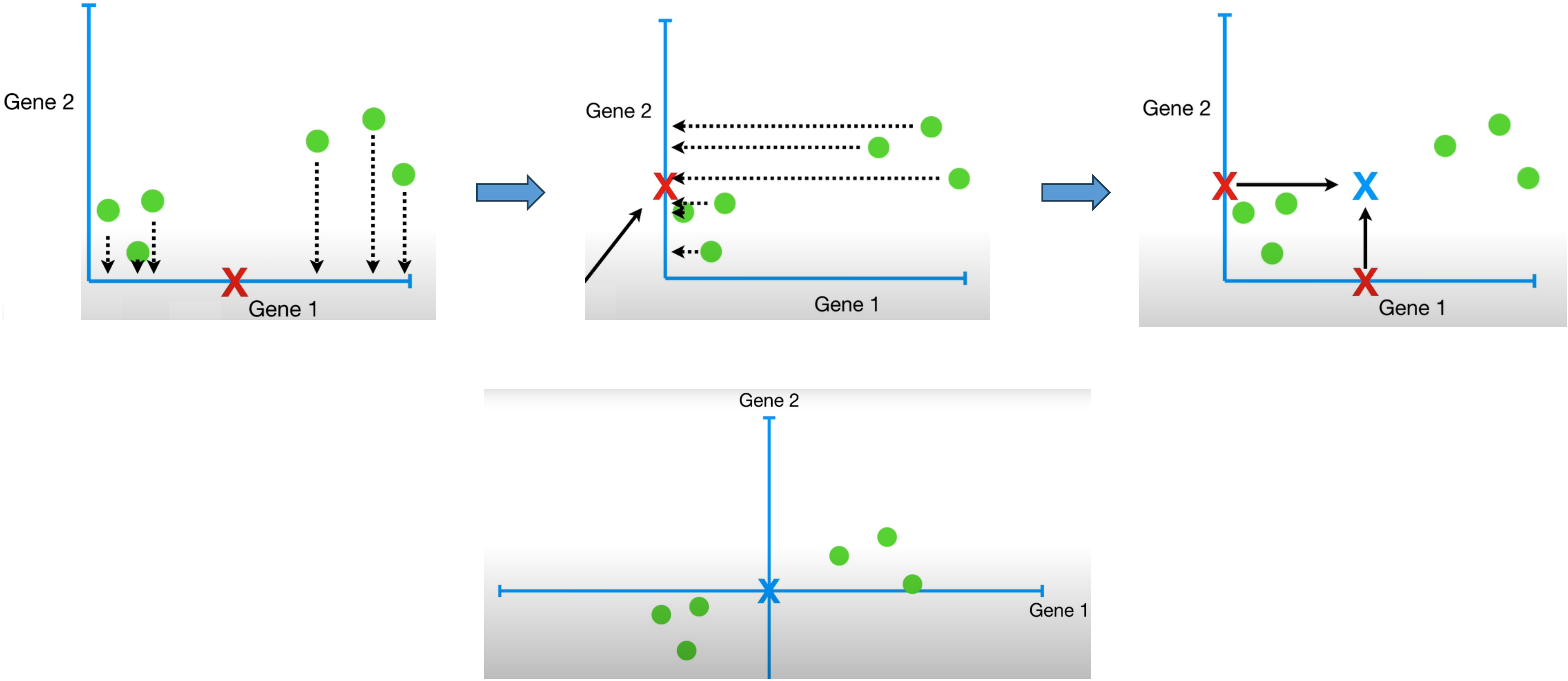
	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1



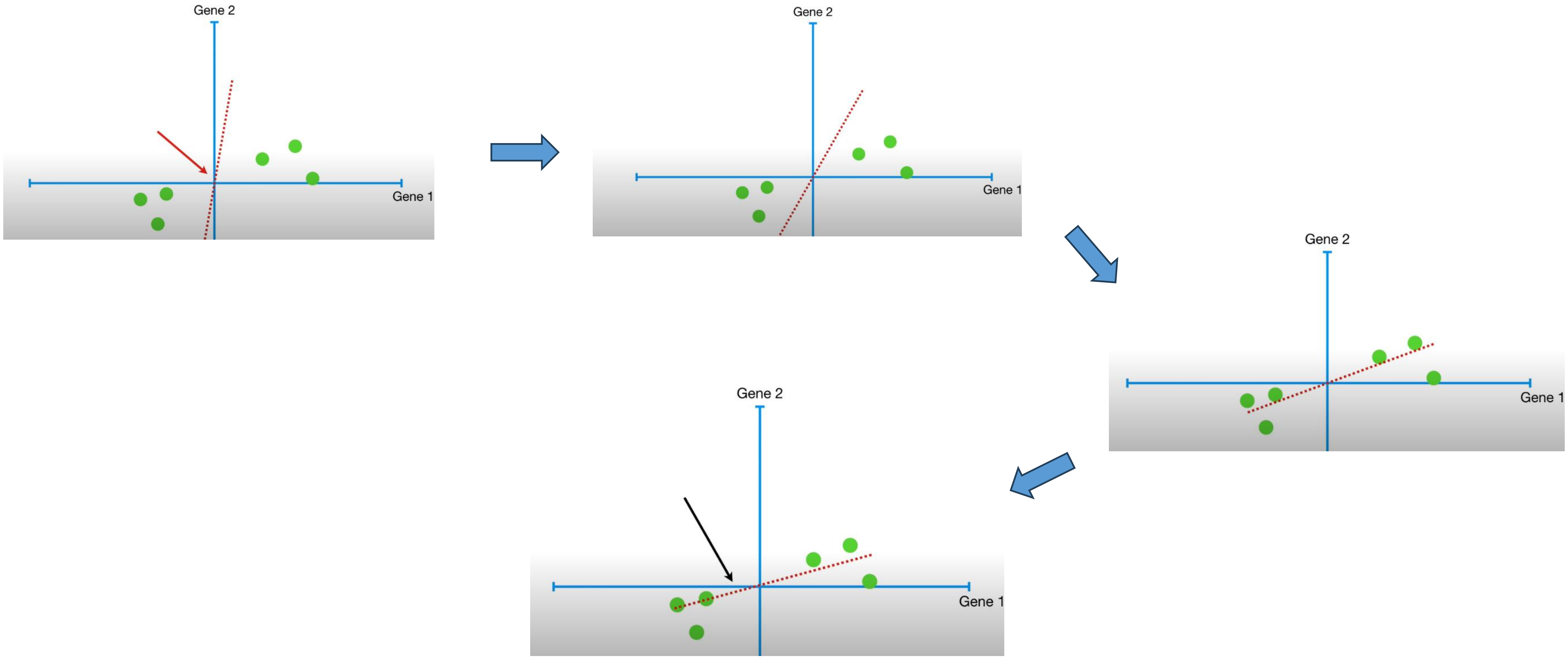
Temel Bileşenler Analizi (PCA)



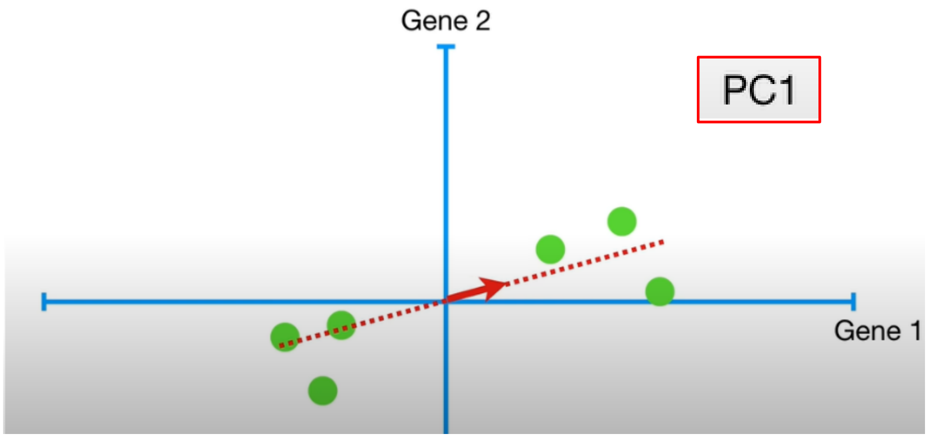
Temel Bileşenler Analizi (PCA)

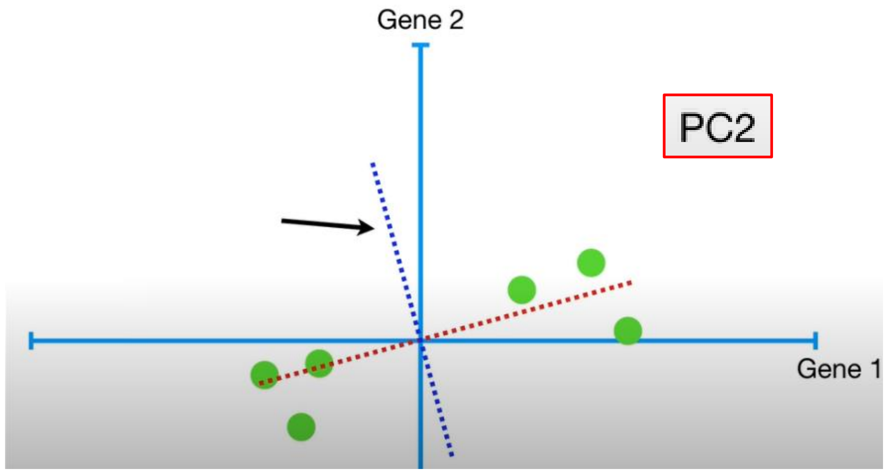


Temel Bileşenler Analizi (PCA)

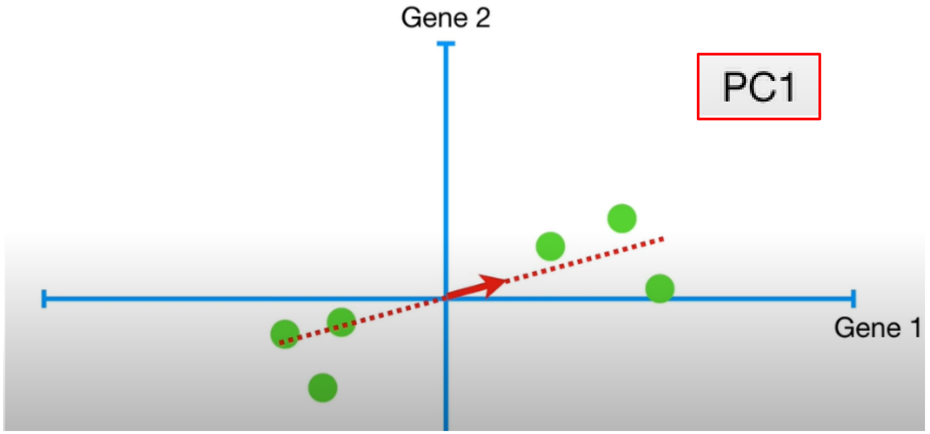


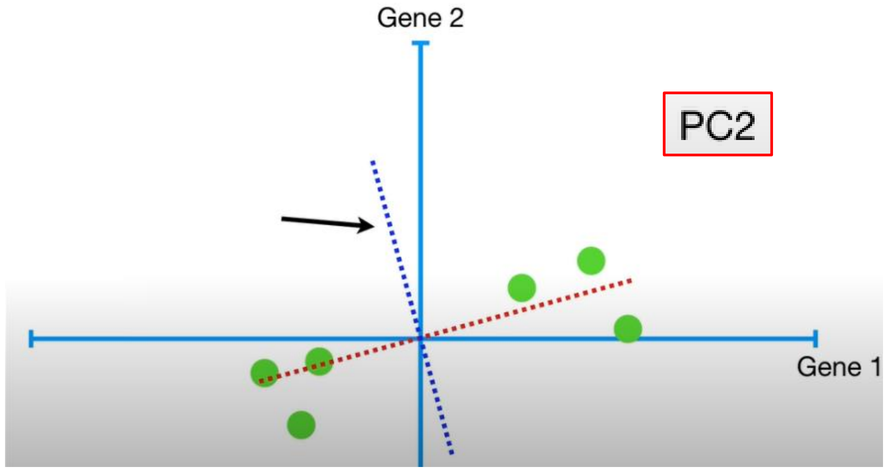
Temel Bileşenler Analizi (PCA)

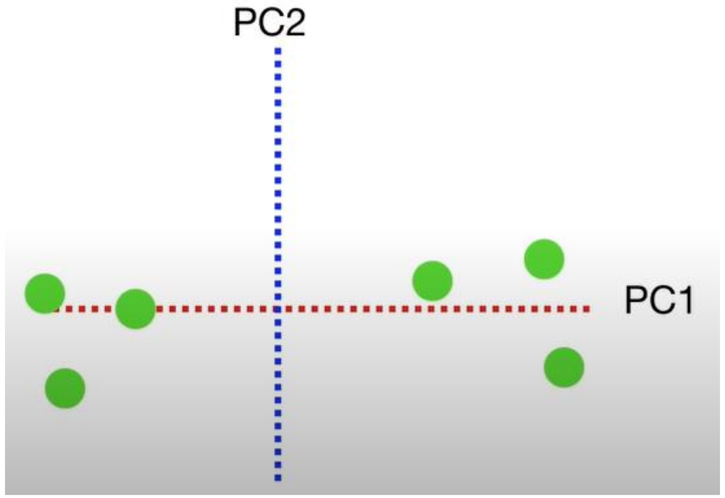




Temel Bileşenler Analizi (PCA)







Temel Bileşenler Analizi (PCA)

Matematiksel Model:

PCA genel olarak 5 temel adımdan oluşur:

- Verileri hazırlama
- Kovaryans/Korelasyon matrisini oluşturma
- Kovaryans/Korelasyon matrisinin özdeğerlerini ve özvektörlerini hesaplama
- Temel bileşenleri (Principal Components) seçme
- Yeni verisetini hesaplama

Temel Bileşenler Analizi (PCA)

Örnek:

Feature 1	Feature 2
4	2
6	3
13	5

Her bir özelliğin ortalamasını çıkaralım.

$$\mu_{\text{Feature 1}} = \frac{4+6+13}{3} = 7.67, \quad \mu_{\text{Feature 2}} = \frac{2+3+5}{3} = 3.33$$

Merkezlenmiş veri seti:

$$X_{\text{merkezlenmiş}} = X - \mu = \begin{bmatrix} 4 & 2 \\ 6 & 3 \\ 13 & 5 \end{bmatrix} - \begin{bmatrix} 7.67 & 3.33 \\ 7.67 & 3.33 \\ 7.67 & 3.33 \end{bmatrix} = \begin{bmatrix} -3.67 & -1.33 \\ -1.67 & -0.33 \\ 5.33 & 1.67 \end{bmatrix}$$

Temel Bileşenler Analizi (PCA)

Kovaryans matrisi şu şekilde hesaplanır:

$$\text{Kovaryans Matrisi} = \frac{1}{n-1}(X_{\text{merkezlenmiş}}^{\top} X_{\text{merkezlenmiş}})$$

1. **Transpoze ve çarpım**:

$$X_{\text{merkezlenmiş}}^{\top} = \begin{bmatrix} -3.67 & -1.67 & 5.33 \\ -1.33 & -0.33 & 1.67 \end{bmatrix}$$

$$X_{\text{merkezlenmiş}}^{\top} X_{\text{merkezlenmiş}} = \begin{bmatrix} -3.67 & -1.67 & 5.33 \\ -1.33 & -0.33 & 1.67 \end{bmatrix} \begin{bmatrix} -3.67 & -1.33 \\ -1.67 & -0.33 \\ 5.33 & 1.67 \end{bmatrix} = \begin{bmatrix} 47.33 & 14.33 \\ 14.33 & 2.33 \end{bmatrix}$$

2. **Kovaryans matrisi**:

$$\text{Kovaryans Matrisi} = \frac{1}{3-1} \cdot \begin{bmatrix} 47.33 & 14.33 \\ 14.33 & 2.33 \end{bmatrix} = \begin{bmatrix} 23.67 & 7.17 \\ 7.17 & 1.17 \end{bmatrix}$$

Temel Bileşenler Analizi (PCA)

Özdeğer denklemi:

$$\det(\text{Kovaryans Matrisi} - \lambda I) = 0$$

$$\det \begin{bmatrix} 23.67 - \lambda & 7.17 \\ 7.17 & 1.17 - \lambda \end{bmatrix} = 0$$

$$(23.67 - \lambda)(1.17 - \lambda) - (7.17)^2 = 0$$

Bu denklemi çözelim:

$$\lambda^2 - 24.84\lambda + 16.32 = 0$$

Çözüm:

$$\lambda_1 = 24.71, \quad \lambda_2 = 0.13$$

Temel Bileşenler Analizi (PCA)

Özvektörler:

1. $\lambda_1 = 24.71$ için:

$$\begin{bmatrix} 23.67 - 24.71 & 7.17 \\ 7.17 & 1.17 - 24.71 \end{bmatrix} v = 0$$

Bu çözümden:

$$v_1 = \begin{bmatrix} 0.97 \\ 0.26 \end{bmatrix}$$

2. $\lambda_2 = 0.13$ için:

$$v_2 = \begin{bmatrix} -0.26 \\ 0.97 \end{bmatrix}$$

Temel Bileşenler Analizi (PCA)

Özvektörleri kullanarak veriyi yeni uzaya dönüştürelim:

$$X_{\text{yeni}} = X_{\text{merkezlenmiş}} \cdot [v_1 \; v_2]$$

$$X_{\text{yeni}} = \begin{bmatrix} -3.67 & -1.33 \\ -1.67 & -0.33 \\ 5.33 & 1.67 \end{bmatrix} \begin{bmatrix} 0.97 & -0.26 \\ 0.26 & 0.97 \end{bmatrix}$$

$$X_{\text{yeni}} = \begin{bmatrix} -3.81 & 0.35 \\ -1.74 & 0.02 \\ 5.55 & -0.38 \end{bmatrix}$$

Yeni özellik uzayı:

$$X_{\text{PCA}} = \begin{bmatrix} -3.81 & 0.35 \\ -1.74 & 0.02 \\ 5.55 & -0.38 \end{bmatrix}$$

Bu, PCA ile indirgenmiş veri setidir. İlk bileşen (v_1) en fazla varyansı açıklayan bileşendir.

Temel Bileşenler Analizi (PCA)

En büyük varyansı açıklayan özvektör (v_1) yönünde veri setini tek boyuta indiriyoruz.

Veri setini v_1 doğrultusunda projekte edelim:

$$X' = \begin{bmatrix} 4 & 2 \\ 6 & 3 \\ 13 & 5 \end{bmatrix} \begin{bmatrix} 0.97 \\ 0.26 \end{bmatrix} = \begin{bmatrix} 3.09 \\ 4.05 \\ 9.44 \end{bmatrix}$$