

The background is a light gray gradient with several realistic water droplets of various sizes scattered across it. The droplets have highlights and shadows, giving them a three-dimensional appearance. The text is centered and rendered in a bold, purple, sans-serif font with a slight drop shadow.

BÖLÜM 13

DOĞRUSAL BAĞIMLILIK

GİRİŞ

- Regresyonda doğrusal bağımlılık sorunu, bağımsız değişkenlerin en az bir doğrusal fonksiyonu sıfıra çok yakın olduğunda ortaya çıkar. (Teknik olarak, doğrusal fonksiyonda aynı doğrultuda olan vektörlerin kümesi tam olarak sıfırdır. Doğrusal bağımlılık sorunu üzerine genel tartışmalarda, “aynı doğrultu” terimi çoğu kez sadece yaklaşık olarak sıfır olan doğrusal fonksiyonların uygulanmasında kullanılır. Bu metinde bu kural takip edilecektir.)
- Bu yakın-tekillik birçok yolla ortaya çıkarılabilir.

TABLO 13.1. Doğrusal bağımlılığı gösteren yapay veri seti için bağımsız değişkenlerin korelasyon matrisi.

	X_1	X_2	X_3
X_1	1.000	.996	.290
X_2	.996	1.000	.342
X_3	.290	.342	1.000

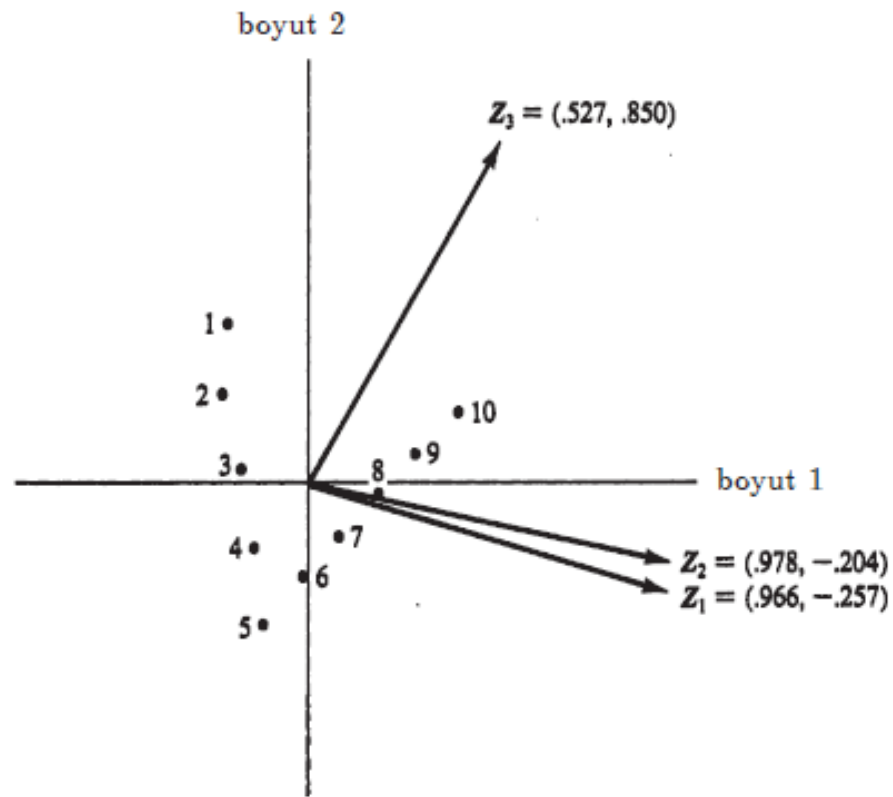
X-UZAYININ YAPISINI ANLAMA

- Her bir deęişkenin kareler toplamı bir olsun diye ölçeklendirilen merkezileştirilmiş ve ölçeklendirilmiş bağımsız deęişkenlerin çarpımlar ve kareler toplamlar $Z'Z$ matrisi, X-uzayının yapısını anlamak için iyi bir başlangıç noktasıdır. (Eđer bağımsız deęişkenler rassal deęişkenler ise bu korelasyon matrisidir ve kolaylık sağlaması açısından X ler sabit olsalar bile korelasyon matrisi olarak adlandırılacaktır)
- Matrisin köşegendişı elemanları X-uzayında ilgili merkezileştirilmiş ve ölçeklendirilmiş vektörler arasındaki açıların kosinüsleridir.
- 1.0 ve -1.0'e yakın deęerler neredeyse aynı doęrultudaki vektörleri gösterir, yaklaşık 0 deęerleri neredeyse dik vektörleri gösterir.

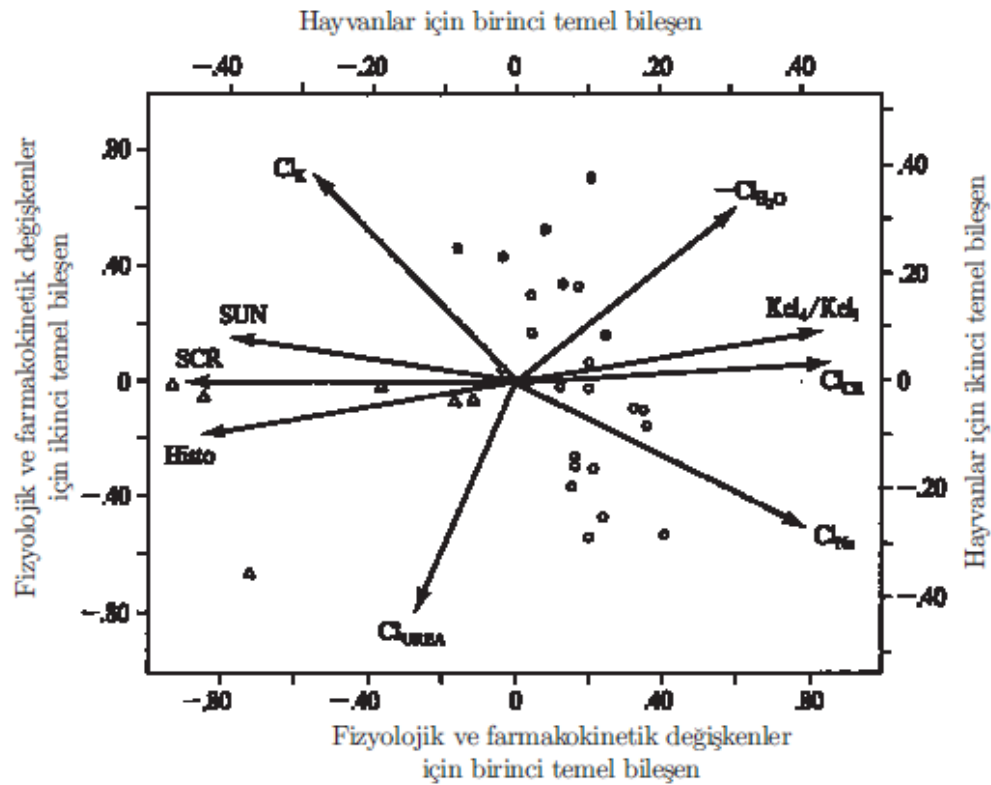
TABLO 13.2. Yapay veri seti için bağımsız değişkenlerin korelasyon matrisinin özdeğerleri ve özvektörleri.

Özdeğer	Özvektörler
$\lambda_1=2.166698$	$v'_1 = (.65594 \quad .66455 \quad .35793)$
$\lambda_2 = .830118$	$v'_2 = (-.28245 \quad -.22365 \quad .93285)$
$\lambda_3 = .002898$	$v'_3 = (.69998 \quad -.71299 \quad .04100)$

Z'lerin doğrusal fonksiyonlarını tanımlar ve en küçük ayrılığı gösterir. Bu özel doğrusal fonksiyonlar eğer biri mevcut ise doğrusal bağımlılık problemine neden olur.



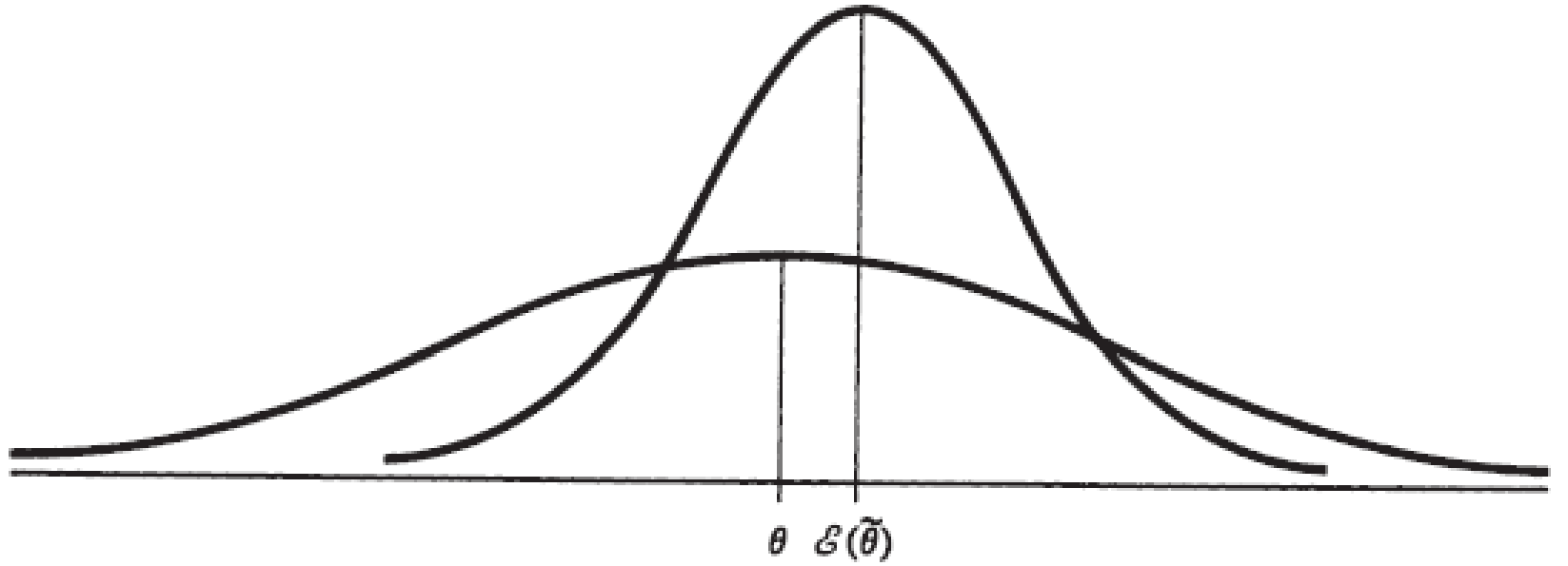
ŞEKİL 13.1. Örnek 13.3. için ilk iki temel bileşen boyutlarının Gabriel ikili grafiği



ŞEKİL 13.2. Dönüştürülmüş fizyolojik verinin ikili grafik, Kel oranı ve böbreği tamamen alınmamış 23 ve 6 adet kontrol hayvanı için histopatolojik indeks. İlk iki boyut tüm matris içindeki yayılımın % 76'ını açıklamaktadır. Üçgenler toksisite gelişen köpekleri, açık daireler böbrek ameliyatı (nefrektomi) olan fakat toksisite gelişmemiş köpekleri ve kapalı daireler kontrol hayvanlarını temsil etmektedir. (İzin ile kullanıldı.)

YANLI REGRESYON

- Regresyon katsayılarının en küçük kareler tahmin edicileri en iyi doğrusal yansız tahmincilerdir.
- Hem verinin doğrusal fonksiyonu olan hem de tahmin edilen parametre için yansız olası tüm tahmin ediciler içinde en küçük kareler tahmin edicileri en küçük varyansa sahiptir.
- Fakat doğrusal bağımlılığın varlığı durumunda bu en küçük varyans kabul edilemez büyüklükte olabilir.
- En küçük kareler tahmin edicinin yansız olması koşulunda bir gevşeklik sağlarsak doğrusal bağımlılık varlığında mümkün tahmin edicilerin daha geniş seti dikkate alındığında daha iyi özellikleri sağlayan biri bulunabilir.



ŞEKİL 13.3. Yanlı tahmin edicinin yansız tahmin ediciye göre daha küçük hata kareler ortalamasına sahip olduğunun gösterimi

TEMEL BİLEŞEN REGRESYONU

- Temel bileşen regresyonu, doğrusal bağımlılık problemine neden olan X -uzayının boyutları fikrinden yok etme bakış açısı ile doğrusal bağımlılık problemine yaklaşır.
- Bu, kavramsal olarak, modelden 'ye anlamlı bilgi katmada bağımsız değişkenin yetersiz yayılıma sahip olduğunda dışlanması ile benzerdir.
- Fakat temel bileşen regresyonunda boyut indirgeme fikri tek bir bağımsız değişkenden çok değişkenlerin doğrusal kombinasyonu ile tanımlanır.

TABLO 13.1. *Temel bileşenler için regresyon katsayılarının tahminleri, bunların standart hataları ve her biri için nitelenebilir kareler toplamı*

<i>Temel</i>	<i>Regresyon</i>		
<i>Bileşen</i>	<i>Katsayıları</i>	<i>Standart</i>	<i>Kareler</i>
<i>j</i>	$\hat{\gamma}_j$	<i>Hatalar</i>	<i>Toplamı^a</i>
1	2.3473	.598	11.940**
2	3.0491	.967	7.718**
3	19.7132	16.361	1.126

TABLO 13.4. Tüm temel bileşenleri kullanarak ($g = 3$) ve üçüncü temel bileşen ($g = 2$) ihmal edilerek regresyon katsayılarının temel bileşen regresyon tahminleri.

Ölçekli Değişken Z_j	g Temel Bileşen Kullanılmış İken Regresyon Katsayıları		Hata Kareler Ortalaması
	$g = 3$	$g = 2$	
1	14.480 (11.464)	.678 (0.478)	14.83
2	-13.182 (11.676)	0.878 (0.452)	15.34
3	4.493 (1.144)	3.685 (0.927)	1.16

DOĐRUSAL BAĐIMLILIK ÜZERİNE GENEL YORUMLAR

Dođrusal bađımlılıđın varlıđında izlenen yol dođrusal bađımlılıđın kaynađı ve yapısı ve regresyon analizinin amacına bađlıdır. Eđer regresyon analizi sadece bađımlı deđiřkenin kestirimini amaçlarsa, verideki yakıntekilliliđin varlıđı belli çok önemli kořullar olduđu sürece ciddi problemlere neden olmaz:

1. Veride görünen dođrusal bađımlılık -uzayının iliřkili yapısının yansımasıdır. Bu örnekleme sürecinde insandan ve verideki uç noktalardan kaynaklanmak zorunda deđildir. [Mason ve Gunst (1985) uç noktalarla uyarılmıř dođrusal bađımlılıđın tespiti ve etkilerini tartıřmıřlardır.
2. X-uzayının korelasyon yapısının tutarlı kalması için sistem veri üretildiđi zamanki gibi aynı şekilde çalıřmaya devam eder. Bu, kestirilen noktalar örnek -uzayında olsa bile regresyon eřitliđinin bazı sistem deđiřikliklerine tepkiyi kestirimde kullanılmadıđını iřaret eder.