

TANI TESTLERİNİN BİRLEŞTİRİLMESİ ve EN UYGUN TEST KOMBİNASYONUNUN BULUNMASI



TURCOSA

Yrd.Doç.Dr.Gökmen ZARARSIZ

Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri
Turcosa Analitik Çözümler Ltd Şti, Kayseri

gokmenzararsiz@hotmail.com

14-15 Nisan 2018, Gazi Üniversitesi, ANKARA

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

SUNUM İÇERİĞİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

- Tanı Testleri
- Tanı Testlerinin Birleştirilmesi
- Ve & Veya Kuralları
 - Ve Kuralı
 - Veya Kuralı
- Seri ve Paralel Yöntemler
 - Seri Yöntem
 - Paralel Yöntem
 - Birleşik Testin Performansı
- Yeni Değişkenlerin Oluşturulması
- Çok Değişkenli Yaklaşımlar
 - Lojistik Regresyon
 - Doğrusal Diskriminant Analizi
 - İleri İstatistiksel Yöntemler
 - Birleşik Testin Performansı
- Kaynaklar



TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

TANI TESTLERİ

- Tıpta tanı koymak üzere kullanılan ve yararlanılan testlere “tanı testleri” adı verilir.
- Sonucu kesin olmayan testlerdir.
- %100 doğru sonuç veren altın standart testler referans alınarak, performans ölçüleri (duyarlılık, seçicilik, vs.) yardımıyla tanı testlerinin performansları belirlenir ve birbirleriyle karşılaştırılırlar.
- Geçerliliği ve kullanımı açısından altın standart teste göre az maliyet, az risk veya kolay uygulama gibi avantajları olmalıdır.

TANI TESTLERİNİN BİRLEŞTİRİLMESİ

- Tanı testlerinin yalnızca bir kısmı hem yüksek duyarlılık, hem de yüksek seçicilik performansına sahiptir.
- Bu nedenle hastalar bazen iki ya da daha fazla teste tabi tutulurlar.
- Testleri uygun biçimde birleştirmek, ayrı ayrı uygulamaya göre daha iyi tanısal sonuçlar verebilir.
- Bu testleri birleştirmede “trade-off” adı verilen “bir şeyi kazanmak için başka bir şeyden fedakarlık etme” durumu vardır.
- “Trade-off” genelde birleşik testin duyarlılık ve seçicilik ölçümleri için kullanılır.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

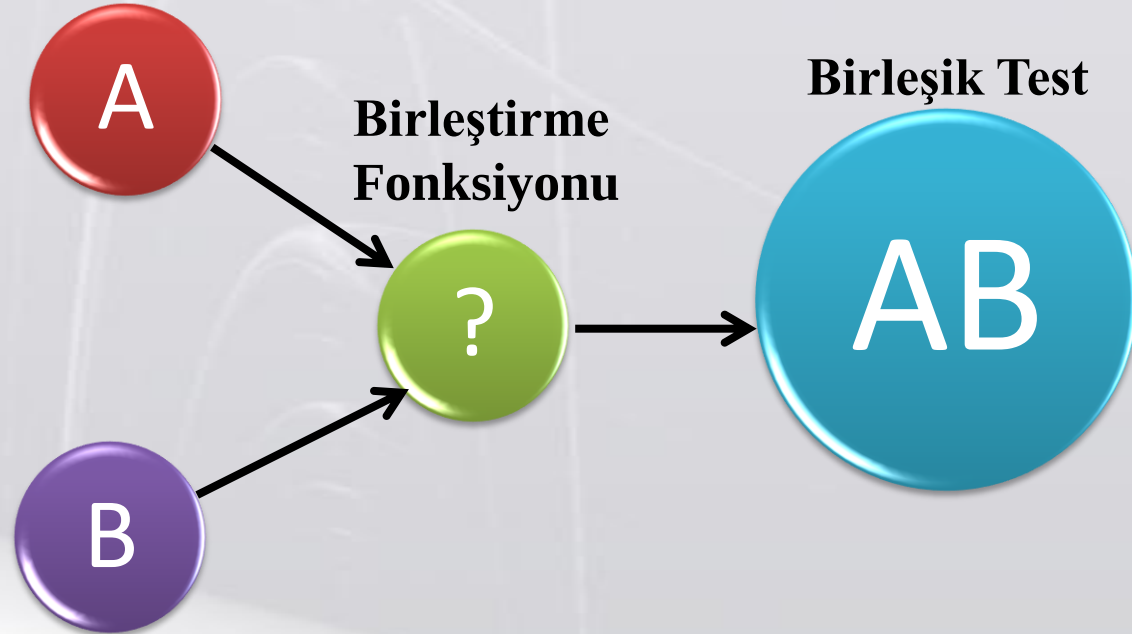
ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

TANI TESTLERİNİN BİRLEŞTİRİLMESİ

- Tanı testlerinin bileşkeleri ile iyi tanısal sonuçlar elde etmek için, uygun birleştirme fonksiyonun seçimi büyük önem taşımaktadır.

Test Bileşenleri



VE KURALI

- “Ve kuralı”na göre, iki önermenin en az biri yanlış ise, birleşik önerme de yanlıştır.

p	q	$p \wedge q$
0	0	0
0	1	0
1	0	0
1	1	1

p: Ankara Türkiye’nin başkentidir (Doğru: $p=1$).

q: 8 asal bir sayıdır (Yanlış: $p=0$).

$p \wedge q$: Ankara Türkiye’nin başkentidir ve 8 asal bir sayıdır (Yanlış: $p \wedge q=0$).

$$p \wedge q = 1 \wedge 0 = 0$$



A ve B olayları bağımsız ise:

$$P(A \cap B) = P(A) \cdot P(B)$$

VE KURALI

- Tanı testlerinin birleştirilmesinde, iki testten en az biri negatif sonuç verdiğinde, birleşik testin sonucu da negatiftir.
- Kişiye pozitif tanı koymanın tek koşulu iki testin de pozitif sonuç vermesi koşuludur.
- İki test birleştirilerek oluşturulan birleşik testin **seçiciliğini arttırarak** kişilerin hastalıklarının kanıtlanması amaçlanır.

A testi	B testi	Birleşik AB Testi
-	-	-
-	+	-
+	-	-
+	+	+

VEYA KURALI

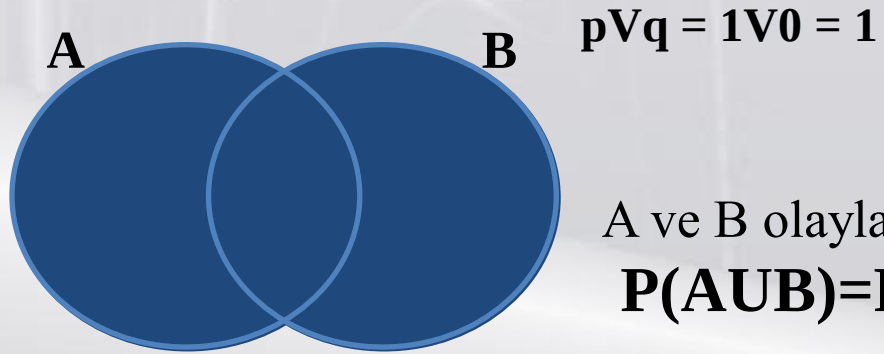
- “Veya kuralı”na göre, iki önermenin en az biri doğru ise, birleşik önerme de doğrudur.

p	q	$p \vee q$
0	0	0
0	1	1
1	0	1
1	1	1

p: Ankara Türkiye’nin başkentidir (Doğru: $p=1$).

q: 8 asal bir sayıdır (Yanlış: $p=0$).

$p \vee q$: Ankara Türkiye’nin başkentidir veya 8 asal bir sayıdır (Doğru: $p \vee q=1$).



A ve B olayları bağımsız ise:

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$$

VEYA KURALI

- Tanı testlerinin birleştirilmesinde, iki testten en az biri pozitif sonuç verdiğinde, birleşik testin sonucu da pozitifdir.
- Kişiye negatif tanı koymanın tek koşulu iki testin de negatif sonuç vermesi koşuludur.
- İki test birleştirilerek oluşturulan birleşik testin **duyarlılığını arttırarak** kişilerin sağlıklılığının kanıtlanması amaçlanır.

A testi	B testi	BirleşikAB Testi
-	-	-
-	+	+
+	-	+
+	+	+

SERİ ve PARALEL YÖNTEMLER

- Tanı testleri genelde iki sonuçludur (“H+, H-”, “riskli, risksiz”, “inflamasyon var, inflamasyon yok”, ...vs.)
- Hastalıklara ilişkin tanıların belirlenmesi için uygulanan birden fazla test “paralel” veya “seri” olarak uygulanır.
- **Paralel testler**, birden fazla tanı testinin aynı anda uygulandığı ve sonuçlarının eş zamanlı değerlendirildiği testlerdir.
- **Seri testler**, birinci testin sonuçlarına göre ikinci testin uygulanıp uygulanmayacağını karar verildiği testlerdir.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

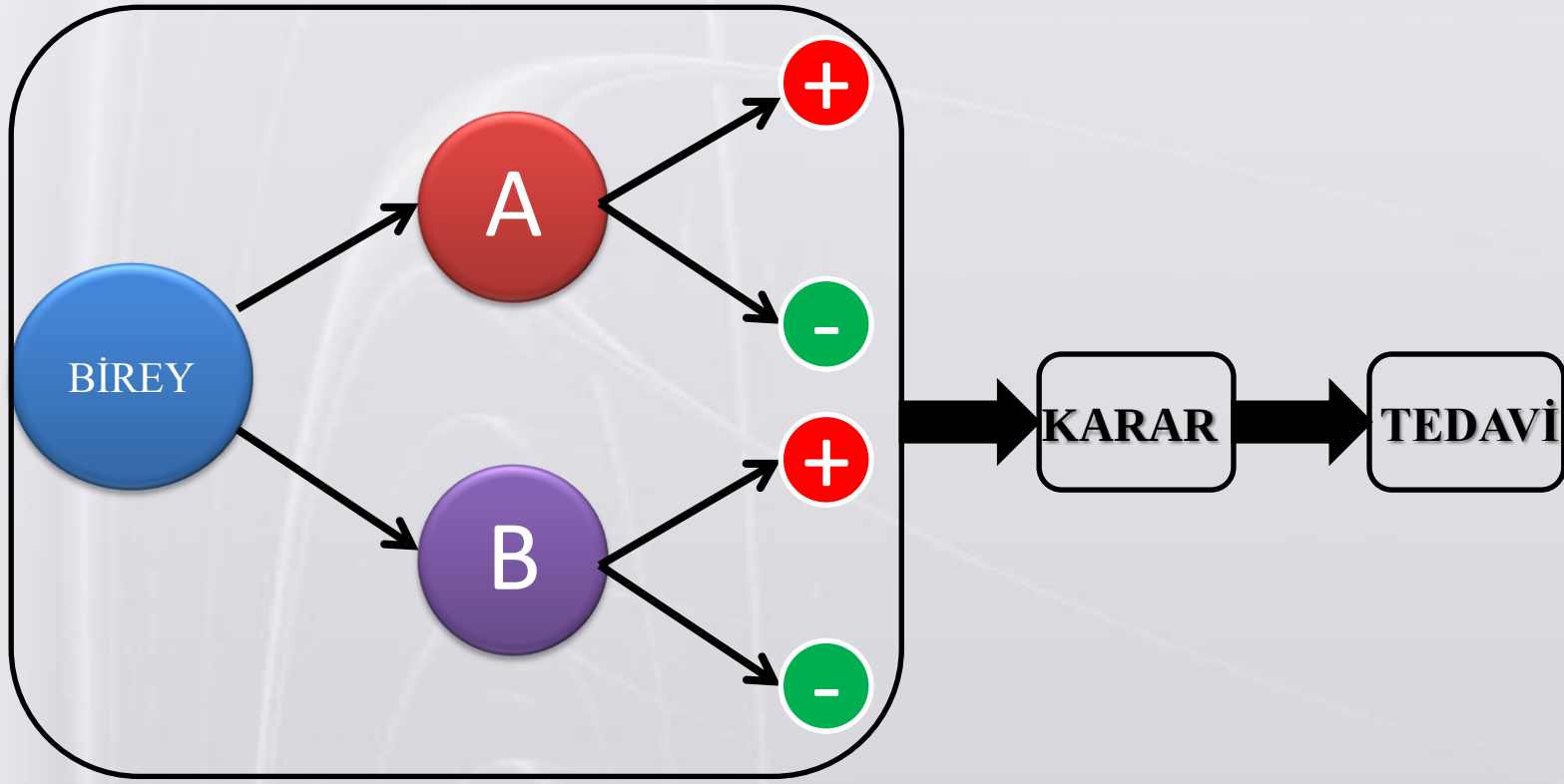
YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



PARALEL TESTLER



- **Avantajı**, hasta tanısı konmuş bireyler için hemen tedavinin başlatılabilmesidir. Çünkü, testler eş zamanlı yapılır ve bireyler ikinci testin sonucu için beklemek zorunda kalmazlar
- **Dezavantajı**, yüksek maliyetli olmasıdır. Çünkü, bireylerin tümü, tüm testlere tabi tutulur.

PARALEL TESTLER

- Paralel testlerde “ve kuralı” uygulanarak birleşik testin seçiciliği artırılabilir.
- Her iki test sonucu pozitif olan hastalar, hasta kabul edilir.
- Testlerden birinin dahi negatif çıkması sonucu birey sağlam kabul edilir.
- Birleşik testin duyarlılık ve seçicilikleri aşağıdaki formüller ile hesaplanır.

$$DUY_{AB}=DUY_A \cdot DUY_B$$

$$SEÇ_{AB}=SEÇ_A+SEÇ_B-(SEÇ_A \cdot SEÇ_B)$$

$$DUY_{AB} < DUY_A, \quad DUY_{AB} < DUY_B, \quad SEÇ_{AB} > SEÇ_A, \quad SEÇ_{AB} > SEÇ_B$$



PARALEL TESTLER

- Paralel testlerde “veya kuralı” uygulanarak birleşik testin duyarlılığı artırılabilir.
- Testlerden birinin sonucu pozitif olan hastalar, hasta kabul edilir.
- Testlerden ikisinin de sonucunun negatif çıkması sonucu birey sağlam kabul edilir.
- Birleşik testin duyarlılık ve seçicilikleri aşağıdaki formüller ile hesaplanır.

$$DUY_{AB}=DUY_A+DUY_B-(DUY_A.DUY_B)$$

$$SEC_{AB}=SEC_A.SEC_B$$

$$DUY_{AB}>DUY_A, \quad DUY_{AB}>DUY_B, \quad SEC_{AB}<SEC_A, \quad SEC_{AB}<SEC_B$$

PARALEL TESTLER

Örnek:

İnvaziv mantar enfeksiyonunu belirlemede hücre kültürünün altın standart olarak kullanıldığı bilinmekte ve β -D-Glukan ile PCR testleri paralel olarak uygulanarak bu hastalığı belirlemede hücre kültürüne alternatif olup olmayacağı araştırılmaktadır. Test sonuçları aşağıda verilmiştir.

β -D-Glukan	Hücre Kültürü		
	+	-	Toplam
+	13	14	27
-	7	16	23
Toplam	20	30	50

$$DUY_{\beta} = 13/20 = 0,65$$

$$SEÇ_{\beta} = 16/30 = 0,53$$

PCR	Hücre Kültürü		
	+	-	Toplam
+	9	11	20
-	11	19	30
Toplam	20	30	50

$$DUY_{PCR} = 9/20 = 0,45$$

$$SEÇ_{PCR} = 19/30 = 0,63$$

PARALEL TESTLER

Bu iki testin sonuçlarını birleştirip birleşik bir test geliştirmek istediğimizde “ve-veya kuralları”ndan faydalanılabilir.

“Ve kuralı”nı uyguladığımızda;

$$\begin{aligned} DUY_{\beta-PCR} &= DUY_{\beta} \cdot DUY_{PCR} \\ SEÇ_{\beta-PCR} &= SEÇ_{\beta} + SEÇ_{PCR} - (SEÇ_{\beta} \cdot SEÇ_{PCR}) \\ DUY_{\beta-PCR} &= 0,65 \cdot 0,45 = 0,29 \\ SEÇ_{\beta-PCR} &= 0,53 + 0,63 - (0,53 \cdot 0,63) = 0,83 \end{aligned}$$

“Veya kuralı”nı uyguladığımızda;

$$\begin{aligned} DUY_{\beta-PCR} &= DUY_{\beta} + DUY_{PCR} - (DUY_{\beta} \cdot DUY_{PCR}) \\ SEÇ_{\beta-PCR} &= SEÇ_{\beta} \cdot SEÇ_{PCR} \\ DUY_{\beta-PCR} &= 0,65 + 0,45 - (0,65 \cdot 0,45) = 0,81 \\ SEÇ_{\beta-PCR} &= 0,53 \cdot 0,63 = 0,33 \end{aligned}$$



PARALEL TESTLER

- Genel sonuçlar aşağıdaki gibidir:

Test	DUY	SEÇ
β -D-Glukan	0,65	0,53
PCR	0,45	0,63
β &PCR (ve kuralı)	0,29	0,83
β &PCR (veya kuralı)	0,81	0,33

- İnvaziv mantar enfeksiyonunu belirlemede, birleşik testin pozitif sonuçlarından emin olmak istendiğinde “ve kuralı”nı kullanarak %29 duyarlılık ve %83 seçicilik elde edilir.
- İnvaziv mantar enfeksiyonunu belirlemede, birleşik testin negatif sonuçlarından emin olmak istendiğinde “veya kuralı”nı kullanarak %81 duyarlılık ve %33 seçicilik elde edilir.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

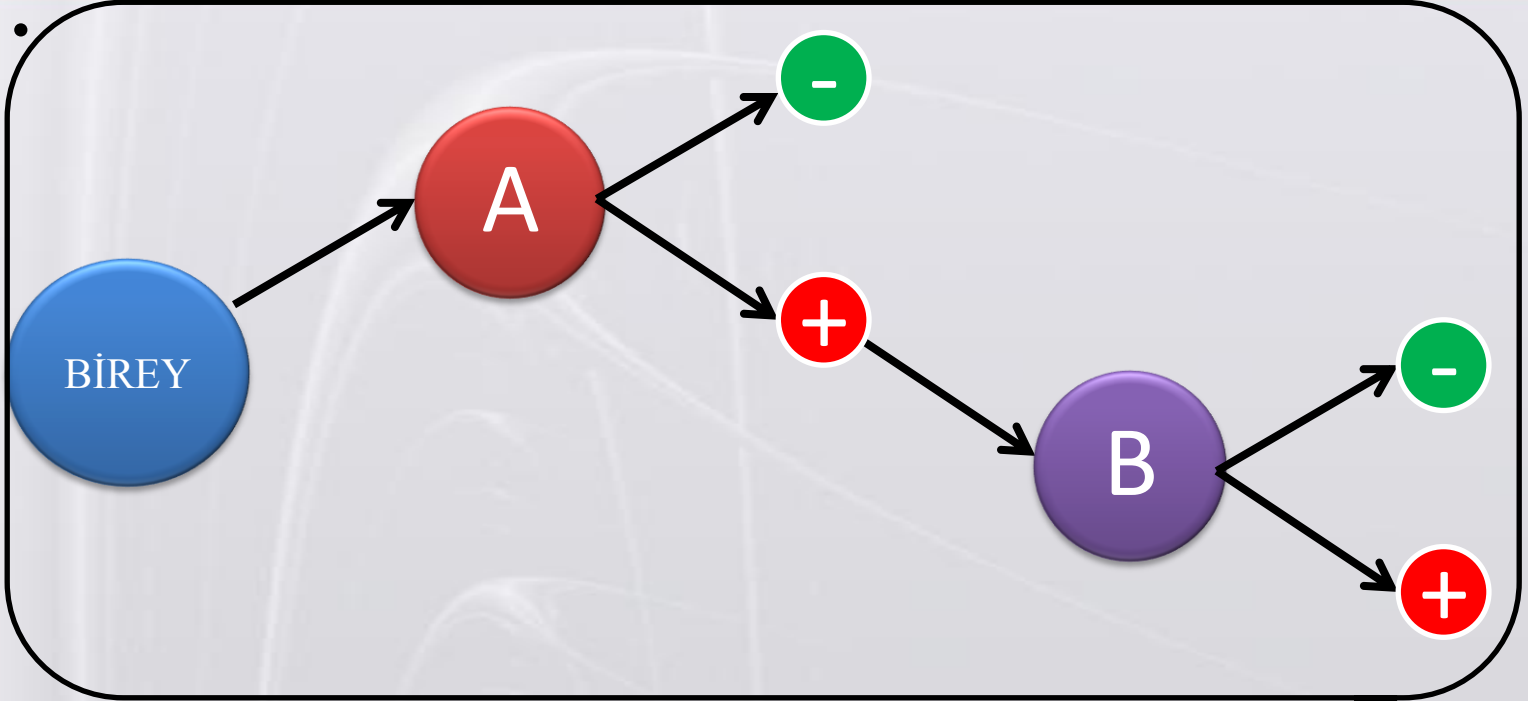
YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

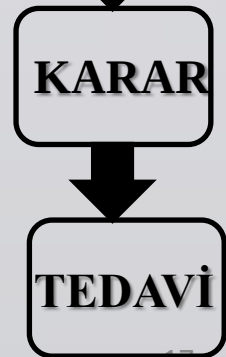


SERİ TESTLER



•**Avantajı**, az maliyetli olmasıdır. Çünkü, bireylerin bir kısmı yalnızca bir teste tabi tutulur.

•**Dezavantajı**, tedavi süresinin gecikmesidir. Çünkü, ilk testte tanısı konamayan bireyler ikinci testin sonuçlarını beklemek zorundadır.



SERİ TESTLER

- Seri testlerde “ve-veya kuralları”, paralel testte olduğu gibi uygulanır.
- Duyarlılık ve seçicilik hesaplamaları benzerdir.
- Tek farklılık, ikinci testin uygulandığı gözlem sayısının birinci teste bağlı olarak daha az olmasıdır.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

SERİ TESTLER

Örnek:

Hepatit C hastalarında fibrozis derecesinin belirlenmesinde histopatolojik görüntülerin altın standart olarak kullanıldığı bilinmekte ve HCV-RNA testi ile karaciğer biyopsisi testleri seri olarak uygulanarak bu hastalığı belirlemede histopatolojik değerlendirmeye alternatif olup olmayacağı araştırılmaktadır. Test sonuçları aşağıda verilmiştir.

HCV-RNA	Histopatolojik Değ.		
	+	-	Toplam
+	18	7	25
-	12	63	75
Toplam	30	70	100

$$\text{DUY}_{\text{HCV}} = 18/30 = 0,60$$
$$\text{SEÇ}_{\text{HCV}} = 63/70 = 0,90$$

Karaciğer biyopsisi	Histopatolojik Değ.		
	+	-	Toplam
+	14	4	18
-	4	3	7
Toplam	18	7	25

$$\text{DUY}_{\text{BiY}} = 14/18 = 0,78$$
$$\text{SEÇ}_{\text{BiY}} = 3/7 = 0,43$$

SERİ TESTLER

Bu iki testin sonuçlarını birleştirip bir test geliştirmek istediğimizde “ve-veya kuralları”ndan faydalanılabilir.

“Ve kuralı”nı uyguladığımızda;

$$DUY_{HCV-BİY} = DUY_{HCV} \cdot DUY_{BİY}$$

$$SEÇ_{HCV-BİY} = SEÇ_{HCV} + SEÇ_{BİY} - (SEÇ_{HCV} \cdot SEÇ_{BİY})$$

$$DUY_{HCV-BİY} = 0,60 \cdot 0,78 = 0,47$$

$$SEÇ_{HCV-BİY} = 0,90 + 0,43 - (0,90 \cdot 0,43) = 0,94$$

“Veya kuralı”nı uyguladığımızda;

$$DUY_{HCV-BİY} = DUY_{HCV} + DUY_{BİY} - (DUY_{HCV} \cdot DUY_{BİY})$$

$$SEÇ_{HCV-BİY} = SEÇ_{HCV} \cdot SEÇ_{BİY}$$

$$DUY_{HCV-BİY} = 0,60 + 0,78 - (0,60 \cdot 0,78) = 0,91$$

$$SEÇ_{HCV-BİY} = 0,90 \cdot 0,43 = 0,39$$

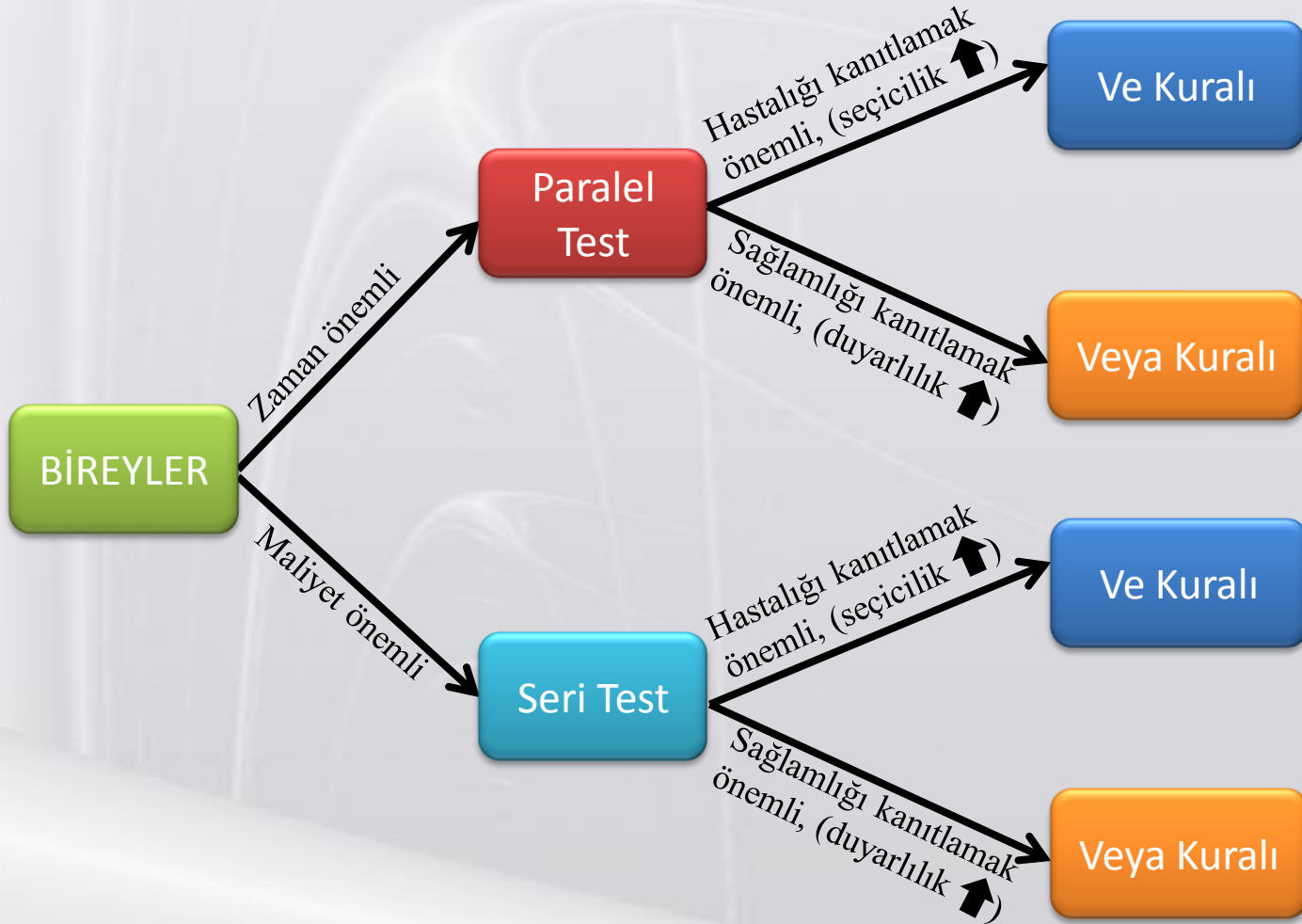
SERİ TESTLER

- Genel sonuçlar aşağıdaki gibidir:

Test	DUY	SEÇ
HCV-RNA	0,60	0,90
Karaciğer Biyopsisi	0,78	0,43
HCV&BİY (ve kuralı)	0,47	0,94
HCV&BİY (veya kuralı)	0,91	0,39

- Hepatit C hastalarında fibrozis derecesini belirlemede, birleşik testin pozitif sonuçlarından emin olmak istendiğinde “ve kuralı”nı kullanarak %47 duyarlılık ve %94 seçicilik elde edilir.
- Hepatit C hastalarında fibrozis derecesini belirlemede, birleşik testin negatif sonuçlarından emin olmak istendiğinde “veya kuralı”nı kullanarak %91 duyarlılık ve %39 seçicilik elde edilir.

PARALEL / SERİ? VE / VEYA?



TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

BİRLEŞİK TESTİN PERFORMANSI

- Birleşik testin performansını değerlendirmek için hem duyarlılık, hem de seçiciliği dikkate alan ölçüler kullanılabilir.
- Pozitif olabilirlik oranı: $LR^+ = DUY / (1 - SEÇ)$
- Negatif olabilirlik oranı: $LR^- = (1 - DUY) / SEÇ$
- Pozitif durum ile ilgilenildiğinde,
 $LR^+_{AB} > LR^+_A$ veya $LR^+_{AB} > LR^+_B$ ise birleşik test tercih edilmelidir.
 $LR^+_{AB} < LR^+_A$ veya $LR^+_{AB} < LR^+_B$ ise testler ayrı ayrı uygulanmalıdır.
- Negatif durum ile ilgilenildiğinde,
 $LR^-_{AB} < LR^-_A$ veya $LR^-_{AB} < LR^-_B$ ise birleşik test tercih edilmelidir.
 $LR^-_{AB} > LR^-_A$ veya $LR^-_{AB} > LR^-_B$ ise testler ayrı ayrı uygulanmalıdır.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



BİRLEŞİK TESTİN PERFORMANSI

- | Test | DUY | SEÇ | LR+ | LR- |
|----------------------------|------|------|------|------|
| β -D-Glukan | 0,65 | 0,53 | 1,38 | 0,66 |
| PCR | 0,45 | 0,63 | 1,22 | 0,59 |
| β &PCR (ve kuralı) | 0,29 | 0,83 | 1,71 | 0,86 |
| β &PCR (veya kuralı) | 0,81 | 0,33 | 1,21 | 0,58 |

- | Test | DUY | SEÇ | LR+ | LR- |
|-----------------------|------|------|------|------|
| HCV-RNA | 0,60 | 0,90 | 6,00 | 0,44 |
| Karaciğer Biyopsisi | 0,78 | 0,43 | 1,37 | 0,51 |
| HCV&BİY (ve kuralı) | 0,47 | 0,94 | 8,83 | 0,56 |
| HCV&BİY (veya kuralı) | 0,91 | 0,39 | 1,49 | 0,23 |

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

YENİ DEĞİŞKENLERİN OLUŞTURULMASI

- Sürekli sonuçlu testler sonuçları uygun matematiksel işlemler ile birleştirilip yeni değişkenler oluşturulabilir ve bu değişkenler ile hastalıklara daha iyi tanı konabilir.
- Örneğin, diyabet hastalığı için kullanılan HOMA-IR (insülin direnci) testi vücutta bir insülin direnci olup olmadığını anlamaya yarayan bir testdir.
- HOMA-IR değeri insülin ve glukoz değerleri ile elde edilen bir değerdir ($HOMA-IR = (insülin * glukoz) / 405$).
- Son yıllarda hematoloji, gastroentroloji ve üroloji gibi birçok tıp alanında hastalıkların tanısında kullanılan N/L değeri aslında kandaki nötrofil sayılarının lenfosit sayılarına oranlanması ile elde edilen bir değerdir.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



YENİ DEĞİŞKENLERİN OLUŞTURULMASI

Polish Journal of Veterinary Sciences Vol. 15, No. 3 (2012), 483-486

DOI 10.2478/v10181-012-0074-7

Original article

Increased AST/ALT ratio in azotaemic dogs infected with *Babesia canis*

W. Zygnier¹, O. Gójska-Zygnier², L.J. Norbury³, H. Wędrychowicz^{1,3}

¹ Division of Parasitology and Parasitic Diseases, Department of Preclinical Sciences, Faculty of Veterinary Medicine, Warsaw University of Life Sciences – SGGW, Ciszewskiego 8, 02-786 Warsaw, Poland

² Center of Small Animal Health Clinic Multiwet, Gagarina 5, 00-753 Warsaw, Poland

³ W. Stefański Institute of Parasitology, Twarda 51/55, 00-818 Warsaw, Poland

Abstract

The AST/ALT ratio was estimated in 182 dogs infected with *Babesia canis*. Among these dogs 65 had anaemia and 68 were azotaemic. Student's *t* test was used to compare means of the AST/ALT ratio in anaemic and non-anaemic dogs, and in azotaemic and non-azotaemic dogs ($p < 0.05$). The differences in AST/ALT ratio between anaemic (1.52 $\pm$ 1.15) and non-anaemic (1.76 $\pm$ 1.34) dogs were statistically insignificant ($p = 0.23$), however, the comparison of AST/ALT ratio between azotaemic (2.68 $\pm$ 1.52) and non-azotaemic (1.08 $\pm$ 0.53) dogs revealed a significantly higher value of this index in azotaemic dogs ($p = 0.00$). The present results suggest that kidney injury contributed to increased AST activity in these dogs.

Key words: AST/ALT ratio, azotaemia, *Babesia canis*, canine babesiosis, kidney injury, dogs

Low clinical relevance of the nonalcoholic fatty liver disease activity score (NAS) in predicting fibrosis progression

January 2012, Vol. 47, No. 1, Pages 108-115 (doi:10.3109/00365521.2011.634024)

Matthias Ekstedt¹, Lennart E. Franzén², Ulrik L. Mathiesen³ & Stenbjørn

Kechagias⁴

Abstract

Background/Aims. The nonalcoholic fatty liver disease (NAFLD) activity score (NAS) is a newly proposed system to grade the necroinflammatory activity in liver biopsies of NAFLD patients. This study evaluates the usefulness of the NAS in predicting clinical deterioration and fibrosis progression in NAFLD. **Methods.** One hundred and twenty-nine patients with biopsy-proven NAFLD were included in a long-term histological follow-up study. Clinical course and change in fibrosis stage were compared between nonalcoholic steatohepatitis (NASH), "borderline NASH," and "not NASH" patients. Significant fibrosis progression was defined as progression of more than one fibrosis stage or development of end-stage liver disease during follow-up. **Results.** Eighty-eight patients accepted reevaluation and 68 underwent repeat liver biopsy. Mean time between biopsies was 13.8 $\pm$ 1.2 years (range 10.3–16.3). At baseline, NASH was diagnosed in 2 (1.6%) patients, and at follow-up, in 1 (1.5%) patient. A trend toward higher baseline NAS was seen in patients ($n = 7$) who developed end-stage liver disease (3.1 $\pm$ 0.9 vs. 2.2 $\pm$ 1.0; $p = 0.050$). Baseline NAS was associated with progressive disease in a univariate binary logistic regression analysis ($p = 0.024$), but no difference was seen in the multivariate analysis including the NAS, portal inflammation, and perisinusoidal fibrosis. Moreover, 18% of patients without NASH progressed significantly in fibrosis stage. **Conclusions.** The ability of the NAS to predict progression of NAFLD is poor. The clinical usefulness of the score is limited due to the significant overlap in clinical development between NAS score groups. To use the NAS as endpoint in treatment trial is not justified.

Neutrophil Lymphocyte Ratio as a Predictor of Stroke

Serhat Tokgoz, MD,* Mehmet Kayrak, MD,† Zehra Akpınar, MD,*
Abdullah Seyithanoğlu, MS,* Figen Güney, MD,* and Betigül Yürüten, MD*

Background: The aim of this study is to investigate the relationship of the neutrophil to lymphocyte ratio (NLR) with short-term mortality in acute stroke. **Methods:** This retrospective study included 255 patients with acute cerebral infarction who presented within 24 hours of symptom onset. A hemogram from peripheral venous blood samples was taken at the time of admission. The NLR was calculated as the ratio of neutrophils to lymphocytes. Duration of follow-up was defined as 60 days. **Results:** Seventy-one of 255 patients died during the follow-up period. The median NLR was significantly increased among the mortality group compared with the survival group (median 11.50, interquartile ratio [IQR] 10.40 vs median 3.79, IQR 4.72; $P = .001$). In our multivariate Cox regression model, NLR > 5.0 (hazard ratio [HR] 3.30; 95% confidence interval [CI] 1.35-8.07), National Institutes of Health Stroke Scale score (HR 1.11; 95% CI 1.07-1.16), glucose values at admission (HR 1.007; 95% CI 1.002-1.011), and history of coronary artery disease (HR 2.49; 95% CI 1.26-4.92) were predictors of short-term mortality. The sensitivity for short-term mortality when the NLR was > 5 was 83.10%, and the specificity was 62.00%. The positive predictive value of a NLR > 5 was 45.7%, and negative predictive value was 90.50%. A strong linear association between NLR and National Institutes of Health Stroke Scale score was also observed ($r = 0.64$; $P = .001$). In addition, the NLR was higher in both the atherosclerotic and cardioembolic stroke subgroups than the lacunar infarct subgroup (6.5 [IQR 7.2], 7.5 [IQR 8.9], and 3.20 [IQR 3.50], respectively; $P = .001$). **Conclusions:** The NLR at the time of hospital admission may be a predictor of short-term mortality in acute stroke patients. Because of the routine use and inexpensive nature of hemogram analysis, the NLR should be investigated in future prospective, randomized controlled trials investigating acute stroke. **Key Words:** Mortality—National Institutes of Health Stroke Scale—neutrophil lymphocyte ratio—stroke. © 2013 by National Stroke Association

Early pregnancy maternal serum PAPP-A and urinary protein-creatinine ratio as predictive markers of pregnancy induced hypertension.

Nasrin B, Fatema N, Jebunnessa F, Shamim S, Chowdhury SB, Ali L.

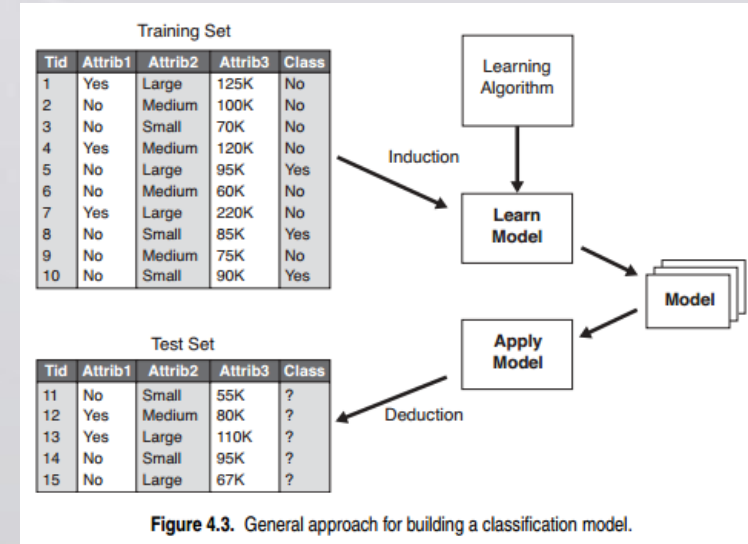
Department of Obstetrics and Gynaecology, Bangabandhu Sheikh Mujib Medical University (BSMMU), Shahbagh, and Square Hospital, Dhaka, Bangladesh.

Abstract

Pregnancy induced hypertension (PIH) is a major complication of pregnancy and is associated with high maternal and perinatal morbidity and mortality. The aim of this study was to investigate the possible causal association of PIH with maternal serum PAPP-A and urinary protein-creatinine ratio (UPCr) as well as to evaluate the usefulness of these two variables as predictive markers of PIH. A total 200 women of 8-16 weeks of pregnancy were enrolled in this study. All the patients were followed up till delivery for the development of pregnancy induced hypertension. Thirty patients were lost in the follow up, 3 (1.76%) developed preeclampsia (PE) and 14 (8.23%) gestational hypertension (GH). By a nested case-control design the 17 pregnancy induced hypertension cases were compared with 48 Controls with normal pregnancy outcome. Maternal serum PAPP-A was significantly lower in the pregnancy induced hypertension group compared to Control [mIU/ml, median (range) 1.8 (0.70-4.1) vs. 5.45 (2.7-10), $p < 0.001$]. UPCR was significantly higher in the pregnancy induced hypertension group compared to Control (mg/mmol, mean $\pm$ SD, 6.66 $\pm$ 1.56 vs. 4.75 $\pm$ 0.96, $p < 0.001$). When tested as a predictive marker of pregnancy induced hypertension the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of maternal serum PAPP-A in the lowest 25th percentile were 82%, 95%, 87% and 93%. At 75th percentile the sensitivity, specificity, PPV and NPV of UPCR were 52%, 85%, 56% and 83% respectively.

ÇOK DEĞİŞKENLİ YAKLAŞIMLAR

- Tanı testleri sonuçlarının birleştirilmesinde çok değişkenli istatistiksel ve veri madenciliği yaklaşımları da kullanılabilir.
- Bu yaklaşımlar altın standart testlerin sonuçlarını belirlemede, kullandıkları sınıflandırma tekniklerine ve birleştirme fonksiyonlarına göre farklılık gösterirler.
- Bu yaklaşımlardan istatistikte en çok kullanılanlar lojistik regresyon ve diskriminant analizi yaklaşımlarıdır.
- Lojistik regresyon, diskriminant analizine göre varsayım açısından daha esnek bir yöntemdir.



LOJİSTİK REGRESYON

- Bağımlı değişken iki ya da ikiden fazla kategoriye sahip nitel bir değişkendir.
- En sık kullanılanı ikili (binary) lojistik regresyon analizidir.
- Bağımlı değişkenin farklı veri yapıları için nominal ve sıralı lojistik regresyon analizleri de vardır.
- Bağımsız değişkenlerin türü nitel veya nicel olabilir.
- İkili lojistik regresyon analizlerinde bağımlı değişkenin 1 değerini alması olasılığı kestirilir.
- Bu da tanı testlerinde hastalığı kestirmek veya hastalıkla ilgili pozitif bir durumu kestirmek anlamına gelir.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



LOJİSTİK REGRESYON

- Bu olasılığın kestiriminde doğrusal regresyon analizleri kullanıldığında, olasılık kestirimleri x'lerin değerlerine bağlı olarak doğrusal olarak değişeceğinden [0,1] aralığı dışında elde edilebilir.

- Ayrıca bağımlı değişken binomiyal dağılıma uygun olduğundan varyans kestirimi doğrusal regresyondan farklılık gösterir ve sabit varyans varsayımı ihmal edilmiş olur.

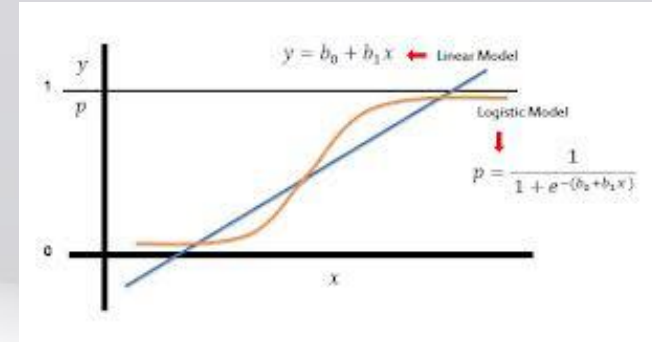
- x ile $\pi(x)$ arasındaki ilişki S biçiminde olduğu için, bu ilişkiyi modellemede lojistik model kullanılabilir.

- Doğrusal model:

$$\pi(x) = P(Y=1 | X=x) = \beta_0 + \beta_1 x$$

- Lojistik model:

$$\pi(x) = P(Y = 1 | X = x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$



LOJİSTİK REGRESYON

- Lojistik model, bağımlı değişkenin odds'u biçiminde ifade edildiğinde ve logaritması alındığında lojit elde edilir.

- Lojit ile bağımsız değişkenler arasında doğrusal bir ilişki vardır. Dolayısıyla lojit $(-\infty, +\infty)$ aralığında değer alabilir.

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x$$

- Lojit ile bağımsız değişkenler arasındaki bu özellik, klinik açıdan anlamlı yorumlar elde edilmesini sağlar.
- **Örn.** bel/boy oranı 0,5'in üzerinde olanlar, 0,5'in altında olanlara göre 3 kat daha fazla obezite riskine sahiptir.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

LOJİSTİK REGRESYON

- Çoklu lojistik regresyon analizinde aşağıdaki model kullanılır:

$$\pi(x) = P(Y = 1 | X = x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}$$

- Örneğin, iki tanı testi sonuçlarını birleştirmek istediğimizde aşağıdaki gibi bir model kullanırız.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}$$

- β parametrelerinin kestiriminde genelde en çok olabilirlik yönteminden, elde edilen parametre değerlerinin anlamlılığı için olabilirlik oranı, Wald ve skor testlerinden faydalanılır.
- Model oluşturmadan önce, bağımlı ve bağımsız değişkenlerin kodlamaları büyük önem taşımaktadır.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



LOJİSTİK REGRESYON

Örnek:

Çölyak hastalığını belirlemede ince bağırsak biyopsisi altın standart olarak kullanılmaktadır. Şeker testi (ST) ve serum antigliadin antikor testinin (IgAAGA) çölyak tanısı üzerinde alternatif olup olmayacağı ile ilgili birtakım çalışmalar yapılmıştır. Bardella ve ark. (1991) çalışmalarında lojistik regresyon kullanarak IgAAGA ve şeker testlerini birleştirmek istemişler ve aşağıdaki bulguları elde etmişlerdir.

IgAAGA	Biyopsi		
	+	-	Toplam
+	17	1	18
-	9	33	42
Toplam	26	34	60

ST	Biyopsi		
	+	-	Toplam
+	25	9	34
-	1	25	26
Toplam	26	34	60

LOJİSTİK REGRESYON

- Lojistik regresyon uygulandığında aşağıdaki model elde edilir:

$$\pi(x) = \frac{e^{-3,5312+3,6098IgAAGA+3,7784ST}}{1 + e^{-3,5312+3,6098IgAAGA+3,7784ST}}$$

- Örneğin, IgAAGA testi 25'in üzerinde sonuç verdiğinde ve ST testi pozitif sonuç verdiğinde:

$$\pi(x) = \frac{e^{-3,5312+3,6098*1+3,7784*1}}{1 + e^{-3,5312+3,6098*1+3,7784*1}} = 0,979$$

- Yani, IgAAGA testi 25'in üzerinde sonuç verdiğinde ve ST testi pozitif sonuç verdiğinde lojistik regresyon eşitliği kullanılırsa bireyin çölyak hastalığı olma olasılığı 0,979'dur.
- Benzer yorumlar IgAAGA ve ST testlerinin diğer kombinasyonları için de yapılabilir.

LOJİSTİK REGRESYON

- Lojistik regresyon modeli sonucu elde edilen sınıflandırma (tanı testlerinde olumsuzluk) tablosu aşağıdaki gibi elde edilir:

IgAAGA+ ST	Biyopsi		Toplam
	+	-	
+	25	10	11
-	1	24	49
Toplam	26	34	60

- Genel sonuçlar ise aşağıdaki gibi elde edilir:

Test	DUY	SEÇ	PKD	NKD	LR+
IgAAGA	0,65	0,97	0,94	0,79	22,2
ST	96,2	73,5	73,5	96,2	3,6
IgAAGAST+ST	96,2	70,6	71,4	96,0	47,3

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



LOJİSTİK REGRESYON

Örnek:

Karaciğer yağlanması belirlemede biyopsinin altın standart olarak kullanıldığı bilinmekte ve MPV (Mean platelet volume) kan değerinin, obezite durumunun (normal, kilolu, obez) ve AST/ALT oranının (<1 , ≥ 1) alternatif tanı yöntemleri olup olamayacakları araştırılmaktadır. Acaba lojistik regresyon analizi kullanılarak tanı testleri birleştirilse daha iyi bir performans elde edilebilir mi?

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

LOJİSTİK REGRESYON

$$\pi(kar.yag.) = \frac{e^{-5,395+1,504(Ob1)+1,974(Ob2)-2,936(AST/ALT)+0,559(MPV)}}{1+e^{-5,395+1,504(Ob1)+1,974(Ob2)-2,936(AST/ALT)+0,559(MPV)}}$$

- Obez olan, AST/ALT oranı 1'den küçük ve MPV değeri 11 olan bir bireyin karaciğer yağlanması olması olasılığı:

$$\pi(kar.yag.) = \frac{e^{-5,395+1,504(0)+1,974(1)-2,936(0)+0,559(11)}}{1+e^{-5,395+1,504(0)+1,974(1)-2,936(0)+0,559(11)}} = 0,939$$

⇒ Birey karaciğer yağlanması olabilir.

Test	DUY	SEÇ	LR+
Obezite	0,94	0,63	2,58
AST/ALT	0,52	0,78	2,36
MPV	0,41	0,85	2,78
L.Reg	0,93	0,76	3,80

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

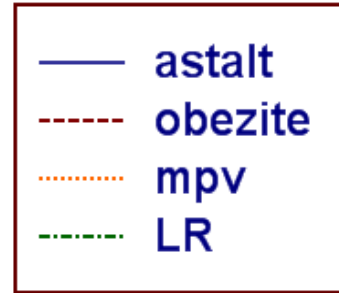
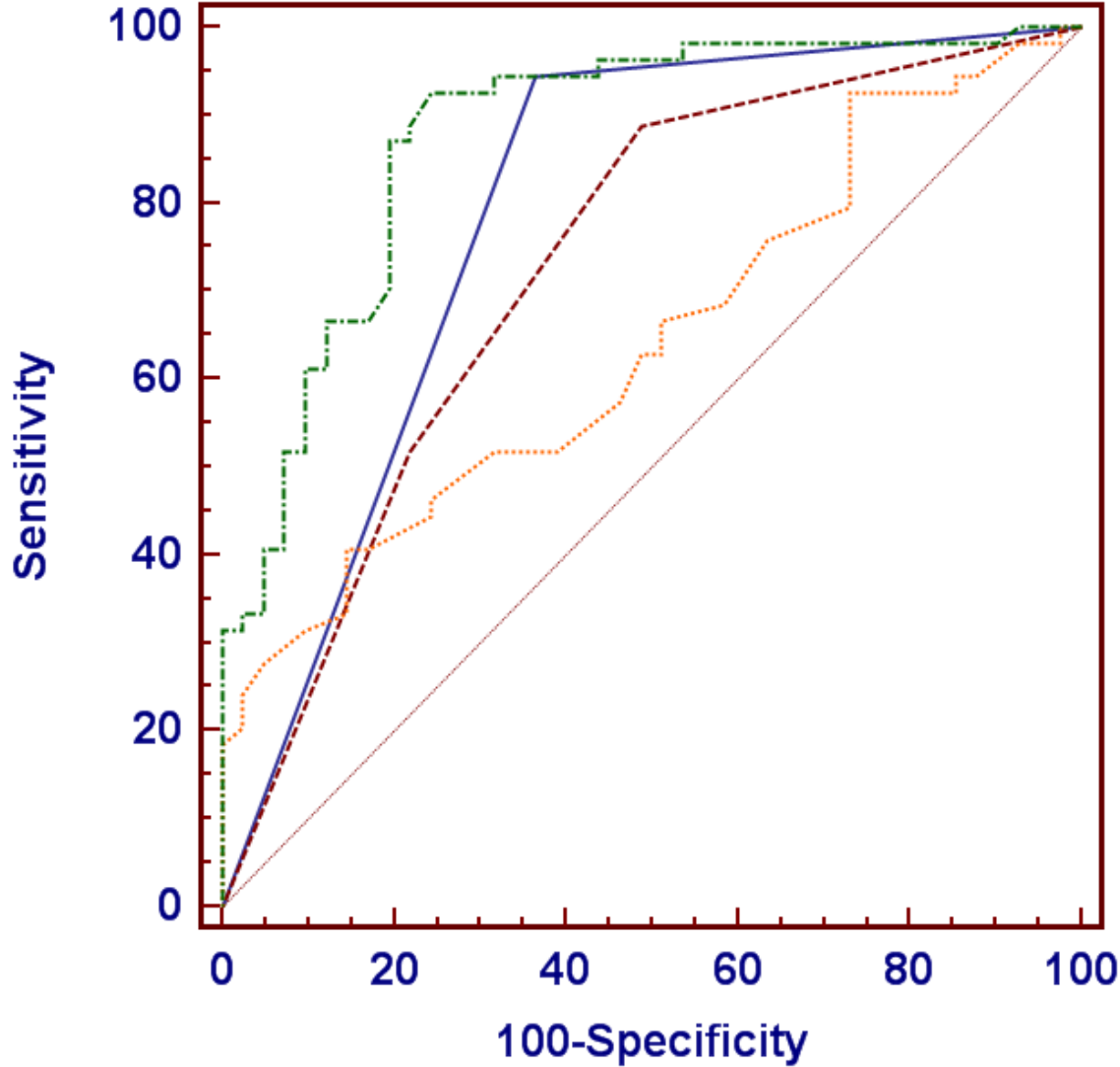
YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



LOJİSTİK REGRESYON



TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

LOJİSTİK REGRESYON

Örnek:

Çocuk ve ergenlerde abdominal obeziteyi belirlemede bel çevresi ve bel/boy çevresi oranları altın standart olarak kullanılmaktadır. Her ne kadar bu ölçümler kolay olarak elde edilse de etkili alternatiflerinin oluşturulması amacı ile beden kitle endeksi, kol yağ alanı gibi ölçümler kullanılmıştır. Kayseri’de yaşamakta olan 6-17 yaşları arasındaki 2621 erkek çocuğa ait tanımlayıcı istatistikler aşağıda verilmiştir. Acaba diskriminant analizi kullanılarak beden kitle endeksi ve kol yağ alanı ölçümleri birleştirilerek daha iyi bir tanı performansı elde edilebilir mi?

Değişken	Abdominal Obezite Yok (n=2358)		Abdominal Obezite Var (n=263)	
	Ort±SD	min-maks	Ort±SD	min-maks
BKİ (kg/m ²)	18,70±2,58	12,59-30,00	24,32±3,58	14,61-35,20
KYA(cm ²)	8,44±4,00	3,03-35,63	16,84±8,09	3,61-43,95

LOJİSTİK REGRESYON

$$\pi(abd.obz.) = \frac{e^{-13,279+0,479(BKI)+0,079(KYA)}}{1+e^{-13,279+0,479(BKI)+0,079(KYA)}}$$

- Beden kitle endeksi 18 kg/m² ve kol yağ alanı 8cm² olan bir çocuğun abdominal obezite olması olasılığı:

$$\pi(abd.obz.) = \frac{e^{-13,279+0,479(18)+0,079(8)}}{1+e^{-13,279+0,479(18)+0,079(8)}} = 0,018$$

⇒ Bireyde abdominal obezite olma olasılığı çok düşüktür.

Test	DUY	SEÇ	LR+
BKI	0,84	0,82	4,53
KYA	0,83	0,76	3,50
L.Reg	0,80	0,85	5,34

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

LOJİSTİK REGRESYON

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

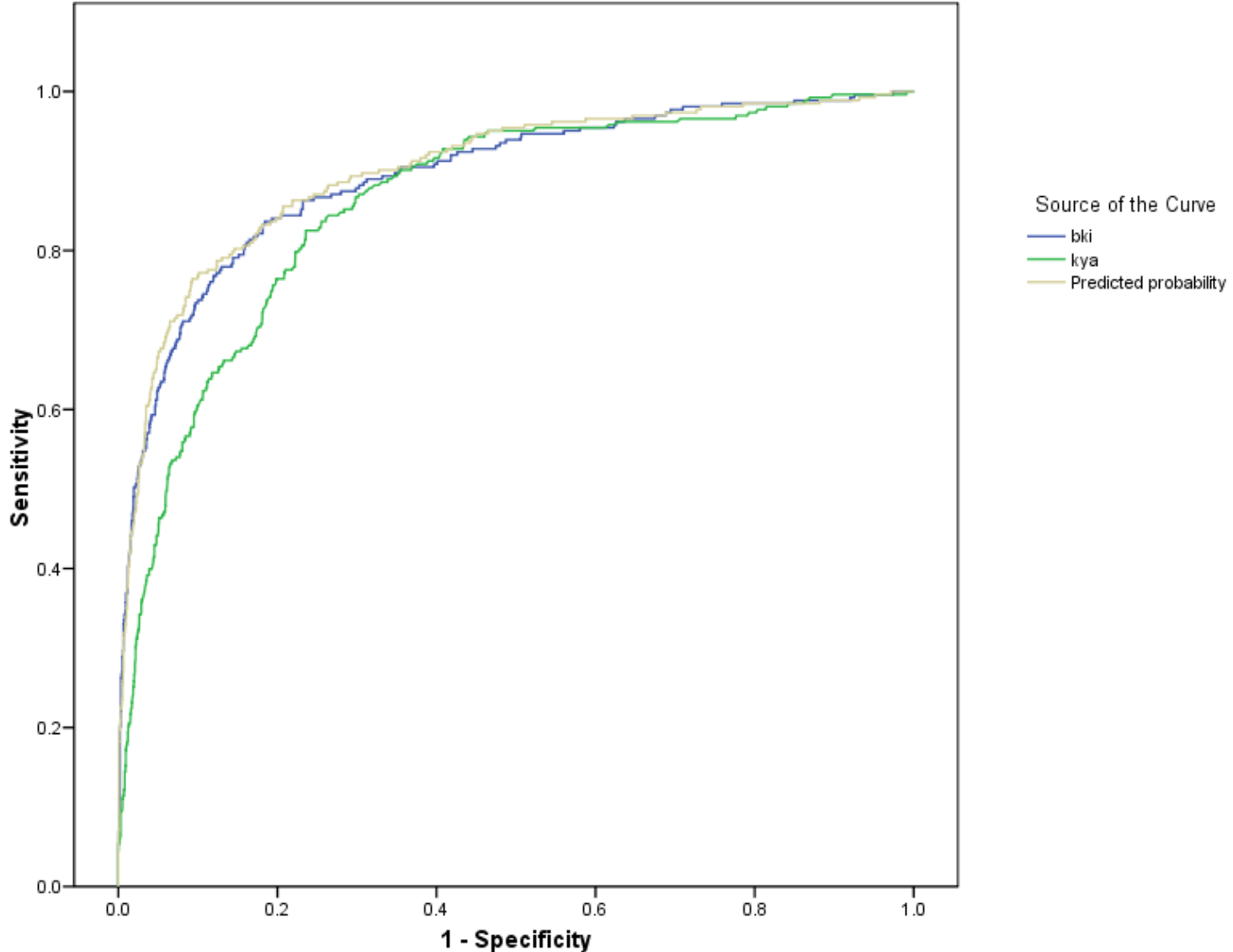
VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



DOĞRUSAL DİSKRİMİNANT ANALİZİ

- Amaçları,
 - 1) grupların ayrımını sağlayan ve bağımsız değişkenlerin doğrusal bir bileşeni olan diskriminant fonksiyonunu bulmak,
 - 2) hangi gruba ait olduğu bilinmeyen bir birimin uygun diskriminant fonksiyonu yardımıyla doğru grubuna en az hata ile atanmasını sağlamak,
 - 3) değişkenlerin grup ayrımlarındaki önem sıralarını belirleme.
- MANOVA ile arasında büyük benzerlikler vardır.
- Uygulanması için gerekli bazı varsayımlar vardır:
 - 1) Bağımsız değişkenlerin çok değişkenli normal dağılıma uygunluk göstermesi,
 - 2) Bağımsız değişkenlerin varyans-kovaryans matrislerinin homojen olması.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

DOĞRUSAL DİSKRİMİNANT ANALİZİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

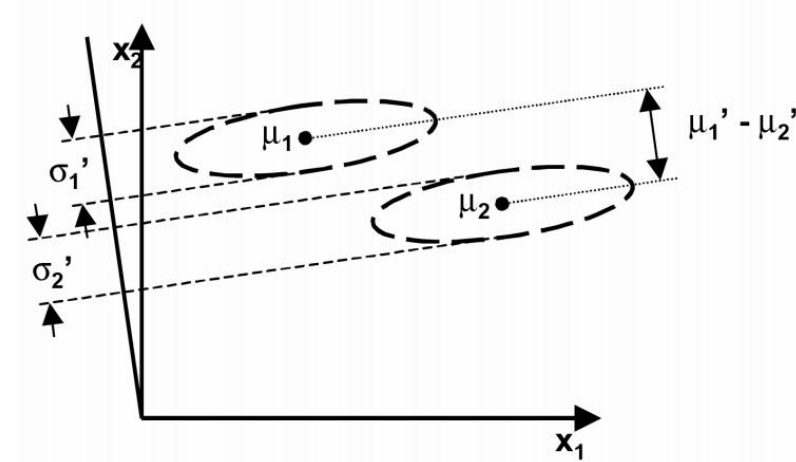
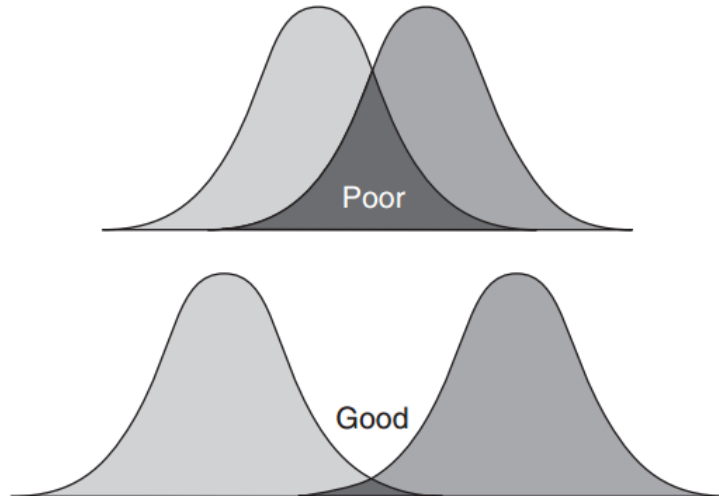
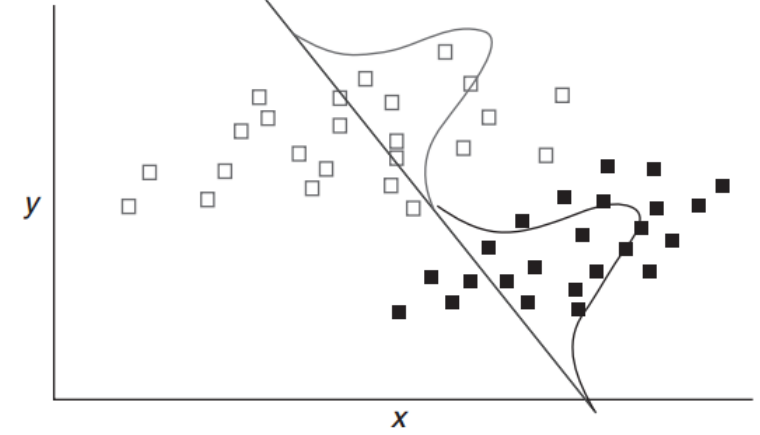
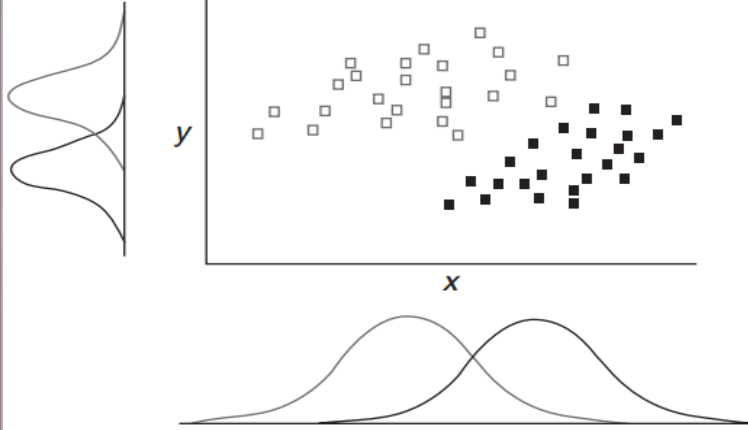
VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



DOĞRUSAL DİSKRİMİNANT ANALİZİ

- Diskriminant fonksiyonu:

$$Y = V_1X_1 + V_2X_2 + \dots + V_pX_p \quad (\text{matris formunda } Y = V'X)$$

$V_i (i : 1, 2, \dots, p)$ özvektörler

- En iyi diskriminant fonksiyonunu bulmak için gruplar arası varyansın gruplar içi varyansa oranını maksimum yapacak özvektörlerin elde edilmesi gerekir.
- Bu amaçla toplam varyans gruplar arası ve grup içi olmak üzere ikiye bölünür: $T = B + W$

$$B = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})' \quad W = \sum_{i=1}^k (n_i - 1)S_i$$

$\bar{X}_i$: i. gruba ilişkin ortalama vektörü

$\bar{X}$: genel ortalama vektörü

n_i : i. grubun gözlem sayısı

S_i : i. gruba ilişkin varyans-kovaryans matrisi

DOĞRUSAL DİSKRİMİNANT ANALİZİ

- W ve B matrisleri elde edildikten sonra $W^{-1}B$ matrisi bulunur.
- Daha sonra $|W^{-1}B - \lambda I| = 0$ eşitliği ile λ özdeğerleri bulunur.
- Her özdeğere karşılık bir özvektör elde edilir ($(W^{-1}B - \lambda I)V = 0$).
- Ortak grup içi varyans-kovaryans matrisi hesaplanır: $S_{wg} = \frac{W}{n-k}$
- M_j , p bağımsız değişkenin ortalamasına ilişkin kolon matrisi için C_j sınıflandırma fonksiyonu katsayıları ve c_{j0} sabiti bulunur:

$$C_j = S_{wg}^{-1} M_j \quad c_{j0} = -\frac{1}{2} C_j M_j$$

DOĞRUSAL DİSKRİMİNANT ANALİZİ

- Fisher ayırma fonksiyonu aşağıdaki biçimde yazılır:

$$C_j = c_{j0} + \sum_{i=1}^p c_{ji} x_i$$

- Her grup için bir tane ayırma fonksiyonu elde edilir.
- Bağımsız değişkenleri bilinen bir gözlem ayırma skoru en büyük olan gruba atanır.
- Değişkenlerin ayırım gücü Wilks Lamda (Λ) istatistiği ile F veya ki-kare dağılımları yardımıyla test edilir.

$$\Lambda = \frac{|W|}{|W + B|}$$

DOĞRUSAL DİSKRİMİNANT ANALİZİ

Örnek:

Çocuk ve ergenlerde abdominal obeziteyi belirlemede bel çevresi ve bel/boy çevresi oranları altın standart olarak kullanılmaktadır. Her ne kadar bu ölçümler kolay olarak elde edilse de etkili alternatiflerinin oluşturulması amacı ile beden kitle endeksi, kol yağ alanı gibi ölçümler kullanılmıştır. Kayseri’de yaşamakta olan 6-17 yaşları arasındaki 2621 erkek çocuğa ait tanımlayıcı istatistikler aşağıda verilmiştir. Acaba diskriminant analizi kullanılarak beden kitle endeksi ve kol yağ alanı ölçümleri birleştirilerek daha iyi bir tanı performansı elde edilebilir mi?

Değişken	Abdominal Obezite Yok (n=2358)		Abdominal Obezite Var (n=263)	
	Ort±SD	min-maks	Ort±SD	min-maks
BKİ (kg/m ²)	18,70±2,58	12,59-30,00	24,32±3,58	14,61-35,20
KYA(cm ²)	8,44±4,00	3,03-35,63	16,84±8,09	3,61-43,95

DOĞRUSAL DİSKRİMİNANT ANALİZİ

- Fisher ayırma fonksiyonları:

$$AO(-) = -26,778 + 3,022(BKI) - 0,513(KYA)$$

$$AO(+) = -42,012 + 3,597(BKI) - 0,287(KYA)$$

Beden kitle endeksi 18 kg/m² ve kol yağ alanı 8cm² olan bir çocuğun abdominal obezite olması olasılığı:

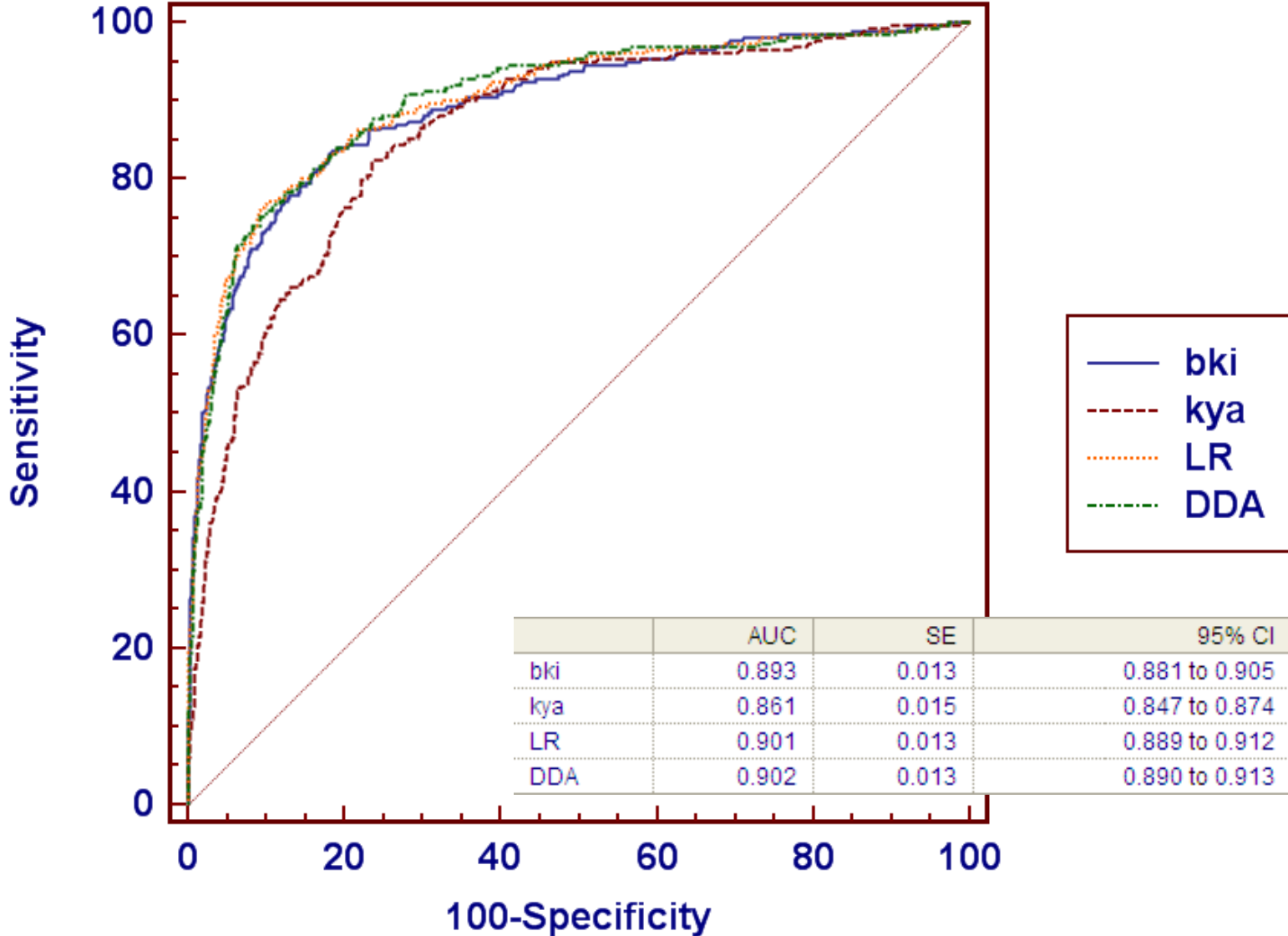
$$AO(-) = -26,778 + 3,022(18) - 0,513(8) = 23,514$$

$$AO(+) = -42,012 + 3,597(18) - 0,287(8) = 20,438$$

⇒ 23,514 > 20,438 olduğu için bireyde abdominal obezite olmama durumu daha olasıdır.

Test	DUY	SEÇ	LR+
BKI	0,84	0,82	4,53
KYA	0,83	0,76	3,50
L.Reg	0,80	0,85	5,34
DDA	0,81	0,84	5,11

DOĞRUSAL DİSKRİMİNANT ANALİZİ



TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

DİĞER DİSKRİMİNANT ANALİZLERİ

- Çok değişkenli normal dağılım varsayımının sağlandığı, varyans-kovaryans matrisleri eşitliği varsayımının sağlanmadığı durumda:
Quadratic discriminant analysis
- Her iki varsayımın sağlanmadığı durumda:
Mixture discriminant analysis
Flexible discriminant analysis
- Değişken sayısının, gözlem sayısından fazla olduğu yüksek boyutlu veri yapılarında:
Penalized discriminant analysis
Regularized discriminat analysis
- Kesikli count türünde değişken yapılarında:
Poisson linear discriminant analysis

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

İLERİ İSTATİSTİKSEL YÖNTEMLER

- Diskriminant analizi istatistiksel olarak güçlü bir test olsa da, çok değişkenli normal dağılım, varyans-kovaryans matrislerinin homojenliği gibi katı varsayımları bulunmaktadır.
- Lojistik regresyon analizi her ne kadar varsayımdan bağımsız gibi gözükse de lojit ile bağımsız değişkenler arası doğrusallığın sağlanması ve çoklu bağlantı probleminin modelde sorun yaratmayacağını incelemesi gereklidir.
- Sınıflandırma problemlerinin çözümüne yönelik birçok ileri istatistiksel yöntemler geliştirilmiştir.

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

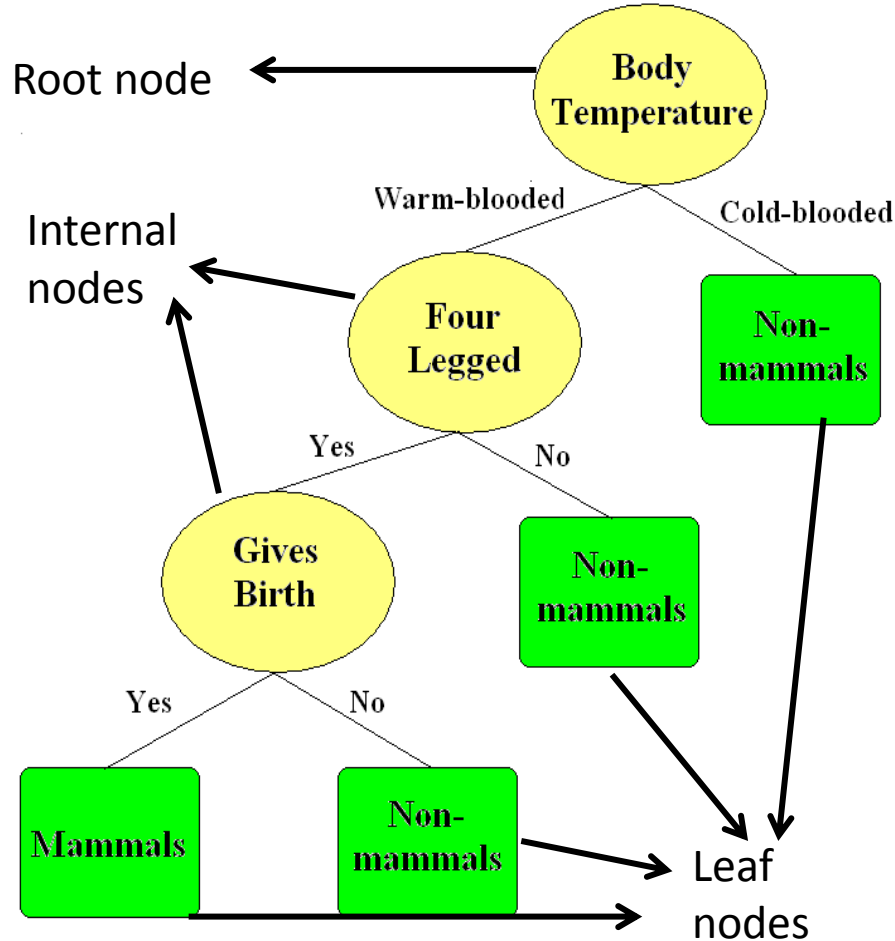
SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

KARAR AĞAÇLARI



- Kullanımları ve yorumlamaları kolaydır.
- Veri seti karar ağaçları ile eğitildikten sonra öğrenme kuralları elde edilir ve öğrenme fonksiyonları ağaç yapıları ile gösterilir.
- Önce ağaç oluşturulur, daha sonra budanır.
- En uygun karar ağacının belirlenmesi için farklı algoritmalar geliştirilmiştir. (CHAID, CART, ID3, C4.5, vs.)

KARAR AĞAÇLARI

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

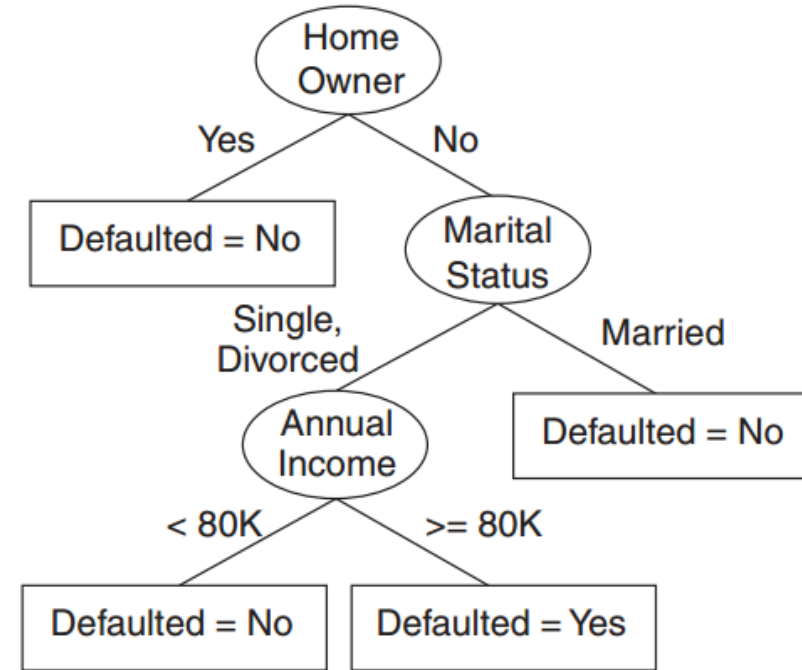
SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

	binary	categorical	continuous	class
Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



KURALA DAYALI SINIFLANDIRMA YÖNTEMLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
human	warm	yes	no	no	mammals
python	cold	no	no	no	reptiles
salmon	cold	no	no	yes	fishes
whale	warm	yes	no	yes	mammals
frog	cold	no	no	sometimes	amphibians
komodo	cold	no	no	no	reptiles
bat	warm	yes	yes	no	mammals
pigeon	warm	no	yes	no	birds
cat	warm	yes	no	no	mammals
leopard shark	cold	yes	no	yes	fishes
turtle	cold	no	no	sometimes	reptiles
penguin	warm	no	no	sometimes	birds
porcupine	warm	yes	no	no	mammals
eel	cold	no	no	yes	fishes
salamander	cold	no	no	sometimes	amphibians
gila monster	cold	no	no	no	reptiles
platypus	warm	no	no	no	mammals
owl	warm	no	yes	no	birds
dolphin	warm	yes	no	yes	mammals
eagle	warm	no	yes	no	birds

R1: (Give Birth = no) $\wedge$ (Can Fly = yes) $\rightarrow$ Birds

R2: (Give Birth = no) $\wedge$ (Live in Water = yes) $\rightarrow$ Fishes

R3: (Give Birth = yes) $\wedge$ (Blood Type = warm) $\rightarrow$ Mammals

R4: (Give Birth = no) $\wedge$ (Can Fly = no) $\rightarrow$ Reptiles

R5: (Live in Water = sometimes) $\rightarrow$ Amphibians

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
hawk	warm	no	yes	no	?
grizzly bear	warm	yes	no	no	?

The rule R1 covers a hawk => Bird

The rule R3 covers the grizzly bear => Mammal



k-EN YAKIN KOMŞULUK ALGORİTMASI

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

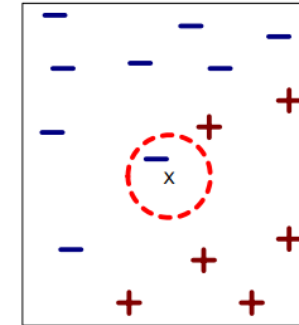
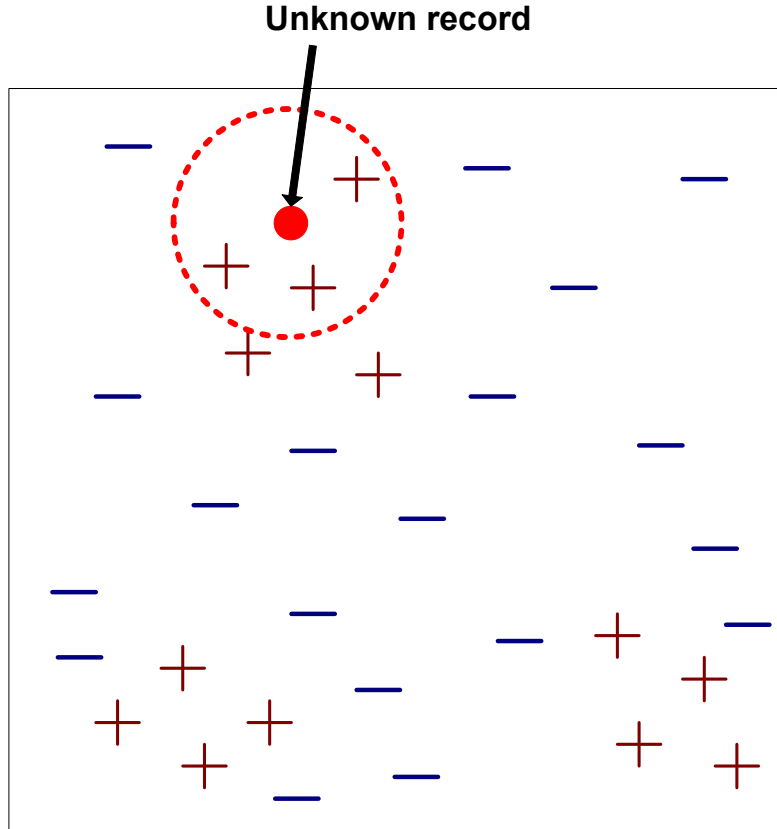
VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

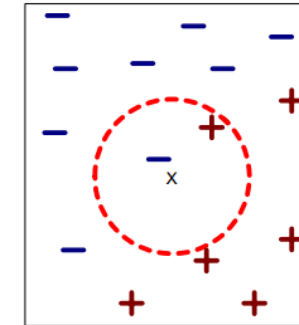
YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

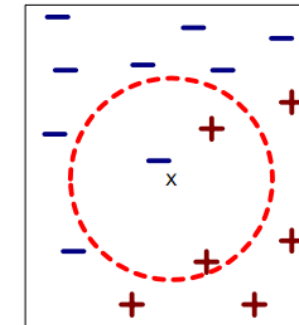
KAYNAKLAR



(a) 1-nearest neighbor



(b) 2-nearest neighbor



(c) 3-nearest neighbor

NAİVE BAYES

- Bayes kuralına dayalı sınıflandırma yaparlar.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

A: attributes

M: mammals

N: non-mammals

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

$$P(A|M)P(M) > P(A|N)P(N)$$

=> Mammals

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

DESTEK VEKTÖR MAKİNELERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

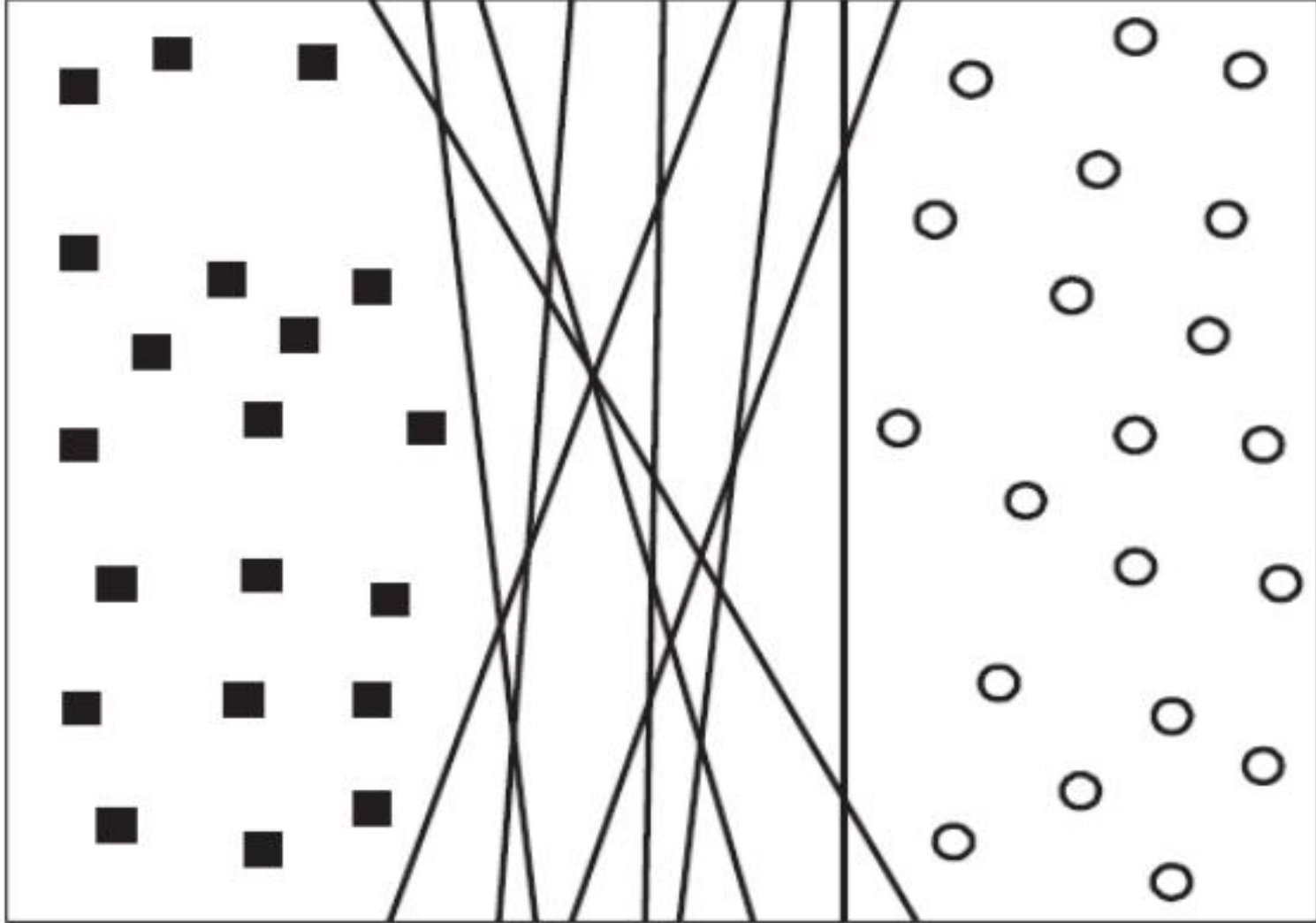
VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

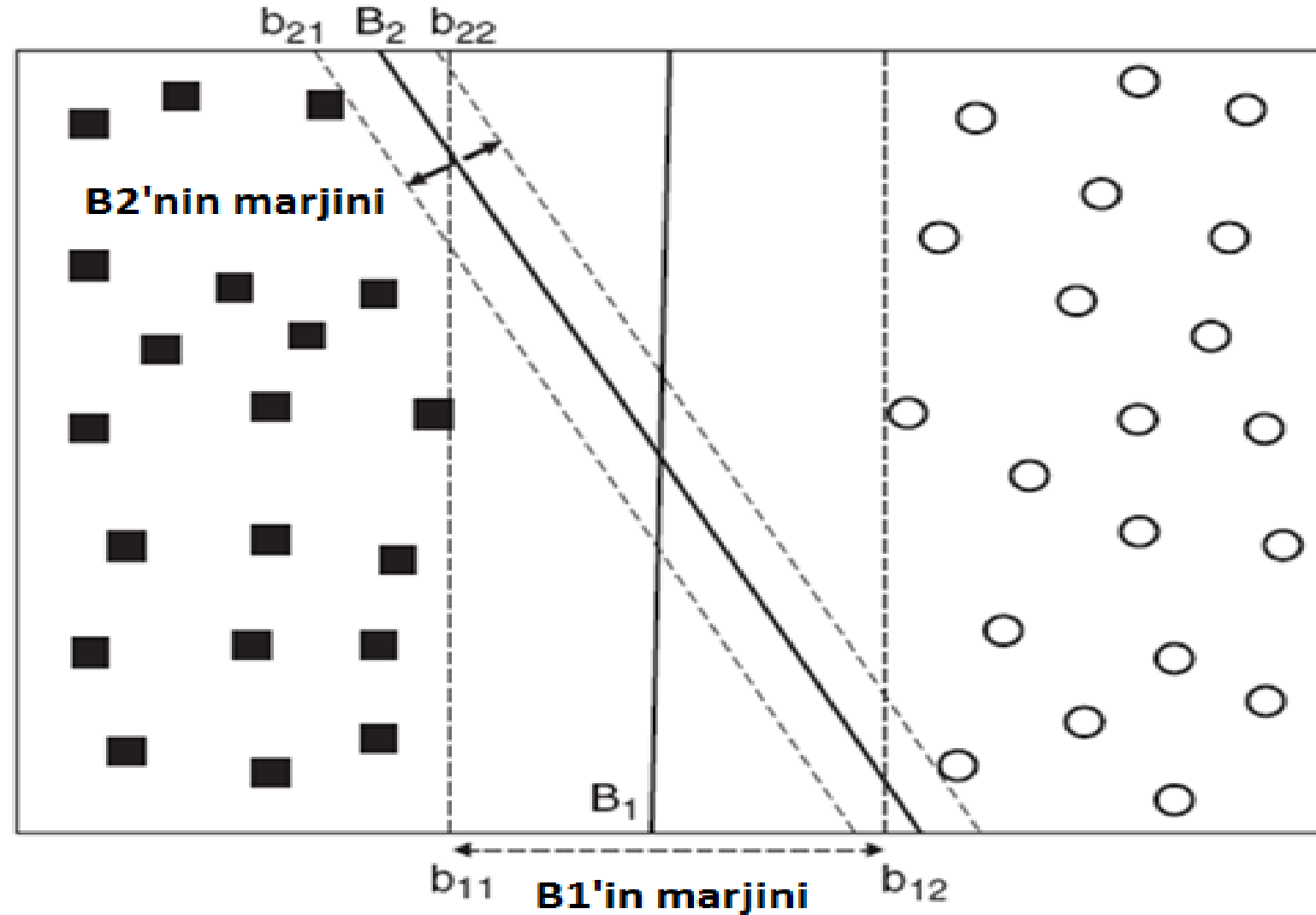
YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



DESTEK VEKTÖR MAKİNELERİ



TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



DESTEK VEKTÖR MAKİNELERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

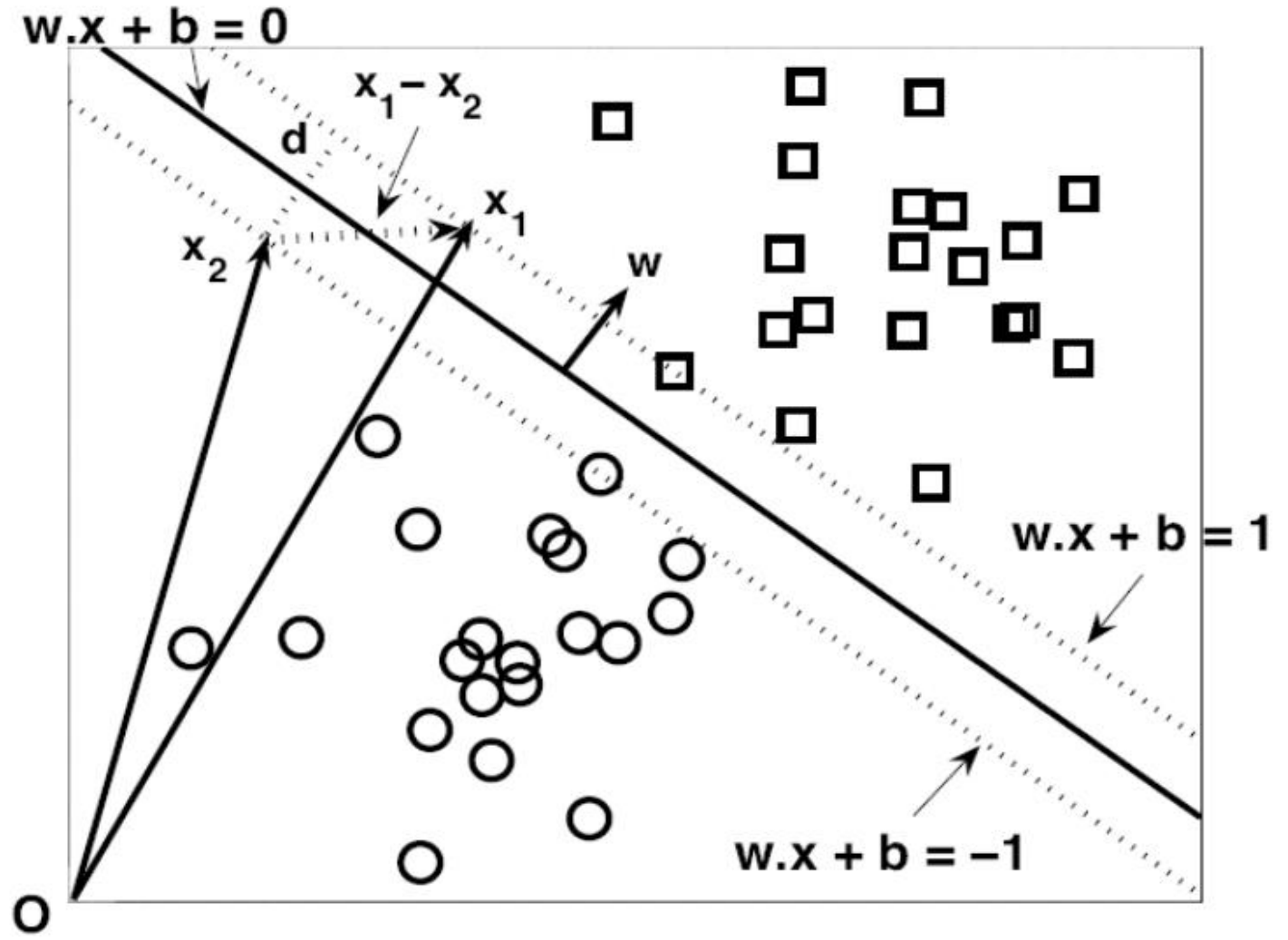
VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

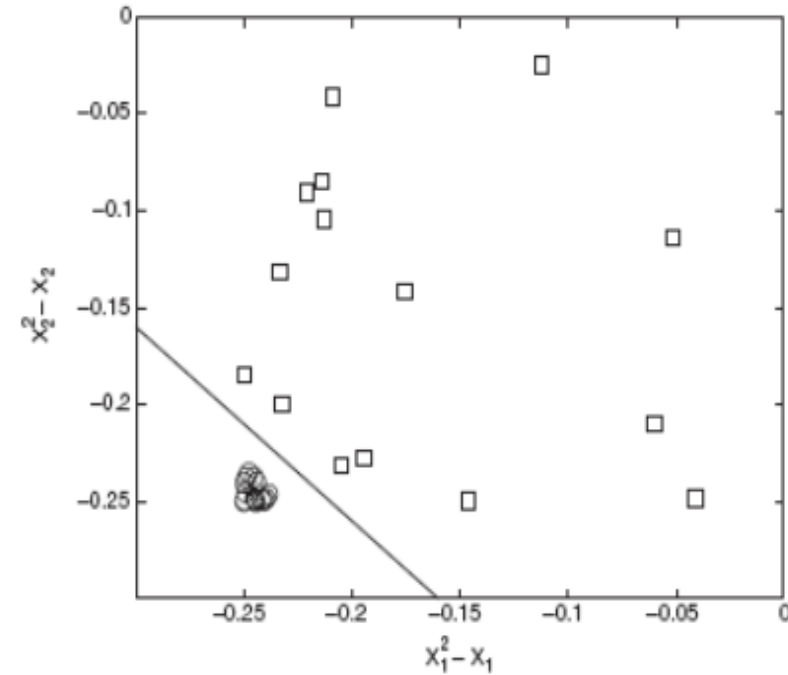
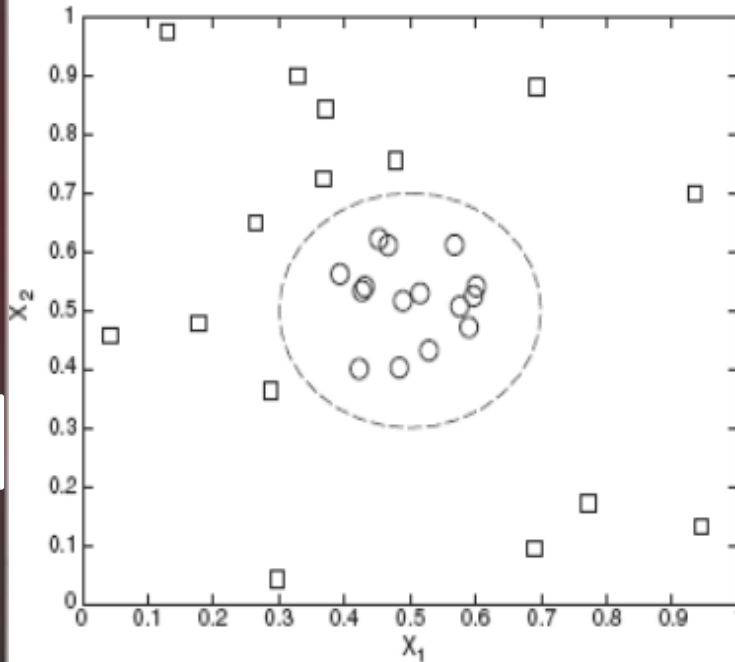
ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



DESTEK VEKTÖR MAKİNELERİ

- Doğrusal ayrılamayan durumlarda karar sınırı nasıl belirlenecek?



DESTEK VEKTÖR MAKİNELERİ

- Dönüşüm için ne tür bir ölçekleme fonksiyonunu kullanılacak?
- Yüksek boyutlu değişken uzayında optimizasyon maliyeti
- Çekirdek hilesi

Doğrusal

$$K(x_i, x_j) = x_i \cdot x_j$$

Polinomiyal

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$$

Radial Tabanlı Fonksiyon (RBF)

$$K(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)$$

Sigmoid

$$K(x_i, x_j) = \tanh(k(x_i \cdot x_j) - c)$$

YAPAY SİNİR AĞLARI

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

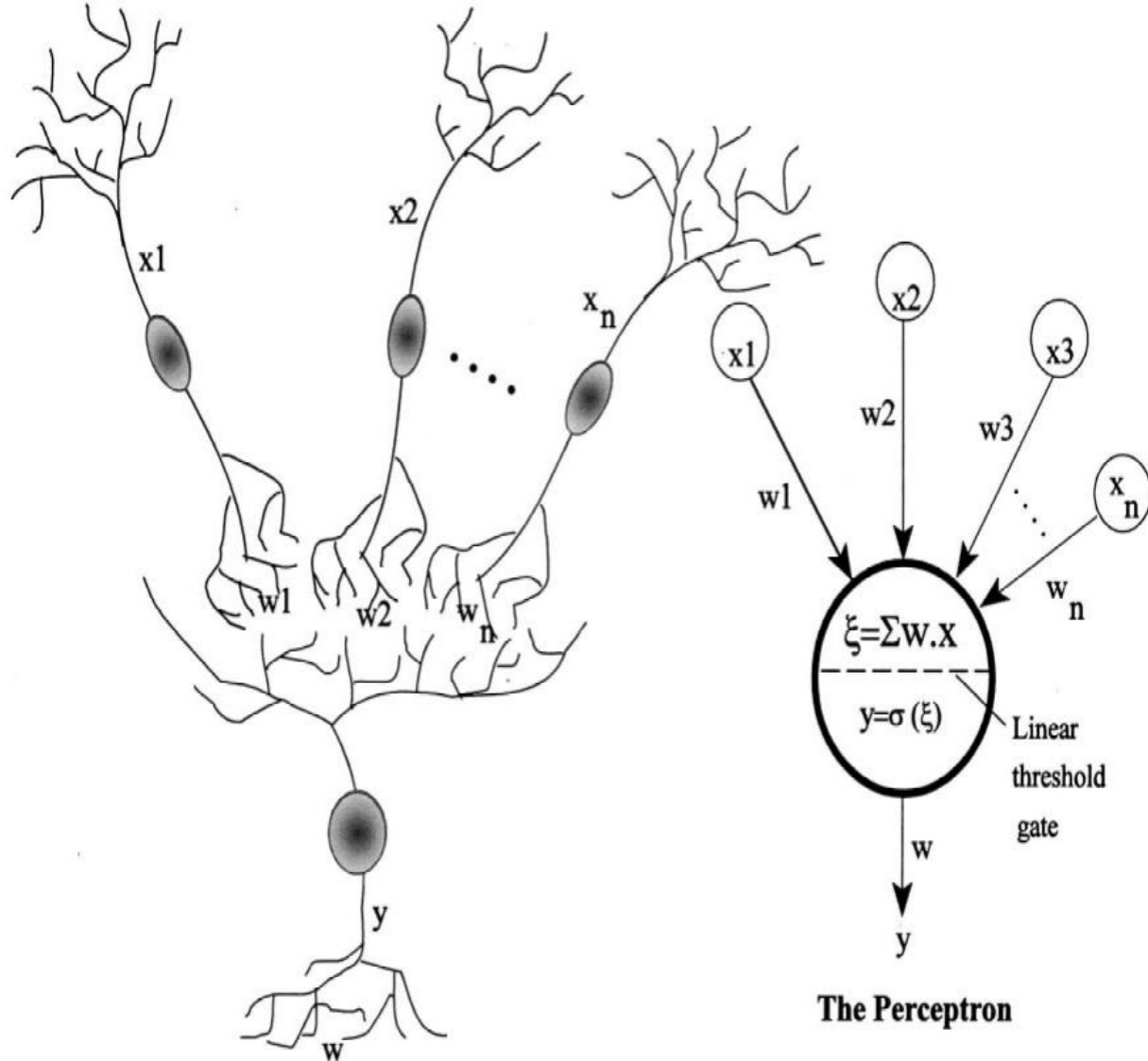
VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



YAPAY SİNİR AĞLARI

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

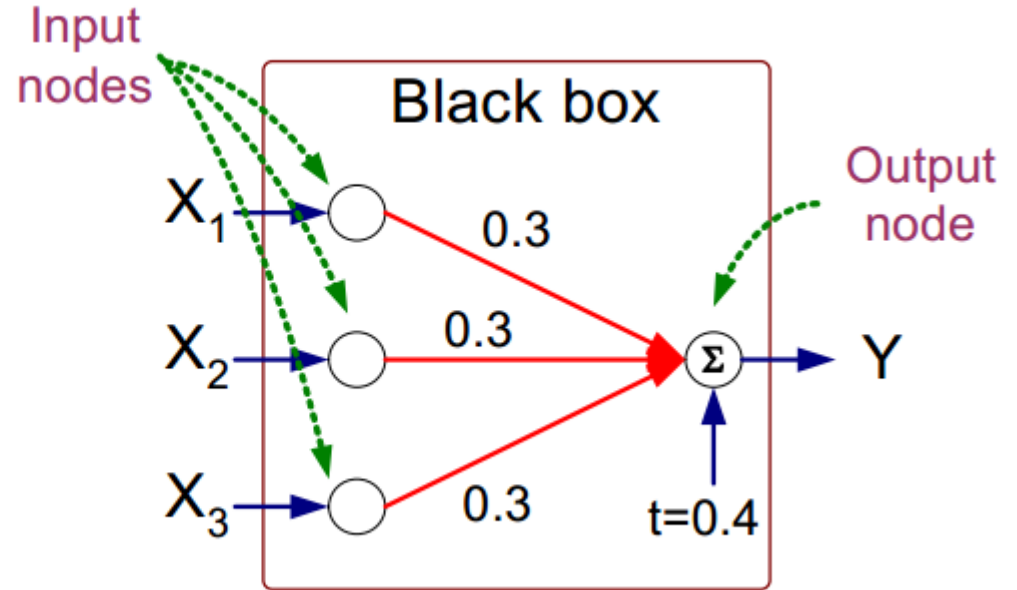
SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

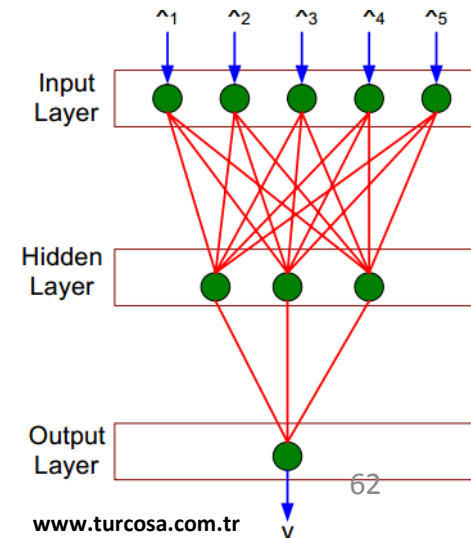
ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

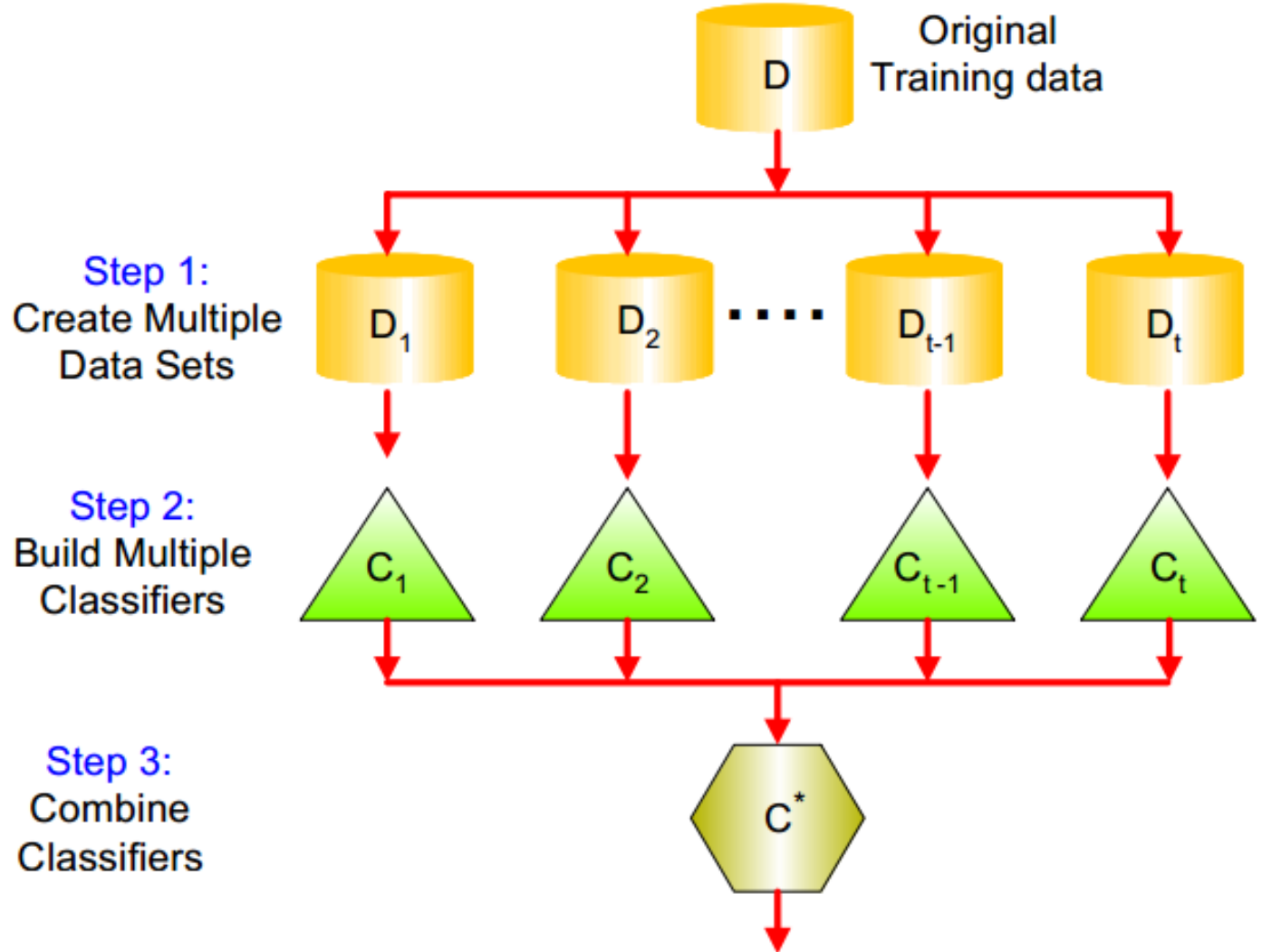
X_1	X_2	X_3	Y
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	1
0	0	0	0



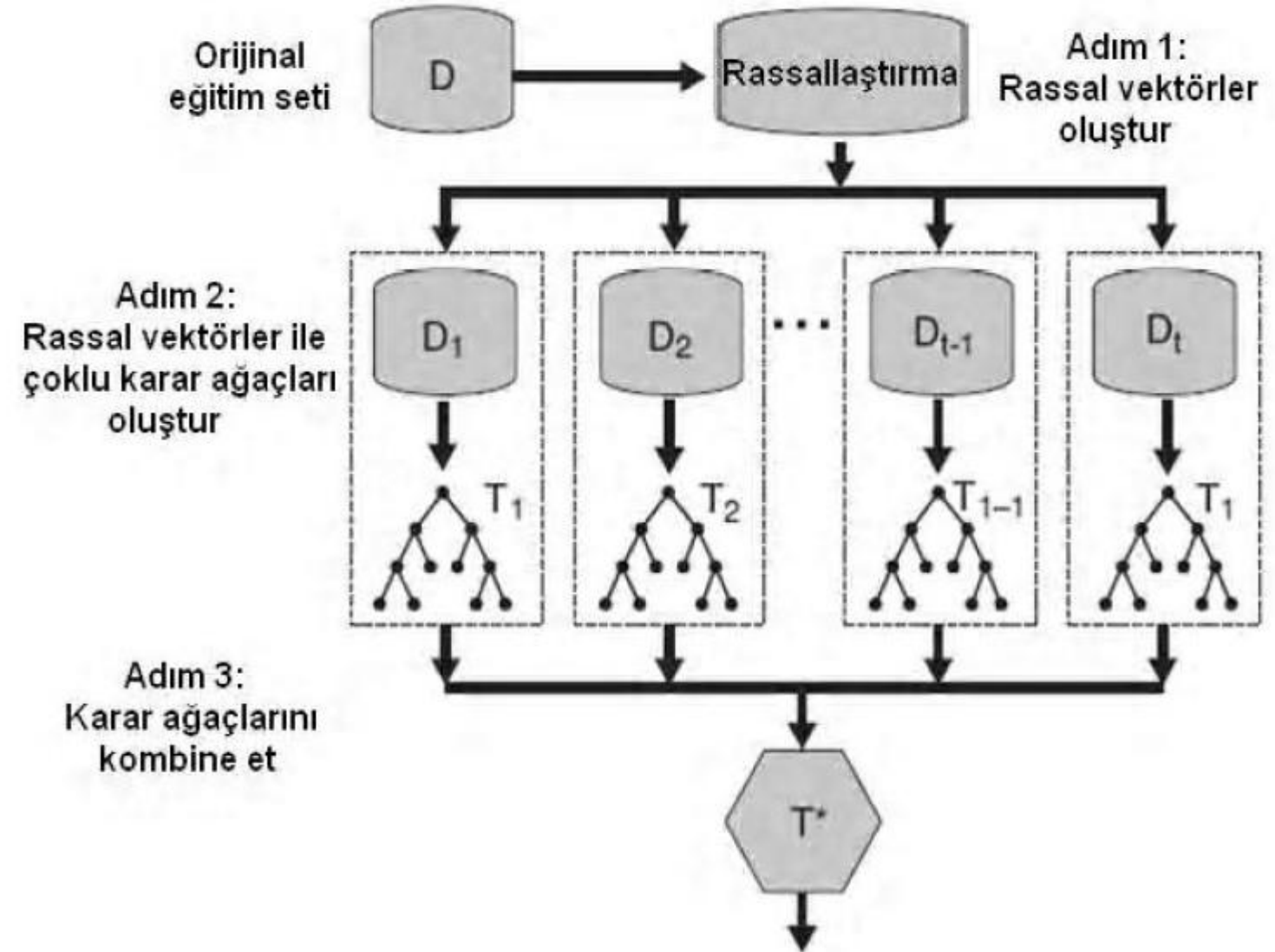
$$\hat{y} = \begin{cases} 1, & \text{eger } 0,3x_1 + 0,3x_2 + 0,3x_3 - 0,4 > 0; \\ 0, & \text{eger } 0,3x_1 + 0,3x_2 + 0,3x_3 - 0,4 < 0; \end{cases}$$



ENSEMBLE (TOPLU) YÖNTEMLER



RANDOM FOREST YÖNTEMİ



UYGULAMA

Örnek:

Akut karınlı hastalarda karın ağrısının cerrahi bir nedenden mi, cerrahi olmayan bir nedenden mi olduğunun belirlenmesi ve hekimlere ameliyat verme kararına destek olması amacıyla lökosit değeri ve D-dimer düzeyleri kullanılmaktadır. Haziran 2006 ile Kasım 2007 tarihleri arasında Kayseri’de prospektif olarak yapılan bir çalışmada travmaya bağlı olmayan akut karın nedeniyle acil servise başvuran 225 hastanın (115-, 110+) lökosit ve D-dimer düzeyleri ölçülmüştür. Acaba cerrahi tanıda çok değişkenli yöntemler lökosit ve D-dimer düzeylerine alternatif olabilir mi?

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

UYGULAMA

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

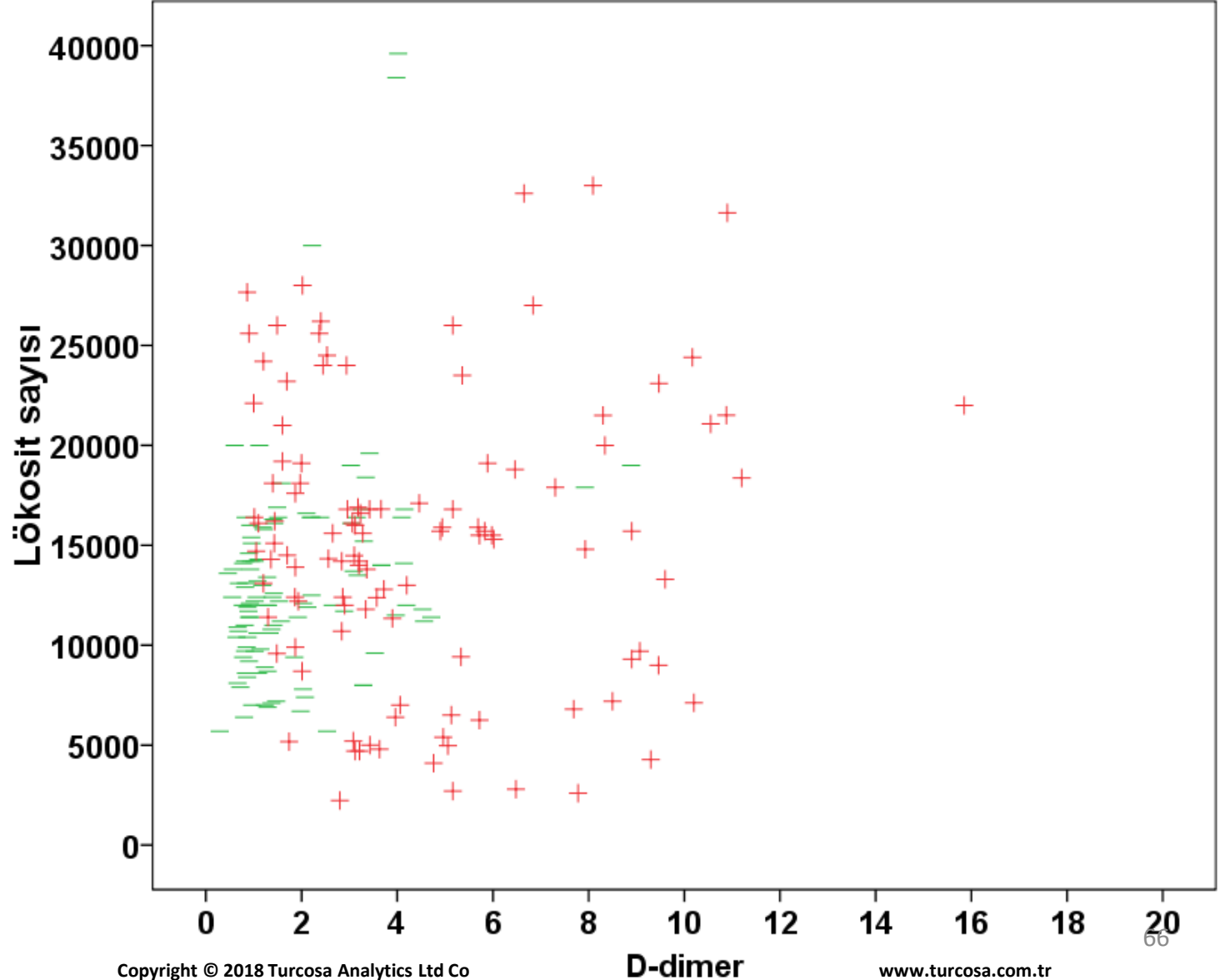
VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR



TANI TESTLERİNİN
BİRLEŞTİRİLMESİ ve
EN UYGUN TEST
KOMBİNASYONUNUN
BULUNMASI

GÖKMEN ZARARSIZ

TANI TESTLERİ

TANI TESTLERİNİN
BİRLEŞTİRİLMESİ

VE&VEYA
KURALLARI

SERİ ve PARALEL
YÖNTEMLER

YENİ
DEĞİŞKENLERİN
OLUŞTURULMASI

ÇOK DEĞİŞKENLİ
YAKLAŞIMLAR

KAYNAKLAR

UYGULAMA

Test	DSO	AUC	DUY	SEÇ	PKD	NKD	LR+
D-dimer	0,738	0,815	0,736	0,739	0,746	0,730	2,82
Lökosit	0,649	0,616	0,582	0,713	0,641	0,660	2,03
Lojistik Reg.	0,719	0,821	0,655	0,783	0,655	0,783	3,02
Diskriminant A.	0,707	0,833	0,600	0,809	0,750	0,679	3,14
Naïve Bayes	0,750	0,848	0,736	0,765	0,736	0,765	3,13
k-En Yakın K.	0,774	0,863	0,727	0,817	0,727	0,817	3,97
CART	0,738	0,762	0,727	0,748	0,727	0,748	2,88
C4.5	0,698	0,770	0,791	0,609	0,791	0,609	2,02
Random Forest	0,742	0,852	0,655	0,826	0,655	0,826	3,76
DVM (Radyal T.)	0,764	0,856	0,773	0,757	0,773	0,757	3,18

Her ölçüm için, en iyi performanslar kalın karakterler ile ifade edilmiştir.

KAYNAKLAR

1. Su JQ, Liu JS (1993) Lineer Combination of Diagnostic Markers, Journal of The American Statistical Association Volume 88, Issue 424,1350-1355,1993
2. Pepe MS, Thompson ML (2000) Combining Diagnostic Tests to Increase Accuracy, Biostatistics, 1(2), 123-140
3. Zhang J (2012) Making diagnoses with multiple tests under no gold standard, The University of Iowa, Doktora Tezi
4. Jing Q, Zhang B (2012) Best Combination of Multiple Diagnostic Tests For Screening Purposes, Statist. Med 29, 2905–2919
5. Korkmaz S. (2012) Tanı Testlerinin Birlestirilmesi ve En Uygun Test Kombinasyonunun Bulunması, Hacettepe Üniversitesi, Ders Sunumu
6. Macaskill P, Walter SD, Irwig L, et al. (2002) Assessing the gain in diagnostic performance when combining two diagnostic tests. Stat Med 21, 2527–2546
7. Zhou XH, Obuchowski NA, McClish DK (2002) Statistical methods in diagnostic medicine. New York: Wiley & Sons
8. Tan PN, Steinbach M, Kumar V (2006). Introduction to Data Mining. Pearson Education, Boston
9. Akyıldız H, Akcan A, Öztürk A, et al. (2008) D-dimer as a predictor of the need for laparotomy in patients with unclear non-traumatic acute abdomen. A preliminary study. Scand Journ of Clin Lab Inv 68(7), 612-617