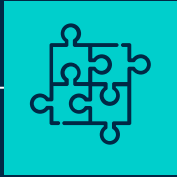


veri bilimi 101

veri hazırlama

Murat Öztürkmen
*Veri Bilimi, Tanı
İktisat Doktora, Y.T.Ü.*

içerik



01

Standardizasyon
ve varyans
ölçekleme



02

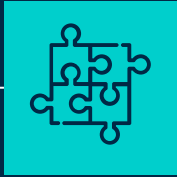
Doğrusal-
olmayan
dönüşümler



03

Normalizasyon

içerik



04

Kategorik
değişken
kodlama



05

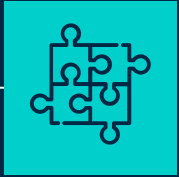
Ayrıştırma



06

Kayıp
değerlerin
doldurulması

içerik

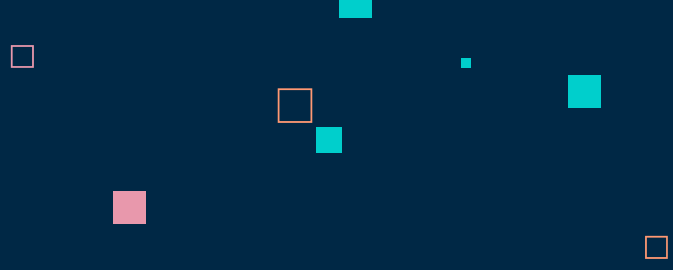




07

Polinomyal
değişken
oluşturma



■ Standardizasyon ve varyans ölçekleme

- 
- ❑ Standardizasyon birçok yapay öğrenme modeli için ortak bir gerekliliktir.
 - ❑ Tekil değişkenler standart normal dağılıma uymuyorsa modelde kötü davranış sergileyebilirler
(sıfır ortalamalı ve birim varyanslı Gaussyen dağılım).
 - ❑ Standardizasyon için yapılabilecekler: min-max ölçekleme, kernel merkezileştirme vb.
- 

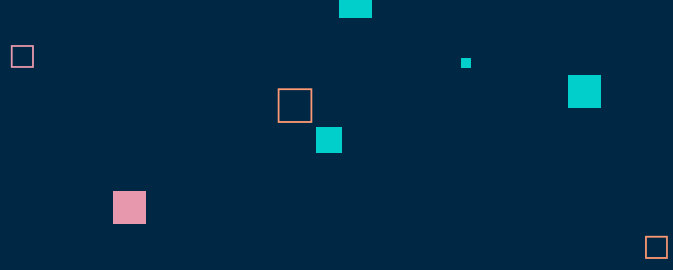
The background is a dark blue gradient. It features several thin, vertical white lines of varying lengths scattered across the frame. Interspersed among these lines are small squares in three colors: light blue, light orange, and light pink. Some squares are solid, while others are outlined. The overall aesthetic is modern and minimalist.

Doğrusal olmayan dönüşümler


- ❑ Kantil dönüşümleri
- ❑ Güç dönüşümleri

Normalizasyon

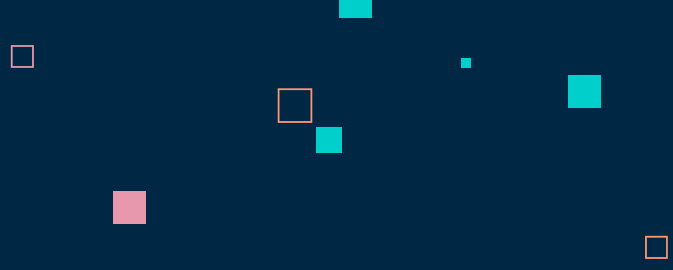

The background is a dark blue field decorated with an abstract pattern of geometric elements. It includes numerous small squares in solid colors (pink, orange, teal) and as thin white outlines. Additionally, there are several thin, vertical white lines of varying lengths scattered across the composition, creating a modern, minimalist aesthetic.

A collection of small squares in various colors (cyan, pink, orange) arranged in a scattered pattern in the top right corner of the slide.

❑ Normalleştirme, tek tek numunelerin birim normuna sahip olacak şekilde ölçeklenmesi işlemidir.

A small cluster of squares in orange and cyan colors located in the bottom left corner of the slide.

Kategorik deęiřkenlerin kodlanması

- 
- A collection of small squares in various colors (cyan, pink, orange) arranged in a scattered pattern in the top right corner of the slide.
- ☐ Etiket kodlama veya sıralı kodlama
 - ☐ 1-0 kodlama
 - ☐ Aylak kodlama
 - ☐ Etki kodlama
 - ☐ Hash kodlama
 - ☐ İkili kodlama
 - ☐ Hedef kodlama
- 
- A small cluster of squares in orange and cyan colors located in the bottom left corner of the slide.

Ayrıştırma


The background is a dark blue field decorated with a pattern of small squares and thin vertical lines. The squares are in three colors: light blue, pink, and orange. Some squares are solid, while others are outlined. The vertical lines are thin and white, extending from the top edge of the image towards the bottom. The overall aesthetic is modern and minimalist.

- Ayrıklaştırma (nicelleştirme veya gruptama olarak da bilinir), sürekli özellikleri ayrı değerlere bölmek için bir yol sağlar.

Kayıp deęerlerin doldurulması

- ❑ Hiç bir şey yapılmayabilir
- ❑ Ortalama/medyan değerlerle doldurma
- ❑ En sık veri veya sıfır/sabit varyanslı veri ile doldurma
- ❑ K-en yakın komşu
- ❑ Zincirli denklem ile çok değişkenli atama (multivariate imputation by chained equation) (ayrıntı için:
<https://www.jstatsoft.org/article/view/v045i03/v45i03.pdf>)

Polinomyal değişken oluşturma

The background is a dark blue gradient. It is decorated with a pattern of small squares in teal, orange, and pink, and thin white vertical lines of varying lengths, creating a modern, geometric aesthetic.

- ❑ Modelleme için değişkenler arasına biraz kompleksite eklenebilir.
- ❑ Doğrusal olmayan ilişkiler modellenenebilir.

sorusu olan?

Murat Öztürkmen

teşekkürler



mozturkmen



homodigitus

