

VERİM

Veri Analitiği
Araştırma ve Uygulama Merkezi

Bayesçi İstatistik

Sinan Yıldırım

· Sabancı ·
Universitesi

MÜHENDİSLİK VE
DOĞA BİLİMLERİ
FAKÜLTESİ

Bayesçi istatistik

- Veri: y
- İstatistiksel model: $p(y | x)$
- Model parametresi: x

sonsal dağılım
(son yargı:
bir sonraki durum
için ön yargı)

\propto öncül dağılım
(ön yargı)

\times olabilirlik
(veriden gelen
bilgi)

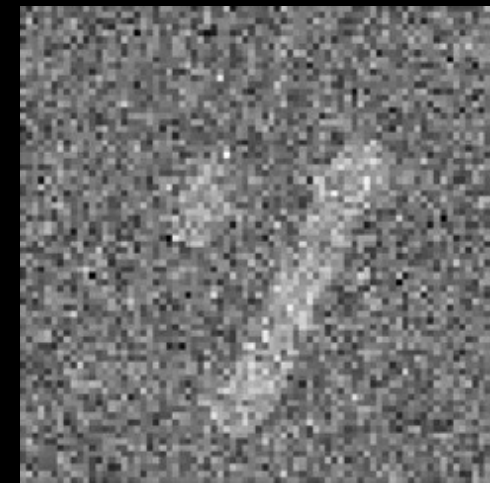
$$p(x | y) = \frac{p(x)p(y | x)}{p(y)}$$

*Bayesçi istatistikte
sonuç temel olarak
bir olasılık dağılımıdır.*

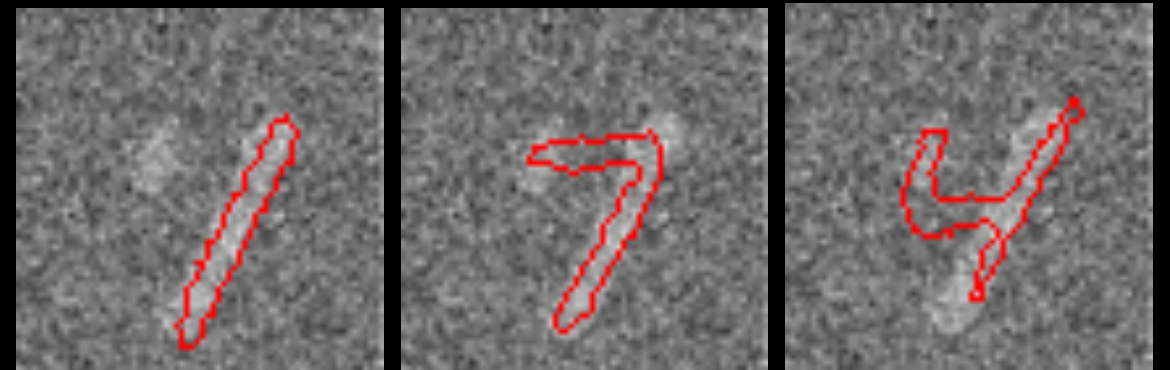
Örnek: İmgelerin sınıflandırılması

- y : sınıflandırılacak imge
- $p(y | x)$: imgenin sınıfa bağlı istatistiksel modeli
- x : İmgenin ait olduğu sınıf.

$$p(x | y) = \frac{p(x)p(y | x)}{p(y)}$$



**Hedef: Eğitim setinden
üretilen öncül bilgiyi
ve imgeyi kullanarak
imgenin ait olduğu sınıfı bulmak.**



Örnek: Arama verisinde anomali tespiti

- y : abonelerin arama verileri (akan veri)
- $p(y | x)$: değişim noktası modeli
- x : Kullanıcı davranışındaki değişikliklerin zamanları.

$$p(x | y) = \frac{p(x)p(y | x)}{p(y)}$$

**Hedef: Gerçek zamanda
sonsal dağılımı
yeni veriyle güncellemek.**

Örnek çalışma: Tensör ayrıştırma

Çalışma ekibi: Taylan Cemgil (BÜ), Burak Kurutmaz (BÜ), Melih Barsbey (BÜ), Umut Şimşekli

- Y : Veri tensörü $Y \approx X_1 \times X_2 \times X_3$
- $p(Y | X = (X_1, X_2, X_3))$
- $X = (X_1, X_2, X_3)$:Çarpan tensörler

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

**Hedef: veri tensörünü
anlamalı bileşenlere ayırmak.**

Zorluklar

- Sonsal dağılımın kapalı formu yok.

- $p(y|x)$ hesaplanamıyor. $p(x|y) = \frac{p(x)p(y|x)}{p(y)}$

- ★ Model karmaşık olabilir -

- ★ Örneğin, sıradizinsel modeller, çizge modeller, durum-uzay modelleri

$$p(y|x) = \int p(y|z, x) dz$$

- ★ Veri çok büyük, $p(y|x)$ fazla zaman alabilir.

$$p(y|x) = \prod_{i=1}^n p(y_i|x), \quad n \gg 1$$

Monte Carlo: Hesaplanamayan sonsal dağılımlardan *örnekleme* yapılır.

