

Model Seçimi ve Karşılaştırması Nedir?

Çalışma Kağıdı

Model seçimi ve karşılaştırması, belirli bir veri kümesi için en iyi açıklayıcı veya tahmin edici güce sahip olan istatistiksel modeli belirleme ve bu modellerin performansını objektif kriterlere göre değerlendirme tekniklerini içerir.

Sorular

1. Aşağıdakilerden hangisi model seçimi için kullanılan bir bilgi kriteridir?

- A) R-kare
- B) AIC
- C) Standart Sapma
- D) Ortalama

2. Bir modelin eğitim verisini çok iyi öğrenip yeni verilere genelleme yapamaması durumuna ne ad verilir?

- A) Eksik Uyum (Underfitting)
- B) Aşırı Uyum (Overfitting)
- C) Doğruluk
- D) Kesinlik

3. Çapraz doğrulama tekniği, modelin hangi özelliğini değerlendirmek için kullanılır?

- A) Eğitim verisine uyumunu
- B) Hesaplama hızını
- C) Genelleme performansını
- D) Parametre sayısını

4. Bir ev fiyatları veri setinde, evin büyüklüğü ile fiyat arasındaki ilişkiyi modellemek istiyorsunuz. Doğrusal regresyon mu yoksa polinom regresyon mu daha uygun olur?

5. Bir müşteri segmentasyonu çalışmasında K-Ortalamalar kümeleme mi yoksa Hiyerarşik kümeleme mi kullanmalısınız?

6. Bir zaman serisi verisi için ARIMA modeli mi yoksa Üstel Düzeltme modeli mi daha iyi performans gösterir?

7. Tanımla: Model Seçimi Nedir?

8. Tanımla: AIC (Akaike Bilgi Kriteri) Nedir?

9. Tanımla: BIC (Bayes Bilgi Kriteri) Nedir?

Cevap Anahtarı

1. B) AIC — AIC (Akaike Bilgi Kriteri) ve BIC (Bayes Bilgi Kriteri) model seçimi için yaygın olarak kullanılan bilgi kriterleridir. R-kare modelin açıklama gücünü gösterir, standart sapma ve ortalama ise veri setinin temel istatistikleridir.
2. B) Aşırı Uyum (Overfitting) — Aşırı Uyum (Overfitting), modelin eğitim verisindeki gürültüyü bile ezberlemesi ve bu nedenle yeni, görünmeyen verilere uyum sağlayamaması durumudur.
3. C) Genelleme performansını — Çapraz doğrulama, modelin bilinmeyen verilere ne kadar iyi uyum sağlayacağını, yani genelleme performansını tahmin etmek için kullanılır.
4. 1. Veri noktalarını çizerek ilişkinin doğrusal olup olmadığını görsel olarak inceleyin. 2. Doğrusal regresyon modelini kurun ve artıkların dağılımını analiz edin. 3. Polinom regresyon modellerini (örn. ikinci derece) deneyin ve R-kare gibi uyum iyiliği ölçütlerini karşılaştırın. 4. Bilgi kriterlerini (AIC, BIC) kullanarak daha basit model lehine karar verin.
5. 1. K-Ortalamlar için önceden küme sayısını belirleyin ve Hiyerarşik kümeleme için dendrogramı inceleyerek potansiyel küme sayılarını belirleyin. 2. Her iki yöntemin sonuçlarını karşılaştırın; örneğin, küme içi varyansı minimize etme veya küme dışı varyansı maksimize etme gibi kriterlere bakın. 3. Bulguların iş mantığına uygunluğunu değerlendirin.
6. 1. Her iki model için de veriyi eğitim ve test setlerine ayırın. 2. Her model için uygun parametreleri belirleyin (örn. ARIMA için p, d, q değerleri). 3. Eğitim seti üzerinde modelleri eğitin ve tahmin hatalarını (örn. MSE, MAE) hesaplayın. 4. Test seti üzerinde modellerin genelleme performansını karşılaştırın ve en düşük hata oranına sahip modeli seçin.
7. Veriye en uygun istatistiksel modeli belirleme sürecidir.
8. Bir modelin veriye uyumunu ve karmaşıklığını dengeleyen bir ölçüttür. Daha düşük değerler tercih edilir.
9. AIC'ye benzer, ancak daha karmaşık modelleri daha fazla cezalandıran bir bilgi kriteridir. Daha düşük değerler tercih edilir.

Bounlu

Tüm kartlar, adım adım çözümler ve AI hoca desteği Notek uygulamasında.
Sınav tarihlerini Promy otomatik hatırlatıcıya çevirir.