

Çapraz Doğrulama Teknikleri Nedir?

Çalışma Kağıdı

Çapraz doğrulama, veri setini birden fazla alt kümeye ayırarak modelin performansını farklı veri parçaları üzerinde değerlendiren bir yöntemdir. Bu sayede modelin aşırı uyum (overfitting) yapıp yapmadığı anlaşılır.

Sorular

- Aşağıdakilerden hangisi çapraz doğrulama tekniklerinin temel amacıdır?
 - Modelin eğitim süresini kısaltmak
 - Modelin bilinmeyen verilere genelleme yeteneğini tahmin etmek
 - Veri setindeki aykırı değerleri bulmak
 - Modelin karmaşıklığını azaltmak
- Bir veri seti 10 eşit parçaya bölünüp her parça sırayla test için kullanılırken geri kalanı eğitim için kullanılıyorsa, bu hangi çapraz doğrulama tekniğidir?
 - Leave-One-Out Çapraz Doğrulama
 - Rastgele Alt Örnekleme
 - 10-Kat Çapraz Doğrulama
 - Stratifiye K-Kat Çapraz Doğrulama
- Çapraz doğrulama, aşırı uyum (overfitting) sorununu tespit etmede nasıl yardımcı olur?
 - Eğitim verileri üzerindeki hatayı azaltarak
 - Test verileri üzerindeki hatayı eğitim verileri üzerindeki hatadan anlamlı derecede yüksek bularak
 - Modelin parametre sayısını artırarak
 - Veri setini tek bir eğitim ve test kümesine ayırarak
- K-Kat Çapraz Doğrulama (K-Fold Cross-Validation) nasıl çalışır?
- Leave-One-Out Çapraz Doğrulama (LOOCV) nedir ve ne zaman tercih edilir?
- Rastgele Alt Örnekleme Çapraz Doğrulama (Random Subsampling) nasıl işler?
- Tanımla: Çapraz Doğrulama Temel Amacı Nedir?
- Tanımla: Aşırı Uyum (Overfitting) Nedir?
- Tanımla: K-Kat Çapraz Doğrulama'da 'K' Ne Anlama Gelir?

Cevap Anahtarı

1. B) Modelin bilinmeyen verilere genelleme yeteneğini tahmin etmek — Çapraz doğrulama, modelin eğitim verileri dışındaki yeni verilere ne kadar iyi uyum sağlayacağını (genelleme yeteneğini) tahmin etmek için kullanılır.
2. C) 10-Kat Çapraz Doğrulama — Veri setinin K eşit parçaya bölünüp, her bir parçanın sırayla test kümesi olarak kullanılması K-Kat Çapraz Doğrulama olarak adlandırılır. Burada K=10'dur.
3. B) Test verileri üzerindeki hatayı eğitim verileri üzerindeki hatadan anlamlı derecede yüksek bularak — Eğer model eğitim verilerinde çok iyi performans gösterirken, çapraz doğrulama ile farklı test kümelerinde performansı önemli ölçüde düşüyorsa, bu aşırı uyumun bir işaretidir.
4. 1. Veri seti K eşit boyutta alt kümeye (katlara) bölünür.,2. Her bir kat sırayla test kümesi olarak kullanılırken, geri kalan K-1 kat eğitim kümesi olarak kullanılır.,3. Her iterasyonda model eğitilir ve test kümesi üzerinde hata hesaplanır.,4. Elde edilen K hata değerinin ortalaması, modelin genel performansını gösterir.
5. 1. LOOCV, K-Kat Çapraz Doğrulama'nın özel bir durumudur, burada K, veri setindeki gözlem sayısıdır (n).,2. Her gözlem bir kez test kümesi olarak kullanılır ve geri kalan n-1 gözlem eğitim kümesi olarak kullanılır.,3. Bu yöntem, özellikle küçük veri setlerinde modelin performansını tahmin etmek için kullanılır.,4. Hesaplama maliyeti yüksek olabilir.
6. 1. Veri seti rastgele olarak bir eğitim kümesi ve bir test kümesi olarak ikiye ayrılır.,2. Model eğitim kümesi üzerinde eğitilir ve test kümesi üzerinde değerlendirilir.,3. Bu işlem M kez tekrarlanır ve elde edilen M performans ölçümünün ortalaması alınır.,4. Veri setinin tamamı her iterasyonda kullanılmayabilir.
7. Bir modelin bilinmeyen verilere genelleme yapabilme yeteneğini tahmin etmek.
8. Modelin eğitim verilerini ezberlemesi, ancak yeni verilere iyi performans gösterememesi durumu.
9. 'K', veri setinin bölündüğü alt küme (kat) sayısıdır.

Bounlu

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