

Zaman serilerinde veri sızıntısını önlemek

Rastgele ayrılan sensör satırları neden gerçek hayattan daha iyi sonuç verir?

Kısaca

- Komşu ölçümler benzer koşulları taşıdığı için, aynı arıza olayının pencereleri hem eğitim hem test kümesine düşerse model olayı önceden görür.
- Ölçekleme ve eksik değer doldurma gibi ön işlem adımları yalnızca eğitim verisinden öğrenilmeli, test verisinin ortalaması bunlara katılmamalıdır.
- Model, pencere ve alarm eşiği doğrulama kümesinde seçilmeli ve son test kümesi bu kararların hiçbirinde kullanılmamalıdır.

01 / Satırlar bağımsız değildir

Bir makinenin art arda alınan iki ölçümü genellikle benzer koşulları taşır. Aynı arıza olayının komşu pencereleri hem eğitim hem teste düşerse model olayı önceden görmüş olur. Veri satırı sayısı büyüse bile bağımsız arıza sayısı küçük kalabilir. Raporun paydasını satır, olay ve ekipman olarak ayrı yazın.

02 / Zamanı ve ekipmanı birlikte düşünün

Gelecek ayı tahmin edecekseniz geçmişle eğitip sonraki dönemle test edin. Yeni makinelerde çalışmayı hedefliyorsanız bazı makineleri tamamen ayırın. Pencere uzunluğu ile tahmin ufku sınırları geçiyorsa araya boşluk bırakın. Tek bir ayırım, hem yeni makine hem yeni mevsim başarısını aynı anda kanıtlamayabilir.

03 / Ön işlem de öğrenir

Ortalama, standart sapma, eksik değer doldurma ve özellik seçimi yalnızca eğitim verisinden öğrenilmelidir. Testin ortalamasını ölçeklemeye katmak geleceği eğitime taşır. Doğrulama katları içinde yeniden kurulan bir ön işlem hattı kullanın. Zaman damgasının yanında bakım sonrası girilen bir sonuç alanının modele sızmadığını kontrol edin.

04 / Son test kümesini koruyun

Modeli, pencereyi ve alarm eşiğini doğrulama kümesinde seçin. Son test kümesi bu kararların kaynağı olmasın. Sonuç yetersizse geliştirmeye devam edebilirsiniz, ancak aynı testi tekrar tekrar kullanmanın bağımsız değerlendirme özelliğini zayıflattığını raporlayın. Yeni bir zaman aralığı ile tekrar doğrulayın.

Uygulayın

Veri hazırlama şemanızı Flow Maker ile çizin: zaman ayırımı, eğitimde ön işlem, doğrulamada eşik seçimi, son test.

Kaynaklar ve çevrimiçi rehber

Güncel çevrimiçi rehberi açın

scikit-learn: common pitfalls

scikit-learn: cross-validation

Önerilen pilot adımları saha doğruluk iddiası içermez. Sentetik örnekler gerçek tesis ölçümü değildir.