

Bolu Abant İzzet Baysal Üniversitesi
MÜHF
Bilgisayar Mühendisliği Bölümü
2023-2024 Yılı Güz Dönemi
Veri Madenciliği Ödevi #2

Soru: 1

Birkaç farklı makine öğrenimi şemasını başlatmak için WEKA workbench'in kullanımına aşina olunması:

(Yardım dökümanları için bkz. <http://www.cs.waikato.ac.nz/ml/weka/>).

Hava durumu verilerini (`weather.numeric.arff`) analiz etmek için varsayılan ayarlarla ZeroR, OneR, NaiveBayes ve J48 olmak üzere dört öğrenme şeması kullanın. Test seçenekleri için, önce "Eğitim setini kullan"ı seçin, ardından tekrar test edin, ancak varsayılan %66 yüzde bölünmesini kullanarak "Yüzde Bölme"yi seçin.

Sonuçlarınızı rapor haline getirin. Her şema ve kullanılan eğitim setinin her boyutu için model yüzde hata oranını verin. Bu, dört sınıflandırıcı model vereceğiniz anlamına gelir (her şema için bir tane). Her model için iki veri kümesi rapor edin. Her küme, doğru sınıflandırılan örneklerin sayısını (ve yüzdesini), yanlış sınıflandırılan örneklerin sayısını (ve yüzdesini) ve bir karışıklık matrisini içerecektir. Ağaç algoritmaları için, ya ağacın metin temsili ya da ağacın bir görüntüsünü ekleyin.

Sonuç raporu örneği (meme kanseri veri setine göre):

Test modu: eğitim verilerini değerlendirilmesi

=== Sınıflandırıcı modeli (tam eğitim seti) ===

ZeroR, sınıf değerini tahmin eder: tekrarlama olmayan olaylar

Doğru Sınıflandırılmış Örnekler	201	70.2797 %
---------------------------------	-----	-----------

Yanlış Sınıflandırılmış Örnekler	85	29.7203 %
----------------------------------	----	-----------

=== Karışıklık Matrisi ===

a b <-- olarak sınıflandırılır

201 0 | a = yinelenmeyen olaylar

85 0 | b = tekrarlama olayları

Bolu Abant İzzet Baysal Üniversitesi
MÜHF
Bilgisayar Mühendisliği Bölümü
2023-2024 Yılı Güz Dönemi
Veri Madenciliği Ödevi #2

Bir ağaç içeren bir sonuç raporu örneği (meme kanseri veri setine göre):

Test modu: bölünmüş %66,0 eğitim, kalan test

=== Sınıflandırıcı modeli (tam eğitim seti) ===

J48 budanmış ağaç

düğüm-başı = evet

| kötü niyetlilik a derecesi = 1: tekrarlanan olaylar (1.01/0.4)

| kötü niyetlilik derecesi = 2: tekrarlanmayan olaylar (26.2/8.0)

| kötü niyetlilik derecesi = 3: tekrarlanan olaylar (30.4/7.4)

düğüm-başı = hayır: tekrarlanmayan olaylar (228.39/53.4)

Doğru Sınıflandırılmış Örnekler 66 68.0412 %

Yanlış Sınıflandırılmış Örnekler 31 31.9588 %

=== Karışıklık Matrisi ===

a b <-- olarak sınıflandırılır

56 8 | a = yinelenmeyen olaylar

23 10 | b = tekrarlama olayları

Sonuçlarınızı analiz edin. Sonuçlara (hata oranlarına) dayanarak, oynayıp oynamayacağınızı belirlerken bu sınıflandırıcılardan hangisine güvenme olasılığınız daha yüksektir? Cevabınızı açıklayın/gerektelendirin. Açıklamanıza, eğitim seti verilerini kullanırken ve eğitim için ayrı bir yüzde kullanırken, şemanın kendisinin genel güçlü/zayıf yönlerinin yanı sıra doğruluk hakkında akıl yürütmeyi de eklemelisiniz.

Raporu teslimi:

1) Tüm 8 çalıştırmadan elde ettiğiniz sonuçları kopyalayıp tek bir sonuç dosyasına yapıştırın (yalnızca yukarıdaki örneklerde belirtilen bilgilere ihtiyacınız vardır).

A. 1. ZeroR, 2. OneR, 3. NaiveBayes ve 4. J48 olmak üzere dört öğrenme şeması olmak üzere 4 öğrenme şeması /modeli —> Test seçenekleri için, önce "Eğitim setini kullan"ı seçin,

B. 5. ZeroR, 6. OneR, 7. NaiveBayes ve 8. J48 olmak üzere dört öğrenme şeması olmak üzere 4 öğrenme şeması /modeli —> ardından tekrar test edin, ancak varsayılan %66 yüzde bölünmesini kullanarak "Yüzde Bölme"yi seçin.

2) Analizinizi sonuç dosyasından ayrı bir belgeye yazın.

3) Sistemde belirtilen son gönderim saatinden önce her iki dosyayı da sistemden gönderin.