Sınıflandırma Algoritmaları Karşılaştırma ve Özet

Bölüm 3'te 7 sınıflandırma modelini öğrendiniz. Daha önce, 2. Bölüm'de Regresyon için sorduğunuz aşağıdaki soruları bu bölüm sonunda da soruyor olabilirsiniz:

Her modelin artıları ve eksileri nelerdir?

Problemim için hangi modeli seçeceğimi nasıl bilebilirim?

Bu modellerin her birini nasıl geliştirebilirim?

Yeniden bu soruların her birine tek tek cevap verelim:

1. Her modelin artıları ve eksileri nelerdir? Burada her sınıflandırma modelinin tüm artılarını ve eksilerini veren bir özet sayfası bulabilirsiniz (buraya tıklayınız).

2. Problemim için hangi modeli seçeceğimi nasıl bilebilirim? Regresyon modelleri için olduğu gibi, probleminizin doğrusal mı yoksa doğrusal olmayan özellikte mi olduğunu anlamanız gerekir. Bunu Bölüm 10 - Model Seçimi dersinde öğreneceksiniz. Sonra: Eğer probleminiz doğrusal ise, Lojistik Regresyon veya SVM'yi kullanabilirsiniz. Eğer probleminiz lineer değil ise, K-NN, Naif Bayes, Karar Ağacı veya Rastgele Ormanı'nı kullanabilirsiniz. Peki hangisini her durumda seçmeliyim? Bunu Bölüm 10'da Model Seçimi k-Katlama Çapraz Doğrulama ile daha detaylı anlatıyor olacağız.

Algoritmalara farklı problemler açısından bakıldığında şunları kullanmayı tercih ederbilirsiniz: - Tahminlerinizi olasılıklarına göre sıralamak istediğinizde Lojistik Regresyon veya Naive Bayes. Örneğin, müşterilerinizi belirli bir ürünü satın alma olasılığı en yüksek olanından en düşük olanına doğru sıralamak istiyorsanız. Bu sayede, pazarlama kampanyalarınızı hedeflenmiş müşterilere yönlendirebilirsiniz. Ve tabi ki bu tür bir iş problemi için, eğer probleminiz doğrusal ise Lojistik Regresyonunu, ve probleminiz doğrusal değilse Naive Bayes'i kullanabilirsiniz. - Müşterilerinizin hangi segmente ait olduğunu tahmin etmek istediğinizde SVM. Segmentler, örneğin, her türlü segment olabilir. Örneğin daha önce kümeleme ile tanımladığınız bazı pazar segmentleri. - Model sonuçlarınızı net bir şekilde yorumlamak istediğinizde Karar Ağacı - Rastgele Orman gibi yöntemleri yine örneğin sadece daha az yorumlama ihtiyacına sahip yüksek performanslı algoritmalar arıyorsanız kullanabilirsiniz.

3. Bu modellerin her birini nasıl geliştirebilirim? 2. Bölüm'deki ile aynı cevabı vermemiz gerekecek: Bölüm 10 - Model Seçiminde, ayarlarınızı yaparak, modelinizin performansını iyileştirmenize bakacağız, yine bu bölümün altında, Parametre Ayarlamaya ayrılmış ikinci bölümü bulacaksınız. Büyük olasılıkla zaten her modelin iki tür parametreden oluştuğunu fark etmişsinizdir: öğrenilen parametreler, örneğin Doğrusal Regresyondaki katsayılar ve hiperparametreler. Hiperparametreler öğrenilmeyen ve model denklemlerde sabit değerleri olan parametrelerdir. Örneğin, lambda regülasyon parametresi veya penaltı parametresi C hiperparametrelerdir. Şimdiye kadar bu hiperparametrelerin varsayılan değerini kullandık ve modeliniz daha yüksek performans göstersin diye bunların optimum değerleri için araştırma yapmadık. En uygun değerlerini bulmak, Parametre Ayarının tam olarak ne olduğuyla ilgilidir. Bu nedenle, model performansınızı geliştirmek ve bazı parametreler ayarlamayı yapmakla ilgilenenler için, doğrudan Bölüm 10 - Model Seçimine atlamak iyi bir fikir olabilir.

Şimdi Bölüm 3'ü tamamladığınız için tebrikler ve yolculuğun bir sonraki bölümüne geçelim: