

Random Forests

Dataset	# cases	# vars	# classes	CART	Bagged CART	Random Forests
Waveform	300	21	3	29.1	19.3	17.2
Breast Cancer	699	9	2	5.9	3.7	2.9
Ionosphere	351	34	2	11.2	7.9	7.1
Diabetes	768	8	2	25.3	23.9	24.2
Glass	214	9	6	30.4	23.6	20.6

Leo Breiman (2001) "Random Forests", Machine Learning, 45, 5-32.

Birçok ağaçtan oluşan bir orman büyütün. (R varsayılan 500'dür)

Her ağacı eğitim verilerinden bağımsız bir önyükleme örneğinde * büyütün.

Her düğümde:

1. M olası değişkenlerin tümü arasından rastgele m değişken seçin (her düğüm için bağımsız olarak).
2. Seçilen m değişkenleri üzerindeki en iyi bölmeyi bulun.

Ağaçları maksimum derinliğe kadar büyütün (sınıflandırma).

Yeni veriler için tahminler almak için ağaçları oylayın / ortalayın.

* Değiştirme ile rastgele örnek N vaka. **CART**'ın

birçok avantajını devralın:

- Hem regresyon hem de sınıflandırma problemlerine uygulanabilir. **Evet.**
- Kategorik öngörücüleri doğal bir şekilde ele alın. **Evet.**
- Büyük sorunlar için bile hesaplama açısından basit ve hızlı takılması. **Evet.**
- Resmi dağıtım varsayımları yoktur (parametrik olmayan). **Evet.**
- Son derece doğrusal olmayan etkileşimleri ve sınıflandırma sınırlarını kaldırabilir. **Evet.**
- Otomatik değişken seçimi. **Evet. Ancak değişken öneme de ihtiyaç var.**
- Yedek değişkenler aracılığıyla ~~eksik değerleri ele alır.~~ **Yakınlıkları kullanma.**
- Ağaç küçükse yorumlanması çok kolaydır. **HAYIR!**

