

Hata Türleri

Fuat Can Beylunioğlu

December 24, 2017

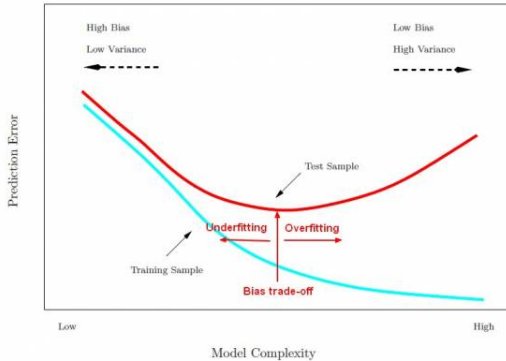
ML ve İstatistik Farkları

- ▶ ML, istatistiğe göre daha pragmatik motivasyona sahiptir
- ▶ İstatistikteki temel hedef modelin sağlıklı olması iken ML hata payını düşürmeyi hedefler
- ▶ İstatistik varsayımlarının karşılanması örneklem üzerinde hazırlanan modelin popülasyonu sorunsuz temsil edeceği anlamına gelmektedir.
- ▶ Öte yandan ML yöntemleri çoğunlukla büyük verilerle çalışır ve iyi bir yöntemle beraber veri büyüdükçe kesinlik de artacaktır.

ML ve İstatistik Farkları (devam)

- ▶ İstatistikten farklı olarak ML modeli geliştirilirken örneklem - popülasyon değil, train set - test set baz alınır.
- ▶ Model train set üzerinde geliştirilir ancak başarısı test set üzerinde ölçülür
- ▶ Aksi takdirde train set üzerinde modelin karmaşıklığı (complexity) arttıkça tüm veriler üzerinden hatasızca geçen bir model yaratılabilir.
- ▶ Ancak bu model başka bir veri kümesi üzerinde başarılı olmayabilir.
- ▶ Modelin derinliği arttıkça train error düşer ancak test error bir seviyeden sonra artmaya başlar

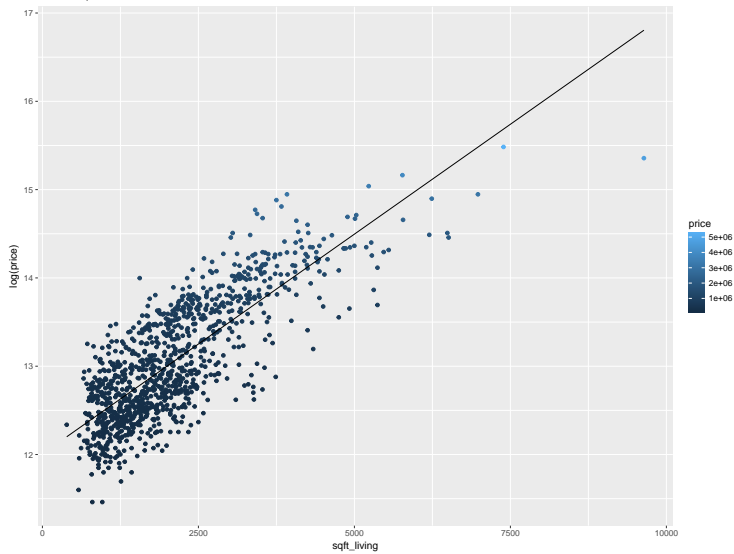
Hata Çeşitleri ve Aşırı Uyum (Overfitting)



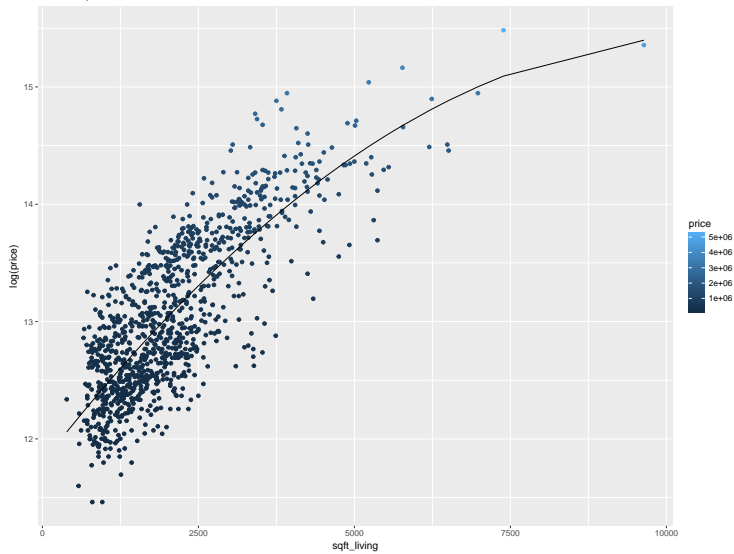
Kaynak:

https://gerardnico.com/wiki/_detail/data_mining/model_complexity_error_training_test.jpg?id=data_mining%3Aoverfitting

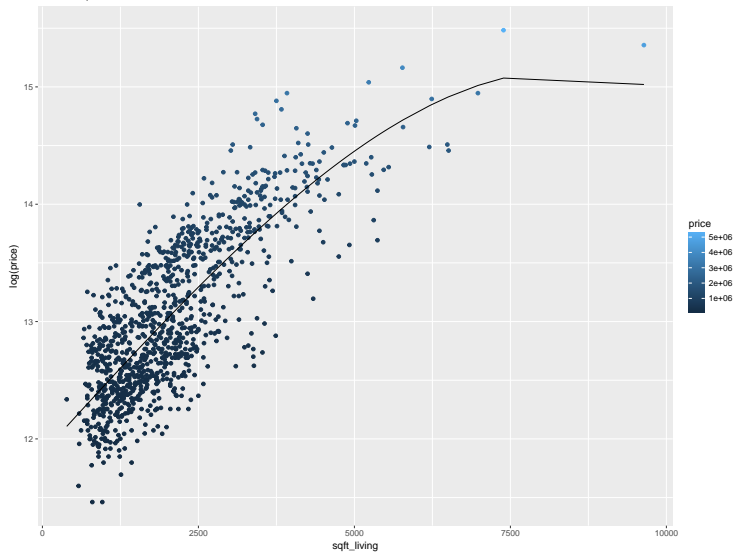
HP vs Sqft, d=1



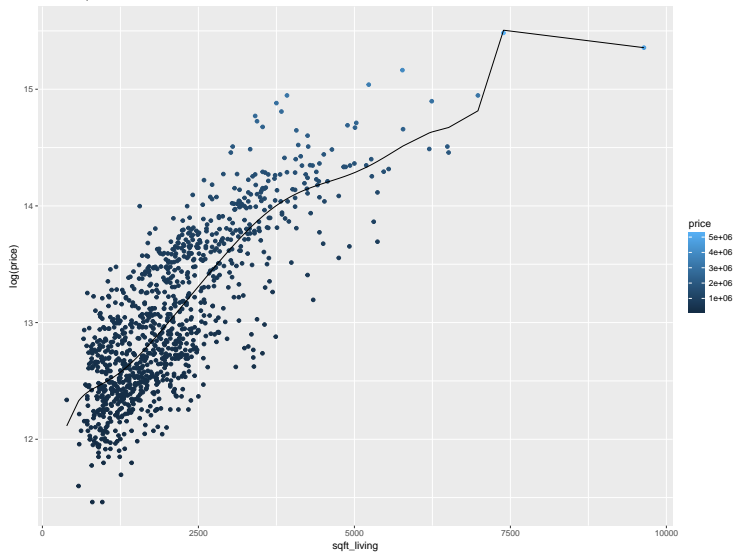
HP vs Sqft, d=2



HP vs Sqft, d=3



HP vs Sqft, d=10



Hatanın 3 Kaynağı

- ▶ Noise (Beyaz Gürültü)
- ▶ Yanlılık (Bias): Modelin gerçekleşen verilerden farklılığı
- ▶ Değişiklik (Variance): Farklı train setler üzerinde elde edilen modellerin değişikliği

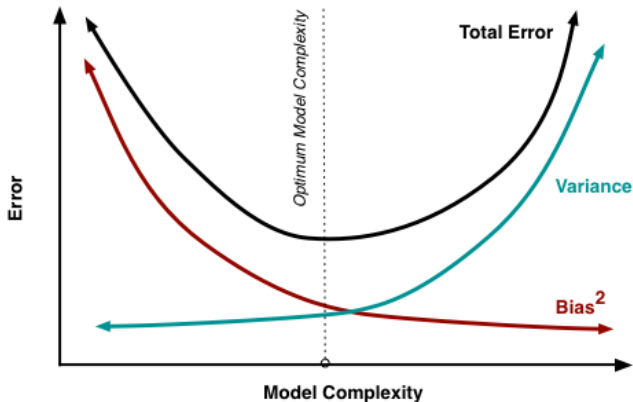
Hata ve Model Derinliği

- ▶ Beyaz gürültü tamamen rassal olup azaltılamaz
- ▶ Yanlılık, model derinliği arttıkça azalır
- ▶ Değişiklik model derinliği arttıkça artar

Öte yandan:

- ▶ Gözlem sayısı arttıkça bias azalır
- ▶ Train error artar, True error azalır

Hata Çeşitleri ve Aşırı Uyum (Overfitting)



Kaynak: <http://scott.fortmann-roe.com/docs/BiasVariance.html>

Sorunun Çözümü

