

Day 15 – ML Theory

Bias–Variance Tradeoff

1. Giriş

Makine öğrenmesinde bir modelin başarısı yalnızca eğitim verisindeki doğruluğuna bağlı değildir. Asıl amaç, **görülmemiş (test / gerçek hayat) verilerde de iyi performans gösterebilen** bir model geliştirmektir.

Bu noktada karşımıza çıkan temel kavram: **Bias–Variance Tradeoff** (Önyargı–Varyans Dengesi)

2. Bias (Önyargı) Nedir?

Bias, modelin gerçek veri dağılımını yeterince öğrenememesi durumudur.

- Model **çok basittir**
- Veri içindeki karmaşık ilişkileri yakalayamaz
- Hem train hem test hatası yüksektir

🔗 Tipik belirtiler:

- Underfitting
- Linear model ile non-linear problemi çözmeye çalışma
- Özellik (feature) sayısının yetersiz olması

🔗 Örnek:

- Gerçek ilişki eğrisel iken sadece doğrusal regresyon kullanmak
-

3. Variance (Varyans) Nedir?

Variance, modelin eğitildiği veriye aşırı duyarlı olmasıdır.

- Model **çok karmaşıktır**
- Gürültüyü (noise) öğrenir
- Train hatası çok düşük, test hatası yüksektir

🔗 Tipik belirtiler:

- Overfitting
- Çok derin decision tree
- Çok fazla feature, az veri

🔗 Örnek:

- Her veri noktasını ezberleyen bir model
-

4. Underfitting vs Overfitting

Durum	Bias	Variance	Train Error	Test Error
Underfitting	Yüksek	Düşük	Yüksek	Yüksek
Ideal Model	Dengeli	Dengeli	Düşük	Düşük
Overfitting	Düşük	Yüksek	Çok düşük	Yüksek

5. Bias–Variance Tradeoff Mantığı

- Model basitleştikçe → **Bias artar, Variance azalır**
- Model karmaşıktıkça → **Bias azalır, Variance artar**

Amaç:

Bias ve Variance arasında optimum dengeyi bulmak

Bu denge, modelin gerçek hayatta en iyi performansı göstermesini sağlar.

6. Bias Yüksekse Ne Yapılır?

Bias problemi varsa:

- Daha karmaşık model kullan
- Feature engineering yap
- Polynomial / non-linear modeller dene
- Regularization gücünü azalt ($\lambda \downarrow$)

Örnek:

- Linear Regression → Random Forest / XGBoost

7. Variance Yüksekse Ne Yapılır?

Variance problemi varsa:

- Daha fazla veri ekle
- Modeli sadeleştir
- Regularization ekle (L1 / L2)
- Feature sayısını azalt
- Cross-validation kullan

Örnek:

- Derin tree → max_depth sınırla

8. Regularization ile İlişkisi

Regularization, variance'ı kontrol altına almak için kullanılır.

- **L1 (Lasso)**: feature selection yapar
- **L2 (Ridge)**: ağırlıkları küçültür
- **ElasticNet**: L1 + L2

🔗 Regularization $\uparrow \rightarrow$ Bias \uparrow , Variance \downarrow

9. Bias–Variance ve Model Seçimi

Model	Bias	Variance
Linear Regression	Yüksek	Düşük
Logistic Regression	Orta	Düşük
Decision Tree (deep)	Düşük	Yüksek
Random Forest	Düşük	Orta
Gradient Boosting	Düşük	Orta–Yüksek
XGBoost (tuned)	Düşük	Dengeli

10. Gerçek Projelerde Kullanımı

Bu kavramlar özellikle şu durumlarda kritik olur:

- Train–test farkı açıklanırken
- Model neden kötü genelliyor sorusunda
- Hyperparameter tuning yaparken
- Feature engineering kararlarında

🔗 Örnek:

- Train accuracy %95, test accuracy %70
→ **Variance problemi**
-

11. Mülakat için Kısa Özet

Bias, modelin çok basit olup gerçek ilişkileri yakalayamamasıdır.
Variance ise modelin veriye aşırı uyum sağlayarak genelleme yapamamasıdır.
İyi bir model, bias ve variance arasında dengede olmalıdır.