

1. Strategi untuk Mengatasi Underfitting pada Linear Regression atau Decision Tree

Underfitting terjadi ketika model tidak mampu menangkap kompleksitas data. Untuk mengatasi hal ini, terdapat dua pendekatan utama:

- **Transformasi atau Penambahan Fitur (Feature Engineering):**
Teknik ini mencakup pembuatan fitur baru dari fitur yang sudah ada (misalnya interaksi antar variabel, polynomial features, atau log-transformasi). Pendekatan ini meningkatkan fleksibilitas model linear tanpa mengganti algoritma. Efek terhadap *bias-variance tradeoff* adalah penurunan bias (karena model menjadi lebih representatif terhadap data), namun berisiko menaikkan varians bila fitur baru terlalu kompleks atau berlebihan.
- **Mengganti ke Model yang Lebih Kompleks:**
Misalnya, mengganti linear regression menjadi Random Forest atau Gradient Boosting Regressor. Model-model ini memiliki kapabilitas untuk menangkap pola yang non-linear secara otomatis. Strategi ini dapat secara signifikan menurunkan bias, namun juga meningkatkan varians, sehingga perlu diseimbangkan dengan teknik validasi silang dan regularisasi.

2. Alternatif Loss Function untuk Regresi: MAE dan Huber Loss

Selain Mean Squared Error (MSE), terdapat dua fungsi kerugian yang umum digunakan untuk regresi:

- **Mean Absolute Error (MAE):**
MAE menghitung rata-rata dari selisih absolut antara nilai aktual dan prediksi. Kelebihannya adalah robust terhadap outlier karena tidak menghitung kuadrat dari error. Namun, kelemahannya adalah kurang sensitif terhadap error kecil, sehingga bisa kurang efektif saat membutuhkan penalti yang tinggi terhadap prediksi buruk.
- **Huber Loss:**
Huber loss menggabungkan kelebihan dari MSE dan MAE. Untuk error kecil, ia bertindak seperti MSE (sensitif), dan untuk error besar, berperilaku seperti MAE (robust). Oleh karena itu, Huber loss lebih cocok digunakan ketika terdapat outlier namun masih ingin mempertahankan sensitivitas terhadap kesalahan kecil. Huber loss ideal untuk data yang memiliki distribusi target non-normal atau mengandung noise.

3. Metode Penilaian Pentingnya Fitur tanpa Mengetahui Nama Fitur

Terdapat beberapa metode untuk menilai pentingnya fitur, meskipun tanpa mengetahui nama variabelnya:

- **Koefisien pada Model Linear:**
Dalam regresi linear, besarnya koefisien (dengan asumsi data sudah dinormalisasi) dapat digunakan untuk mengukur kontribusi masing-masing fitur terhadap prediksi. Namun, koefisien bisa bias jika terjadi multikolinearitas atau skala fitur tidak seragam.
- **Feature Importance berdasarkan Impurity Reduction (seperti dalam Decision Tree):**

Metode ini menghitung seberapa besar suatu fitur berkontribusi dalam menurunkan impurity (misalnya Gini atau entropy). Meskipun populer, metode ini cenderung memberikan skor tinggi kepada fitur dengan lebih banyak nilai unik (high cardinality) dan bisa bias terhadap skala.

4. Desain Eksperimen untuk Pemilihan Hyperparameter Optimal

Untuk memilih hyperparameter optimal, pendekatan seperti **Grid Search** atau **Random Search** dengan validasi silang (cross-validation) dapat digunakan. Misalnya:

- **SGDRegressor:** tuning *learning rate* untuk memastikan konvergensi yang stabil.
- **Decision Tree:** tuning *max_depth* untuk mengontrol kompleksitas model.

Tradeoff yang harus dianalisis:

- **Komputasi:** Grid search sangat mahal secara waktu, terutama dengan banyak parameter.
- **Stabilitas Pelatihan:** Learning rate yang terlalu besar dapat menyebabkan pelatihan tidak stabil.
- **Generalisasi:** Hyperparameter yang terlalu kompleks (misalnya depth terlalu tinggi) bisa menyebabkan overfitting.

Solusi terbaik adalah memilih kombinasi parameter yang memaksimalkan kinerja validasi namun tetap mempertahankan kesederhanaan model.

5. Penanganan Pola Non-Linear dan Heteroskedastisitas pada Linear Regression

Jika residual plot menunjukkan adanya pola non-linear dan variansi residual yang tidak konstan (heteroskedastisitas), maka beberapa langkah yang dapat diambil adalah:

- **Transformasi Data:**
Terapkan transformasi seperti log, square root, atau Box-Cox pada variabel target atau fitur. Ini dapat membantu menstabilkan variansi dan menangkap pola non-linear.
- **Mengubah Model:**
Gunakan model regresi non-linear seperti polynomial regression atau ganti ke model yang lebih fleksibel seperti Random Forest, Gradient Boosting, atau Support Vector Regression.
- **Gunakan Weighted Regression:**
Jika heteroskedastisitas tetap muncul, gunakan regresi tertimbang (Weighted Least Squares) agar error yang besar tidak mendominasi pelatihan.

Jika kamu ingin penjelasan versi ini dalam bentuk slide, ringkasan visual, atau format essay lebih panjang, aku juga bisa bantu. Mau?