

Analisis Model Regresi - Tugas Essay

1. Strategi Mengatasi Underfitting pada Linear Regression dan Decision Tree

Underfitting terjadi ketika model terlalu sederhana untuk menangkap pola data. Pada Linear Regression dan Decision Tree, dua pendekatan yang dapat digunakan adalah:

a. Transformasi Fitur (Feature Transformation): - Untuk Linear Regression, penambahan fitur polinomial (Polynomial Features) memungkinkan model menangkap hubungan non-linear. - Contoh: Menambahkan fitur kuadrat atau interaksi antar fitur. - **Dampak pada Bias-Variance Tradeoff:** Menurunkan bias karena model menjadi lebih fleksibel, tetapi risiko varians meningkat (overfitting jika derajat polinomial terlalu tinggi).

b. Penggantian Model ke Algoritma Lebih Kompleks: - Misalnya mengganti Linear Regression ke Random Forest Regressor atau Gradient Boosting. - Decision Tree juga bisa diubah menjadi ensemble seperti Bagging atau Boosting. - **Dampak pada Bias-Variance Tradeoff:** Bias turun karena model lebih kompleks, varians bisa dikontrol lewat regularisasi atau pruning.

2. Alternatif Loss Function untuk Regresi

a. Mean Absolute Error (MAE): - MAE mengukur rata-rata absolut selisih antara nilai prediksi dan aktual. - **Kelebihan:** Lebih robust terhadap outlier karena tidak mengkuadratkan error. - **Kelemahan:** Tidak peka terhadap besar error (outlier berpengaruh kecil). - **Cocok:** Dataset dengan outlier atau distribusi non-Gaussian.

b. Huber Loss: - Kombinasi MAE dan MSE; menggunakan kuadrat error untuk error kecil dan absolut error untuk error besar. - **Kelebihan:** Seimbang antara sensitivitas MSE dan robust MAE. - **Kelemahan:** Memerlukan hyperparameter delta (threshold) yang harus disetel. - **Cocok:** Dataset dengan outlier sedang (tidak ekstrem) dan ingin kompromi antara MAE dan MSE.

3. Metode Pengukuran Pentingnya Fitur

a. Koefisien Regresi (untuk Linear Regression): - Menunjukkan seberapa besar pengaruh fitur terhadap target. - **Prinsip:** Semakin besar koefisien (positif/negatif), semakin besar pengaruh fitur. - **Keterbatasan:** Hanya untuk hubungan linear, sensitif terhadap multikolinearitas.

b. Feature Importance dari Tree-Based Model: - Berdasarkan impurity reduction (seperti Gini atau Entropy pada Decision Tree). - **Prinsip:** Semakin sering fitur digunakan untuk split, semakin penting. - **Keterbatasan:** Bias terhadap variabel dengan lebih banyak kategori atau skala besar.

4. Desain Eksperimen Hyperparameter Tuning

- a. Teknik Grid Search atau Randomized Search:** - Melakukan cross-validation untuk setiap kombinasi hyperparameter (misal max_depth di Decision Tree).
- b. Analisis Tradeoff:** - **Komputasi:** Grid Search mahal waktu jika parameter banyak. - **Stabilitas Pelatihan:** Randomized Search lebih cepat, risiko missing optimal. - **Generalisasi:** Cross-validation mencegah overfitting ke data latih.
- c. Alternatif:** Bayesian Optimization untuk efisiensi lebih baik.

5. Analisis Residual Plot pada Linear Regression

Jika residual plot menunjukkan pola non-linear atau heteroskedastisitas:

- a. Transformasi Data:** - Log, square root, atau Box-Cox transform untuk stabilisasi variansi.
- b. Ubah Model:** - Gunakan Polynomial Regression atau model non-linear seperti Decision Tree atau SVR.
- c. Regularisasi:** - Gunakan Ridge/Lasso Regression untuk mengurangi variansi berlebih.
- d. Heteroskedastisitas:** - Terapkan metode robust regression (misalnya Huber Regressor).