

Praktikum Fisika Komputasi

MACHINE LEARNING REGRESI LINEAR DAN POLINOMIAL

Senin, 25 November 2024

Mutiara Rachmatul Fajriyah (1227030024)

- Kode Program Pertama

Program ini menggunakan *library* Python scikit-learn untuk membangun model regresi linier sederhana. Tujuannya adalah memprediksi nilai keluaran (output) berdasarkan nilai masukan (input) menggunakan data yang diberikan. Disini *library* numpy digunakan untuk mengolah data numerik (array), dan *LinearRegression* digunakan untuk membuat model regresi linier. Kemudian x adalah data input (variabel bebas), berbentuk array 2 dimensi karena diperlukan oleh model, dan y adalah data output (variabel target), yaitu hasil yang diinginkan untuk diprediksi. Kemudian kita membuat model regresi linier dengan *LinearRegression*, dan melatih (*fit*) model menggunakan data input x dan output y . Lalu menghitung skor R^2 (koefisien determinasi) untuk mengevaluasi seberapa baik model sesuai dengan data. Selanjutnya melakukan prediksi data input baru yaitu 6, dan modelnya menggunakan `regr.predict()` untuk memprediksi output berdasarkan input tsb. Kemudian hasil prediksinya ditampilkan. Dan hasilnya untuk inputnya 6 dan outputnya 7.0 yang artinya jika masukan bernilai 6, maka model memprediksi keluaran bernilai 7.0.

- Kode Program Kedua

Program ini menggunakan beberapa *library*, seperti *PolynomialFeatures* untuk mengubah data menjadi bentuk polinomial, *linear_model* untuk membuat model regresi linier, dan *numpy* untuk mengolah data numerik. Data input x diberikan dalam bentuk array 2 dimensi, berisi angka-angka seperti [0], [4], [8], dan seterusnya, sementara data output y adalah hasil yang diinginkan untuk setiap nilai x . Dalam data ini, output mengikuti pola kuadrat, seperti $y = x^2 + c$. Kemudian, nilai yang ingin diprediksi adalah 12, yang didefinisikan dalam array `predict`. Agar sesuai dengan regresi polinomial, data input x dan `predict` diubah menggunakan *PolynomialFeatures(degree=2)* sehingga mencakup nilai asli, nilai pangkat dua, dan bias (konstanta). Hasilnya, data diubah menjadi format seperti [1, x , x^2]. Setelah transformasi, model regresi linier dibuat menggunakan *LinearRegression()* dan dilatih dengan data input yang sudah diubah (polinomial) dan output y menggunakan metode `.fit()`. Terakhir, program menggunakan metode `.predict()` untuk memprediksi output berdasarkan nilai input 12 yang sudah diubah menjadi polinomial. Output hasil prediksi ditampilkan, misalnya 15, yang berarti untuk input 12, model memprediksi hasilnya adalah 15. Program ini sederhana tetapi efektif untuk mempelajari pola data kuadrat menggunakan regresi polinomial.

- Kode Program ketiga

Pertama, data input X berupa nilai 1 sampai 10, dan output Y mengikuti pola $Y = X^2 + 2X + 1$, menghasilkan nilai yang terus meningkat seperti [3, 7, 13, ..., 111]. Data ini dibagi menjadi dua bagian: **data latih (80%)** untuk membangun model dan **data uji (20%)** untuk mengevaluasi performa model, menggunakan `train_test_split`. Model pertama adalah **regresi linear**, yang mencoba menemukan garis lurus terbaik yang mendekati pola data. Model kedua adalah **regresi polinomial derajat 2**, yang lebih kompleks dan dapat menangkap pola data kuadrat dengan lebih baik. Untuk regresi polinomial, data input X diubah menggunakan `PolynomialFeatures(degree=2)` menjadi bentuk yang mencakup nilai X, X^2 , dan bias. Setelah kedua model dilatih pada data latih, prediksi dilakukan untuk seluruh data input (baik linear maupun polinomial). Performa kedua model dievaluasi dengan menghitung **Mean Squared Error (MSE)** pada data uji. MSE menunjukkan seberapa besar rata-rata kesalahan prediksi model dibandingkan dengan nilai aktual. Hasilnya, **MSE untuk regresi linear lebih tinggi** dibandingkan regresi polinomial, menunjukkan bahwa model polinomial lebih cocok untuk data kuadrat.

Hasil MSE

MSE Regresi Linear: Hasilnya lebih besar karena model linear tidak mampu mengikuti pola data kuadrat dengan baik.

MSE Regresi Polynomial Degree 2: Nilainya lebih kecil karena model polinomial mampu menangkap pola data dengan lebih baik.

Grafik

Grafik yang dihasilkan menunjukkan perbandingan antara model dan data asli:

- Titik biru: Data latih (digunakan untuk melatih model).
- Titik hijau: Data uji (digunakan untuk mengevaluasi model).
- Garis merah: Hasil prediksi regresi linear, berupa garis lurus.
- Garis kuning: Hasil prediksi regresi polinomial degree 2, berupa kurva yang sesuai dengan pola kuadrat.

Dari grafik, terlihat bahwa garis merah (linear) tidak dapat mengikuti pola data secara akurat, sedangkan garis kuning (polinomial) menyesuaikan data dengan lebih baik.