Regresi Polinomial

Regresi Polinomial adalah bentuk regresi yang menggunakan polinom untuk memodelkan hubungan antara variabel independen dan dependen. Ini memungkinkan kita untuk menangkap hubungan non-linear yang tidak dapat dimodelkan dengan regresi linear.

# Mengapa Regresi Polinomial?

Regresi linear mengasumsikan hubungan linear variabel independen dan dependen. Namun, dalam banyak kasus, hubungan tersebut mungkin non-linear. Misalnya, hubungan antara usia dan pendapatan mungkin berbentuk kurva, di mana pendapatan meningkat dari variabel independen.

# Persamaan Regresi Polinomial:

y = b0 + b1x + b2x^2 + ... + bnx^n + e

di mana:

y: variabel dependen

x: variabel independen

B0, b1, …, bn: koefisien regresi

n: derajat polinomial

e: error

# Derajat Polinom

Derajat polinom menentukan kompleksitas model. Polinom derajat 1 adalah model linear, polinom derajat 2 adalah kuadratik, polinom derajat 3 adalah model kubik, dan seterusnya.

# Overfitting dalam Regresi Polinomial

Penting untuk berhati-hati dalam memilih derajat polinom. Jika derajat terlalu tinggi, model dapat overfitting, yaitu terlalu menyesuaikan diri dengan data latih dan tidak dapat menggeneralisasi dengan baik pada data baru.

# Penerapan Regresi Polinomial dengan Scikit-Learn

Scikit-learn tidak memiliki kelas khusus untuk regresi polinomial, tetapi kita dapat menggunakan PolynomialFeature untuk membuat fitur polinomial dan kemudian menggunakan LinearRegression untuk melakukan regresi.

**Tugas 22:**

Lanjutkan dengan dataset “Advertising”

**Tugas**

1. **Membuat Fitur Polinomial:**

* Gunakan PolynomialFeature dari scikit-learn untuk membuat fitur polinomial derajat 2 dari variabel TV.
* Gabungkan fitur polinomial ini dengan fitur asli lainnya (Radio dan Newspaper).

1. **Membuat Model Regresi Polinomial:**

* Gunakan LinearRegression dari scikit-learn untuk membuat model regresi.
* Gunakan fitur polinomial dan fitur asli lainnya sebagai input, dan Sales sebagai target.
* Latih model pada data latih.

1. **Visualisasi:**

* Buat scatter plot antara TV dan Sales.
* Tambahkan garis regresi polinomial pada plot

**Contoh Kode (Scikit-Learn dan Matplotlib):**

from sklearn.preprocessing import PolynomialFeatures

from sklearn.linear\_model import LinearRegression

# ... (kode untuk membaca dan mempersiapkan dataset Advertising)

# Membuat fitur polinomial

poly = PolynomialFeatures(degree=2)

X\_train\_poly = poly.fit\_transform(X\_train[['TV']])

X\_test\_poly = poly.transform(X\_test[['TV']])

# Menggabungkan fitur polinomial dengan fitur lain

X\_train\_poly = np.concatenate((X\_train\_poly, X\_train[['Radio', 'Newspaper']]), axis=1)

X\_test\_poly = np.concatenate((X\_test\_poly, X\_test[['Radio', 'Newspaper']]), axis=1)

# Membuat model regresi polinomial

model = LinearRegression()

model.fit(X\_train\_poly, y\_train)

# Visualisasi

plt.scatter(df['TV'], df['Sales'], color='blue')

plt.plot(df['TV'], model.predict(poly.fit\_transform(df[['TV']])), color='red')

plt.title('Regresi Polinomial (TV vs Sales)')

plt.xlabel('TV')

plt.ylabel('Sales')

plt.show()

**Catatan:**

* Coba bereksperimen dengan derajat polinom yang berbeda (misalnya, 3 atau 4) dan lihat bagaimana pengaruhnya terhadap model.
* Anda juga dapat mencoba membuat fitur polinomial dari variabel Radio atau Newspaper.

**Tugas**

1. **Membuat Fitur Polinomial**

**from sklearn.preprocessing import PolynomialFeatures**

**from sklearn.linear\_model import LinearRegression**

**import numpy as np**

**# Membuat fitur polinomial**

**poly = PolynomialFeatures(degree=2)**

**X\_train\_poly = poly.fit\_transform(X\_train[:, 0].reshape(-1, 1)) # Ubah menjadi array**

**X\_test\_poly = poly.transform(X\_test[:, 0].reshape(-1, 1)) # Ubah menjadi array**

1. **Membuat Model Regresi Polinomial**

**# Membuat model regresi polinomial**

**model = LinearRegression()**

**model.fit(X\_train\_poly, y\_train)**

1. **Visualisasi**

**# Visualisasi**

**# Membuat data untuk visualisasi dengan range TV yang lebih halus**

**X\_grid = np.arange(min(df['TV']), max(df['TV']), 0.1)**

**X\_grid = X\_grid.reshape(-1, 1)**

**# Visualisasi**

**plt.scatter(df['TV'], df['Sales'], color='blue')**

**plt.plot(X\_grid, model.predict(poly.fit\_transform(X\_grid)), color='red')**

**plt.title('Regresi Polinomial (TV vs Sales)')**

**plt.xlabel('TV')**

**plt.ylabel('Sales')**

**plt.show()**