

Polynomial Regression

Brandewa Pandu A. (2343110

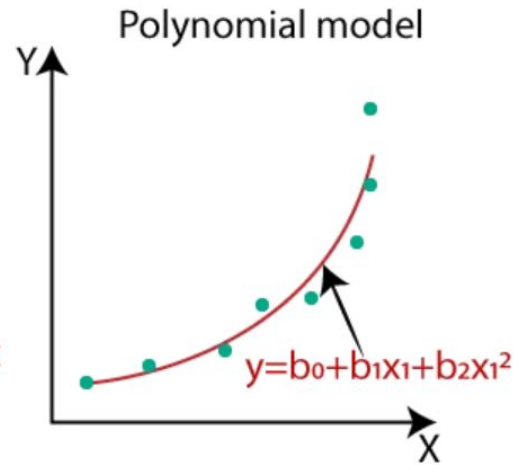
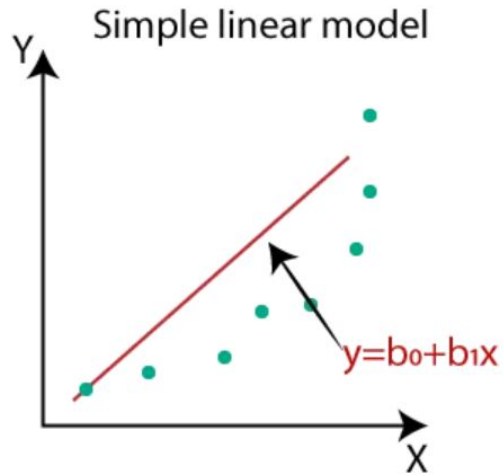
Rizqi Agus S. (2343110

Satria Bagus A.I (2343110

Polynomial Regression

Polynomial Regression adalah salah satu jenis regresi yang memodelkan hubungan antara variabel X dan Y menggunakan persamaan polinomial (berpangkat). Metode ini digunakan ketika pola data bersifat melengkung, bukan linear.

Regresi polinomial digunakan ketika titik-titik dalam data yang tidak bisa ditangkap oleh Model Linear Regression dan regresi linear tidak bisa menggambarkan hasil dengan baik dan jelas.



Terdapat contoh perbedaan antara linear regresi dengan polinomial. dimana Model dari linear regresi tidak bisa menangkap data yang tidak berdekatan dengan garis merah. Sebaliknya untuk Model Polinomial dapat menangkap model dengan baik dan jelas.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Penjelasan rumus

- **Y** = Variabel yang ingin diprediksi
- **X** = Variabel Independen
- **B0** = Intersep (Konstanta)
- **B1, B2, B3** = Koefisien untuk setiap pangkat X
- **H** = Derajat Polinomial
- **E** = Selisih

Cost Function

Cost function disebut juga dengan **MSE** Dimana (Mean Squared Error) bertujuan untuk mengukur seberapa besar total kesalahan (Error) dari model kita. Cost Function Dirumuskan sebagai Berikut

$$J = \frac{1}{n} \sum (pred - (b_0 + b_1x + b_2x^2 + b_3x^3 + \dots))^2$$

Artinya:

- Cost function polynomial regression menghitung rata-rata kesalahan antara prediksi model (kurva) dan data asli. Semua error dijumlahkan, kemudian dirata-rata, lalu dikuadratkan agar jelas seberapa jauh model meleset.

Hipotesa Function

Jika Anda menggunakan polinomial **orde 2** (seperti yang Anda tunjukkan sebelumnya):

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

Jika Anda menggunakan polinomial **orde 3**:

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

Fungsi Hipotesa (h_0) Adalah formula yang digunakan untuk membuat model untuk membuat prediksi. Orde adalah persamaan garis kurva. Orde adalah pangkat tertinggi dari persamaan variabel (independen) dalam persamaan regresi.

Fungsi hipotesa, dilambangkan sebagai $h_0(x)$ atau $h(x)$, adalah representasi matematis dari model yang kita coba pelajari. Ini adalah persamaan yang akan kita gunakan untuk memprediksi nilai Y berdasarkan X .

Contoh Kasus

Prediksi harga tanah berdasarkan luas tanah.

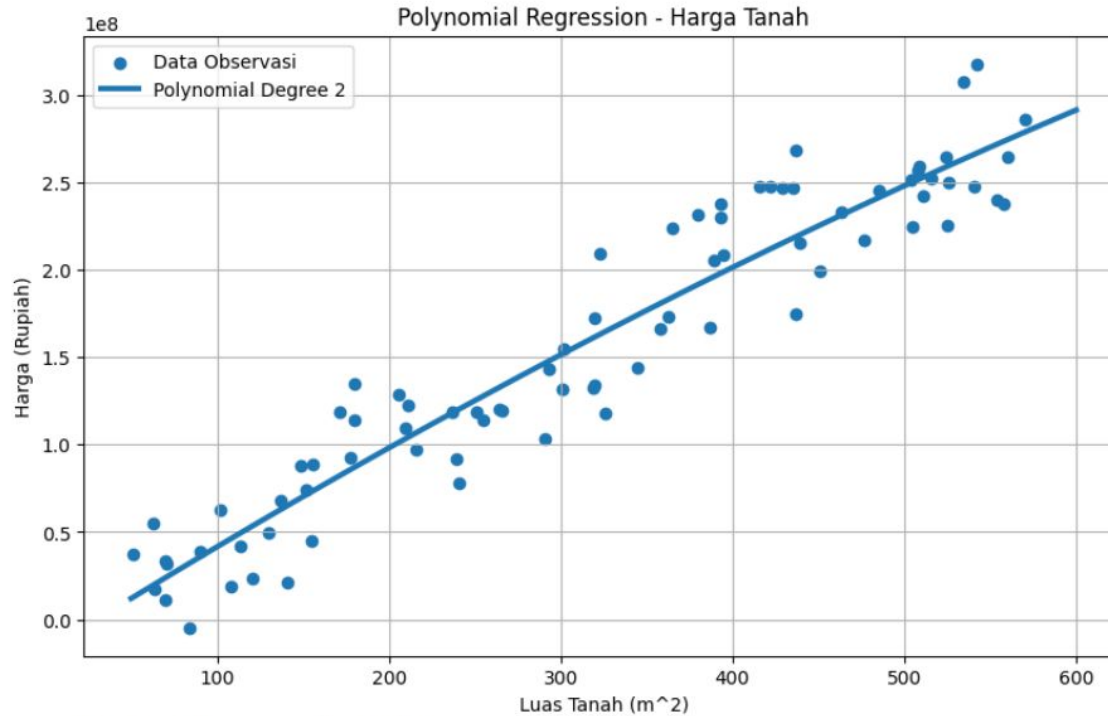
- Karena data sering tidak linear, kita menggunakan polynomial regression.

Harga rumah (dipengaruhi luas, lokasi, umur bangunan)

Pengaruh **dosis obat** vs efek samping

Hasil

- Degree = 2
MSE = 670241545760843.2
RMSE = 25889023.65406705
 $R^2 = 0.9064495245427585$



Luas 100 m² → Prediksi harga ≈ Rp 41,548,402
Luas 200 m² → Prediksi harga ≈ Rp 98,098,277
Luas 300 m² → Prediksi harga ≈ Rp 151,343,074
Luas 500 m² → Prediksi harga ≈ Rp 247,917,434

Penjelasan Hasil

Kita menggunakan Polynomial Regression derajat 2

MSE = Mean Squared Error

Nilainya besar karena data harga tanah dalam satuan Rupiah sangat besar

RMSE = 25 juta

rata-rata error model adalah sekitar \pm Rp 25 juta dari harga sebenarnya.

$R^2 = 0.906 = 90.6\%$

Model mampu menjelaskan 90.6% variasi harga tanah berdasarkan luas tanah.

Titik-titik biru (data observasi) menunjukkan data asli harga tanah.

Terlihat bahwa semakin besar luas tanah, harga semakin meningkat — namun tidak selalu naik secara lurus, ada sedikit kelengkungan

Garis biru menggambarkan prediksi terbaik dari model Polynomial Regression.

Model ini mencoba mengikuti pola data yang tidak sepenuhnya linear.

Tujuan Polynomial Regression

1. Memodelkan Hubungan Non-Linier

Ini adalah tujuan paling fundamental. Ketika Anda memplot data dan melihat pola yang melengkung (misalnya, berbentuk parabola atau kurva S), Regresi Polinomial memungkinkan Anda untuk **menangkap dan merepresentasikan kelengkungan** ini.

- **Regresi Linier:** Memaksakan garis lurus ($Y = B_0 + B_1X$) pada data.
- **Regresi Polinomial:** Menggunakan kurva ($Y = B_0 + B_1X + B_2X^2 + \dots$) untuk mengikuti pola data yang sebenarnya.

2. Peningkatan Akurasi Model

Dengan menggunakan kurva yang lebih fleksibel, Regresi Polinomial seringkali menghasilkan **kesalahan prediksi (error) yang lebih kecil** dibandingkan model linier pada data yang non-linier. Ini berarti model Anda memiliki **kekuatan penjelas (explanatory power)** yang lebih baik untuk data *training*.

3. Transformasi ke Model Linier

Meskipun bertujuan memodelkan hubungan non-linier, tujuan teknisnya adalah untuk **mengubah masalah non-linier menjadi masalah Regresi Linier Berganda**.

- Dengan mengubah fitur X menjadi fitur baru seperti X^2 , X^3 , dan seterusnya, kita sebenarnya dapat menggunakan algoritma Regresi Linier standar (seperti *Least Squares* atau *Gradient Descent*) untuk menemukan koefisien optimal. Model tersebut tetap **linier dalam parameter** (B).

Kapan Polinomial Digunakan

Polinomial digunakan ketika Hubungan antara X dan Y tidak linear (melengkung). Hal ini terjadi ketika garis membentuk sebagai berikut:

- Pola U
- Garis Kurva naik
- Garis Kurva Turun
- Pola Parabola

Contoh untuk studi kasus nya pada Harga tanah vs Luas Tanah, Dimana harga makin naik garis kurva juga kan melngkung.

Kapan Polinomial Tidak Digunakan

Regresi Polinomial tidak digunakan ketika memiliki banyak derajat. Dimana derajat terlalu tinggi **B1, B2, B3x2 (kuadrat)**, Hal ini menyebabkan over fitting.

Data sangat kompleks atau sangat besar Jika datanya ratusan ribu baris atau pola ekstrem kompleks (misal pola zig-zag harian), polynomial regression tidak memadai.

Kesimpulan

Regresi Polinomial adalah alat statistik dan pembelajaran mesin yang merupakan **ekstensi dari Regresi Linier Berganda**, yang dirancang khusus untuk memodelkan **hubungan non-linier (melengkung)** antara variabel independen (X) dan variabel dependen (Y).