Pengenalan Pola/ Pattern Recognition

Regresi Linier

Imam Cholissodin S.Si., M.Kom.

Pokok Pembahasan

1. Regresi Linier
   - ✓ Pengertian Regresi Linier
   - ✓ Model Regresi Linier
   - ✓ Visualisasi Regresi Linier
2. Case Study

Regresi Linier

- Regresi adalah membangun model untuk memprediksi nilai dari data masukan yang diberikan.
- Prediksi berbeda dengan klasifikasi (dalam *machine learning*, klasifikasi dianggap sebagai salah satu jenis dari prediksi).
- Klasifikasi digunakan untuk memprediksi label kelas/kategori.
- Metode utama untuk melakukan prediksi : Membangun model regresi yaitu dengan mencari hubungan antara satu atau lebih variabel independen atau prediktor (X) dengan variabel dependen atau respon (Y).
- Macam-Macam Analisis Regresi :
    - Linear and multiple regression.
    - Non-linear regression (neural networks, support vector machines).
    - Other regression methods : generalized linear model, Poisson regression, log-linear models, regression trees.

Model Regresi Linier

- Model Regresi Linier Sederhana :

  $y = \alpha + \beta x + \varepsilon$ *atau* $\mu_{y|x} = \alpha + \beta x$

- Dimana y adalah variabel respon (dependent), atau variabel yang ingin kita prediksi, x adalah variabel prediktor (independen) dan ε adalah variabel tingkat kesalahan yang merupakan satu-satunya komponen acak dalam model regresi.
- Keterangan :
  - $\mu_{y|x}$ adalah mean dari y dengan syarat x yang telah ditentukan (atau nilai x telah diketahui sebelumnya), $\mu_{y|x}$ disebut juga sebagai rata-rata bersyarat y.
  - α adalah titik potong garis regresi pada sumbu koordinat.
  - β adalah besarnya gradien/kemiringan garis regresi.

Model Regresi Linier

- Model Umum Regresi Linier :

$$Y = b_0 + b_1 X + \varepsilon$$

- Keterangan :
  - b_0 dan b_1 adalah parameter yang akan ditentukan nilainya untuk membangun persamaan regresi.
  - X telah diketahui sebelumnya dan bernilai konstan.
  - Deviasi/Penyimpangan nilai ε bersifat independent dan berdistribusi Normal $\varepsilon \sim N(0,\sigma^2)$.
  - Nilai-nilai parameter regresi b_0 dan b_1 tidak diketahui sebelumnya. Kita perkirakan nilai mereka dengan menghitung dari data yang ada.
  - b_1 menunjukkan tingkat perubahan untuk setiap kenaikan nilai X.
- Hasil Estimasi Persamaan Regresi :

$$\hat{y} = b_0 + b_1 x$$

Model Regresi Linier

- Estimasi Persamaan Umum Regresi :

$$\hat{y} = b_0 + b_1 x$$

- Menghitung *sum of squared errors* (SSE):

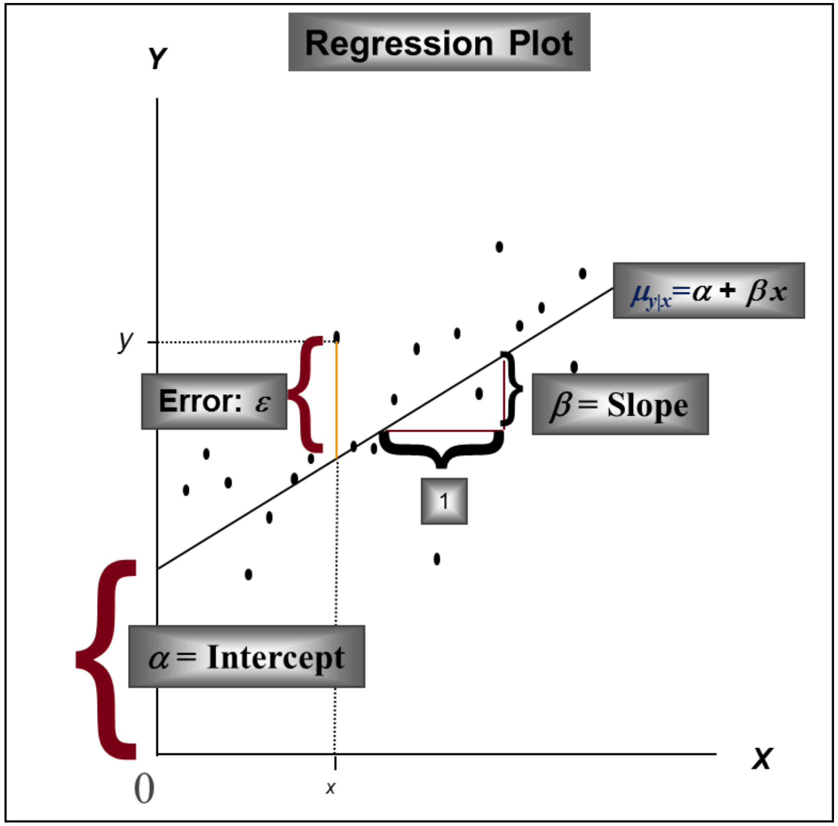
$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y - b_0 - b_1 x\right)^2$$

- Metode *least squares*/ kuadrat terkecil memberikan kita hasil estimasi "terbaik" untuk kita set pada data sampel.
- Metode *least squares* / kuadrat terkecil memilih nilai-nilai b_0 dan b_1 untuk meminimalkan *sum of squared errors* (SSE).

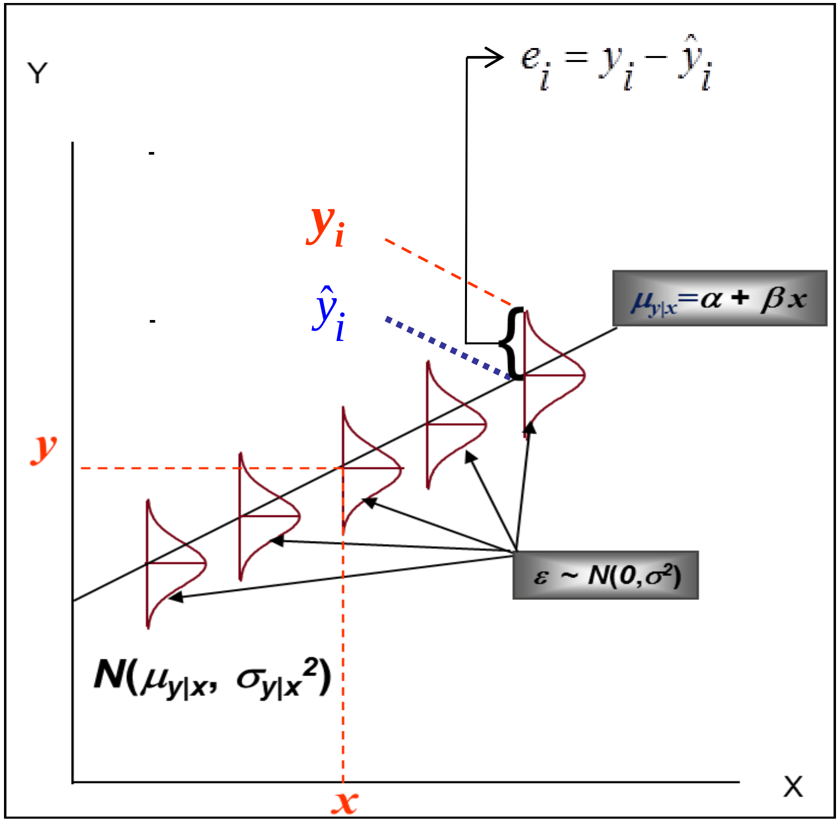
$$b_1 = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n\sum x^2 - \left(\sum x\right)^2} \qquad b_0 = \bar{y} - b_1\bar{x} = \frac{\left(\sum y - b_1 \sum x\right)}{n}$$

Visualisasi Regresi Linier

- Plot Regresi :



- Plot Error Regresi :



Contoh Studi Kasus

- Perhatikan data Biaya iklan yang digunakan (X) dan hubungannya dengan Tingkat penjualan (Y) diberikan dalam dataset berikut :

No	X	Y
1	41	1250
2	54	1380
3	63	1425
4	54	1425
5	48	1450
6	46	1300
7	62	1400
8	61	1510
9	64	1575
10	71	1650

Tentukan persamaan Regresinya!

Contoh Studi Kasus

- Penyelesaian :
  - Mengestimasi least squares/kuadrat terkecil dari koefisien Regresi :

$$n = 10 \qquad \sum x = 564 \qquad \sum x^2 = 32604$$

$$\sum y = 14365 \qquad \sum xy = 818755$$

$$b_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{10(818755) - (564)(14365)}{10(32604) - (564)^2} = 10.8$$

$$b_0 = \frac{\left(\sum \mathrm{y} - \mathrm{b}_1 \sum \mathrm{x}\right)}{\mathrm{n}} = 1436.5 - 10.8(56.4) = 828$$

$$\hat{y} = b_0 + b_1 x$$

  - Hasil Estimasi Persamaan Regresinya adalah :

$$\hat{\mathrm{y}} = 828 + 10.8\mathrm{x}$$

Ini berarti bahwa jika biaya iklan meningkat sebesar $ 1, maka kita akan mendapati tingkat penjualan naik $ 10.8

Selesai