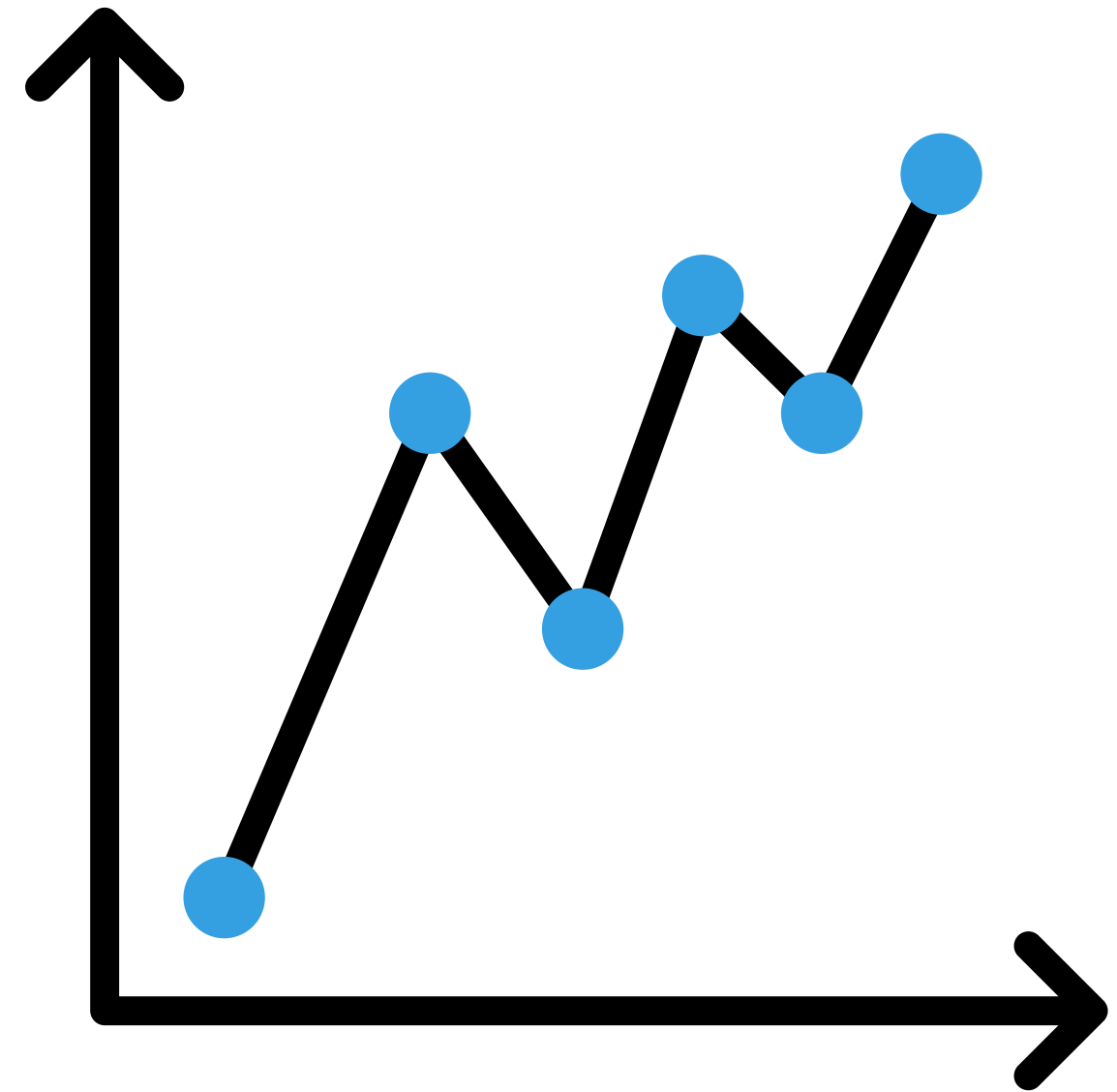


Kelompok 1

# LOGISTIC REGRESSION



# Apa Itu Logistic Regression?

Logistic Regression adalah metode untuk memprediksi probabilitas suatu data masuk ke kelas 1 atau kelas 0.

Metode ini digunakan untuk klasifikasi biner, bukan regresi angka.

# Hipotesa Function

Hipotesa function ( $h(x)$ ) pada Logistic Regression adalah fungsi yang mengubah hasil fungsi linear ( $z$ ) menjadi probabilitas menggunakan fungsi sigmoid. Nilai probabilitas ini digunakan untuk menentukan apakah data termasuk kelas 0 atau 1.

**Rumus:**

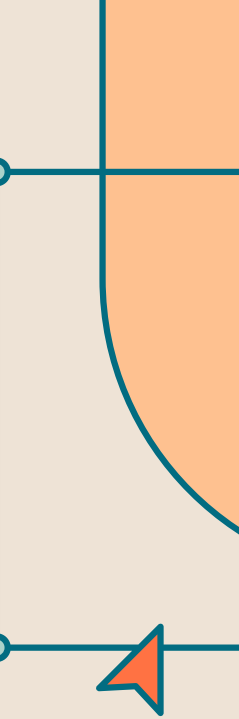
$$h(x) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

$h(x)$  = prediksi probabilitas

$\sigma$  = fungsi sigmoid

$z$  = fungsi linear

$e$  = konstanta euler



# Fungsi Linear (z)

Sebelum masuk ke fungsi sigmoid, logistic regression menghitung nilai z menggunakan fungsi linear:

**Rumus:**

$$z = w \cdot x + b$$

$w$  = bobot

$x$  = fitur

$b$  = bias

## Laboratorium Logistic Regression

Eksperimen bagaimana Bobot (w), Bias (b), dan Threshold mengubah nasib.

### RUMUS DI BALIK LAYAR

$$P = \text{Sigmoid}(w \cdot x + b)$$

Langkah 1: Skor Linear (z)

$$z = (2 \times 3) + 0 = 6.00$$

Langkah 2: Bedah Sigmoid (P)

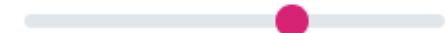
a. $e^{-6.00}$	= 0.0025
b. $1 + 0.0025$	= 1.0025
c. $1 / 1.0025$	= 0.9975

### DECISION BOUNDARY

Batas Lulus saat  $P > 0.50$

Terjadi saat  $x = 0.00$

Input (x): Jam Belajar 3



Geser ini untuk mensimulasikan siswa yang berbeda.

Bobot (w): "Ngefek-nya" 2



Kecuraman kurva (Sensitivitas).

Bias (b): Titik Awal 0



Geser Kurva ke Kiri/Kanan.

Threshold (Batas Lulus) 50%



Standar kelulusan. Rendah = Mudah Lulus.



Sesuai pada slide awal tadi, untuk menghitung Hipotesa Function menggunakan rumus :

$$h(x) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

Kita butuh nilai z . Maka :

$$z = w \cdot x + b$$

Dari gambar disamping dilakukanlah perhitungan untuk mencari fungsi linear di samping kiri grafik, dengan cara substitusi dari variabel input (x), bobot (w), dan bias (b).

Singkatnya, didapatkanlah nilai sebesar 6.00 untuk nilai z-nya, Itu berarti fungsi linear  $(z) = 6,00$  seperti pada **Langkah 1**.

Langkah selanjutnya, substitusikanlah nilai z ke dalam fungsi  $h(x)$ , seperti yang dijelaskan dalam perhitungan **Langkah 2**.

Didapatlah  $h(x) = 0,9975$ . Itu berarti peluang/probabilitas yang dihasilkan adlaah sebesar 0,9975

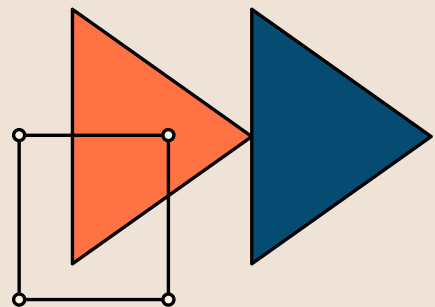
Pada gambar, threeshold diatur sebesar 50% (0,5), maka  $0,9975 > 0,5$ .

**Maka keputusan = 1 (LULUS)**

# Cost Function

Cost Function adalah rumus untuk mengukur seberapa besar kesalahan (error) antara prediksi model dengan label sebenarnya.

Semakin kecil nilai cost, semakin baik model memprediksi.



$$\text{Cost} = -[ y \cdot \log(\mathbf{h(x)}) + (1 - y) \cdot \log(1 - \mathbf{h(x)}) ]$$

Bagian Kiri (Jika  $y=1$ )    Bagian Kanan (Jika  $y=0$ )

Keterangan:

1.  $y$  = Label asli (kebenaran) dari data.
2.  $h(x)$  = Nilai prediksi model (output dari fungsi sigmoid).
3.  $\log$  = Digunakan untuk memberikan hukuman yang besar saat prediksi jauh dari nilai aslinya.

# Dua Skenario Perhitungan Label

Skenario untuk Data LULUS (True / 1).

$$Cost = -\log(h(x))$$

Sesuai dengan hasil perhitungan pada siswa lulus tadi.  
Karena Lulus ( $y = 1$ ), maka :

$$Cost = -\log(h(x))$$

Masukkan angkanya:

$$Cost = -\log(0.9975)$$

(Menggunakan kalkulator logaritma natural / ln)

$$Cost = -(-0.0025)$$

$$Cost = 0.0025$$

Skenario untuk Data TIDAK LULUS (False / 0).

$$Cost = -\log(1 - h(x))$$

Kita bikin simulasi tidak lulus, pakai nilai 0,25

$$Cost = -\log(1 - h(x))$$

Masukkan angkanya:

$$Cost = -\log(1 - 0.25)$$

$$Cost = -\log(0.75)$$

(Buka kalkulator, tekan `log` untuk 0.75, hasilnya -0.125)

$$Cost = -(-0.125)$$

$$Cost = 0.125$$

# Terbentuklah Cost Function dari 2 Skenario

## Menghitung Rata-Rata Error (Cost Function)

### LANGKAH 1: KUMPULKAN ERROR INDIVIDU

- |                                |        |
|--------------------------------|--------|
| 1. Error Mahasiswa A (Lulus) : | 0.0025 |
| 2. Error Mahasiswa B (Gagal) : | 0.1250 |

### LANGKAH 2: JUMLAHKAN (SIGMA $\Sigma$ )

$$\text{Total Error} = 0.0025 + 0.1250 = 0.1275$$

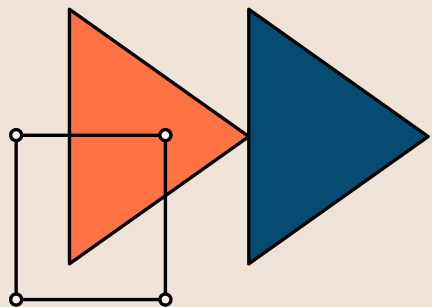
### LANGKAH 3: BAGI JUMLAH DATA (1/M)

$$\text{Rata-rata} = 0.1275 \div 2 \text{ Orang}$$

NILAI COST FUNCTION (J)

**0.06375**

Kesimpulan: Error Sangat Kecil (Model Pintar)





# STUDI KASUS: PREDIKSI KELULUSAN MAHASISWA

## DESKRIPSI MASALAH

Memprediksi apakah seorang mahasiswa akan LULUS atau  
TIDAK LULUS berdasarkan:

- Fitur 1 ( $x_1$ ): Jam Belajar per Minggu
- Fitur 2 ( $x_2$ ): Nilai Ujian Tengah Semester (UTS)
- Target ( $y$ ): Status Kelulusan (0 = Tidak Lulus, 1 = Lulus)



**Kapan Harus  
Digunakan?**

# Kapan Menggunakan & Menghindari

## ✓ SANGAT DISARANKAN

**Klasifikasi Biner** : Saat target hanya punya dua opsi (Ya/Tidak, Lulus/Gagal).

**Linear Separable** : Data bisa dipisahkan garis lurus secara kasar.

**Butuh Probabilitas** : Output berupa % peluang (misal: 80% risiko).

**Data Sederhana** : Dataset kecil hingga menengah.

## ✗ SEBAIKNYA HINDARI

**Pola Rumit (Non-Linear)** : Data melingkar atau spiral.

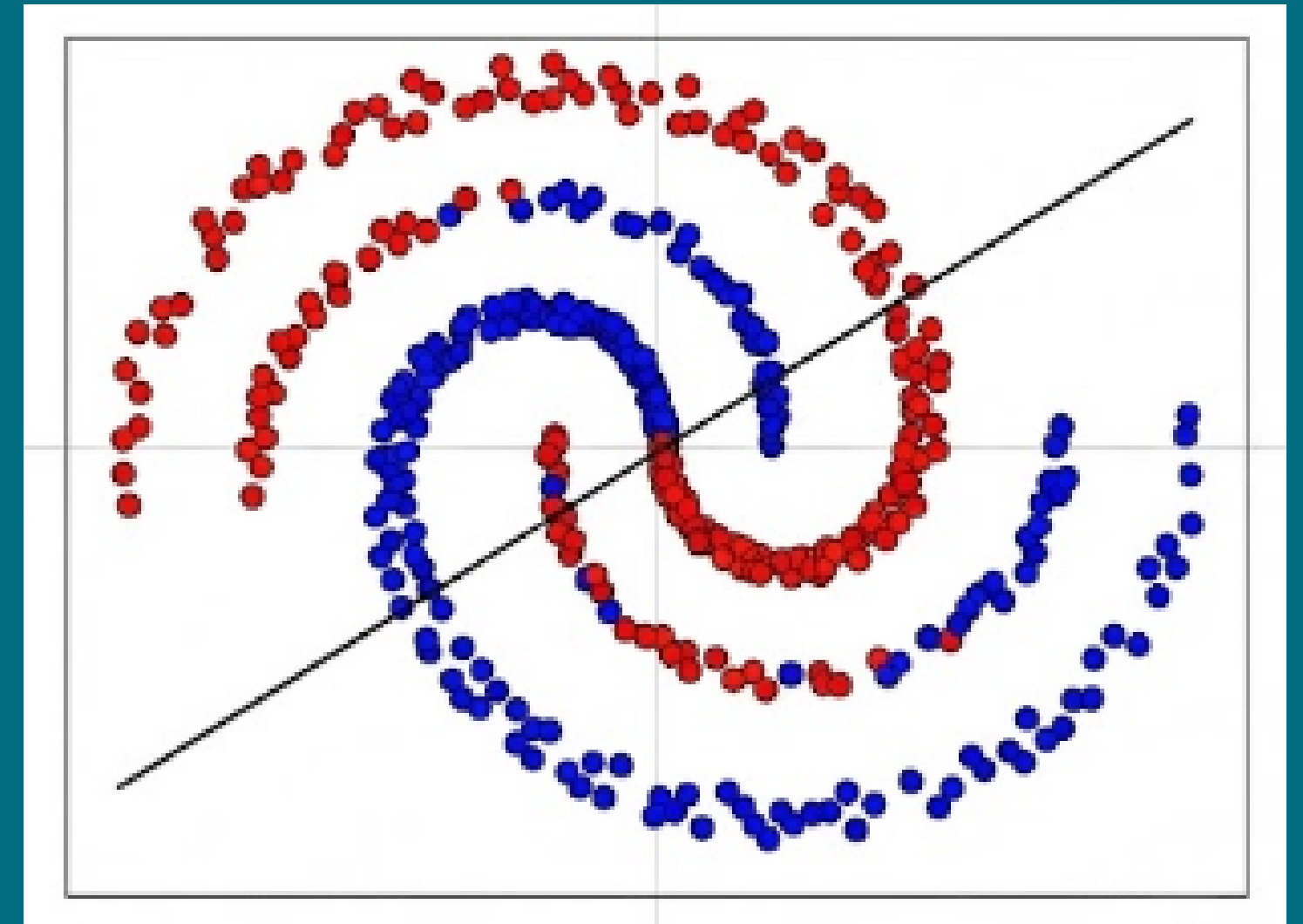
**Banyak Missing Values** : Tidak tahan data kosong.

**Multikolinearitas** : Fitur input saling berhubungan erat.

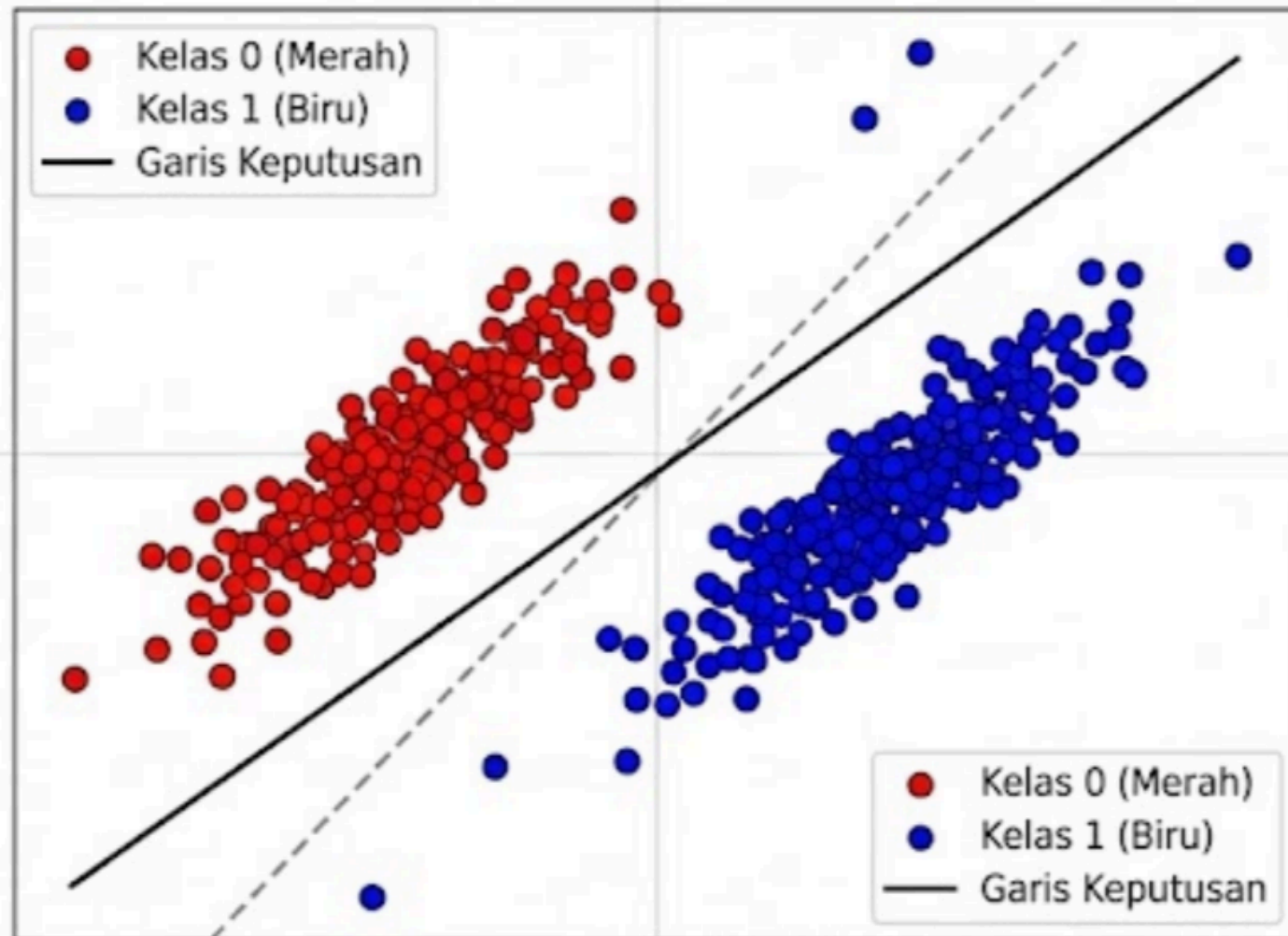
# Batasan Linearitas

## Mengapa Gagal di Data Kompleks?

Logistic Regression bekerja dengan menarik satu garis lurus. Lihat gambar di samping. Garis lurus tidak mungkin memisahkan titik biru dan merah pada pola spiral.



# Analisis Outlier



## APAKAH RENTAN? YA!!!

Outlier punya efek "magnet".

Karena Logistic Regression ingin meminimalkan total error, satu data ekstrem (outlier) yang salah prediksi akan menghasilkan penalti error yang sangat besar.

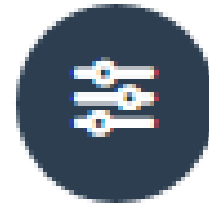
Akibatnya, garis keputusan akan tertarik paksa ke arah outlier, merusak akurasi data normal lainnya.

# Strategi Mitigasi



## Hapus Outlier

Gunakan metode IQR atau Z-Score untuk membuang data ekstrem sebelum training.



## Regularisasi

Gunakan L1/L2 Regularization untuk membatasi bobot agar tidak terlalu sensitif.



## Robust Scaler

Scaling data menggunakan median/kuartil yang lebih tahan terhadap outlier.