



Normalisasi Data

Ali Ridho Barakbah

Knowledge Engineering Research Group

Soft Computing Laboratory

Department of Information and Computer Engineering

Electronic Engineering Polytechnic Institute of Surabaya

Normalization

- adalah proses penskalaan nilai atribut dari data sehingga bisa jatuh pada range tertentu.
- Contoh: Misalnya berkenaan dengan pencatatan tingkat kematian penduduk di Indonesia per bulannya berdasarkan jenis umur. Secara sederhana, disana ada 3 dimensi data, yaitu bulan (1-12), umur (0-150 misalnya), dan jumlah kematian (0-jutaan). Kalau kita bentangkan range masing-masing dimensi, maka kita akan mendapatkan ketidak-seimbangan range pada dimensi yang ketiga, yaitu jumlah kematian.

Normalization methods

- Min-Max
- Z-Score
- Decimal Scaling
- Sigmoidal
- Softmax

Normalization method (Min-Max)

- Min-Max merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli.
- Rumus:
$$\text{newdata} = (\text{data} - \text{min}) * (\text{newmax} - \text{newmin}) / (\text{max} - \text{min}) + \text{newmin}$$
- Keuntungan dari metode ini adalah keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses normalisasi. Tidak ada data bias yang dihasilkan oleh metode ini. Kekurangannya adalah jika ada data baru, metode ini akan memungkinkan terjebak pada "out of bound" error.

Normalization method (Z-Score)

- Z-score merupakan metode normalisasi yang berdasarkan mean (nilai rata-rata) dan standard deviation (deviasi standar) dari data.
- Rumus:
$$\text{newdata} = (\text{data} - \text{mean}) / \text{std}$$
- Metode ini sangat berguna jika kita tidak mengetahui nilai aktual minimum dan maksimum dari data.

Normalization method (Decimal Scaling)

- Metode ini melakukan normalisasi dengan menggerakkan nilai desimal dari data ke arah yang diinginkan.

- Rumus:

$$\text{newdata} = \text{data} / 10^i$$

dimana i adalah nilai integer untuk menggerakkan nilai desimal ke arah yang diinginkan.

Normalization method (Sigmoidal)

- Sigmoidal normalization melakukan normalisasi data secara nonlinier ke dalam range -1 - 1 dengan menggunakan fungsi sigmoid.

- Rumus:

$$\text{newdata} = (1 - e^{(-x)}) / (1 + e^{(-x)})$$

dimana:

$$x = (\text{data} - \text{mean}) / \text{std}$$

e = nilai eksponensial (2,718281828)

- Metode ini sangat berguna pada saat data-data yang ada melibatkan data-data outlier.

Normalization method (Softmax)

- Metode ini merupakan pengembangan transformasi secara linier. Output range-nya adalah 0-1.

- Rumus:

$$\text{newdata} = 1/(1+e^{(-\text{transfdata})})$$

dimana:

$$\text{transfdata} = (\text{data}-\text{mean})/(x*(\text{std}/(2*3.14)))$$

x = respon linier di deviasi standar