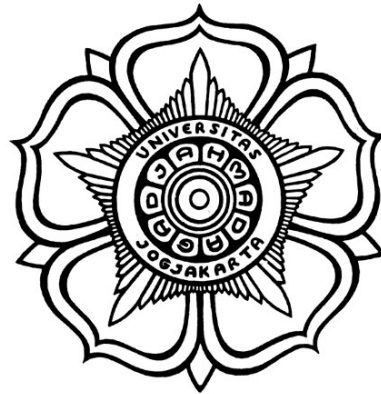


LAPORAN TUGAS AKHIR I

**PEMBELAJARAN REPRESENTASI SPASIAL-TEMPORAL
BERBASIS JARINGAN SARAF GRAF UNTUK OPTIMASI
PEMBOBOTAN PADA MODEL REGRESI DENGAN
KOEFSISIEN BERVARIASI DAN REGRESI LOKAL
TERBOBOTI**



Kurniawan Chandra Wijaya
22/497908/PA/21466

PEMBIMBING : Prof. Dr. Abdurakhman, S.Si., M.Si.

**PROGRAM SARJANA PROGRAM STUDI STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA**

2025

DAFTAR ISI

| | |
|--|----------|
| DAFTAR ISI | 2 |
| BAB I PENDAHULUAN | 2 |
| 1.1. Latar Belakang | 2 |
| 1.2. Pembatasan Masalah | 2 |
| 1.3. Tujuan dan Manfaat Penelitian | 2 |
| 1.4. Tinjauan Pustaka | 2 |
| 1.5. Metodologi Penelitian | 2 |
| 1.6. Sistematika Penulisan | 3 |
| BAB II LANDASAN TEORI | 4 |
| 2.1. Aljabar Matriks dan Tensor | 4 |
| 2.1.1. Ruang Vektor dan Matriks | 4 |
| 2.1.2. Sifat-Sifat Matriks | 10 |
| 2.1.3. Invers Matriks | 18 |
| 2.1.4. Determinan dan <i>Rank</i> | 21 |
| 2.1.5. Nilai Eigen dan Vektor Eigen | 27 |
| 2.1.6. Tensor | 29 |
| 2.1.7. Dekomposisi Matriks dan Tensor | 34 |
| 2.1.8. Diferensial Vektor dan Matriks | 42 |
| 2.2. Regresi Spasial dan Spasial-Temporal | 44 |
| 2.2.1. Regresi Linear Biasa atau <i>Ordinary Least Squares</i> (OLS) | 44 |
| 2.2.2. Regresi Terboboti | 50 |
| 2.2.3. Model Koefisien Bervariasi | 54 |
| 2.2.4. Regresi dengan Dependensi Spasial dan Spasial Temporal | 57 |
| 2.2.5. Regresi dengan Heterogenitas Spasial dan Spasial Temporal | 60 |
| 2.3. Jaringan Saraf Tiruan | 65 |
| 2.3.1. Sejarah dan Model Dasar | 66 |
| 2.3.2. Arsitektur Feedforward Multilapis | 66 |
| 2.3.3. Fungsi Aktivasi | 67 |
| 2.3.4. Pembelajaran dengan Propagasi Mundur | 70 |
| 2.3.5. Jaringan Saraf Tiruan sebagai Aproksimasi Universal | 70 |
| 2.4. Jaringan Saraf Graf | 72 |
| 2.4.1. Dasar Graf dan Laplacian | 73 |
| 2.4.2. Konvolusi pada Graf | 74 |
| 2.4.3. Kerangka Message Passing | 74 |
| 2.4.4. Arsitektur Umum dalam GNN | 75 |
| 2.4.5. Kekuatan Representasi GNN | 77 |

| | |
|--|------------|
| 2.5. Pembelajaran Representasi | 79 |
| 2.5.1. Konsep Dasar Representasi | 80 |
| 2.5.2. Representasi Klasik | 80 |
| 2.5.3. Representasi dalam Pembelajaran Mendalam | 81 |
| 2.5.4. Landasan Teoretis Representasi | 82 |
| 2.5.5. Isu dan Permasalahan dalam Pembelajaran Representasi | 83 |
| 2.6. Pembelajaran Semi-Terawasi | 83 |
| 2.6.1. Konsep Dasar Pembelajaran Semi-Terawasi | 84 |
| 2.6.2. Kerangka Transduktif dan Induktif dalam Pembelajaran Semi-Terawasi | 84 |
| 2.6.3. Asumsi-Asumsi Fundamental pada Pembelajaran Semi-Terawasi | 85 |
| 2.6.4. Metode-Metode Klasik Pembelajaran Semi-Terawasi | 86 |
| 2.6.5. Pendekatan Pembelajaran Semi-Terawasi pada Pembelajaran Mendalam | 87 |
| 2.6.6. Landasan Teoretis Pembelajaran Semi-Terawasi | 89 |
| BAB III RENCANA PENELITIAN | 90 |
| 3.1. Rencana Penelitian | 90 |
| 3.2. Rencana Pengambilan Data | 90 |
| 3.3. Rencana Jadwal Penelitian | 91 |
| BAB IV METODOLOGI PENELITIAN | 92 |
| 4.1. Literatur Optimasi Pembobotan pada Model Regresi Spasial | 92 |
| 4.1.1. <i>Geographically Weighted Artificial Neural Network</i> (GWANN). | 92 |
| 4.1.2. <i>Spatial and Attribute Neural Network Weighted Regression</i> (SAN-NWR). | 92 |
| 4.1.3. <i>Spatial Regression Graph Convolutional Neural Networks</i> (SR-GCNN). | 93 |
| 4.1.4. <i>Spatio-Temporal Graph Convolutional Networks</i> (ST-GCN) | 93 |
| 4.1.5. <i>Geographically Neural Network Weighted Regression</i> (GNNWR). | 94 |
| 4.2. <i>Graph Neural Networks Geographically and Temporally Varying Coefficients</i> (GNN-GTVC). | 96 |
| 4.2.1. Formulasi Umum GNN-GTVC dan GNN-GTWR | 97 |
| 4.2.2. Arsitektur GNN untuk Pembelajaran Bobot | 97 |
| 4.2.3. Penentuan Bobot berdasarkan Hasil Representasi | 98 |
| 4.2.4. Penentuan Bobot berdasarkan Hasil Representasi | 98 |
| 4.2.5. Fungsi Objektif dan Proses Pelatihan. | 100 |
| DAFTAR PUSTAKA | 103 |

BAB I

PENDAHULUAN

1.1. Latar Belakang

Latar belakang tidak boleh terlalu panjang. Perhatikan pemenggalan kata, cetak miring, dan EYD. Pemenggalan kata di template ini enggak bisa otomatis sesuai ya, jadi kalau ada yang kurang sesuai, bisa disesuaikan secara manual. Ada 2 opsi, opsi satu ubah kata/kalimat, dan opsi dua dipenggal mandiri pakai ”-”.

1.2. Pembatasan Masalah

Isinya ya batasan masalah kamu, biar skripsi mu enggak merambat ke mana-mana.

1.3. Tujuan dan Manfaat Penelitian

Tujuan dan manfaat dari penulisan skripsi ini adalah sebagai berikut:

1. Tujuan 1
2. Tujuan 2
3. Tujuan 3

1.4. Tinjauan Pustaka

Berikan pengantar terlebih dahulu. Berikut merupakan beberapa literatur yang digunakan sebagai acuan dan referensi dalam penulisan skripsi ini.

1.5. Metodologi Penelitian

Metodologi penelitian yang digunakan pada penulisan skripsi ini adalah studi literatur dan studi kasus. Studi literatur dilakukan dengan mengumpulkan infor-

masi dari buku, jurnal, artikel, skripsi, *paper*, dan sumber-sumber lainnya dari internet atau media cetak yang berkaitan dengan topik skripsi penulis. Informasi yang diperoleh digunakan sebagai acuan dan referensi dalam Studi kasus dilakukan untuk Algoritma yang digunakan pada penelitian ini adalah Data yang digunakan sebagai studi kasus merupakan data sekunder berupa Proses analisis dan komputasi dilakukan menggunakan software

1.6. Sistematika Penulisan

Pada penyusunan skripsi ini, penulis mengacu pada sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, pembatasan masalah, tujuan dan manfaat penelitian, tinjauan pustaka, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi tentang dasar-dasar teori yang digunakan untuk membahas permasalahan yang ada dalam penelitian ini. Bab ini mencakup dasar teori tentang

BAB III XXXXX

Bab ini menjelaskan mengenai

BAB IV STUDI KASUS

Bab ini menjelaskan mengenai

BAB V PENUTUP

Bab ini berisi kesimpulan yang didapat berdasarkan hasil analisis dan saran yang dapat digunakan untuk penelitian selanjutnya.

BAB II

LANDASAN TEORI

Pada bab ini akan dibahas mengenai konsep, teori, dan referensi yang menjadi dasar dalam melakukan analisis suatu metode statistik. Konsep dasar yang dibahas pada bab ini adalah aljabar matriks dan tensor, regresi spasial dan spasial-temporal, jaringan saraf tiruan, jaringan saraf graf, pembelajaran representasi, dan pembelajaran semi-terawasi.

2.1. Aljabar Matriks dan Tensor

Aljabar linear merupakan salah satu fondasi matematis terpenting dalam statistika, ekonometrika, serta pembelajaran mesin. Banyak metode regresi spasial maupun spasial-temporal, termasuk model berbasis jaringan saraf graf atau *graph neural networks* (GNN), dapat diformulasikan dalam kerangka operasi matriks dan tensor. Oleh karena itu, pemahaman yang kuat mengenai struktur vektor, matriks, dan tensor serta sifat-sifat aljabarnya diperlukan sebelum membahas regresi spasial-temporal berbobot.

2.1.1. Ruang Vektor dan Matriks

Pembahasan aljabar linear umumnya berangkat dari konsep lapangan atau *field* sebagai struktur aljabar dasar tempat bilangan berlaku. Dari lapangan, dibangun ruang vektor, kemudian matriks sebagai representasi transformasi linear, dan tensor sebagai generalisasi multidimensi.

Definisi 2.1.1 (Lapangan, (Lang, 1987)) Misalkan K adalah subhimpunan dari bilangan kompleks \mathbb{C} , K disebut sebagai lapangan apabila memenuhi kondisi berikut.

- (a) Apabila x, y adalah elemen dari K , maka $x + y$ dan xy juga merupakan elemen dari K .

(b) Apabila $x \in K$, maka $-x$ juga merupakan elemen dari K . Lebih lanjut, jika $x \neq 0$, maka x^{-1} merupakan elemen dari K .

(c) 0 dan 1 merupakan elemen dari K .

Berdasarkan definisi di atas, dapat diperhatikan bahwa \mathbb{R} dan \mathbb{C} merupakan lapangan.

Contoh 2.1.2 Apabila dinotasikan \mathbb{Q} sebagai himpunan bilangan rasional, yaitu

$$\mathbb{Q} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0 \right\}, \quad (2.1)$$

maka dapat dinyatakan bahwa \mathbb{Q} membentuk suatu lapangan. Misalkan $\frac{a}{b}, \frac{c}{d} \in \mathbb{Q}$ dengan $a, b, c, d \in \mathbb{Z}$ dan $b, d \neq 0$. Jelas bahwa

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \in \mathbb{Q}, \quad (2.2)$$

sehingga sifat tertutup terhadap penjumlahan, sebagai salah satu aksioma lapangan, terpenuhi. Lebih lanjut, untuk setiap $\frac{a}{b} \in \mathbb{Q}$ berlaku pula bahwa $-\frac{a}{b} \in \mathbb{Q}$. Selain itu, jika $a \neq 0$, maka invers perkalian $\frac{a}{b}$ diberikan oleh

$$\left(\frac{a}{b} \right)^{-1} = \frac{b}{a} \in \mathbb{Q}. \quad (2.3)$$

Dengan demikian, keberadaan invers aditif dan invers perkalian juga terjamin. Akhirnya, apabila dipilih $\frac{a}{b} = \frac{0}{1}$ dan $\frac{a}{b} = \frac{1}{1}$, diperoleh bahwa elemen identitas penjumlahan 0 dan identitas perkalian 1 termasuk dalam \mathbb{Q} . Di sisi lain, dapat dengan mudah diperiksa bahwa \mathbb{Z} bukan merupakan lapangan, karena apabila $n \in \mathbb{Z}$ dan $n \neq 0$, jelas bahwa $n^{-1} \notin \mathbb{Z}$, kecuali untuk kasus $n = 1$ atau $n = -1$.

Dalam ruang vektor dan matriks, dikenal juga istilah sublapangan. Apabila K dan L keduanya merupakan lapangan, serta dimisalkan bahwa $K \subseteq L$, maka K dapat disebut sebagai sublapangan atau *subfield* dari L . Elemen-elemen dari lapangan disebut sebagai bilangan atau skalar. (Lang, 1987)

Definisi 2.1.3 (Ruang Vektor, (Lang, 1987)) Suatu ruang vektor V atas suatu lapangan K adalah sebuah himpunan objek yang dapat dijumlahkan dan dikalikan dengan elemen-elemen dari K , sedemikian rupa sehingga hasil penjumlahan dua elemen dari V merupakan elemen V kembali, dan hasil perkalian sebuah elemen V dengan sebuah elemen dari K juga merupakan elemen dari V . Selain itu, sifat-sifat berikut dipenuhi:

- (a) $\forall u, v, w \in V$ berlaku $(u + v) + w = u + (v + w)$.
- (b) $\exists 0 \in V$ sedemikian sehingga $\forall u \in V$ berlaku $u + 0 = u$.
- (c) $\forall u \in V, \exists -u \in V$ sedemikian sehingga $u + (-u) = 0$.
- (d) $\forall u, v \in V$ berlaku $u + v = v + u$.
- (e) $\forall c \in K$ dan $\forall u, v \in V$ berlaku $c(u + v) = cu + cv$.
- (f) $\forall a, b \in K$ dan $\forall v \in V$ berlaku $(a + b)v = av + bv$.
- (g) $\forall a, b \in K$ dan $\forall v \in V$ berlaku $(ab)v = a(bv)$.
- (h) $\forall u \in V$ berlaku $1 \cdot u = u$, dengan 1 adalah elemen identitas pada K .

Dalam mendefinisikan ruang vektor, lapangan tempat ruang vektor berada harus didefinisikan dengan spesifik, misalnya \mathbb{C}^n merupakan ruang vektor atas \mathbb{C} , tetapi \mathbb{R}^n bukan ruang vektor atas \mathbb{C} . \mathbb{R}^n merupakan ruang vektor atas \mathbb{R} .

Definisi 2.1.4 (Ruang Matriks, (Lang, 1987)) Misalkan K merupakan lapangan dan $m, n \geq 1 \in \mathbb{Z}$. Matriks adalah suatu larik atau array dari bilangan-bilangan dalam K yang dinotasikan sebagai

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Notasi matriks pada definisi di atas dapat dipersingkat dengan menulis (a_{ij}) , $i = 1, 2, \dots, m$ dan $j = 1, 2, \dots, n$. Matriks tersebut merupakan matriks $m \times n$ yang berarti bahwa matriks tersebut memiliki m -baris dan n -kolom. Sebagai contoh, kolom pertama dari matriks tersebut adalah

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$$

dan baris pertamanya adalah $(a_{11}, a_{12}, \dots, a_{1n})$. Nilai a_{ij} disebut sebagai komponen dari matriks.

Dua buah matriks dikatakan sama apabila keduanya memiliki ukuran yang sama dan elemen-elemen pada posisi yang bersesuaian juga sama. Oleh karena itu, jika $\mathbf{A} = (a_{ij})$ dan $\mathbf{B} = (b_{ij})$, maka $\mathbf{A} = \mathbf{B}$ apabila $a_{ij} = b_{ij}$, untuk setiap i, j .

Definisi 2.1.5 (Kebebasan Linear, (Lang, 1987)) Misalkan V ruang vektor atas lapangan K dan misalkan pula $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ elemen-elemen dari V , maka $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ dikatakan bebas linear jika dan hanya jika kapan pun a_1, a_2, \dots, a_n merupakan bilangan sehingga

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n = \mathbf{0}, \quad (2.4)$$

maka $a_i = 0, \forall i = 1, 2, \dots, n$.

Vektor $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ dikatakan bergantung linear terhadap K apabila terdapat elemen-elemen a_1, a_2, \dots, a_n di K yang semuanya tidak sama dengan nol, sedemikian sehingga Persamaan (2.4) terpenuhi.

Contoh 2.1.6 Misalkan $V = K^n$ dan vektor $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ didefinisikan sebagai

berikut.

$$\begin{aligned}\mathbf{v}_1 &= (1, 0, \dots, 0) \\ &\vdots \\ \mathbf{v}_n &= (0, 0, \dots, 1).\end{aligned}$$

Vektor-vektor $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ dikatakan bebas linear. Misalkan a_1, a_2, \dots, a_n merupakan bilangan-bilangan sedemikian sehingga Persamaan (2.4) terpenuhi. Sebab

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n = (a_1, a_2, \dots, a_n), \quad (2.5)$$

dapat disimpulkan bahwa $a_i = 0, \forall i = 1, 2, \dots, n$.

Definisi 2.1.7 (*Span* atau *Jangkauan Linear*, (Axler, 2024)) Misalkan V ruang vektor atas lapangan K dan $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\} \subseteq V$. *Span* dari S , ditulis $\text{span}(S)$, adalah himpunan semua kombinasi linear dari elemen-elemen S , yaitu

$$\text{span}(S) = \left\{ \sum_{i=1}^m a_i \mathbf{v}_i : a_i \in K \right\}. \quad (2.6)$$

Dengan kata lain, $\text{span}(S)$ adalah subruang terkecil dari V yang memuat S . Di sisi lain, span dari himpunan kosong $\{\}$ didefinisikan sebagai $\{0\}$.

Contoh 2.1.8 Pada \mathbb{R}^2 , ambil $S = \{(1, 0), (0, 1)\}$, maka

$$\text{span}(S) = \{a(1, 0) + b(0, 1) : a, b \in \mathbb{R}\} = \mathbb{R}^2. \quad (2.7)$$

Hal ini berarti dua vektor standar membentang atau menjangkau (*spanning*) seluruh bidang \mathbb{R}^2 . Pada \mathbb{R}^3 , ambil $S = \{(1, 0, 0), (0, 1, 0)\}$, maka

$$\text{span}(S) = \{a(1, 0, 0) + b(0, 1, 0) : a, b \in \mathbb{R}\} = \{(x, y, 0) : x, y \in \mathbb{R}\}, \quad (2.8)$$

yaitu bidang xy dalam \mathbb{R}^3 .

Definisi 2.1.9 (Basis, (Axler, 2024)) Suatu basis V adalah himpunan vektor-vektor di V yang bebas linear dan menjangkau (spanning) V .

Contoh 2.1.10 Perhatikan vektor-vektor $\mathbf{v}_1 = (1, 2, -4)$ dan $\mathbf{v}_2 = (7, -5, 6)$ di dalam \mathbb{R}^3 . Jelas bahwa \mathbf{v}_1 dan \mathbf{v}_2 bebas linear, sebab tidak terdapat skalar $c \in \mathbb{R}$ sehingga $\mathbf{v}_2 = c\mathbf{v}_1$. Namun, himpunan $\{\mathbf{v}_1, \mathbf{v}_2\}$ hanya terdiri dari dua vektor, sehingga $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ paling jauh merupakan subruang berdimensi 2, yaitu sebuah bidang melalui titik asal di \mathbb{R}^3 . Sebab $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \subsetneq \mathbb{R}^3$, maka himpunan ini tidak dapat dijadikan basis dari \mathbb{R}^3 .

Teorema 2.1.11 (Karakterisasi Basis, (Axler, 2024)) Sebuah list dari vektor-vektor $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ merupakan basis dari V jika dan hanya jika $\forall \mathbf{v} \in V$ dapat dituliskan secara unik dalam bentuk

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n, \quad (2.9)$$

dengan $a_1, a_2, \dots, a_n \in \mathbb{R} \cup \mathbb{C}$.

Bukti. Pertama andaikan bahwa $\mathbf{v}_1, \dots, \mathbf{v}_n$ adalah basis dari V . Sebab $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ menjangkau V , maka untuk setiap $\mathbf{v} \in V$ terdapat skalar $a_1, \dots, a_n \in F$ sehingga Persamaan (2.9) terpenuhi. Untuk menunjukkan bahwa representasi tersebut unik, misalkan terdapat skalar $c_1, \dots, c_n \in F$ dengan

$$\mathbf{v} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n. \quad (2.10)$$

Dengan mengurangkan Persamaan (2.10) dari Persamaan (2.9) akan diperoleh

$$0 = (a_1 - c_1)\mathbf{v}_1 + \dots + (a_n - c_n)\mathbf{v}_n. \quad (2.11)$$

Sebab $\mathbf{v}_1, \dots, \mathbf{v}_n$ bebas linear, maka setiap koefisien harus nol, yaitu $a_k - c_k = 0$ untuk $k = 1, \dots, n$. Dengan demikian $a_1 = c_1, \dots, a_n = c_n$, sehingga representasi (2.9) bersifat unik. Hal ini menyelesaikan pembuktian arah pertama.

Sebaliknya, andaikan setiap $\mathbf{v} \in V$ dapat dituliskan secara unik dalam ben-

tuk

$$\mathbf{v} = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n, \quad (2.12)$$

maka jelas $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ menjangkau V . Untuk membuktikan bahwa $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ bebas linear, andaikan terdapat $a_1, \dots, a_n \in F$ sehingga

$$0 = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n. \quad (2.13)$$

Sebab representasi (2.12) bersifat unik, khususnya untuk $\mathbf{v} = 0$, maka harus berlaku $a_1 = \cdots = a_n = 0$. Jadi $\mathbf{v}_1, \dots, \mathbf{v}_n$ bebas linear. Dengan demikian $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ adalah basis V . ■

Definisi 2.1.12 (Ruang Vektor Berdimensi-Hingga (Axler, 2024)) Suatu ruang vektor V atas medan F dikatakan berdimensi-hingga apabila terdapat suatu basis $\{v_1, \dots, v_n\}$ dari V yang terdiri atas sejumlah hingga (n buah) vektor. Dengan kata lain, V berdimensi-hingga jika V dapat dijangkau oleh suatu himpunan vektor berukuran hingga.

Definisi 2.1.13 (Dimensi (Axler, 2024)) Dimensi dari suatu ruang vektor berdimensi-hingga V adalah banyaknya elemen dalam setiap basis V . Sebab semua basis dari V memiliki panjang yang sama, maka bilangan ini terdefinisi dengan baik. Dimensi V dilambangkan dengan $\dim V$.

Contoh 2.1.14 Jika $U = \{(x, x, y) \in \mathbb{R}^3 : x, y \in \mathbb{R}\}$, maka $\dim U = 2$ karena $(1, 1, 0), (0, 0, 1)$ merupakan basis dari U .

2.1.2. Sifat-Sifat Matriks

Seluruh bagian ini bekerja di atas suatu lapangan K (umumnya $K = \mathbb{R}$ atau $K = \mathbb{C}$). Untuk $m, n \geq 1$, himpunan seluruh matriks $m \times n$ berelemen di K dilambangkan dengan $\text{Mat}_{m \times n}(K)$. Dengan penjumlahan dan perkalian skalar yang didefinisikan komponen-demi-komponen, $\text{Mat}_{m \times n}(K)$ membentuk ruang vektor atas lapangan K (Axler, 2024).

Definisi 2.1.15 (Kesamaan, Penjumlahan, dan Perkalian dengan Skalar) Untuk

$\mathbf{A} = (a_{ij}), \mathbf{B} = (b_{ij}) \in \text{Mat}_{m \times n}(K)$ dan $c \in K$, berlaku beberapa hal berikut.

- (a) $\mathbf{A} = \mathbf{B}$ jika dan hanya jika $a_{ij} = b_{ij}$ untuk semua i, j ;
- (b) $(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$ (didefinisikan hanya bila ukuran sama); dan
- (c) $(c\mathbf{A})_{ij} = c a_{ij}$.

Definisi 2.1.16 (Perkalian Matriks, (Axler, 2024)) Jika $\mathbf{A} \in \text{Mat}_{m \times n}(K)$ dan $\mathbf{B} \in \text{Mat}_{n \times p}(K)$, maka hasil kali $\mathbf{AB} \in \text{Mat}_{m \times p}(K)$ didefinisikan oleh

$$(\mathbf{AB})_{ik} = \sum_{j=1}^n a_{ij} b_{jk} \quad (1 \leq i \leq m, 1 \leq k \leq p). \quad (2.14)$$

Contoh 2.1.17 Perkalian matriks berukuran 3×2 dan 2×4 :

$$\underbrace{\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}}_{3 \times 2} \underbrace{\begin{pmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{pmatrix}}_{2 \times 4} = \underbrace{\begin{pmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{pmatrix}}_{3 \times 4}. \quad (2.15)$$

Proposisi 2.1.18 (Sifat-Sifat Elementer Perkalian Matriks, (Axler, 2024)) Untuk ukuran yang sesuai, berlaku:

- (a) sifat asosiatif, yaitu $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$;
- (b) sifat distributif, yaitu $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ dan $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$;
- (c) sifat identitas, yaitu $\mathbf{I}_m \mathbf{A} = \mathbf{A}$ dan $\mathbf{A} \mathbf{I}_n = \mathbf{A}$ untuk $\mathbf{A} \in \text{Mat}_{m \times n}(K)$; dan
- (d) dapat terjadi $\mathbf{AB} \neq \mathbf{BA}$.

Bukti. Di bawah ini adalah pembuktian untuk keempat sifat elementer perkalian matriks.

(a) Untuk $(\mathbf{AB})\mathbf{C}$, berlaku

$$\begin{aligned}
 ((\mathbf{AB})\mathbf{C})_{ik} &= \sum_{\ell=1}^p (\mathbf{AB})_{i\ell} c_{\ell k} \\
 &= \sum_{\ell=1}^p \left(\sum_{j=1}^n a_{ij} b_{j\ell} \right) c_{\ell k} \\
 &= \sum_{j=1}^n \sum_{\ell=1}^p a_{ij} b_{j\ell} c_{\ell k}.
 \end{aligned} \tag{2.16}$$

Di sisi lain, untuk $\mathbf{A}(\mathbf{BC})$ berlaku

$$\begin{aligned}
 (\mathbf{A}(\mathbf{BC}))_{ik} &= \sum_{j=1}^n a_{ij} (\mathbf{BC})_{jk} \\
 &= \sum_{j=1}^n a_{ij} \left(\sum_{\ell=1}^p b_{j\ell} c_{\ell k} \right) \\
 &= \sum_{j=1}^n \sum_{\ell=1}^p a_{ij} b_{j\ell} c_{\ell k}.
 \end{aligned} \tag{2.17}$$

Keduanya merupakan matriks yang sama.

(b) Untuk $\mathbf{A}(\mathbf{B} + \mathbf{C})$, berlaku

$$\begin{aligned}
 (\mathbf{A}(\mathbf{B} + \mathbf{C}))_{ik} &= \sum_{j=1}^n a_{ij} (b_{jk} + c_{jk}) \\
 &= \sum_{j=1}^n a_{ij} b_{jk} + \sum_{j=1}^n a_{ij} c_{jk} \\
 &= (\mathbf{AB})_{ik} + (\mathbf{AC})_{ik}.
 \end{aligned} \tag{2.18}$$

Kasus $(\mathbf{A} + \mathbf{B})\mathbf{C}$ serupa.

(c) Untuk \mathbf{I}_m merupakan matriks identitas berukuran $m \times m$, berlaku

$$(\mathbf{I}_m \mathbf{A})_{ik} = \sum_{j=1}^m (\mathbf{I}_m)_{ij} a_{jk} = \sum_{j=1}^m \delta_{ij} a_{jk} = a_{ik}, \tag{2.19}$$

dan

$$(\mathbf{A}\mathbf{I}_n)_{ik} = \sum_{j=1}^n a_{ij}(\mathbf{I}_n)_{jk} = \sum_{j=1}^n a_{ij}\delta_{jk} = a_{ik}, \quad (2.20)$$

dengan

$$\delta_{ij} = \begin{cases} 1, & \text{jika } i = j, \\ 0, & \text{jika } i \neq j. \end{cases} \quad (2.21)$$

Dapat disimpulkan bahwa $\mathbf{IA} = \mathbf{A}$.

- (d) Sifat ini akan dibuktikan dengan *counterexample*. Pertimbangkan perkalian matriks berikut.

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (2.22)$$

Apabila urutan dibalik, akan memberikan matriks $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, sehingga kontradiksi apabila $\mathbf{AB} = \mathbf{BA}$ untuk semua \mathbf{A} dan \mathbf{B} .

■

Definisi 2.1.19 (Perkalian Hadamard, (Horn & Johnson, 2012)) Jika diberikan $\mathbf{A} = (a_{ij}) \in \text{Mat}_{m \times n}$ dan $\mathbf{B} = (b_{ij}) \in \text{Mat}_{m \times n}$, maka perkalian Hadamard atau perkalian Schur dari \mathbf{A} dan \mathbf{B} merupakan perkalian elemen per elemen dari matriks tersebut atau dapat dituliskan sebagai berikut.

$$\mathbf{A} \odot \mathbf{B} = (a_{ij}b_{ij}) \in \text{Mat}_{m \times n}. \quad (2.23)$$

Notasi perkalian Hadamard \odot dalam beberapa literatur dapat dituliskan juga dengan \circ .

Contoh 2.1.20 Misalkan

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}.$$

Maka hasil perkalian Hadamard adalah

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} 1 \cdot 7 & 2 \cdot 8 & 3 \cdot 9 \\ 4 \cdot 10 & 5 \cdot 11 & 6 \cdot 12 \end{bmatrix} = \begin{bmatrix} 7 & 16 & 27 \\ 40 & 55 & 72 \end{bmatrix}.$$

Definisi 2.1.21 (*Transpose*) Jika $\mathbf{A} \in \text{Mat}_{m \times n}(K)$, maka transpose dari \mathbf{A} , dilambangkan sebagai \mathbf{A}^\top , dengan $\mathbf{A}^\top \in \text{Mat}_{n \times m}(K)$ didefinisikan oleh $(\mathbf{A}^\top)_{ji} = a_{ij}$.

Contoh 2.1.22 Salah satu contoh *transpose* dari matriks adalah

$$\mathbf{A} = \begin{pmatrix} -5 & 2 & 4 \\ 3 & 6 & -2 \end{pmatrix} \iff \mathbf{A}^\top = \begin{pmatrix} -5 & 3 \\ 2 & 6 \\ 4 & -2 \end{pmatrix}, \quad (2.24)$$

sedangkan untuk vektor adalah

$$\mathbf{a} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \iff \mathbf{a}^\top = (2, -3, 1). \quad (2.25)$$

Notasi untuk *transpose* pada vektor \mathbf{a} dapat juga dinotasikan dengan \mathbf{a}' . Di sisi lain, untuk skalar di lapangan K atau $c \in K$, $c^\top = c$.

Teorema 2.1.23 (*Sifat-Sifat Transpose*) Untuk \mathbf{A}, \mathbf{B} berukuran sesuai dan $c \in K$, berlaku:

- (a) $(\mathbf{A}^\top)^\top = \mathbf{A}$;
- (b) $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$;
- (c) $(c\mathbf{A})^\top = c\mathbf{A}^\top$; dan
- (d) $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$.

Bukti. Berikut ini adalah bukti dari keempat sifat *transpose* tersebut.

(a) Elemen-elemen dari $(\mathbf{A}^\top)^\top$ akan sama dengan elemen-elemen dari \mathbf{A} , karena

$$((\mathbf{A}^\top)^\top)_{ij} = a_{ij} = (\mathbf{A})_{ij}. \quad (2.26)$$

(b) Elemen-elemen dari $(\mathbf{A} + \mathbf{B})^\top$ adalah $((\mathbf{A} + \mathbf{B})^\top)_{ij} = a_{ji} + b_{ji}$ yang merupakan elemen-elemen dari $\mathbf{A}^\top + \mathbf{B}^\top$.

(c) Elemen-elemen dari $(c\mathbf{A})^\top$ adalah ca_{ji} yang merupakan elemen-elemen dari $c\mathbf{A}^\top$.

(d) Dengan menuliskan $c_{ik} = (\mathbf{AB})_{ik} = \sum_j a_{ij}b_{jk}$ akan didapatkan

$$\begin{aligned} ((\mathbf{AB})^\top)_{ki} &= c_{ik} \\ &= \sum_j a_{ij}b_{jk} \\ &= \sum_j (\mathbf{B}^\top)_{kj}(\mathbf{A}^\top)_{ji} \\ &= (\mathbf{B}^\top \mathbf{A}^\top)_{ki}. \end{aligned} \quad (2.27)$$

■

Definisi 2.1.24 (Matriks Simetris, Diagonal, dan Identitas) Berikut ini adalah beberapa definisi terkait matriks simetris, diagonal, dan identitas.

(a) $\mathbf{A} \in \text{Mat}_{n \times n}(K)$ disebut simetris jika $\mathbf{A}^\top = \mathbf{A}$.

(b) \mathbf{A} disebut diagonal jika $a_{ij} = 0$ untuk $i \neq j$.

(c) \mathbf{I}_n adalah matriks identitas berukuran n , dengan elemen diagonal 1 dan selainnya 0.

Contoh 2.1.25 Matriks $\mathbf{A} = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 10 & -7 \\ 4 & -7 & 9 \end{pmatrix}$ adalah matriks simetris karena $\mathbf{A}^\top = \mathbf{A}$.

Definisi 2.1.26 (Matriks Idempoten, (Dhrymes, 2013)) Suatu matriks persegi \mathbf{A} dikatakan idempoten jika dan hanya jika

$$\mathbf{A}\mathbf{A} = \mathbf{A}^2 = \mathbf{A}. \quad (2.28)$$

Definisi 2.1.27 (Jejak atau Trace) Untuk $\mathbf{A} = (a_{ij}) \in \text{Mat}_{n \times n}(K)$, jejak atau trace didefinisikan sebagai

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}. \quad (2.29)$$

Teorema 2.1.28 (Sifat-Sifat Trace, (Axler, 2024)) Untuk ukuran yang sesuai, berlaku:

- (a) $\text{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha \text{tr}(\mathbf{A}) + \beta \text{tr}(\mathbf{B})$;
- (b) $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, lebih umum $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA})$; dan
- (c) jika \mathbf{B} invertibel, maka $\text{tr}(\mathbf{B}^{-1}\mathbf{AB}) = \text{tr}(\mathbf{A})$.

Bukti. Berikut ini adalah bukti-bukti untuk ketiga sifat di atas.

(a)

$$\begin{aligned} \text{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) &= \sum_i (\alpha a_{ii} + \beta b_{ii}) \\ &= \alpha \sum_i a_{ii} + \beta \sum_i b_{ii} \\ &= \alpha \text{tr}(\mathbf{A}) + \beta \text{tr}(\mathbf{B}). \end{aligned} \quad (2.30)$$

(b) Untuk dua faktor,

$$\begin{aligned}
 \text{tr}(\mathbf{AB}) &= \sum_i (\mathbf{AB})_{ii} \\
 &= \sum_i \sum_j a_{ij} b_{ji} \\
 &= \sum_j \sum_i b_{ji} a_{ij} \\
 &= \sum_j (\mathbf{BA})_{jj} \\
 &= \text{tr}(\mathbf{BA}).
 \end{aligned} \tag{2.31}$$

Untuk tiga faktor,

$$\begin{aligned}
 \text{tr}(\mathbf{ABC}) &= \sum_i (\mathbf{ABC})_{ii} \\
 &= \sum_i \sum_j \sum_k a_{ij} b_{jk} c_{ki} \\
 &= \sum_j \sum_k \sum_i b_{jk} c_{ki} a_{ij} \\
 &= \text{tr}(\mathbf{BCA}).
 \end{aligned} \tag{2.32}$$

(c) Dengan sifat (b), dapat diketahui bahwa

$$\begin{aligned}
 \text{tr}(\mathbf{B}^{-1}\mathbf{AB}) &= \text{tr}(\mathbf{ABB}^{-1}) \\
 &= \text{tr}(\mathbf{AI}) \\
 &= \text{tr}(\mathbf{A}).
 \end{aligned} \tag{2.33}$$

■

Definisi 2.1.29 (Matriks Semi-Definit Positif) Misalkan \mathbf{A} merupakan suatu matriks persegi berdimensi m dan \mathbf{x} merupakan vektor dengan m elemen. \mathbf{A} dikatakan

semi-definit positif jika dan hanya jika untuk semua vektor \mathbf{x} berlaku

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0. \quad (2.34)$$

Matriks \mathbf{A} dikatakan definit positif jika dan hanya jika untuk \mathbf{x} tak kosong berlaku

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0. \quad (2.35)$$

Definisi 2.1.30 (Matriks Definit Negatif) *Misalkan matriks \mathbf{A} merupakan matriks persegi berukuran n , maka \mathbf{A} dikatakan definit negatif jika dan hanya jika $-\mathbf{A}$ definit positif. Hal yang sama juga berlaku untuk semi-definit negatif.*

Beberapa literatur menggunakan notasi \succeq untuk menunjukkan matriks semidefinit positif, seperti $\mathbf{A} \succeq 0$ berarti $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0, \forall \mathbf{x}$.

2.1.3. Invers Matriks

Sesuai kerangka Axler, konsep invers dipahami melalui operator linear dan representasinya dalam suatu basis. Sebuah matriks bujur sangkar merepresentasikan operator linear $T : K^n \rightarrow K^n$, dan invers matriks berkorespondensi dengan invers operator linear T^{-1} ketika T bijektif.

Definisi 2.1.31 (Invers Matriks, (Axler, 2024)) *Misalkan $\mathbf{A} \in K^{n \times n}$. Matriks \mathbf{A} disebut invertibel (atau non-singular) apabila terdapat matriks $\mathbf{B} \in K^{n \times n}$ sedemikian sehingga*

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n, \quad (2.36)$$

dengan \mathbf{I}_n matriks identitas berukuran $n \times n$. Matriks \mathbf{B} disebut invers dari \mathbf{A} dan dinotasikan \mathbf{A}^{-1} .

Teorema 2.1.32 (Keunikan Invers) *Jika \mathbf{A} invertibel, maka \mathbf{A}^{-1} tunggal.*

Bukti. Jika \mathbf{B} dan \mathbf{C} keduanya memenuhi $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$ dan $\mathbf{AC} = \mathbf{CA} = \mathbf{I}_n$, maka

$$\mathbf{B} = \mathbf{BI}_n = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{I}_n \mathbf{C} = \mathbf{C}. \quad (2.37)$$



Proposisi 2.1.33 (Sifat-Sifat Invers, (Axler, 2024)) Misalkan $\mathbf{A}, \mathbf{B} \in K^{n \times n}$ invertibel dan $\alpha \in K$. Berlaku:

- (a) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$;
- (b) $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$; pada $K = \mathbb{C}$ juga berlaku $(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*$;
- (c) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$;
- (d) Jika $\mathbf{A} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k)$ dengan setiap \mathbf{B}_i invertibel (ukuran mungkin berbeda), maka

$$\mathbf{A}^{-1} = \text{diag}(\mathbf{B}_1^{-1}, \dots, \mathbf{B}_k^{-1}).$$

Salah satu cara atau algoritma perhitungan invers matriks adalah dengan invers Gauss-Jordan. Jika \mathbf{A} invertibel, maka terdapat barisan operasi baris elementer (matriks elementer $\mathbf{E}_k \cdots \mathbf{E}_1$) sehingga $\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} = \mathbf{I}$. Dengan demikian,

$$\mathbf{A}^{-1} = \mathbf{E}_k \cdots \mathbf{E}_1. \quad (2.38)$$

Secara praktis, dapat dilakukan lakukan reduksi baris teraugmentasi:

$$[\mathbf{A} \mid \mathbf{I}] \xrightarrow{\text{operasi baris}} [\mathbf{I} \mid \mathbf{A}^{-1}].$$

Contoh 2.1.34 Ambil

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Dengan melakukan eliminasi Gauss–Jordan pada $[\mathbf{A} \mid \mathbf{I}]$ diperoleh

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right] \xrightarrow{R_1 \leftarrow R_1 - \frac{1}{2}R_2} \left[\begin{array}{ccc|ccc} \frac{3}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right]$$

$$R_2 \leftarrow R_2 - R_3 \xrightarrow{\quad} \left[\begin{array}{ccc|ccc} \frac{3}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 1 & 1 & -1 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right] \xrightarrow{R_3 \leftarrow R_3 - R_2} \left[\begin{array}{ccc|ccc} \frac{3}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 1 & 1 & -1 & 0 & 1 & -1 \\ -1 & 0 & 3 & 0 & -1 & 2 \end{array} \right]$$

$$R_1 \leftarrow R_1 + \frac{1}{2}R_3 \xrightarrow{\quad} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 0 & 1 & -1 \\ -1 & 0 & 3 & 0 & -1 & 2 \end{array} \right] \xrightarrow{R_2 \leftarrow R_2 - R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & -1 & 2 & -2 \\ -1 & 0 & 3 & 0 & -1 & 2 \end{array} \right]$$

$$R_3 \leftarrow R_3 + R_1 \xrightarrow{\quad} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & -1 & 2 & -2 \\ 0 & 0 & 4 & 1 & -2 & 3 \end{array} \right] \xrightarrow{R_3 \leftarrow \frac{1}{4}R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & -1 & 2 & -2 \\ 0 & 0 & 1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{array} \right]$$

$$R_1 \leftarrow R_1 - R_3, R_2 \leftarrow R_2 + 2R_3 \xrightarrow{\quad} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{4} & -\frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{array} \right].$$

Maka

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{pmatrix}.$$

Contoh 2.1.35 (Invers 2×2 dengan Rumus Klasik) Untuk $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ dengan

$ad - bc \neq 0$,

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (2.39)$$

Rumus ini konsisten dengan Gauss–Jordan dan dapat dipandang sebagai kasus khusus *adjoint* ketika $n = 2$.

2.1.4. Determinan dan Rank

Pada seluruh bagian ini, V menyatakan ruang vektor berdimensi hingga atas suatu lapangan K dan $\mathcal{L}(V, W)$ himpunan semua pemetaan linear $V \rightarrow W$.

Definisi 2.1.36 (Pemetaan Linear, (Axler, 2024)) Misalkan V, W ruang vektor atas suatu lapangan K . Suatu pemetaan $T : V \rightarrow W$ disebut pemetaan linear jika untuk semua $\mathbf{u}, \mathbf{v} \in V$ dan $\lambda \in K$ berlaku

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}), \quad T(\lambda \mathbf{v}) = \lambda T(\mathbf{v}). \quad (2.40)$$

Notasi $\mathcal{L}(V, W)$ menyatakan himpunan semua pemetaan linear dari V ke W . Jika $V = W$, maka $\mathcal{L}(V) = \mathcal{L}(V, V)$ adalah himpunan semua operator linear pada V .

Contoh 2.1.37 Berikut ini adalah beberapa contoh pemetaan linear.

(a) Jika $V = W = \mathbb{R}^2$, maka pemetaan

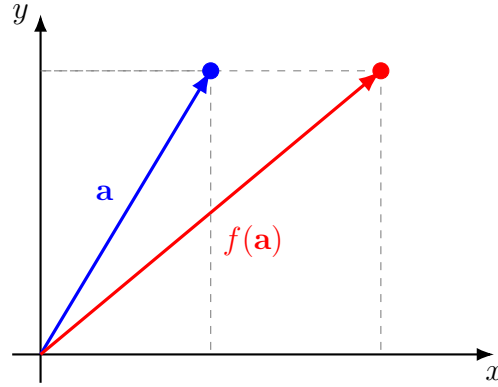
$$T(\mathbf{x}, \mathbf{y}) = (\mathbf{x} + 2\mathbf{y}, 3\mathbf{x} - \mathbf{y}) \quad (2.41)$$

adalah elemen $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$.

(b) Jika $V = \mathbb{R}^3$ dan $W = \mathbb{R}^2$, maka setiap $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ dapat direpresentasikan oleh matriks 2×3 , misalnya

$$T(x_1, x_2, x_3) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \quad (2.42)$$

Transformasi linear dapat divisualisasikan sebagai pemetaan yang membawa suatu vektor ke vektor lain dengan cara yang teratur (tanpa membengkokkan ruang). Sebagai contoh, Gambar 2.1 berikut memperlihatkan fungsi linear $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dengan $f(x, y) = (2x, y)$ yang meregangkan komponen x dari setiap vektor dengan faktor 2. Vektor \mathbf{a} (biru) dipetakan ke $f(\mathbf{a})$ (merah).



Gambar 2.1 Ilustrasi transformasi linear $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dengan $f(x, y) = (2x, y)$.
(Sumber: Dokumen penulis)

Definisi 2.1.38 (*Null Space, Range, dan Rank, (Axler, 2024)*) Untuk $T \in \mathcal{L}(V, W)$, null space dan range didefinisikan sebagai

$$\text{null } T := \{\mathbf{v} \in V : T\mathbf{v} = 0\}, \quad \text{range } T := \{T\mathbf{v} : \mathbf{v} \in V\} \subseteq W.$$

Jika V berdimensi-hingga, rank T didefinisikan oleh $\text{rank } T := \dim(\text{range } T)$. Dalam notasi matriks, untuk $\mathbf{A} \in K^{m \times n}$ yang merepresentasikan T relatif terhadap basis yang dipilih, $\text{rank}(\mathbf{A}) := \text{rank}(T) = \dim(\text{range } T)$.

Contoh 2.1.39 Misalkan

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \end{pmatrix} \in \mathbb{R}^{2 \times 3}, \quad T : \mathbb{R}^3 \rightarrow \mathbb{R}^2, \text{ dan } T(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

(a) Sistem homogen $\mathbf{A}\mathbf{x} = 0$ memberi

$$x_2 = -4t, \quad x_1 = 5t, \quad x_3 = t, \quad (t \in \mathbb{R}),$$

sehingga

$$\text{null } T = \text{span}\{(5, -4, 1)^\top\} \quad \text{dan} \quad \dim(\text{null } T) = 1. \quad (2.43)$$

(b) Kolom pertama dan kedua bebas linear, maka

$$\text{range } T = \text{span}\{(1, 0)^\top, (2, 1)^\top\} \quad (2.44)$$

sehingga $\dim(\text{range } T) = 2$.

Contoh sebelumnya menyebutkan tentang Teorema Rank-Nulitas yang merupakan teorema fundamental dalam pemetaan linear. Lebih lanjut, secara formal, teorema tersebut dirumuskan sebagai berikut.

Teorema 2.1.40 (Teorema Fundamental Pemetaan Linear atau Rank–Nulitas, (Axler, 2024)) *Jika V berdimensi-hingga dan $T \in \mathcal{L}(V, W)$, maka*

$$\dim V = \dim(\text{null } T) + \dim(\text{range } T). \quad (2.45)$$

Contoh 2.1.41 Berdasarkan Contoh 2.1.39, diketahui bahwa $\dim(\text{null } T) = 1$ dan $\dim(\text{range } T) = 2$. Konsisten dengan Rank–Nulitas, $1 + 2 = 3 = \dim \mathbb{R}^3$.

Definisi 2.1.42 (Bentuk Multilinear Anti-Simetri, (Axler, 2024)) *Misalkan V ruang vektor atas lapangan K . Suatu fungsi $\omega : V^n \rightarrow K$ disebut bentuk n -linear jika linear pada setiap argumennya. Bentuk n -linear ω disebut anti-simetri atau alternating jika*

$$\omega(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = -\omega(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n) \quad (2.46)$$

untuk setiap pertukaran dua argumen $\mathbf{v}_i, \mathbf{v}_j \in V$. Konsekuensinya, jika ada dua argumen sama, maka $\omega(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n) = 0$.

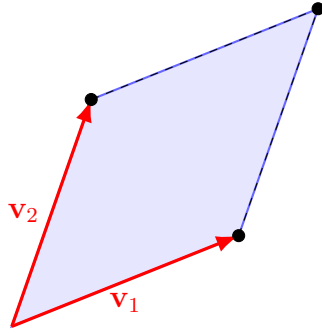
Contoh 2.1.43 Pada $V = \mathbb{R}^2$, fungsi

$$\omega((x_1, y_1), (x_2, y_2)) = x_1 y_2 - x_2 y_1 \quad (2.47)$$

adalah bentuk bilinear anti-simetri, karena

$$\omega((x_2, y_2), (x_1, y_1)) = x_2 y_1 - x_1 y_2 = -\omega((x_1, y_1), (x_2, y_2)). \quad (2.48)$$

Nilainya sama dengan luas terorientasi jajaran genjang yang dibentuk oleh kedua vektor.



Gambar 2.2 Ilustrasi bentuk bilinear anti-simetri $\omega(v_1, v_2) = x_1 y_2 - x_2 y_1$ pada \mathbb{R}^2 .
(Sumber: Dokumen penulis)

Luas dari jajar genjang pada Gambar 2.2 adalah

$$|\mathbf{v}_1| |\mathbf{v}_2| \cdot \sin(\theta),$$

dengan $\cos(\theta) = \frac{\mathbf{v}_1 \mathbf{v}_2}{|\mathbf{v}_1| |\mathbf{v}_2|}$. Perhatikan bahwa $\sin(\theta) = \sqrt{1 - \cos^2(\theta)}$, maka

$$\begin{aligned} |\mathbf{v}_1| |\mathbf{v}_2| \sin(\theta) &= \sqrt{|\mathbf{v}_1|^2 |\mathbf{v}_2|^2 - (x_1 x_2 + y_1 y_2)^2} \\ &= \sqrt{x_1^2 x_2^2 + x_1^2 y_2^2 + y_1^2 x_2^2 + y_1^2 y_2^2 - (x_1^2 x_2^2 + 2x_1 x_2 y_1 y_2 + y_1^2 y_2^2)} \\ &= \sqrt{x_1^2 y_2^2 + y_1^2 x_2^2 - 2x_1 x_2 y_1 y_2} \\ &= \sqrt{(x_1 y_2 - x_2 y_1)^2} \\ &= |x_1 y_2 - x_2 y_1|. \end{aligned} \quad (2.49)$$

Persamaan terakhir menunjukkan bahwa luas jajar genjang pada Gambar 2.2 adalah $|x_1 y_2 - x_2 y_1| = |\omega(v_1, v_2)|$.

Contoh 2.1.43 memberikan intuisi penting tentang determinan. Fungsi bilinear anti-simetri $\omega(v_1, v_2) = x_1 y_2 - x_2 y_1$ tidak lain adalah determinannya pada \mathbb{R}^2 . Secara

geometri, determinan mengukur luas terorientasi dari jajar genjang yang dibentang oleh dua vektor. Nilai absolut $|\det|$ memberikan ukuran luas (atau volume pada dimensi lebih tinggi), sedangkan tanda determinan membedakan orientasi. Jika \mathbf{v}_1 ke \mathbf{v}_2 berputar berlawanan arah jarum jam, maka $\det(\cdot) > 0$ dan sebaliknya jika searah jarum jam, maka $\det(\cdot) < 0$.

Definisi 2.1.44 (Determinan, (Axler, 2024)) Misalkan V berdimensi $n \geq 1$. Ruang semua bentuk n -linear beranti-simetri pada V memiliki dimensi 1. Untuk setiap $T \in \mathcal{L}(V)$, determinan T , ditulis $\det T$, adalah satu-satunya skalar di K dengan sifat

$$\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = (\det T) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n), \quad (2.50)$$

untuk setiap bentuk n -linear beranti-simetri ω pada V . Jika \mathbf{A} adalah matriks $n \times n$ yang merepresentasikan T pada suatu basis V , maka determinan matriks didefinisikan oleh $\det(\mathbf{A}) := \det(T)$.

Contoh 2.1.45 Ambil $V = \mathbb{R}^2$ dan operator linear $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ yang direpresentasikan oleh matriks

$$\mathbf{A} = \begin{pmatrix} 3 & 5 \\ 7 & 2 \end{pmatrix}.$$

Untuk dua vektor $\mathbf{v}_1 = (2, 1)^\top$ dan $\mathbf{v}_2 = (1, 2)^\top$, bentuk bilinear anti-simetri standar di \mathbb{R}^2 adalah

$$\omega(\mathbf{v}_1, \mathbf{v}_2) = \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3. \quad (2.51)$$

Sekarang apabila dilihat bayangan kedua vektor tersebut,

$$T\mathbf{v}_1 = \mathbf{A}\mathbf{v}_1 = (11, 16)^\top, \quad T\mathbf{v}_2 = \mathbf{A}\mathbf{v}_2 = (13, 11)^\top.$$

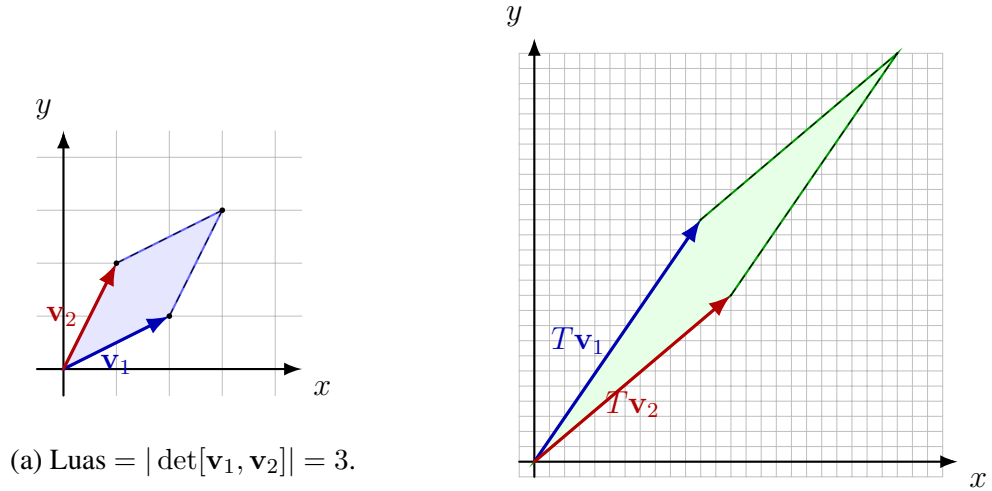
sehingga

$$\omega(T\mathbf{v}_1, T\mathbf{v}_2) = \det \begin{pmatrix} 11 & 13 \\ 16 & 11 \end{pmatrix} = 11 \cdot 11 - 13 \cdot 16 = -87. \quad (2.52)$$

Dengan demikian, sesuai definisi determinan operator,

$$\omega(T\mathbf{v}_1, T\mathbf{v}_2) = (\det T) \omega(\mathbf{v}_1, \mathbf{v}_2) \implies \det T = -87/3 = -29.$$

Sebab $\det(\mathbf{A}) := \det(T)$, diperoleh $\det(\mathbf{A}) = -29$, hasil ini konsisten dengan rumus determinan matriks 2×2 yang sudah dikenal, yaitu $\det \mathbf{A} = 6 - 35 = -29$.



(a) Luas = $|\det[\mathbf{v}_1, \mathbf{v}_2]| = 3$.

(b) Luas = $|\det[T\mathbf{v}_1, T\mathbf{v}_2]| = 87$.

Gambar 2.3 Perbandingan luas parallelogram sebelum dan sesudah transformasi A .
(Sumber: Dokumen penulis)

Berdasarkan ilustrasi di atas, terlihat bahwa $\det(\mathbf{A})$ memiliki makna geometris sebagai skala luas jajar genjang hasil transformasi oleh A .

Teorema 2.1.46 (Sifat Dasar Determinan, (Axler, 2024)) Untuk $S, T \in \mathcal{L}(V)$ dan skalar $c \in K$ berlaku:

- (a) $\det \mathbf{I}_V = 1$;
- (b) $\det(ST) = \det S \cdot \det T$; dan
- (c) T invertibel jika dan hanya jika $\det T \neq 0$.

Jika $\mathbf{A}, \mathbf{B} \in K^{n \times n}$ merepresentasikan S, T , maka $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ dan \mathbf{A} invertibel jika dan hanya jika $\det(\mathbf{A}) \neq 0$.

2.1.5. Nilai Eigen dan Vektor Eigen

Dalam aljabar linear, salah satu konsep penting adalah *eigenvalues* dan *eigenvectors* yang memberikan wawasan tentang bagaimana suatu operator linear bertindak pada vektor tertentu tanpa mengubah arahnya.

Definisi 2.1.47 (Vektor Eigen dan Nilai Eigen, (Axler, 2024, Lang, 1987)) Misalkan $T : V \rightarrow V$ adalah operator linear pada ruang vektor V atas \mathbb{C} .

(a) Suatu vektor $\mathbf{v} \in V$, $\mathbf{v} \neq \mathbf{0}$, disebut sebagai vektor eigen dari T jika terdapat skalar $\lambda \in \mathbb{C}$ sehingga

$$T(\mathbf{v}) = \lambda \mathbf{v}. \quad (2.53)$$

(b) Skalar λ yang memenuhi persamaan di atas disebut nilai eigen dari T .

Secara intuitif, vektor eigen adalah vektor yang tidak berubah arah setelah dikenakan transformasi T , melainkan hanya mengalami skala sebesar λ .

Contoh 2.1.48 Pertimbangkan $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dengan

$$T(x, y) = (2x, 3y). \quad (2.54)$$

(a) Untuk vektor $\mathbf{v} = (1, 0)$, berlaku $T(1, 0) = (2, 0) = 2(1, 0)$, sehingga \mathbf{v} adalah vektor eigen dengan nilai $\lambda = 2$.

(b) Untuk vektor $\mathbf{w} = (0, 1)$, berlaku $T(0, 1) = (0, 3) = 3(0, 1)$, sehingga \mathbf{w} adalah vektor eigen dengan nilai $\lambda = 3$.

Teorema 2.1.49 (Keberadaan Nilai Eigen pada Ruang Vektor Kompleks, (Axler, 2024)) Jika V adalah ruang vektor berdimensi-hingga atas \mathbb{C} dan $T : V \rightarrow V$ operator linear, maka T memiliki setidaknya satu nilai eigen.

Bukti. Misalkan $\dim V = n$. Pilih basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ dari V . Dalam basis ini, T direpresentasikan oleh suatu matriks $\mathbf{A} \in \text{Mat}_n(\mathbb{C})$. Selanjutnya, definisikan polinomial karakteristik

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}_n). \quad (2.55)$$

Polinomial ini berderajat n dalam variabel λ , dengan koefisien dalam \mathbb{C} . Teorema dasar aljabar menyatakan bahwa setiap polinomial kompleks nonkonstan memiliki akar. Sebab $\deg p_{\mathbf{A}}(\lambda) = n \geq 1$, maka terdapat $\lambda_0 \in \mathbb{C}$ sehingga

$$p_{\mathbf{A}}(\lambda_0) = \det(\mathbf{A} - \lambda_0 \mathbf{I}_n) = 0. \quad (2.56)$$

Jika $\det(\mathbf{A} - \lambda_0 \mathbf{I}_n) = 0$, maka matriks $\mathbf{A} - \lambda_0 \mathbf{I}_n$ tidak invertibel. Hal ini berarti terdapat $\mathbf{v} \neq \mathbf{0}$ dalam \mathbb{C}^n dengan

$$(\mathbf{A} - \lambda_0 \mathbf{I}_n)\mathbf{v} = \mathbf{0}. \quad (2.57)$$

Persamaan di atas ekuivalen dengan $\mathbf{A}\mathbf{v} = \lambda_0 \mathbf{v}$. Jadi, \mathbf{v} adalah vektor eigen tak-nol dari \mathbf{A} , dan λ_0 adalah nilainya. Sebab \mathbf{A} merepresentasikan operator T , maka kesimpulan $\mathbf{A}\mathbf{v} = \lambda_0 \mathbf{v}$ identik dengan

$$T(\mathbf{v}) = \lambda_0 \mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}. \quad (2.58)$$

Dengan demikian, T memiliki setidaknya satu nilai eigen di \mathbb{C} . ■

Definisi 2.1.50 (Subruang Invariant (Axler, 2024)) Subruang $U \subseteq V$ disebut *invariant terhadap T* jika $T(U) \subseteq U$. Jika \mathbf{v} adalah vektor eigen dengan nilai λ , maka $\text{span}\{\mathbf{v}\}$ adalah subruang invariant.

Contoh 2.1.51 Pertimbangkan matriks rotasi

$$\mathbf{R} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Jika $\theta = 0$, semua vektor adalah vektor eigen dengan $\lambda = 1$. Jika $\theta = \pi$, semua vektor adalah vektor eigen dengan $\lambda = -1$. Jika $\theta \notin \{0, \pi\}$, maka R tidak memiliki nilai eigen real, tetapi memiliki nilai eigen kompleks $e^{i\theta}$ dan $e^{-i\theta}$.

2.1.6. Tensor

Suatu vektor \mathbf{v} memiliki elemen v_i dengan satu indeks dan suatu matriks \mathbf{M} memiliki elemen M_{ij} dengan dua indeks. Suatu tensor merupakan generalisasi dari vektor dan matriks yang memiliki $d \in \mathbb{N}$ indeks. Pembahasan klasik mengenai tensor dibahas dalam aljabar multilinear oleh Greub (1978). Pengenalan tentang tensor diawali dengan konsep pemetaan multilinear.

Definisi 2.1.52 (Pemetaan Bilinear, (Greub, 1978)) Misalkan tiga ruang vektor E, F , dan G . Definisikan suatu pemetaan

$$\varphi : E \times F \rightarrow G.$$

φ disebut *bilinear* apabila *linear terhadap masing-masing argumen ketika yang lain dianggap tetap, yaitu*

(a) *linear terhadap argumen pertama atau*

$$\varphi(\lambda x_1 + \mu x_2, y) = \lambda \varphi(x_1, y) + \mu \varphi(x_2, y), \quad x_1, x_2 \in E, y \in F, \quad (2.59)$$

dengan $\lambda, \mu \in \Gamma$ merupakan skalar; dan

(b) *linear terhadap argumen kedua atau*

$$\varphi(x, \lambda y_1 + \mu y_2) = \lambda \varphi(x, y_1) + \mu \varphi(x, y_2), \quad x \in E, y_1, y_2 \in F. \quad (2.60)$$

Apabila $G = \Gamma$, maka φ dikatakan sebagai *fungsi bilinear*.

Contoh 2.1.53 Perkalian dalam atau *dot product* merupakan salah satu contoh dari fungsi bilinear. Ambil $E = F = \mathbb{R}^2$ dan $G = \mathbb{R}$. Definisikan

$$\varphi(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \cdot \mathbf{y}, \quad (2.61)$$

yaitu *dot product*. Jika $\mathbf{x} = (x_1, x_2)$ dan $\mathbf{y} = (y_1, y_2)$, maka

$$\varphi(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2. \quad (2.62)$$

Dapat ditunjukkan bahwa fungsi tersebut linear terhadap \mathbf{x} dan linear terhadap \mathbf{y} , karena

$$\begin{aligned} \varphi(\lambda \mathbf{x}^{(1)} + \mu \mathbf{x}^{(2)}, \mathbf{y}) &= \sum_{i=1}^n \left(\lambda x_i^{(1)} + \mu x_i^{(2)} \right) y_i \\ &= \sum_{i=1}^n \lambda x_i^{(1)} y_i + \sum_{i=1}^n \mu x_i^{(2)} y_i \\ &= \lambda \varphi(\mathbf{x}^{(1)}, \mathbf{y}) + \mu \varphi(\mathbf{x}^{(2)}, \mathbf{y}). \end{aligned} \quad (2.63)$$

Analog untuk kasus linear terhadap \mathbf{y} .

Contoh 2.1.54 Satu contoh fungsi bilinear lainnya adalah perkalian luar atau *outer product*. Ambil $\mathbf{x} \in \mathbb{R}^m$ dan $\mathbf{y} \in \mathbb{R}^n$. *Outer product* didefinisikan sebagai $\mathbf{x} \otimes \mathbf{y} \in \mathbb{R}^{m \times n}$ dengan elemen $(\mathbf{x} \otimes \mathbf{y})_{ij} = x_i y_j$. Pada kasus dimensi dua, $\mathbf{x} \otimes \mathbf{y} = \mathbf{xy}^\top$. Dapat diperhatikan bahwa

$$(\lambda \mathbf{x}^{(1)} + \mu \mathbf{x}^{(2)}) \otimes \mathbf{y} = \lambda(\mathbf{x}^{(1)} \otimes \mathbf{y}) + \mu(\mathbf{x}^{(2)} \otimes \mathbf{y}), \quad (2.64)$$

karena elemen dari $(\lambda \mathbf{x}^{(1)} + \mu \mathbf{x}^{(2)}) \otimes \mathbf{y}$ adalah $[(\lambda x^{(1)} + \mu x^{(2)}) \mathbf{y}]_{ij} = (\lambda x^{(1)} y + \mu x^{(2)} y)_{ij}$ yang sama dengan elemen dari $\lambda(\mathbf{x}^{(1)} \otimes \mathbf{y}) + \mu(\mathbf{x}^{(2)} \otimes \mathbf{y})$. Analog untuk kasus linear terhadap \mathbf{y} .

Definisi 2.1.55 (Pemetaan Multilinear, (Greub, 1978)) Misalkan terdapat $p + 1$ ruang vektor E_i ($i = 1, 2, \dots, p$) dan G . Suatu pemetaan $\varphi : E_1 \times E_2 \times \dots \times E_p \rightarrow G$ disebut sebagai *linear- p* apabila $\forall i$ ($1 \leq i \leq p$) berlaku

$$\begin{aligned} \varphi(x_1, \dots, x_{i-1}, \lambda x_i + \mu y_i, x_{i+1}, \dots, x_p) &= \lambda \varphi(x_1, \dots, x_i, \dots, x_p) \\ &\quad + \mu \varphi(x_1, \dots, y_i, \dots, y_p), \end{aligned} \quad (2.65)$$

dengan $x_i, y_i \in E$, $\lambda, \mu \in \Gamma$. Jika $G = \Gamma$, maka φ disebut sebagai fungsi linear- p .

Dalam kasus $p = 2$, subruang dari G yang dihasilkan dari vektor-vektor $\varphi(x_1, \dots, x_p)$, $x_i \in E_i$ dinotasikan sebagai $\text{Im } \varphi$.

Definisi 2.1.56 (Sifat Universal, (Greub, 1978)) Misalkan terdapat ruang vektor E dan F dan pemetaan bilinear \otimes dari $E \times F$ ke ruang vektor T . \otimes dapat dikatakan memiliki sifat universal apabila untuk setiap pemetaan bilineaer $\varphi : E \times F \rightarrow H$, terdapat satu pemetaan linear yang unik $f : T \rightarrow H$ sehingga diagram di bawah ini terpenuhi.

$$\begin{array}{ccc} E \times F & \xrightarrow{\varphi} & H \\ \otimes \downarrow & \nearrow f & \\ T & & \end{array}$$

Gambar 2.4 Diagram sifat universal.
(Sumber: Dokumen penulis)

Definisi 2.1.57 (Isomorfisme) Suatu pemetaan linear T disebut isomorfisme apabila kedua kondisi di bawah ini terpenuhi.

- (a) T bersifat satu-satu, yaitu jika $T(\mathbf{x}) = T(\mathbf{y})$, maka $\mathbf{x} = \mathbf{y}$.
- (b) T merupakan fungsi onto, yaitu jika $\mathbf{w} \in W$, maka terdapat $\mathbf{v} \in V$ sedemikian sehingga $T(\mathbf{v}) = \mathbf{w}$.

Dua subruang yang memiliki isomorfisme di atas disebut sebagai isomorfik.

Definisi 2.1.58 (Produk Tensor, (Greub, 1978)) Suatu produk tensor atau tensor product dari dua ruang vektor E dan F adalah pasangan (T, \otimes) , dengan $\otimes : E \times F \rightarrow T$ merupakan pemetaan linear dengan sifat universal. Ruang T yang secara unik ditentukan oleh E dan F hingga isomorfisme disebut sebagai produk tensor dari E dan F , dinotasikan dengan $E \otimes F$.

Definisi 2.1.59 (Ruang Tensor, (Greub, 1978; Hackbusch, 2012)) Misalkan V_1, \dots, V_d ruang vektor atas suatu lapangan K . Ruang tensor orde- d didefinisikan sebagai hasil produk tensor

$$V_1 \otimes V_2 \otimes \dots \otimes V_d,$$

yang dibangun dari kombinasi linear tensor murni

$$\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \dots \otimes \mathbf{v}_d, \quad \mathbf{v}_i \in V_i.$$

Teorema 2.1.60 (Sifat Universal Produk Tensor, (Greub, 1978)) Diberikan pemetaan multilinear

$$\varphi : V_1 \times \dots \times V_d \rightarrow W,$$

maka terdapat pemetaan linear unik $\tilde{\varphi} : V_1 \otimes \dots \otimes V_d \rightarrow W$ sedemikian sehingga

$$\varphi(\mathbf{v}_1, \dots, \mathbf{v}_d) = \tilde{\varphi}(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_d). \quad (2.66)$$

Contoh 2.1.61 Berikut ini adalah contoh untuk beberapa dimensi d .

1. Untuk $d = 1$, diperoleh V_1 , sehingga tensor orde-1 identik dengan vektor.
2. Untuk $d = 2$, diperoleh $V_1 \otimes V_2$, yang isomorfik dengan ruang matriks $\dim(V_1) \times \dim(V_2)$. Matriks adalah tensor orde-2.
3. Untuk $d = 3$ dan $V = \mathbb{R}^n$, setiap $\mathcal{T} \in V^{\otimes 3}$ ditulis

$$\mathcal{T} = \sum_{i,j,k=1}^n t_{ijk} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k,$$

dengan \mathbf{e}_i basis standar dan $t_{ijk} \in \mathbb{R}$.

Definisi 2.1.62 (Basis dan Dimensi Tensor, (Hackbusch, 2012)) Jika $\{\mathbf{e}_i\}_{i=1}^n$ basis V , maka

$$\{ \mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \dots \otimes \mathbf{e}_{i_d} : 1 \leq i_1, \dots, i_d \leq n \}$$

adalah basis dari $V^{\otimes d}$, dan

$$\dim(V^{\otimes d}) = (\dim V)^d. \quad (2.67)$$

Definisi 2.1.63 (Tensor Murni dan Rank, (Hackbusch, 2012)) Sebuah tensor $\mathcal{X} \in V_1 \otimes \cdots \otimes V_d$ disebut murni jika

$$\mathcal{X} = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_d.$$

Tensor umum dapat ditulis sebagai kombinasi linear hingga

$$\mathcal{X} = \sum_{r=1}^R \mathbf{v}_1^{(r)} \otimes \cdots \otimes \mathbf{v}_d^{(r)},$$

dan R minimum yang mungkin disebut rank dari \mathcal{X} .

Contoh 2.1.64 Ambil $V = \mathbb{R}^2$ dengan basis $\mathbf{e}_1, \mathbf{e}_2$. Tensor

$$\mathcal{X} = \mathbf{e}_1 \otimes \mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1 \otimes \mathbf{e}_1$$

memerlukan dua suku murni, sehingga $\text{rank}(\mathcal{X}) = 2$.

Definisi 2.1.65 (Matrisasi dan j -rank, (Hackbusch, 2012)) Misalkan $\mathcal{X} \in \bigotimes_{j=1}^d K^{I_j}$ adalah tensor orde- d dengan ukuran $I_1 \times I_2 \times \cdots \times I_d$. Matrisasi mode- j dari \mathcal{X} , ditulis $M_j(\mathcal{X})$, adalah matriks dalam $K^{I_j \times I_{[j]}}$ dengan $I_{[j]} = \prod_{k \neq j} I_k$, yang diperoleh dengan meratakan tensor sepanjang mode- j . Dengan kata lain, indeks i_j dibiarkan tetap sebagai indeks baris, sedangkan gabungan semua indeks lainnya $(i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_d)$ dipetakan menjadi indeks kolom. Rank dari matriks $M_j(\mathcal{X})$ disebut j -rank dari \mathcal{X} .

Contoh 2.1.66 Ambil tensor orde-3 $\mathcal{T} \in \mathbb{R}^{2 \times 3 \times 4}$. Matrisasi mode-1 menghasilkan matriks $M_1(\mathcal{T}) \in \mathbb{R}^{2 \times 12}$, dengan baris menyatakan dimensi pertama ($i_1 = 1, 2$) dan kolom merupakan penggabungan dari indeks dimensi kedua ($i_2 = 1, 2, 3$) dan ketiga ($i_3 = 1, \dots, 4$). Secara skematis, tensor berbentuk balok $2 \times 3 \times 4$ diubah

menjadi matriks berukuran 2×12 .

Definisi 2.1.67 (Perkalian Kronecker, (Hackbusch, 2012)) Untuk $\mathbf{A} \in \text{Mat}_{m \times n}$ dan $\mathbf{B} \in \text{Mat}_{p \times q}$, perkalian Kronecker dari matriks tersebut dituliskan sebagai $\mathbf{A} \otimes \mathbf{B}$, didefinisikan sebagai

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \in \text{Mat}_{mp \times nq}. \quad (2.68)$$

Contoh 2.1.68 Apabila diambil

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in \mathbb{R}^{2 \times 2}, \quad \mathbf{B} = \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} \in \mathbb{R}^{2 \times 2}.$$

Perkalian Kronecker $\mathbf{A} \otimes \mathbf{B}$ adalah

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1 \cdot \mathbf{B} & 2 \cdot \mathbf{B} \\ 3 \cdot \mathbf{B} & 4 \cdot \mathbf{B} \end{pmatrix} = \begin{pmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{pmatrix}. \quad (2.69)$$

Hasilnya adalah matriks berukuran 4×4 karena \mathbf{A} berukuran 2×2 dan \mathbf{B} juga 2×2 , sehingga ukurannya $2 \cdot 2 \times 2 \cdot 2 = 4 \times 4$.

2.1.7. Dekomposisi Matriks dan Tensor

Dekomposisi matriks merupakan representasi suatu matriks sebagai hasil kali atau gabungan matriks-matriks lain yang bentuknya lebih sederhana dan terstruktur. Seiring berkembangnya analisis data multidimensi, konsep dekomposisi meluas dari matriks ke tensor. Oleh karena itu, pemahaman yang sistematis terhadap dekomposisi matriks dan tensor menjadi esensial dalam penelitian ini. Namun, sebelum membahas masing-masing dekomposisi, bagian berikut memberikan definisi-definisi yang digunakan dalam proses dekomposisi.

Definisi 2.1.69 (Hasil Kali Dalam, (Axler, 2024)) Misalkan V ruang vektor atas F ($F = \mathbb{R}$ atau $F = \mathbb{C}$). Suatu fungsi $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$ disebut hasil kali dalam (inner product) jika untuk semua $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ dan $\lambda \in F$ berlaku

$$(a) \langle \mathbf{v}, \mathbf{v} \rangle \geq 0;$$

$$(b) \langle \mathbf{v}, \mathbf{v} \rangle = 0 \iff \mathbf{v} = \mathbf{0};$$

$$(c) \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle, \text{ dan } \langle \lambda \mathbf{u}, \mathbf{v} \rangle = \lambda \langle \mathbf{u}, \mathbf{v} \rangle; \text{ dan}$$

$$(d) \langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}.$$

Definisi 2.1.70 (Adjoint, (Axler, 2024)) Untuk operator linear $T : V \rightarrow W$ antar ruang berhasil kali dalam, adjoint $T^* : W \rightarrow V$ didefinisikan sebagai pemetaan unik yang memenuhi

$$\langle T\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, T^*\mathbf{w} \rangle \quad (2.70)$$

untuk semua $\mathbf{v} \in V, \mathbf{w} \in W$.

Definisi 2.1.71 (Matriks Unitary dan Ortogonal, (Axler, 2024)) Matriks persegi $\mathbf{Q} \in \mathbb{C}^{n \times n}$ disebut unitary jika $\mathbf{Q}^* \mathbf{Q} = \mathbf{I}$. Jika \mathbf{Q} real, syarat ini setara dengan $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$, dan \mathbf{Q} disebut ortogonal.

Definisi 2.1.72 (Matriks Hermitian, (Axler, 2024)) Matriks $\mathbf{A} \in \mathbb{C}^{n \times n}$ disebut Hermitian jika $\mathbf{A}^* = \mathbf{A}$. Jika \mathbf{A} real, syarat ini setara dengan $\mathbf{A}^\top = \mathbf{A}$ atau simetris.

Definisi 2.1.73 (Dekomposisi Eigen, (Axler, 2024)) Misalkan $\mathbf{A} \in \mathbb{C}^{n \times n}$. Jika \mathbf{A} memiliki n buah vektor eigen bebas linear, maka matriks \mathbf{A} disebut diagonalisabel. Dalam hal ini, terdapat matriks invertibel \mathbf{P} yang kolom-kolomnya merupakan vektor-vektor eigen \mathbf{A} , serta matriks diagonal $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ yang berisi nilai-nilai eigen λ_i pada diagonal utamanya, sehingga

$$\mathbf{A} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^{-1}. \quad (2.71)$$

Faktorisasi ini disebut dekomposisi eigen atau eigendecomposition.

Contoh 2.1.74 Pertimbangkan matriks

$$\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 1 & 2 \end{pmatrix}.$$

Karakteristik polinomial dari \mathbf{A} adalah $\det(\mathbf{A} - \lambda I) = \lambda^2 - 7\lambda + 6$, sehingga diperoleh nilai eigen $\lambda_1 = 6$ dan $\lambda_2 = 1$. Untuk $\lambda_1 = 6$, persamaan $(\mathbf{A} - 6I)\mathbf{v} = 0$ menghasilkan vektor eigen $\mathbf{v}_1 = (4, 1)^\top$. Sementara itu, untuk $\lambda_2 = 1$, persamaan $(\mathbf{A} - I)\mathbf{v} = 0$ memberikan vektor eigen $\mathbf{v}_2 = (-1, 1)^\top$. Dengan demikian, dapat dibentuk matriks

$$\mathbf{P} = \begin{pmatrix} 4 & -1 \\ 1 & 1 \end{pmatrix},$$

yang kolom-kolomnya adalah vektor-vektor eigen \mathbf{A} , serta matriks diagonal

$$\mathbf{\Lambda} = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix},$$

yang berisi nilai-nilai eigen pada diagonal utamanya. Oleh karena itu, \mathbf{A} dapat dituliskan dalam bentuk

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1},$$

yang merupakan dekomposisi eigen dari matriks \mathbf{A} . Dalam hal ini, dekomposisi menunjukkan bahwa transformasi linear yang direpresentasikan oleh \mathbf{A} ekuivalen dengan transformasi basis ke arah vektor-vektor eigen, dilanjutkan dengan transformasi diagonal, kemudian dikembalikan lagi ke basis semula melalui \mathbf{P}^{-1} .

Definisi 2.1.75 (Faktorisasi QR, (Axler, 2024; Hackbusch, 2012)) Untuk setiap $\mathbf{A} \in \mathbb{F}^{m \times n}$, terdapat matriks unitary atau ortogonal \mathbf{Q} dan segitiga atas \mathbf{R} sehingga

$$\mathbf{A} = \mathbf{Q}\mathbf{R}. \quad (2.72)$$

QR dapat dikonstruksi melalui Algoritma Gram–Schmidt atau Householder. Pemakaian Gram–Schmidt dan penyelesaian $\mathbf{A}\mathbf{x} = \mathbf{b}$ dengan $\mathbf{R}\mathbf{x} = \mathbf{Q}^\mathbf{b}$ dibahas di*

(Axler, 2024). Kompleksitas dan varian numeriknya, yaitu pivoted *QR* dan *Householder* dirangkum di (Hackbusch, 2012).

Contoh 2.1.76 Pertimbangkan matriks

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},$$

dengan kolom-kolom $\mathbf{a}_1 = (1, 1, 0)^\top$ dan $\mathbf{a}_2 = (1, 0, 1)^\top$. Prosedur Gram–Schmidt dimulai dengan menormalkan kolom pertama, sehingga

$$\|\mathbf{a}_1\| = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}, \quad \mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} = \frac{1}{\sqrt{2}}(1, 1, 0)^\top.$$

Selanjutnya, kolom kedua diortogonalisasi terhadap \mathbf{q}_1 melalui proyeksi; koefisien proyeksinya adalah

$$r_{12} = \langle \mathbf{q}_1, \mathbf{a}_2 \rangle = \frac{1}{\sqrt{2}}(1 \cdot 1 + 1 \cdot 0 + 0 \cdot 1) = \frac{1}{\sqrt{2}},$$

sehingga vektor residu menjadi

$$\hat{\mathbf{a}}_2 = \mathbf{a}_2 - r_{12}\mathbf{q}_1 = (1, 0, 1)^\top - \frac{1}{2}(1, 1, 0)^\top = \left(\frac{1}{2}, -\frac{1}{2}, 1\right)^\top.$$

Normalisasi residu tersebut menghasilkan

$$\|\hat{\mathbf{a}}_2\| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + 1^2} = \sqrt{\frac{3}{2}}, \quad \mathbf{q}_2 = \frac{\hat{\mathbf{a}}_2}{\|\hat{\mathbf{a}}_2\|} = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)^\top.$$

Dengan demikian, kolom-kolom ortonormal terkumpul dalam

$$\mathbf{Q} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \end{pmatrix}.$$

Matriks segitiga atas \mathbf{R} diperoleh dari koefisien Gram–Schmidt, yakni $r_{11} = \|\mathbf{a}_1\| = \sqrt{2}$, $r_{12} = \langle \mathbf{q}_1, \mathbf{a}_2 \rangle = \frac{1}{\sqrt{2}}$, dan $r_{22} = \|\hat{\mathbf{a}}_2\| = \sqrt{\frac{3}{2}}$, dengan $r_{21} = 0$, sehingga

$$\mathbf{R} = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{3}{2}} \end{pmatrix}.$$

Terlihat bahwa $\mathbf{A} = \mathbf{QR}$ dengan \mathbf{Q} berkolom ortonormal dan \mathbf{R} segitiga atas, sesuai definisi faktorisasi QR.

Definisi 2.1.77 (Dekomposisi Nilai Tunggal atau *Singular Value Decomposition* (SVD), (Axler, 2024; Hackbusch, 2012)) Untuk $\mathbf{A} \in \mathbb{F}^{m \times n}$ terdapat unitary atau ortogonal $\mathbf{U} \in \mathbb{F}^{m \times m}$, $\mathbf{V} \in \mathbb{F}^{n \times n}$, dan diagonal taknegatif $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ sehingga

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \quad (2.73)$$

dengan nilai singular $\sigma_1 \geq \dots \geq \sigma_r > 0$.

Contoh 2.1.78 (SVD pada matriks 2×2) Pertimbangkan matriks

$$\mathbf{A} = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}.$$

Perhitungan SVD diawali dengan membentuk $\mathbf{A}^\top \mathbf{A}$, yaitu

$$\mathbf{A}^\top \mathbf{A} = \begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix}.$$

Matriks simetris ini memiliki polinomial karakteristik

$$\det \left(\begin{pmatrix} 25 - \lambda & 20 \\ 20 & 25 - \lambda \end{pmatrix} \right) = (25 - \lambda)^2 - 400 = 0,$$

sehingga nilai eigennya adalah $\lambda_1 = 45$ dan $\lambda_2 = 5$. Nilai singular σ_i diperoleh

sebagai akar kuadrat positif dari nilai eigen tersebut, yakni

$$\sigma_1 = \sqrt{45} = 3\sqrt{5}, \quad \sigma_2 = \sqrt{5}.$$

Vektor-vektor singular kanan adalah eigenvektor dari $\mathbf{A}^\top \mathbf{A}$. Untuk $\lambda_1 = 45$ persamaan $(\mathbf{A}^\top \mathbf{A} - 45\mathbf{I})\mathbf{v} = \mathbf{0}$ memberikan ruang eigen yang dibentang oleh $(1, 1)^\top$, sedangkan untuk $\lambda_2 = 5$ memberikan ruang eigen yang dibentang oleh $(1, -1)^\top$. Dengan mengortonormalkan keduanya, dapat diambil

$$\mathbf{V} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Selanjutnya, vektor-vektor singular kiri ditentukan dari rumus $\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A} \mathbf{v}_i$. Substitusi menghasilkan

$$\mathbf{u}_1 = \frac{1}{3\sqrt{5}} \mathbf{A} \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{5}} \mathbf{A} \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) = \begin{pmatrix} \frac{3}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} \end{pmatrix},$$

yang keduanya saling ortonormal. Dengan demikian,

$$\mathbf{U} = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \end{pmatrix}, \quad \mathbf{V} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

sehingga faktorisasi nilai tunggal $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$ terbentuk. Interpretasi faktorisasi ini adalah bahwa transformasi linear yang direpresentasikan oleh \mathbf{A} ekuivalen dengan perubahan basis ke arah vektor-vektor singular kanan, diikuti oleh penskalaan diagonal oleh nilai-nilai singular, lalu dilanjutkan dengan pemetaan kembali melalui basis vektor singular kiri.

Definisi 2.1.79 (Dekomposisi Cholesky, (Axler, 2024; Hackbusch, 2012)) Jika $\mathbf{A} \in \mathbb{R}^{n \times n}$ simetris dan positif definit, maka terdapat unik segitiga bawah \mathbf{L} dengan diagonal positif sehingga $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$ (atau $\mathbf{A} = \mathbf{R}^* \mathbf{R}$ dengan \mathbf{R} segitiga atas).

Contoh 2.1.80 (Dekomposisi Cholesky pada matriks 2×2) Pertimbangkan matriks simetris positif definit

$$\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}.$$

Sebab determinan minor utamanya bernilai positif, yakni $\det(4) = 4 > 0$ dan $\det(\mathbf{A}) = 4 \cdot 3 - 2 \cdot 2 = 8 > 0$, maka \mathbf{A} memang positif definit. Dekomposisi Cholesky mencari segitiga bawah $\mathbf{L} = \begin{pmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{pmatrix}$ dengan diagonal positif sehingga $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$. Dengan menyamakan entri pada

$$\mathbf{L}\mathbf{L}^\top = \begin{pmatrix} \ell_{11}^2 & \ell_{11}\ell_{21} \\ \ell_{11}\ell_{21} & \ell_{21}^2 + \ell_{22}^2 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix},$$

diperoleh berturut-turut $\ell_{11}^2 = 4$ sehingga $\ell_{11} = 2$, $\ell_{11}\ell_{21} = 2$ sehingga $\ell_{21} = 1$, dan $\ell_{21}^2 + \ell_{22}^2 = 3$ sehingga $\ell_{22} = \sqrt{2}$. Dengan demikian,

$$\mathbf{L} = \begin{pmatrix} 2 & 0 \\ 1 & \sqrt{2} \end{pmatrix} \quad \text{dan} \quad \mathbf{A} = \mathbf{L}\mathbf{L}^\top.$$

Untuk tensor, dinotasikan tensor orde-3 sebagai $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$. Mode- n product ditulis $\mathcal{X} \times_n \mathbf{A}$, yaitu .

Definisi 2.1.81 (Dekomposisi CP dan Tucker, (Hackbusch, 2012)) Dua pendekatan yang paling umum digunakan untuk mendekomposisi tensor berorde lebih tinggi adalah dekomposisi CP (Canonical Polyadic atau CANDECOMP) dan dekomposisi Tucker atau High-Order Singular Value Decomposition (HOSVD).

Dekomposisi CP bertujuan menuliskan suatu tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ sebagai jumlah terbatas dari tensor dengan rank satu, yaitu

$$\mathcal{X} \approx \sum_{r=1}^R \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r, \quad (2.74)$$

dengan $\mathbf{a}_r \in \mathbb{R}^I$, $\mathbf{b}_r \in \mathbb{R}^J$, dan $\mathbf{c}_r \in \mathbb{R}^K$. Jumlah komponen R disebut rank CP

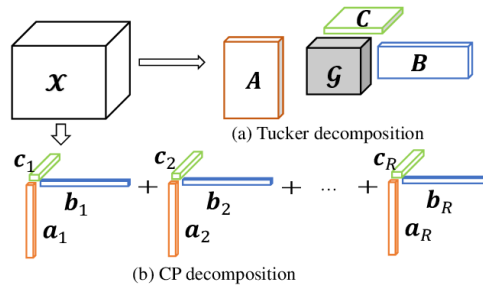
dari tensor, dan dekomposisi ini sering dipandang sebagai generalisasi langsung dari dekomposisi eigen untuk matriks, karena setiap komponen merupakan analogi vektor eigen yang membentuk rank-one tensor.

Sementara itu, dekomposisi Tucker atau HOSVD menuliskan tensor \mathcal{X} dalam bentuk

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}, \quad (2.75)$$

dengan \mathcal{G} merupakan tensor inti yang berukuran lebih kecil dan $\mathbf{A}, \mathbf{B}, \mathbf{C}$ adalah matriks faktor ortonormal. Tidak seperti CP yang memaksa representasi dalam bentuk jumlah rank satu, dekomposisi Tucker memperbolehkan adanya interaksi yang lebih kompleks di dalam tensor inti \mathcal{G} . HOSVD menjamin bahwa matriks faktor diperoleh dari dekomposisi nilai tunggal (SVD) pada matriks hasil pembentangan (unfolding) tensor sepanjang setiap mode. Dekomposisi ini juga dikenal bersifat quasi-optimal, dalam arti bahwa jika dilakukan pemotongan (truncation) pada singular values, maka kesalahan aproksimasi yang diperoleh tetap terkendali secara teoretis.

Perhatikan ilustrasi di bawah ini untuk penjelasan terkait dekomposisi CP dan Tucker secara visual.



Gambar 2.5 Ilustrasi (a) Dekomposisi Tucker dan (b) Dekomposisi CP pada Tensor. (Sumber: Jiang, Huang, Zhao, & Deng, 2020)

Contoh 2.1.82 Sebagai ilustrasi, ambil tensor $\mathcal{X} \in \mathbb{R}^{2 \times 2 \times 2}$ dengan entri

$$\mathcal{X}_{ijk} = \begin{cases} 1, & \text{jika } i = j = k, \\ 0, & \text{selainnya.} \end{cases}$$

Tensor ini hanya bernilai 1 pada posisi diagonal $(1, 1, 1)$ dan $(2, 2, 2)$, sedangkan seluruh elemen lainnya bernilai 0.

Dalam kerangka CP, tensor ini dapat direpresentasikan sebagai jumlah dari dua rank-one tensors:

$$\mathcal{X} \approx \mathbf{e}_1 \otimes \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 \otimes \mathbf{e}_2,$$

dengan $\mathbf{e}_1 = (1, 0)^\top$ dan $\mathbf{e}_2 = (0, 1)^\top$. Setiap komponen di sini langsung menangkap satu “garis diagonal” dari tensor, sehingga rank CP-nya adalah 2. Hal ini mirip dengan menyatakan sebuah matriks diagonal sebagai jumlah dari matriks-matriks *rank* satu.

Dalam kerangka Tucker/HOSVD, prosedur dilakukan dengan membentangkan tensor \mathcal{X} ke dalam tiga matriks *unfolding* sesuai mode-1, mode-2, dan mode-3. Masing-masing matriks kemudian dikenai dekomposisi nilai tunggal (SVD) sehingga diperoleh faktor ortonormal \mathbf{A} , \mathbf{B} , dan \mathbf{C} . Tensor inti \mathcal{G} diperoleh dengan mengalikan \mathcal{X} berturut-turut dengan \mathbf{A}^\top , \mathbf{B}^\top , dan \mathbf{C}^\top pada mode yang sesuai. Dalam contoh ini, faktor-faktor ortonormal yang muncul berkorespondensi langsung dengan basis standar \mathbf{e}_1 , \mathbf{e}_2 , sedangkan inti \mathcal{G} akan berukuran $2 \times 2 \times 2$ dengan entri yang lebih terkonsentrasi pada diagonal.

2.1.8. Diferensial Vektor dan Matriks

Kalkulus matriks adalah perpanjangan dari kalkulus diferensial pada ruang Euclidean ke fungsi yang melibatkan vektor dan matriks sebagai argumen maupun hasil.

Definisi 2.1.83 (Konvensi Turunan pada Vektor dan Matriks, (Dhrymes, 2013))

Misalkan

$$\mathbf{y} = \psi(\mathbf{x}), \quad (2.76)$$

dengan \mathbf{y} dan \mathbf{x} secara berurutan merupakan vektor kolom berukuran m dan n .

Notasi

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j} \right], \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n \quad (2.77)$$

merupakan notasi dari matriks derivatif parsial orde pertama (atau matriks Jacobi-an) dari transformasi \mathbf{x} ke \mathbf{y} sedemikian sehingga baris ke- i berisi turunan parsial dari y_i terhadap elemen dari \mathbf{x} , yaitu

$$\frac{\partial y_i}{\partial x_1}, \frac{\partial y_i}{\partial x_2}, \dots, \frac{\partial y_i}{\partial x_n}. \quad (2.78)$$

Proposisi 2.1.84 *Jika*

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.79)$$

dengan \mathbf{A} merupakan matriks berukuran $m \times n$ yang tidak bergantung pada \mathbf{x} , maka berlaku

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}. \quad (2.80)$$

Proposisi 2.1.85 *Jika*

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.81)$$

dengan \mathbf{y} berukuran $m \times 1$, \mathbf{A} berukuran $m \times n$, \mathbf{x} berukuran $n \times 1$, serta \mathbf{A} dan \mathbf{x} bergantung pada vektor $\boldsymbol{\alpha}$ berukuran r , maka

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\alpha}} = (\mathbf{x}^\top \otimes \mathbf{I}_m) \frac{\partial \mathbf{A}}{\partial \boldsymbol{\alpha}} + \mathbf{A} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}}. \quad (2.82)$$

Proposisi 2.1.86 *Jika*

$$\mathbf{y} = \mathbf{z}^\top \mathbf{A}\mathbf{x}, \quad (2.83)$$

dengan \mathbf{z} berukuran $m \times 1$, \mathbf{A} berukuran $m \times n$, \mathbf{x} berukuran $n \times 1$, serta \mathbf{A} independen dari \mathbf{z} dan \mathbf{x} , maka

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \mathbf{x}^\top \mathbf{A}^\top, \quad \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{z}^\top \mathbf{A}. \quad (2.84)$$

Proposisi 2.1.87 *Jika*

$$\mathbf{y} = \mathbf{x}^\top \mathbf{A}\mathbf{x}, \quad (2.85)$$

dengan \mathbf{x} berukuran $n \times 1$ serta \mathbf{A} berukuran $n \times n$ dan independen dengan \mathbf{x} ,

maka

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top). \quad (2.86)$$

Jika \mathbf{A} adalah matriks simetri, maka

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = 2\mathbf{x}^\top \mathbf{A}. \quad (2.87)$$

Proposisi 2.1.88 Misalkan matriks \mathbf{A} nonsingular yang berukuran $m \times m$ dan bergantung dengan parameter skalar α , maka

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1}. \quad (2.88)$$

Teorema 2.1.89 (Aturan Jejak, , (Dhrymes, 2013)) Untuk matriks dengan ukuran yang sesuai berlaku:

$$\frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{A}\mathbf{X}) = \mathbf{A}^\top, \quad \frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{X}.$$

2.2. Regresi Spasial dan Spasial-Temporal

Analisis regresi merupakan suatu teknik statistik untuk menginvestigasi dan memodelkan hubungan antarvariabel. Aplikasi dari analisis regresi sangat luas dan digunakan dalam hampir semua bidang, seperti teknik atau *engineering*, ilmu fisika dan kimia, ekonomi, manajemen, ilmu biologi, dan sosiologi. Dalam analisis regresi, variabel respons atau variabel dependen (dinotasikan dengan variabel random Y) dimodelkan dengan fungsi dalam variabel penjelas atau variabel independen (dinotasikan dengan variabel random X). (Montgomery, Peck, & Vining, 2012)

2.2.1. Regresi Linear Biasa atau *Ordinary Least Squares* (OLS)

Analisis regresi linear memodelkan hubungan antarvariabel dengan fungsi linear. Apabila terdapat satu variabel prediktor maka model yang terbentuk disebut sebagai regresi linear sederhana, sedangkan apabila terdapat lebih dari satu variabel prediktor maka model yang terbentuk disebut sebagai regresi linear berganda.

Definisi 2.2.1 (Regresi Linear Sederhana, (Montgomery dkk., 2012)) Model regresi linear sederhana merupakan model dengan satu variabel penjelas X yang mempunyai hubungan garis lurus terhadap satu respons Y . Model ini dirumuskan dengan

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \boldsymbol{\varepsilon}, \quad (2.89)$$

dengan intersep β_0 dan koefisien β_1 merupakan konstanta yang tidak diketahui, serta $\boldsymbol{\varepsilon}$ adalah vektor galat acak. \mathbf{y} dan \mathbf{x} berturut-turut adalah vektor acak Y dan X . Parameter β_0 dan β_1 biasa disebut sebagai koefisien-koefisien regresi.

Definisi 2.2.2 (Regresi Linear Berganda, (Montgomery dkk., 2012)) Regresi linear berganda merupakan model regresi yang memiliki lebih dari satu variabel independen. Model ini merupakan generalisasi dari regresi linear sederhana. Vektor respons \mathbf{y} mungkin memiliki hubungan dengan p -buah variabel penjelas, yaitu matriks acak \mathbf{X} yang dibentuk atas p -buah vektor acak $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$, sehingga

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_p \mathbf{x}_p + \boldsymbol{\varepsilon}. \quad (2.90)$$

Model di atas dapat dituliskan dalam notasi matriks dan vektor, yaitu

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad (2.91)$$

atau

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.92)$$

Model OLS mengasumsikan bahwa galat memiliki rata-rata nol dan variansi konstan σ^2 yang tidak diketahui nilainya. Lebih lanjut, model ini juga mengasumsikan bahwa komponen galat tidak memiliki autokorelasi yang berarti bahwa nilai dari suatu galat tidak bergantung pada galat lainnya. Dengan kata lain $\boldsymbol{\varepsilon} \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2)$, untuk suatu σ^2 yang bersifat konstan.

Model regresi linear memandang vektor acak \mathbf{y} sebagai fungsi dari matriks acak \mathbf{X} . Oleh karena itu, rata-rata dari distribusi \mathbf{y} dapat dituliskan sebagai

$$\mathbb{E}[\mathbf{y} \mid \mathbf{X}] = \mathbb{E}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \mathbf{X}\boldsymbol{\beta} \quad (2.93)$$

dan variansinya

$$\text{Var}(\mathbf{y} \mid \mathbf{X}) = \text{Var}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n, \quad (2.94)$$

karena $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ dan $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$. Oleh karena itu, rata-rata dari \mathbf{y} adalah fungsi linear dari \mathbf{X} meskipun variansinya tidak bergantung pada nilai \mathbf{X} .

Definisi 2.2.3 (Asumsi Regresi Linear Klasik, (Greene, 2002; Gujarati & Porter, 2009; Montgomery dkk., 2012)) *Model regresi linear klasik didasarkan pada sejumlah asumsi yang dikenal sebagai asumsi Gauss–Markov. Asumsi-asumsi ini diperlukan agar penduga kuadrat terkecil memiliki sifat optimal sebagai best linear unbiased estimator (BLUE). Asumsi-asumsi tersebut adalah sebagai berikut.*

- (a) *Model bersifat linear dalam parameter, yaitu $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.*
- (b) *Nilai tengah galat adalah nol bersyarat pada \mathbf{X} , yaitu $\mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$.*
- (c) *Matriks desain \mathbf{X} memiliki rank penuh sehingga $\mathbf{X}^\top \mathbf{X}$ nonsingular.*
- (d) *Galat memiliki ragam konstan, $\text{Var}(\varepsilon_i \mid \mathbf{X}) = \sigma^2$ untuk semua i .*
- (e) *Komponen galat tidak berkorelasi, $\text{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) = 0$ untuk $i \neq j$.*
- (f) *Untuk keperluan inferensi, diasumsikan $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.*

Estimasi parameter dalam regresi linear dilakukan dengan metode kuadrat terkecil atau *ordinary least squares* (OLS). Prinsip dasarnya adalah mencari nilai $\boldsymbol{\beta}$ yang meminimumkan jumlah kuadrat galat

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (2.95)$$

Perhatikan bahwa fungsi kuadrat terkecil dari persamaan di atas dapat dituliskan

sebagai

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) = \mathbf{y}^\top \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta. \quad (2.96)$$

Dengan diferensial matriks, gradien terhadap β adalah

$$\nabla_\beta S(\beta) = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \beta. \quad (2.97)$$

Syarat orde pertama meminimumkan S adalah $\nabla_\beta S(\beta) = \mathbf{0}$, sehingga diperoleh

$$\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{y}, \quad (2.98)$$

yang disebut persamaan normal. Selanjutnya, matriks Hessian atau orde kedua adalah

$$\nabla_\beta^2 S(\beta) = 2\mathbf{X}^\top \mathbf{X} \succeq \mathbf{0}, \quad (2.99)$$

sehingga S konveks dalam β . Jika $\mathbf{X}^\top \mathbf{X}$ non-singular atau setara $\text{rank}(\mathbf{X}) = p$, maka Hessian definit positif dan solusi orde pertama unik serta memberikan minimum global.

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.100)$$

Jika $\mathbf{X}^\top \mathbf{X}$ singular, himpunan minimum tidak tunggal. Salah satu solusi adalah solusi *norma-minimum* $\hat{\beta} = \mathbf{X}^+ \mathbf{y}$ dengan \mathbf{X}^+ merupakan *Moore–Penrose pseudoinverse*.

Teorema 2.2.4 (Teorema Gauss–Markov, (Greene, 2002; Montgomery dkk., 2012)) *Di bawah asumsi regresi linear klasik, penduga OLS $\hat{\beta}$ bersifat tak bias, yaitu $\mathbb{E}[\hat{\beta}] = \beta$, dan memiliki kovariansi*

$$\text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.101)$$

Selain itu, $\hat{\beta}$ merupakan best linear unbiased estimator (BLUE), artinya $\hat{\beta}$ memiliki variansi terkecil di antara semua penduga linear tak bias.

Bukti. Penduga OLS dapat dituliskan sebagai

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.102)$$

Substitusi $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ menghasilkan

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\beta + \varepsilon) = \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \varepsilon, \quad (2.103)$$

sehingga rata-rata dari $\hat{\beta}$ adalah

$$\mathbb{E}[\hat{\beta}] = \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[\varepsilon] = \beta, \quad (2.104)$$

yang menunjukkan bahwa $\hat{\beta}$ tak bias. Kovariansnya diperoleh dari

$$\text{Var}(\hat{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Var}(\varepsilon) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.105)$$

Untuk sifat BLUE, dipertimbangkan penduga linear tak bias umum $\tilde{\beta} = \mathbf{C}\mathbf{y}$ dengan $\mathbf{C}\mathbf{X} = \mathbf{I}$. Variansi dari $\tilde{\beta}$ adalah

$$\text{Var}(\tilde{\beta}) = \mathbf{C}\sigma^2\mathbf{I}\mathbf{C}^\top = \sigma^2\mathbf{C}\mathbf{C}^\top. \quad (2.106)$$

Melalui dekomposisi $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top + \mathbf{D}$ dengan $\mathbf{D}\mathbf{X} = \mathbf{0}$, diperoleh

$$\text{Var}(\tilde{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} + \sigma^2 \mathbf{D}\mathbf{D}^\top \succeq \text{Var}(\hat{\beta}), \quad (2.107)$$

yang membuktikan bahwa $\hat{\beta}$ memiliki variansi minimum di antara semua penduga linear tak bias. ■

Teorema 2.2.5 (Kepadanan OLS dan MLE, (Montgomery dkk., 2012)) *Jika asumsi normalitas galat terpenuhi, maka penduga kemungkinan maksimum (PKM) atau maximum likelihood estimation (MLE) untuk β identik dengan penduga kua-*

drat terkecil (OLS), yaitu

$$\hat{\beta}_{MLE} = \hat{\beta}_{OLS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.108)$$

Penduga MLE untuk varians galat adalah

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X} \hat{\beta})^\top (\mathbf{y} - \mathbf{X} \hat{\beta}). \quad (2.109)$$

Bukti. Dengan asumsi $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, fungsi *likelihood* untuk β dan σ^2 adalah

$$L(\beta, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta)\right), \quad (2.110)$$

yang memiliki log-*likelihood*

$$\ell(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta). \quad (2.111)$$

Maksimisasi terhadap β dengan turunan pertama didapatkan

$$\begin{aligned} \frac{\partial \ell(\beta, \sigma^2)}{\partial \beta} &= \frac{2\mathbf{X}^\top}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta) \\ &= \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} \beta \\ \mathbf{0} &= \mathbf{X}^\top \mathbf{y} - \mathbf{X}^\top \mathbf{X} \hat{\beta} \\ \mathbf{X}^\top \mathbf{X} \hat{\beta} &= \mathbf{X}^\top \mathbf{y} \\ \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \end{aligned} \quad (2.112)$$

Perhatikan bahwa turunan kedua terhadap β menghasilkan $-\frac{1}{\sigma^2} (\mathbf{X}^\top \mathbf{X})$. Sebab $\mathbf{X}^\top \mathbf{X} \succeq 0$, maka $\frac{\partial \ell}{\partial \beta} \preceq 0$ yang menjadikan $\hat{\beta}$ memaksimumkan fungsi log-*likelihood*. Oleh karena itu

$$\hat{\beta}_{MLE} = \hat{\beta}_{OLS}. \quad (2.113)$$

Selanjutnya, turunan pertama ℓ terhadap σ^2 akan menghasilkan

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ 0 &= -n\hat{\sigma}^2 + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ \hat{\sigma}^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).\end{aligned}\tag{2.114}$$

■

2.2.2. Regresi Terboboti

Regresi terboboti dibentuk apabila model regresi linear klasik melanggar asumsi homoskedastisitas galat. Heteroskedastisitas terjadi apabila komponen galat regresi tidak memiliki variansi yang konstan. Apabila terjadi heteroskedastisitas, maka Teorema Gauss-Markov tidak berlaku sehingga penduga dari model regresi tidak lagi bersifat BLUE.

Definisi 2.2.6 (Model Regresi Linear dengan Heteroskedastisitas, (Greene, 2002; Gujarati & Porter, 2009; Montgomery dkk., 2012)) *Pertimbangkan model regresi linear*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}, \tag{2.115}$$

dengan struktur ragam-kovarians galat

$$\text{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \boldsymbol{\Omega}, \tag{2.116}$$

dengan $\boldsymbol{\Omega}$ adalah matriks simetri definit-positif berukuran $n \times n$. Model yang memperkenalkan struktur ragam-kovarians galat disebut sebagai model kuadrat terkecil tergeneralisasi atau generalized least squares (GLS).

Kasus umum dari GLS yang sering dipakai adalah kuadrat terkecil terboboti atau *weighted least squares* (WLS), yaitu $\boldsymbol{\Omega} = \text{diag}(\omega_1, \dots, \omega_n)$ yang merupakan heteroskedastisitas diagonal dengan bobot $\mathbf{W} = \boldsymbol{\Omega}^{-1} = \text{diag}(w_1, \dots, w_n)$, $w_i = 1/\omega_i$.

Penduga GLS diperoleh dengan meminimumkan jumlah kuadrat terbobot

$$S_{\mathbf{W}}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (2.117)$$

dengan $\mathbf{W} = \boldsymbol{\Omega}^{-1}$. Diferensiasi terhadap $\boldsymbol{\beta}$ dan disamakan dengan nol memberikan

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} S_{\mathbf{W}}(\boldsymbol{\beta}) &= -2\mathbf{W}\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{karena } \mathbf{W} \text{ simetris}) \\ 0 &= -2\mathbf{X}^\top \mathbf{W}\mathbf{y} - 2\mathbf{X}^\top \mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} \\ \mathbf{X}^\top \mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^\top \mathbf{W}\mathbf{y}, \end{aligned} \quad (2.118)$$

$$(2.119)$$

sehingga pada asumsi $\mathbf{X}^\top \mathbf{W}\mathbf{X}$ nonsingular,

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}\mathbf{y}. \quad (2.120)$$

Penduga ini tak bias jika \mathbf{W} tidak acak atau independen dari $\boldsymbol{\varepsilon}$ dan $\mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$, sebab $\mathbb{E}[\hat{\boldsymbol{\beta}}_{\text{GLS}} \mid \mathbf{X}] = \boldsymbol{\beta}$.

Teorema 2.2.7 (Aitken atau Generalisasi Gauss–Markov, (Greene, 2002; Montgomery dkk., 2012)) Misalkan model regresi linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}, \quad \text{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \boldsymbol{\Omega},$$

dengan $\boldsymbol{\Omega}$ simetris definit-positif, maka penduga

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y}$$

adalah best linear unbiased estimator (BLUE) dari $\boldsymbol{\beta}$.

Bukti. Pertama, perhatikan bahwa setiap penduga linear bagi $\boldsymbol{\beta}$ dapat ditulis dalam bentuk $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ dengan $\mathbf{A} \in \mathbb{R}^{p \times n}$. Agar tak bias, berlaku

$$\mathbb{E}[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}] = \mathbf{A}\mathbb{E}[\mathbf{y} \mid \mathbf{X}] = \mathbf{A}\mathbf{X}\boldsymbol{\beta}.$$

Supaya sama dengan β untuk semua β , syarat yang harus dipenuhi adalah $\mathbf{A}\mathbf{X} = \mathbf{I}_p$. Selanjutnya, kovarians dari $\tilde{\beta}$ diberikan oleh

$$\text{Var}(\tilde{\beta} \mid \mathbf{X}) = \mathbf{A} \sigma^2 \mathbf{\Omega} \mathbf{A}^\top.$$

Di antara semua matriks \mathbf{A} yang memenuhi $\mathbf{A}\mathbf{X} = \mathbf{I}_p$, pilih

$$\mathbf{A}_0 = (\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Omega}^{-1}.$$

Jelas bahwa $\mathbf{A}_0 \mathbf{X} = \mathbf{I}_p$, sehingga \mathbf{A}_0 menghasilkan penduga tak bias. Penduga inilah yang dikenal sebagai *generalized least squares*.

Sekarang, ambil sembarang penduga tak bias lain \mathbf{A} . Matriks \mathbf{A} tersebut dapat ditulis sebagai $\mathbf{A} = \mathbf{A}_0 + \mathbf{D}$ dengan $\mathbf{D}\mathbf{X} = \mathbf{0}$, sebab $\mathbf{A}_0 \mathbf{X} = \mathbf{I}_p$ dan syarat $\mathbf{A}\mathbf{X} = \mathbf{I}_p$ harus tetap dipenuhi. Dengan demikian,

$$\tilde{\beta} = \mathbf{A}\mathbf{y} = \mathbf{A}_0 \mathbf{y} + \mathbf{D}\mathbf{y}.$$

Variansi dari penduga ini adalah

$$\text{Var}(\tilde{\beta} \mid \mathbf{X}) = \text{Var}(\mathbf{A}_0 \mathbf{y} \mid \mathbf{X}) + \sigma^2 \mathbf{D} \mathbf{\Omega} \mathbf{D}^\top.$$

Bagian pertama adalah kovarians dari $\hat{\beta}_{\text{GLS}}$, sedangkan bagian kedua selalu semi definit positif karena $\mathbf{\Omega}$ definit-positif. Hal ini berarti tambahan $\mathbf{D}\mathbf{y}$ hanya menambah variansi tanpa mengurangi bias. Dengan demikian, kovarians dari sembarang penduga linear tak bias selalu lebih besar atau sama dengan kovarians GLS. Oleh karena itu, $\hat{\beta}_{\text{GLS}}$ meminimalkan kovarians di antara semua penduga linear tak bias, sehingga merupakan BLUE. ■

Teorema 2.2.8 (Kepadanan GLS dan MLE di Bawah Normalitas, (Greene, 2002; Montgomery dkk., 2012)) Jika $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{\Omega})$, maka penduga maksimum

likelihood (MLE) untuk β adalah

$$\hat{\beta}_{MLE} = \hat{\beta}_{GLS} = (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{y}, \quad (2.121)$$

dan MLE untuk σ^2 (dengan Ω diketahui) adalah

$$\hat{\sigma}_{MLE}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\beta})^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}). \quad (2.122)$$

Bukti. Di bawah normalitas galat, fungsi *likelihood* untuk β dan σ^2 adalah

$$L(\beta, \sigma^2 \mid \mathbf{y}) = (2\pi\sigma^2)^{-n/2} |\Omega|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta)\right).$$

Hal ini berarti fungsi *log-likelihood* untuk β dan σ^2 adalah

$$\ell(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

Maksimisasi ℓ untuk β dengan turunan pertama menghasilkan

$$\begin{aligned} \frac{\partial \ell(\beta, \sigma^2 \mid \mathbf{y})}{\partial \beta} &= \frac{2}{2\sigma^2} \mathbf{X}^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta) \\ 0 &= \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \hat{\beta} \\ \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \hat{\beta} &= \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{y} \\ \hat{\beta} &= (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{y}, \end{aligned} \quad (2.123)$$

sehingga $\hat{\beta}_{MLE} = \hat{\beta}_{GLS}$. Selanjutnya, turunan terhadap σ^2 memberi $\partial \ell / \partial \sigma^2 = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{y} - \mathbf{X}\hat{\beta})^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}) = 0$, yang menghasilkan rumus $\hat{\sigma}_{MLE}^2$ di atas. ■

Dalam analisis nyata, Ω biasanya jarang diketahui. *Feasible GLS* (FGLS) memperkirakan Ω , misalnya dari residu OLS atau model varian parametrik, lalu menyusun $\widehat{\mathbf{W}} = \widehat{\Omega}^{-1}$ dan menghitung $\hat{\beta}_{FGLS} = (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{y}$. Di bawah kondisi reguler, $\hat{\beta}_{FGLS}$ konsisten dan asimtotik efisien. Namun, untuk sampel kecil, sifat *finite-sample* bergantung pada kualitas spesifikasi $\widehat{\Omega}$.

2.2.3. Model Koefisien Bervariasi

Model koefisien bervariasi merupakan perluasan dari model regresi linear klasik yang memperbolehkan koefisien regresi berubah secara sistematis sesuai dengan karakteristik tertentu, seperti lokasi geografis, waktu, atau variabel lainnya. Model ini merupakan jawaban terhadap keterbatasan model regresi tradisional yang mengasumsikan hubungan konstan antara variabel independen dan dependen di seluruh wilayah observasi. (Fan & Zhang, 2008)

Definisi 2.2.9 (Model Koefisien Bervariasi Umum, (Fan & Zhang, 2008)) *Model koefisien bervariasi dapat diformulasikan sebagai*

$$Y_i = \sum_{k=0}^p \beta_k(z_i) X_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.124)$$

dengan komponen-komponen sebagai berikut:

- (a) Y_i merupakan variabel dependen pada observasi ke- i .
- (b) X_{ik} merupakan variabel independen ke- k pada observasi ke- i , dengan $X_{i0} = 1$.
- (c) $\beta_k(z_i)$ merupakan koefisien yang bervariasi untuk variabel ke- k sebagai fungsi dari vektor karakteristik z_i .
- (d) z_i merupakan vektor variabel yang menentukan variasi koefisien, seperti koordinat spasial (u_i, v_i) .
- (e) ε_i merupakan galat acak dengan $\mathbb{E}[\varepsilon_i] = 0$ dan $\text{Var}(\varepsilon_i) = \sigma^2$.

Pendekatan utama untuk mengestimasi model koefisien bervariasi adalah melalui regresi linear lokal, yang memberikan fleksibilitas dalam menangkap variasi koefisien tanpa perlu menspesifikasi bentuk fungsional secara parametrik.

Definisi 2.2.10 (Regresi Linear Lokal untuk Model Koefisien Bervariasi, (Fan & Zhang, 2008)) *Untuk setiap titik z_0 , koefisien $\beta_k(z_0)$ diestimasi dengan me-*

minimumkan jumlah kuadrat tertimbang lokal:

$$\sum_{i=1}^n K_h(z_i - z_0) \left[Y_i - \sum_{k=0}^p \beta_k X_{ik} \right]^2, \quad (2.125)$$

dengan komponen-komponen:

- (a) $K_h(\cdot)$ adalah fungsi kernel dengan bandwidth h .
- (b) $K_h(u) = K(u/h)/h$ untuk kernel univariat atau $K_h(u) = K(u/h)/h^d$ untuk kernel multivariat dimensi d .
- (c) Fungsi kernel umum yang digunakan adalah Gaussian: $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$.

Teorema 2.2.11 (Estimator Regresi Linear Lokal) *Estimator regresi linear lokal untuk $\beta_k(z_0)$ diberikan oleh:*

$$\hat{\beta}(z_0) = (\mathbf{X}^T \mathbf{W}(z_0) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(z_0) \mathbf{y}, \quad (2.126)$$

dengan $\mathbf{W}(z_0) = \text{diag}\{K_h(z_1 - z_0), \dots, K_h(z_n - z_0)\}$ adalah matriks bobot diagonal.

Teorema 2.2.12 (Pemilihan Bandwidth Optimal, (Fan & Zhang, 2008)) *Bandwidth optimal h dapat dipilih menggunakan metode cross-validation dengan minimumkan:*

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \left[Y_i - \hat{Y}_{-i}(X_i, z_i; h) \right]^2, \quad (2.127)$$

dengan $\hat{Y}_{-i}(X_i, z_i; h)$ adalah prediksi leave-one-out untuk observasi ke- i menggunakan bandwidth h .

Teorema 2.2.13 (Sifat Asimtotik Estimator) *Di bawah kondisi reguler, estimator regresi linear lokal memiliki sifat:*

- (a) **Bias:** $\mathbb{E}[\hat{\beta}_k(z_0)] - \beta_k(z_0) = O(h^2)$ jika $\beta_k(\cdot)$ memiliki turunan kedua yang kontinu.

(b) **Variansi:** $\text{Var}(\hat{\beta}_k(z_0)) = O((nh)^{-1})$.

(c) **MSE Optimal:** Bandwidth optimal yang meminimumkan MSE proporsional dengan $n^{-1/5}$.

Teorema 2.2.14 (Confidence Bands untuk Model Koefisien Bervariasi, (Fan & Zhang, 2008)) Confidence band pointwise untuk $\beta_k(z_0)$ dapat dikonstruksi sebagai:

$$\hat{\beta}_k(z_0) \pm z_{\alpha/2} \sqrt{\hat{\sigma}^2 \mathbf{e}_k^T (\mathbf{X}^T \mathbf{W}(z_0) \mathbf{X})^{-1} \mathbf{e}_k}, \quad (2.128)$$

dengan \mathbf{e}_k adalah vektor unit ke- k dan $\hat{\sigma}^2$ adalah estimator variansi residual.

Teorema 2.2.15 (Pengujian Konstansi Koefisien) Untuk menguji apakah koefisien konstan secara global, dapat digunakan uji berbasis kernel:

$$H_0 : \beta_k(z) = \beta_k \text{ untuk semua } z \quad (2.129)$$

$$H_1 : \beta_k(z) \text{ bervariasi dengan } z \quad (2.130)$$

Statistik uji dapat dikonstruksi menggunakan pendekatan bootstrap atau distribusi asimtotik normal.

Validasi model koefisien bervariasi dapat dilakukan melalui berbagai teknik, termasuk *cross-validation*, *bootstrap*, dan analisis residual. Pemilihan bandwidth yang optimal merupakan aspek kritis yang menentukan keseimbangan antara bias dan variansi estimator.

Contoh 2.2.16 Untuk model dengan satu kovariat yang bervariasi terhadap lokasi z :

$$Y_i = \beta_0(z_i) + \beta_1(z_i)X_i + \varepsilon_i, \quad (2.131)$$

estimasi dilakukan dengan meminimumkan:

$$\sum_{j=1}^n K_h(z_j - z_i) [Y_j - \beta_0 - \beta_1 X_j]^2 \quad (2.132)$$

untuk setiap titik z_i . Hal ini menghasilkan estimator lokal yang adaptif terhadap karakteristik data pada setiap lokasi.

Model koefisien bervariasi menyediakan kerangka kerja yang fleksibel untuk menganalisis heterogenitas spasial dan temporal dalam hubungan statistik, dengan aplikasi yang luas dalam ekonometrika spasial, epidemiologi, dan analisis lingkungan.

2.2.4. Regresi dengan Dependensi Spasial dan Spasial Temporal

Dependensi spasial (*spatial dependence*) atau autokorelasi spasial terjadi ketika nilai suatu variabel di suatu lokasi berkorelasi dengan nilai variabel yang sama di lokasi lain yang berdekatan secara geografis. Konsep ini melanggar asumsi independensi dalam regresi linear klasik dan memerlukan perlakuan khusus dalam pemodelan ekonometrika spasial (Anselin, 1988; LeSage & Pace, 2009).

Definisi 2.2.17 (Dependensi Spasial) Misalkan Y_i merupakan nilai variabel di lokasi i dan Y_j merupakan nilai variabel di lokasi tetangga j . Dependensi spasial terjadi jika

$$\text{Cov}(Y_i, Y_j) \neq 0 \quad \text{untuk lokasi } i \neq j \text{ yang berdekatan} \quad (2.133)$$

yang dapat diukur menggunakan Moran's I statistic:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.134)$$

dengan w_{ij} adalah bobot spasial antara lokasi i dan j .

Model regresi dengan dependensi spasial dapat diklasifikasikan menjadi beberapa bentuk utama berdasarkan dimana dependensi spasial muncul dalam spesifikasi model.

Definisi 2.2.18 (Spatial Autoregressive Model (SAR)) Model SAR atau spatial lag model mengasumsikan bahwa variabel dependen di suatu lokasi dipengaruhi oleh

nilai variabel dependen di lokasi tetangga:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i \quad (2.135)$$

atau dalam bentuk matriks:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.136)$$

dengan ρ adalah parameter autokorelasi spasial dan \mathbf{W} adalah matriks bobot spasial.

Definisi 2.2.19 (Spatial Error Model (SEM)) Model SEM mengasumsikan bahwa dependensi spasial terjadi pada komponen error:

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{W} \mathbf{u} + \boldsymbol{\varepsilon} \quad (2.137)$$

dengan λ adalah parameter autokorelasi spasial pada error dan $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

Definisi 2.2.20 (Spatial Durbin Model (SDM)) Model SDM menggabungkan lag spasial dari variabel dependen dan independen:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{W} \mathbf{X} \boldsymbol{\delta} + \boldsymbol{\varepsilon} \quad (2.138)$$

dengan $\boldsymbol{\delta}$ adalah vektor parameter untuk lag spasial variabel independen.

Untuk kasus spasial temporal, model-model di atas dapat diperluas untuk mengakomodasi dimensi waktu:

Definisi 2.2.21 (Spatio-Temporal Autoregressive (STAR)) Model STAR merupakan perluasan SAR untuk data panel spasial:

$$y_{it} = \rho \sum_{j=1}^n w_{ij} y_{jt} + \sum_{l=1}^p \phi_l y_{it-l} + \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it} \quad (2.139)$$

dengan ϕ_l adalah parameter autoregresif temporal.

Definisi 2.2.22 (*Generalized Spatio-Temporal Autoregressive (GSTAR)*) Model GSTAR mengizinkan heterogenitas dalam parameter spasial:

$$y_{it} = \sum_{j=1}^n \rho_{ij} w_{ij} y_{jt} + \sum_{l=1}^p \phi_{il} y_{it-l} + \varepsilon_{it} \quad (2.140)$$

dengan ρ_{ij} dan ϕ_{il} dapat bervariasi antar lokasi.

Untuk kasus data panel spasial, model-model di atas dapat diperluas untuk mengakomodasi dimensi waktu, menghasilkan model spasial-temporal yang lebih kompleks:

Definisi 2.2.23 (*Spatio-Temporal Autoregressive (STAR)*) Model STAR merupakan perluasan dari model spasial untuk data panel:

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_t + \phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad (2.141)$$

yang menggabungkan lag spasial ($\rho \mathbf{W} \mathbf{y}_t$) dan lag temporal ($\phi \mathbf{y}_{t-1}$).

Definisi 2.2.24 (*Spatio-Temporal Error Model (STEM)*) Jika dependensi spasial-temporal muncul pada komponen error:

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \mathbf{u}_t, \quad \mathbf{u}_t = \lambda \mathbf{W} \mathbf{u}_t + \theta \mathbf{u}_{t-1} + \boldsymbol{\varepsilon}_t \quad (2.142)$$

Definisi 2.2.25 (*Spatio-Temporal Durbin Model (STDM)*) Model yang memasukkan lag spasial dan temporal dari variabel dependen dan independen:

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_t + \phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{W} \mathbf{X}_t \boldsymbol{\delta} + \mathbf{W} \mathbf{X}_{t-1} \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_t \quad (2.143)$$

Model-model ini menyediakan kerangka yang fleksibel untuk menangkap dependensi baik dalam dimensi spasial maupun temporal. Pemilihan model bergantung pada karakteristik data dan hipotesis yang ingin diuji.

Definisi 2.2.26 (*Spatial Autoregressive Model (SAR)*) Jika hanya ada lag spasial

dari variabel dependen, maka

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.144)$$

dengan $\rho \neq 0$, $\boldsymbol{\delta} = \mathbf{0}$, $\lambda = 0$.

Definisi 2.2.27 (Spatial Error Model (SEM)) Jika dependensi spasial muncul hanya pada komponen error, maka

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{W} \mathbf{u} + \boldsymbol{\varepsilon}, \quad (2.145)$$

dengan $\rho = 0$, $\boldsymbol{\delta} = \mathbf{0}$, $\lambda \neq 0$.

Definisi 2.2.28 (Spatial Durbin Model (SDM)) Jika model memasukkan lag spasial dari variabel dependen dan independen, maka

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{W} \mathbf{X} \boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (2.146)$$

dengan $\rho \neq 0$, $\boldsymbol{\delta} \neq \mathbf{0}$, $\lambda = 0$.

Model-model ini dapat diperluas ke dimensi spasial-temporal dengan menggunakan matriks bobot yang mengakomodasi baik dimensi spasial maupun temporal.

2.2.5. Regresi dengan Heterogenitas Spasial dan Spasial Temporal

Heterogenitas spasial (*spatial heterogeneity*) terjadi ketika hubungan statistik antara variabel berbeda di berbagai lokasi geografis. Berbeda dengan dependensi spasial yang menangkap korelasi antar lokasi, heterogenitas spasial menangkap variasi dalam struktur hubungan itu sendiri (Fotheringham, Brunson, & Charlton, 2002a).

Definisi 2.2.29 (Heterogenitas Spasial) Heterogenitas spasial terjadi ketika para-

meter regresi β bervariasi menurut lokasi geografis:

$$\beta = \beta(s_i) = \beta(u_i, v_i) \quad (2.147)$$

dimana (u_i, v_i) adalah koordinat geografis lokasi i . Hal ini mengindikasikan bahwa hubungan struktural yang sama dapat memiliki kekuatan dan arah yang berbeda di lokasi yang berbeda.

Model utama untuk menangkap heterogenitas spasial adalah *Geographically Weighted Regression* (GWR) yang memungkinkan koefisien regresi bervariasi secara kontinyu di ruang geografis.

Definisi 2.2.30 (Geographically Weighted Regression (GWR)) Model GWR memungkinkan koefisien regresi bervariasi menurut lokasi:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (2.148)$$

dimana $\beta_k(u_i, v_i)$ adalah koefisien regresi untuk variabel k di lokasi (u_i, v_i) .

Estimasi parameter GWR dilakukan menggunakan *locally weighted least squares* (LWLS):

Teorema 2.2.31 (Estimator LWLS untuk GWR) Untuk setiap lokasi i , koefisien lokal diestimasi dengan:

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i \mathbf{y} \quad (2.149)$$

dimana \mathbf{W}_i adalah matriks diagonal bobot dengan elemen:

$$w_{ij} = \exp \left(-\frac{d_{ij}^2}{h^2} \right) \quad (2.150)$$

dengan d_{ij} adalah jarak antara lokasi i dan j , dan h adalah bandwidth.

Definisi 2.2.32 (Fungsi Kernel untuk Pembobotan Spasial) Beberapa fungsi ker-

nel yang umum digunakan dalam GWR:

$$\begin{aligned}
 (a) \text{ Gaussian kernel: } w_{ij} &= \exp \left(-\frac{1}{2} \left(\frac{d_{ij}}{h} \right)^2 \right) \\
 (b) \text{ Bi-square kernel: } w_{ij} &= \begin{cases} \left(1 - \left(\frac{d_{ij}}{h} \right)^2 \right)^2 & \text{jika } d_{ij} < h \\ 0 & \text{jika } d_{ij} \geq h \end{cases} \\
 (c) \text{ Tri-cube kernel: } w_{ij} &= \begin{cases} \left(1 - \left(\frac{d_{ij}}{h} \right)^3 \right)^3 & \text{jika } d_{ij} < h \\ 0 & \text{jika } d_{ij} \geq h \end{cases}
 \end{aligned}$$

Pemilihan bandwidth h sangat krusial dalam GWR dan dapat dilakukan melalui cross-validation:

Definisi 2.2.33 (Cross-Validation untuk Bandwidth) *Bandwidth optimal h^* dipilih dengan meminimalkan fungsi cross-validation:*

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \quad (2.151)$$

dimana $\hat{y}_{\neq i}(h)$ adalah prediksi untuk observasi i menggunakan model yang difit tanpa observasi i .

Untuk data panel spasial, GWR dapat diperluas menjadi *Geographically and Temporally Weighted Regression* (GTWR):

Definisi 2.2.34 (Geographically and Temporally Weighted Regression (GTWR)) *Model GTWR memungkinkan koefisien bervariasi baik secara spasial maupun temporal:*

$$y_{it} = \beta_0(u_i, v_i, t) + \sum_{k=1}^p \beta_k(u_i, v_i, t)x_{ikt} + \varepsilon_{it} \quad (2.152)$$

dengan estimator:

$$\hat{\beta}(u_i, v_i, t) = (\mathbf{X}^T \mathbf{W}_{i,t} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_{i,t} \mathbf{y} \quad (2.153)$$

dimana $\mathbf{W}_{i,t}$ menggabungkan bobot spasial dan temporal:

$$w_{ij,t} = \exp \left(-\frac{d_{ij}^2}{h_s^2} - \frac{(t_i - t_j)^2}{h_t^2} \right) \quad (2.154)$$

dengan h_s dan h_t adalah bandwidth spasial dan temporal.

Teorema 2.2.35 (Uji Heterogenitas Spasial) Untuk menguji keberadaan heterogenitas spasial, dapat digunakan uji Breusch-Pagan lokal:

$$BP = \frac{1}{2} \mathbf{f}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{f} \quad (2.155)$$

dimana $\mathbf{f} = \mathbf{X}^T \mathbf{A} \mathbf{e}^2$ dengan \mathbf{A} adalah matriks transformasi spasial dan \mathbf{e}^2 adalah vektor residual kuadrat. Di bawah hipotesis nul homoskedastisitas spasial, statistik BP mengikuti distribusi χ^2 dengan derajat bebas p .

Bukti. Misalkan model regresi linear $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ dengan hipotesis:

$$H_0 : \text{Var}(\varepsilon_i) = \sigma^2 \text{ (homoskedastisitas)} \quad (2.156)$$

$$H_1 : \text{Var}(\varepsilon_i) = \sigma^2 h(\mathbf{z}_i^T \boldsymbol{\alpha}) \text{ (heteroskedastisitas spasial)} \quad (2.157)$$

dimana \mathbf{z}_i adalah vektor karakteristik spasial lokasi i dan $h(\cdot)$ adalah fungsi varian.

Untuk fungsi $h(t) = \exp(t)$, maka $\log[\text{Var}(\varepsilon_i)] = \log(\sigma^2) + \mathbf{z}_i^T \boldsymbol{\alpha}$. Di bawah H_0 , $\boldsymbol{\alpha} = \mathbf{0}$, sehingga

$$\mathbb{E}[\varepsilon_i^2] = \sigma^2 + \mathbf{z}_i^T \boldsymbol{\alpha}. \quad (2.158)$$

Regresi auxiliary menggunakan residual kuadrat OLS:

$$e_i^2 = \gamma_0 + \mathbf{z}_i^T \boldsymbol{\gamma} + v_i \quad (2.159)$$

dengan \mathbf{z}_i dapat berupa koordinat spasial, lag spasial, atau fungsi jarak.

Definisikan $\mathbf{f} = \mathbf{X}^T \mathbf{A} \mathbf{e}^2$ dimana \mathbf{A} adalah matriks yang mengkodekan

struktur spasial. Statistik Breusch-Pagan lokal:

$$BP = \frac{1}{2} \mathbf{f}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{f} \quad (2.160)$$

Di bawah H_0 dan asumsi regularitas:

- (i) $\mathbb{E}[\mathbf{f}] = \mathbf{0}$ karena $\mathbb{E}[e_i^2] = \sigma^2$ konstan
- (ii) $\text{Var}(\mathbf{f}) = 2\sigma^4(\mathbf{X}^T \mathbf{X})$ untuk error normal
- (iii) $\mathbf{f} \xrightarrow{d} \mathcal{N}(\mathbf{0}, 2\sigma^4(\mathbf{X}^T \mathbf{X}))$ secara asimtotik

Oleh karena itu:

$$BP = \frac{1}{2} \mathbf{f}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{f} \quad (2.161)$$

$$= \frac{1}{4\sigma^4} \mathbf{f}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{f} \cdot 2\sigma^4 \quad (2.162)$$

$$\xrightarrow{d} \frac{1}{2} \mathbf{Z}^T \mathbf{Z} \quad (2.163)$$

dimana $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ adalah vektor normal standar p -dimensi.

Karena $\frac{1}{2} \mathbf{Z}^T \mathbf{Z} = \frac{1}{2} \sum_{j=1}^p Z_j^2$ dan $Z_j^2 \sim \chi^2(1)$, maka:

$$BP \xrightarrow{d} \frac{1}{2} \chi^2(p) \quad (2.164)$$

Dengan normalisasi yang tepat, diperoleh $BP \sim \chi^2(p)$ di bawah H_0 . ■

Definisi 2.2.36 (Diagnostik Model GWR) *Beberapa diagnostik penting untuk model GWR:*

(a) **Akaike Information Criterion (AIC):**

$$AIC = 2n \log(\hat{\sigma}) + n \log(2\pi) + n \left\{ \frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})} \right\} \quad (2.165)$$

(b) **Efektif derajat bebas:** $\text{tr}(\mathbf{S})$ dimana \mathbf{S} adalah hat matrix

(c) **Local R-squared:** $R_i^2 = 1 - \frac{SS_{res,i}}{SS_{tot,i}}$ untuk setiap lokasi i

Teorema 2.2.37 (Sifat Statistik Model Heterogenitas Spasial) (a) *Estimator GWR $\hat{\beta}(s_i)$ bersifat tidak bias lokal: $E[\hat{\beta}(s_i)] = \beta(s_i)$ untuk bandwidth tetap.* (b) *Varians estimator GWR umumnya lebih besar dari OLS global: $Var[\hat{\beta}(s_i)] > Var[\hat{\beta}_{OLS}]$.* (c) *Konsistensi memerlukan $h \rightarrow 0$ dan $nh^d \rightarrow \infty$ dimana d adalah dimensi spasial.* (d) *Untuk GTWR, diperlukan kondisi tambahan $h_t \rightarrow 0$ dan $nh_s^d h_t \rightarrow \infty$ untuk konsistensi spasial-temporal.*

Bukti. (a) dan (b): ketakbiasan mengikuti dari linearitas estimator GLS/OLS terprofil dan $\mathbb{E}[\varepsilon] = 0$; konsistensi berasal dari konsistensi $\hat{\rho}$, $\hat{\lambda}$ dan regularitas standar (rank penuh, batas spektral ρ, λ). (c) LWLS: bukti standar teori penaksir lokal-kernel; lihat analisis WLS lokal dengan bobot non-acak, di mana $\mathbb{E}[\hat{\beta}(s_i)] = \beta(s_i)$ dan $Var(\hat{\beta}(s_i)) = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i Var(\mathbf{y}) \mathbf{W}_i \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1}$. ■

2.3. Jaringan Saraf Tiruan

Jaringan Saraf Tiruan (*Artificial Neural Networks*, ANN) merupakan model komputasi yang meniru cara kerja otak biologis dalam mengolah informasi. Konsep neuron buatan pertama kali diperkenalkan oleh McCulloch and Pitts (1943), kemudian dikembangkan menjadi model perceptron oleh Rosenblatt (1958). Perkembangan penting terjadi ketika Rumelhart, Hinton, and Williams (1986) memperkenalkan algoritma *backpropagation*, yang memungkinkan pelatihan jaringan multilapis. Literatur formal dan matematis mengenai ANN dibahas oleh Bishop (1995), Haykin (2009), serta Goodfellow, Bengio, and Courville (2016), sedangkan fondasi teoretis kapasitas aproksimasi ANN dibuktikan oleh Cybenko (1989) dan Hornik (1991).

2.3.1. Sejarah dan Model Dasar

Definisi 2.3.1 (Neuron McCulloch–Pitts (McCulloch & Pitts, 1943)) *Neuron biner didefinisikan sebagai fungsi $y : \{0, 1\}^p \rightarrow \{0, 1\}$ dengan*

$$y = H(\mathbf{w}^\top \mathbf{x} - \theta),$$

di mana $\mathbf{x} = (x_1, \dots, x_p)^\top$ adalah masukan biner, \mathbf{w} bobot, θ ambang batas, dan $H(\cdot)$ adalah fungsi Heaviside.

Definisi 2.3.2 (Perceptron Rosenblatt (Rosenblatt, 1958)) *Model perceptron memetakan $\mathbf{x} \in \mathbb{R}^p$ ke keluaran*

$$y = \varphi(\mathbf{w}^\top \mathbf{x} + b),$$

dengan $\varphi(\cdot)$ fungsi aktivasi step, \mathbf{w} bobot, dan b bias.

Teorema 2.3.3 (Keterbatasan Perceptron (Minsky & Papert, 1969)) *Perceptron tunggal tidak mampu merepresentasikan fungsi yang tidak linear separable, misalnya XOR. Keterbatasan ini mendorong lahirnya jaringan multilapis.*

2.3.2. Arsitektur Feedforward Multilapis

Definisi 2.3.4 (Neuron Tiruan Umum (Haykin, 2009)) *Neuron tiruan dengan masukan $\mathbf{x} \in \mathbb{R}^p$ menghasilkan keluaran*

$$y = \varphi(\mathbf{w}^\top \mathbf{x} + b),$$

dengan $\varphi(\cdot)$ fungsi aktivasi non-linear yang terdiferensiasi.

Definisi 2.3.5 (Arsitektur Feedforward Multilapis (Bishop, 1995)) *Jaringan feedforward dengan L lapisan tersembunyi memiliki representasi*

$$\mathbf{a}^{(l)} = \varphi(\mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}), \quad l = 1, \dots, L,$$

dengan $\mathbf{a}^{(0)} = \mathbf{x}$ sebagai masukan, dan keluaran $\hat{y} = g(\mathbf{a}^{(L)})$.

Contoh 2.3.6 Misalkan $\mathbf{x} = (1, 2)^\top$, bobot $\mathbf{w} = (0.5, 0.25)^\top$, bias $b = 1$, dan $\varphi(z) = \tanh(z)$. Maka keluaran neuron adalah

$$y = \tanh(0.5 \cdot 1 + 0.25 \cdot 2 + 1) = \tanh(2) \approx 0.964.$$

2.3.3. Fungsi Aktivasi

Fungsi aktivasi memperkenalkan non-linearitas pada ANN. Tanpa fungsi aktivasi non-linear, jaringan saraf hanya merepresentasikan transformasi linear berlapis.

Definisi 2.3.7 (Fungsi Sigmoid) *Fungsi sigmoid didefinisikan sebagai*

$$\varphi(z) = \frac{1}{1 + e^{-z}}.$$

Contoh 2.3.8 Untuk $z = 0$, $\varphi(0) = 0.5$. Untuk $z = 2$, $\varphi(2) \approx 0.88$.

Definisi 2.3.9 (Fungsi Tanh) *Fungsi tanh didefinisikan sebagai*

$$\varphi(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}.$$

Contoh 2.3.10 Untuk $z = 0$, $\tanh(0) = 0$. Untuk $z = 1$, $\tanh(1) \approx 0.76$.

Definisi 2.3.11 (Fungsi ReLU (Nair & Hinton, 2010)) *Fungsi ReLU (Rectified Linear Unit) didefinisikan sebagai*

$$\varphi(z) = \max(0, z).$$

Contoh 2.3.12 Jika $z = -3$, maka $\varphi(z) = 0$. Jika $z = 2.5$, maka $\varphi(z) = 2.5$.

Definisi 2.3.13 (Fungsi Leaky ReLU (Maas, 2013)) *Fungsi Leaky ReLU menga-*

tasi masalah dying ReLU dengan

$$\varphi(z) = \begin{cases} z, & z \geq 0, \\ \alpha z, & z < 0, \end{cases}$$

dengan $\alpha \in (0, 1)$ kecil, misalnya 0.01.

Contoh 2.3.14 Jika $z = -3$ dan $\alpha = 0.01$, maka $\varphi(z) = -0.03$. Jika $z = 2$, maka $\varphi(z) = 2$.

Definisi 2.3.15 (Fungsi GELU (Hendrycks & Gimpel, 2016)) *Fungsi Gaussian Error Linear Unit (GELU) adalah*

$$\varphi(z) = z \cdot \Phi(z),$$

dengan $\Phi(z)$ fungsi distribusi normal standar. Aproksimasi praktis:

$$\varphi(z) \approx 0.5z \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}} (z + 0.0447z^3) \right] \right).$$

Contoh 2.3.16 Untuk $z = 1$, $\varphi(1) \approx 0.84$. Untuk $z = -1$, $\varphi(-1) \approx -0.16$.

Lema 2.3.17 (Sifat Fungsi Aktivasi) *Fungsi aktivasi $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ yang digunakan pada jaringan saraf tiruan sebaiknya memenuhi:*

- (i) φ **non-linear**,
- (ii) φ **terdiferensiasi** hampir di semua titik,
- (iii) φ **tidak konstan**,
- (iv) φ memiliki **range memadai**, tidak menyebabkan saturasi total.

Bukti.[Justifikasi Matematis]

- (i) *Non-linearitas.* Jika φ linear, misalnya $\varphi(z) = az+b$, maka sebuah multilayer

perceptron dengan L lapisan menghasilkan

$$f(\mathbf{x}) = \mathbf{W}^{(L)} \varphi(\mathbf{W}^{(L-1)} \varphi(\dots \varphi(\mathbf{W}^{(1)} \mathbf{x}))) = \mathbf{W}' \mathbf{x} + \mathbf{b}',$$

dengan $\mathbf{W}' = \prod_{l=1}^L \mathbf{W}^{(l)}$ dan \mathbf{b}' kombinasi linear dari \mathbf{b} . Jadi komposisi jaringan tetap linear. Akibatnya, jaringan tidak lebih kuat dari regresi linear biasa dan tidak dapat mengaproksimasi fungsi non-linear (misalnya XOR).

- (ii) *Diferensiabilitas*. Algoritma backpropagation memerlukan turunan $\varphi'(z)$ untuk menghitung gradien kerugian L :

$$\frac{\partial L}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} (\mathbf{a}^{(l-1)})^\top, \quad \delta^{(l)} = (\mathbf{W}^{(l+1)})^\top \delta^{(l+1)} \odot \varphi'(z^{(l)}).$$

Jika $\varphi'(z)$ tidak terdefinisi di banyak titik, maka gradien tidak dapat dihitung. Oleh karena itu φ minimal harus terdiferensiasi hampir di semua titik. Fungsi seperti ReLU sah karena nondiferensiabilitas hanya di $z = 0$, himpunan titik berukuran nol.

- (iii) *Non-konstansi*. Jika $\varphi(z) = c$ konstan, maka $\varphi'(z) = 0$ untuk semua z . Substitusi ke formula backprop menghasilkan $\delta^{(l)} = \mathbf{0}$ untuk semua l , sehingga

$$\frac{\partial L}{\partial \mathbf{W}^{(l)}} = \mathbf{0}.$$

Bobot tidak pernah diperbarui, dan proses pelatihan tidak berjalan.

- (iv) *Range yang memadai*. Jika φ cepat jenuh, misalnya sigmoid dengan limit $\varphi(z) \rightarrow 0$ untuk $z \rightarrow -\infty$ dan $\varphi(z) \rightarrow 1$ untuk $z \rightarrow +\infty$, maka untuk $|z|$ besar:

$$\varphi'(z) = \varphi(z)(1 - \varphi(z)) \approx 0.$$

Akibatnya gradien $\delta^{(l)}$ mengecil eksponensial ke arah nol (fenomena *vanishing gradient*). Hal ini membuat pembelajaran lapisan dalam gagal. Untuk menghindari itu, digunakan fungsi dengan range lebih lebar atau gradien yang tidak hilang, seperti ReLU atau Leaky ReLU.



2.3.4. Pembelajaran dengan Propagasi Mundur

Teorema 2.3.18 (Aturan Rantai pada Jaringan) *Jika $y = f^{(L)} \circ f^{(L-1)} \circ \dots \circ f^{(1)}(\mathbf{x})$, maka*

$$\frac{\partial y}{\partial \mathbf{W}^{(l)}} = \frac{\partial f^{(L)}}{\partial f^{(L-1)}} \cdots \frac{\partial f^{(l+1)}}{\partial f^{(l)}} \frac{\partial f^{(l)}}{\partial \mathbf{W}^{(l)}}.$$

Teorema 2.3.19 (Backpropagation) (Goodfellow dkk., 2016; Rumelhart dkk., 1986)) *Gradien fungsi kerugian $L(\hat{y}, y)$ terhadap bobot $\mathbf{W}^{(l)}$ adalah*

$$\frac{\partial L}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} (\mathbf{a}^{(l-1)})^\top,$$

dengan $\delta^{(l)} = (\mathbf{W}^{(l+1)})^\top \delta^{(l+1)} \odot \varphi'(z^{(l)})$.

2.3.5. Jaringan Saraf Tiruan sebagai Aproksimasi Universal

Teorema 2.3.20 (Universal Approximation Theorem (UAT)) (Cybenko, 1989))

Jika $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ adalah fungsi sigmoid kontinu tak-konstan (kontinu, $\lim_{t \rightarrow -\infty} \varphi(t) = \alpha$, $\lim_{t \rightarrow +\infty} \varphi(t) = \beta$ dengan $\alpha \neq \beta$), maka untuk setiap $f \in C([0, 1]^p)$ dan $\varepsilon > 0$ terdapat $m \in \mathbb{N}$, koefisien $c_j \in \mathbb{R}$, bobot $\mathbf{w}_j \in \mathbb{R}^p$, dan bias $b_j \in \mathbb{R}$ sedemikian sehingga

$$g(\mathbf{x}) = \sum_{j=1}^m c_j \varphi(\mathbf{w}_j^\top \mathbf{x} + b_j) \quad \text{memenuhi} \quad \sup_{\mathbf{x} \in [0, 1]^p} |f(\mathbf{x}) - g(\mathbf{x})| < \varepsilon.$$

Bukti. Tuliskan $K = [0, 1]^p$ dan $\mathcal{A} = \text{span}\{\varphi(\mathbf{w}^\top \mathbf{x} + b) : \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}\} \subset C(K)$. Kita buktikan $\overline{\mathcal{A}}^{\|\cdot\|_\infty} = C(K)$. Anggap kontradiksi: ada $f_0 \in C(K) \setminus \overline{\mathcal{A}}$. Dengan teorema Hahn–Banach (pemisahan), ada fungsional linear kontinu non-nol $L : C(K) \rightarrow \mathbb{R}$ yang menganihilasi $\overline{\mathcal{A}}$, yakni $L(g) = 0$ untuk semua $g \in \overline{\mathcal{A}}$, tetapi $L(f_0) \neq 0$. Oleh teorema Riesz–Markov, terdapat ukuran tanda terbatas ν pada K sehingga $L(h) = \int_K h(\mathbf{x}) d\nu(\mathbf{x})$ untuk semua $h \in C(K)$. Karena L menganihilasi

$\overline{\mathcal{A}}$, maka

$$\int_K \varphi(\mathbf{w}^\top \mathbf{x} + b) d\nu(\mathbf{x}) = 0 \quad \text{untuk semua } \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}. \quad (2.166)$$

Klaim kunci: (2.166) mengimplikasikan $\nu \equiv 0$. Untuk membuktikannya, gunakan sifat *discriminatory* dari fungsi sigmoid: karena φ kontinu dengan limit berbeda di $\pm\infty$, maka untuk setiap $s \in \mathbb{R}$ dan setiap $\mathbf{w} \neq \mathbf{0}$, ada parameter $a > 0$ dan $b \in \mathbb{R}$ sedemikian sehingga $\varphi(a(\mathbf{w}^\top \mathbf{x} - s) + b)$ mendekati fungsi langkah $H(\mathbf{w}^\top \mathbf{x} - s)$ secara seragam pada K ketika $a \rightarrow \infty$ (normalisasi nilai batasnya membuat aproksimasi seragam ini valid pada kompak). Dengan Dominated Convergence Theorem, dari (2.166) diperoleh

$$\int_K \mathbf{1}\{\mathbf{w}^\top \mathbf{x} \geq s\} d\nu(\mathbf{x}) = 0 \quad \text{untuk semua } \mathbf{w} \in \mathbb{R}^p, s \in \mathbb{R}.$$

Himpunan indikator setengah-ruang $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} \geq s\}$ membangkitkan σ -aljabar Borel pada \mathbb{R}^p (separasi linear titik-titik oleh hiperbidang); dengan argumen kelas monoton/ π -sistem, ini memaksa $\nu(B) = 0$ untuk semua himpunan Borel $B \subseteq K$, sehingga $\nu \equiv 0$. Konsekuensinya $L \equiv 0$, bertentangan dengan $L(f_0) \neq 0$. Kontradiksi ini menunjukkan $\overline{\mathcal{A}} = C(K)$, sehingga aproksimasi seragam diperoleh. ■

Teorema 2.3.21 (Generalisasi UAT (aktivasi terbatas kontinu tak-konstan) (Hornik, 1991)) *Biarkan $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ kontinu, terbatas, tak-konstan, dan discriminatory (yakni, jika ukuran tanda ν pada K memenuhi $\int_K \varphi(\mathbf{w}^\top \mathbf{x} + b) d\nu(\mathbf{x}) = 0$ untuk semua \mathbf{w}, b , maka $\nu \equiv 0$). Untuk setiap $f \in C([0, 1]^p)$ dan $\varepsilon > 0$ terdapat jaringan satu-lapis $g(\mathbf{x}) = \sum_{j=1}^m c_j \varphi(\mathbf{w}_j^\top \mathbf{x} + b_j)$ dengan $\sup_{\mathbf{x} \in [0, 1]^p} |f(\mathbf{x}) - g(\mathbf{x})| < \varepsilon$.*

Bukti. Pembuktian identik dengan Teorema UAT di atas; satu-satunya properti yang digunakan dari sigmoid adalah *discriminatory*. Karena asumsi teorema menyatakan φ diskriminatoris, maka langkah Hahn–Banach + Riesz–Markov + limit indikator setengah-ruang tetap berlaku, sehingga $\overline{\text{span}\{\varphi(\mathbf{w}^\top \mathbf{x} + b)\}} = C([0, 1]^p)$. Dalam Hornik (1991), kelas fungsi aktivasi yang memenuhi sifat ini mencakup berbagai

fungsi terbatas kontinu non-konstan (mis. \tanh), sehingga hasil aproksimasi universal tetap berlaku. ■

Teorema 2.3.22 (Aproksimasi universal untuk aktivasi tak-terbatas non-polinomial (mis. ReLU)) (Leshno, Lin, Pinkus, & Schocken, 1993)) *Biarkan $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ kontinu dan tidak berupa polinomial pada himpunan berdensitas penuh (non-polinomial a.e.). Maka untuk setiap kompak $K \subset \mathbb{R}^p$ dan setiap $f \in C(K)$, linear span*

$$\mathcal{H}_\sigma = \text{span}\{\sigma(\mathbf{w}^\top \mathbf{x} + b) : \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}\}$$

padat di $C(K)$ (aproksimasi seragam). Secara khusus, $\sigma(z) = \max\{0, z\}$ (ReLU) dan aktivasi tak-terbatas non-polinomial lain (mis. GELU) memenuhi sifat aproksimasi universal.

Bukti. Garis besar argumen mengikuti Leshno dkk. (1993). Arah “jika” (sufficiency): karena σ non-polinomial, himpunan ridge functions $\{\sigma(\mathbf{w}^\top \cdot + b)\}$ tidak terkandung dalam subruang invarian berhingga dimensi yang ditentukan oleh polinomial; menggunakan hasil kepadatan untuk *ridge functions* non-polinomial (melalui teori momen/Fourier), diperoleh kepadatan $\overline{\mathcal{H}_\sigma} = C(K)$ dalam norma supremum. Arah “hanya jika” (necessity): bila σ polinomial derajat d , maka untuk setiap \mathbf{w}, b , komposisi $\sigma(\mathbf{w}^\top \mathbf{x} + b)$ adalah polinomial derajat $\leq d$ dalam variabel \mathbf{x} ; span semua komposisi demikian berada dalam penutupan subruang fungsi polinomial derajat $\leq d$, yang tidak padat di $C(K)$ ketika $p \geq 1$ (teorema Stone–Weierstrass memerlukan derajat tak-terbatas untuk kepadatan). Karena itu, σ harus non-polinomial agar kepadatan berlaku. Contoh ReLU memenuhi hipotesis; maka aproksimasi universal mengikuti. ■

2.4. Jaringan Saraf Graf

Jaringan Saraf Graf (*Graph Neural Networks*, GNN) merupakan kelas arsitektur pembelajaran representasi yang dirancang khusus untuk data terstruktur dalam bentuk graf. Berbeda dengan jaringan saraf tiruan (JST) yang bekerja pada vektor atau matriks, GNN memanfaatkan informasi topologi melalui graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

dengan simpul \mathcal{V} dan sisi \mathcal{E} . Penggunaan GNN telah berkembang pesat, baik dalam teori maupun aplikasi, meliputi sistem rekomendasi, data spasial-temporal, analisis jaringan sosial, dan model ekonometrika spasial (Bronstein, Bruna, LeCun, Szlam, & Vandergheynst, 2017; Scarselli, Gori, Tsoi, Hagenbuchner, & Monfardini, 2009; Wu dkk., 2021).

2.4.1. Dasar Graf dan Laplacian

Definisi 2.4.1 (Graf dan Matriks Ketetanggaan) *Graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan $|\mathcal{V}| = n$ direpresentasikan oleh:*

$$A_{ij} = \begin{cases} 1 & \text{jika } (i, j) \in \mathcal{E}, \\ 0 & \text{lainnya,} \end{cases}$$

disebut matriks ketetanggaan. Matriks derajat didefinisikan sebagai $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$, dengan $d_i = \sum_j A_{ij}$.

Definisi 2.4.2 (Laplacian Graf (Chung, 1997)) *Laplacian graf adalah $\mathbf{L} = \mathbf{D} - \mathbf{A}$. Normalisasi simetris diberikan oleh*

$$\mathbf{L}_{\text{sym}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}.$$

Teorema 2.4.3 (Spektral Teorema Laplacian (Chung, 1997)) *Matriks Laplacian \mathbf{L} adalah simetris dan positif semidefinit, sehingga dapat didekomposisi sebagai*

$$\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top,$$

dengan \mathbf{U} ortogonal dan $\mathbf{\Lambda}$ diagonal berisi nilai eigen $\lambda_1, \dots, \lambda_n \geq 0$. Selalu berlaku $\lambda_1 = 0$ dengan vektor eigen $\mathbf{1}$.

Contoh 2.4.4 Untuk graf rantai tiga simpul,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{D} = \text{diag}(1, 2, 1), \quad \mathbf{L} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}.$$

Nilai eigen $\lambda = \{0, 1, 3\}$, dengan eigenvektor ortogonal.

2.4.2. Konvolusi pada Graf

Definisi 2.4.5 (Transformasi Fourier pada Graf (Shuman, Narang, Frossard, Ortega, & Vandergheynst, 2013)) *Jika $\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^\top$, maka transformasi Fourier graf untuk sinyal $f \in \mathbb{R}^n$ adalah $\hat{f} = \mathbf{U}^\top f$. Konvolusi graf didefinisikan melalui*

$$f *_{\mathcal{G}} g = \mathbf{U}((\mathbf{U}^\top f) \odot (\mathbf{U}^\top g)).$$

Teorema 2.4.6 (GCN sebagai Aproksimasi Polinomial (Kipf & Welling, 2017)) *Filter spektral $\varphi(\Lambda)$ dapat diaproksimasi dengan polinomial Chebyshev orde 1. Dari sini diturunkan propagasi Graph Convolutional Network (GCN):*

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}\right),$$

dengan $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$, $\tilde{\mathbf{D}}$ matriks derajatnya, dan $\mathbf{W}^{(l)}$ bobot parameter.

2.4.3. Kerangka Message Passing

Teorema 2.4.7 (Message Passing Neural Networks (MPNN) (Gilmer, Schoenholz, Riley, Vinyals, & Dahl, 2017)) *Setiap GNN berbasis pesan dapat ditulis dalam bentuk*

$$\mathbf{m}_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} M^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{e}_{ij}), \quad (2.167)$$

$$\mathbf{h}_i^{(l+1)} = U^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{m}_i^{(l+1)}), \quad (2.168)$$

dengan $M^{(l)}$ fungsi pesan dan $U^{(l)}$ fungsi pembaruan.

- Lema 2.4.8 (GNN Umum sebagai Kasus Khusus MPNN)** • Pada GCN, $M^{(l)}(\mathbf{h}_i, \mathbf{h}_j) = \frac{1}{\sqrt{d_i d_j}} \mathbf{W}^{(l)} \mathbf{h}_j$, dan $U^{(l)}$ berupa penjumlahan + aktivasi.
- Pada GAT, $M^{(l)}$ ditimbang dengan koefisien perhatian α_{ij} .
 - Pada GraphSAGE, $M^{(l)}$ diganti dengan operator agregasi AGGREGATE (misalnya mean, max-pooling, atau LSTM).

2.4.4. Arsitektur Umum dalam GNN

Definisi 2.4.9 (Graph Convolutional Networks (GCN) (Kipf & Welling, 2017))

GCN mendefinisikan operasi propagasi lapisan ke- l sebagai

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}\right),$$

dengan:

- $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$ matriks ketetanggaan dengan self-loop,
- $\tilde{\mathbf{D}} = \text{diag}(\sum_j \tilde{A}_{ij})$ matriks derajat,
- $\mathbf{H}^{(l)} \in \mathbb{R}^{n \times d_l}$ representasi simpul pada lapisan l ,
- $\mathbf{W}^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ bobot terlatih,
- $\sigma(\cdot)$ fungsi aktivasi non-linear.

Contoh 2.4.10 Ambil graf dengan $n = 3$ simpul dan sisi $\{(1, 2), (2, 3)\}$:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Matriks derajat $\tilde{\mathbf{D}} = \text{diag}(2, 3, 2)$. Misalkan fitur awal $\mathbf{H}^{(0)} = \mathbf{I}_3$ (fitur identitas),

dan bobot $\mathbf{W}^{(0)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Setelah propagasi satu lapisan:

$$\mathbf{H}^{(1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(0)} \mathbf{W}^{(0)}\right).$$

Hasilnya tiap simpul memperoleh representasi yang merupakan kombinasi rata-rata tetangga terdekat (dengan normalisasi simetris).

Definisi 2.4.11 (Graph Attention Networks (GAT)) (Veličković dkk., 2018)) *GAT mengganti normalisasi derajat dengan mekanisme perhatian. Koefisien perhatian untuk sisi (i, j) adalah*

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_k]))},$$

dengan:

- $\mathbf{h}_i \in \mathbb{R}^d$ vektor fitur simpul i ,
- \mathbf{W} bobot transformasi linear,
- $\mathbf{a} \in \mathbb{R}^{2d'}$ vektor bobot perhatian,
- \parallel operator konkatenasi.

Pembaruan fitur simpul adalah

$$\mathbf{h}'_i = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{W}\mathbf{h}_j\right).$$

Contoh 2.4.12 Ambil graf segitiga $\mathcal{V} = \{1, 2, 3\}$ dengan $\mathcal{E} = \{(1, 2), (2, 3), (3, 1)\}$. Misalkan fitur awal $\mathbf{h}_1 = (1, 0)$, $\mathbf{h}_2 = (0, 1)$, $\mathbf{h}_3 = (1, 1)$, bobot $\mathbf{W} = \mathbf{I}_2$, dan $\mathbf{a} = (1, 1, 1, 1)^\top$. Untuk simpul 1, skor perhatian terhadap simpul 2 dan 3 dihitung,

lalu dinormalisasi dengan softmax menjadi α_{12}, α_{13} . Fitur baru simpul 1:

$$\mathbf{h}'_1 = \sigma(\alpha_{12}\mathbf{h}_2 + \alpha_{13}\mathbf{h}_3).$$

Dengan demikian, GAT memungkinkan simpul memberi bobot berbeda pada tetangga yang berbeda.

Definisi 2.4.13 (GraphSAGE (Hamilton, Ying, & Leskovec, 2017)) *GraphSAGE mendefinisikan pembaruan simpul i sebagai*

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{W}^{(l)} \cdot \text{AGGREGATE}\left(\{\mathbf{h}_i^{(l)}\} \cup \{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\}\right)\right),$$

dengan AGGREGATE fungsi agregasi non-linear yang bisa berupa mean, max-pooling, atau LSTM.

Contoh 2.4.14 Ambil graf rantai 3 simpul. Misalkan $\mathbf{h}_1 = (1, 0)$, $\mathbf{h}_2 = (0, 1)$, $\mathbf{h}_3 = (1, 1)$, dan fungsi agregasi mean. Untuk simpul 2 dengan tetangga $\{1, 3\}$:

$$\text{AGGREGATE}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \frac{1}{3}((1, 0) + (0, 1) + (1, 1)) = \frac{1}{3}(2, 2) = (0.67, 0.67).$$

Setelah transformasi linear dengan $\mathbf{W} = \mathbf{I}_2$ dan aktivasi identitas, diperoleh $\mathbf{h}'_2 = (0.67, 0.67)$.

2.4.5. Kekuatan Representasi GNN

Definisi 2.4.15 (Isomorfisme Graf) Dua graf $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ dan $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ disebut isomorfik jika terdapat bijeksi $\pi : \mathcal{V}_1 \rightarrow \mathcal{V}_2$ sehingga

$$(i, j) \in \mathcal{E}_1 \iff (\pi(i), \pi(j)) \in \mathcal{E}_2.$$

Sebuah fungsi f dikatakan invarian terhadap isomorfisme graf jika $f(\mathcal{G}_1) = f(\mathcal{G}_2)$ untuk setiap graf isomorfik $\mathcal{G}_1 \cong \mathcal{G}_2$.

Definisi 2.4.16 (Tes Weisfeiler–Lehman (WL) 1-dimensi) Algoritma WL satu di-

mensi (color refinement) memperbarui label simpul $h_i^{(t)}$ secara iteratif:

$$h_i^{(t+1)} = \text{HASH}\left(h_i^{(t)}, \{h_j^{(t)} : j \in \mathcal{N}(i)\}\right),$$

dimulai dari label awal $h_i^{(0)}$. Dua graf dikatakan dapat dibedakan oleh tes WL jika multiset label akhirnya berbeda.

Teorema 2.4.17 (Keterbatasan GNN Berbasis Agregasi (Xu, Hu, Leskovec, & Jegelka, 2019)) *GNN berbasis agregasi dengan update umum*

$$h_i^{(l+1)} = \varphi\left(W^{(l)} \cdot \text{AGGREGATE}(\{h_i^{(l)}\} \cup \{h_j^{(l)} : j \in \mathcal{N}(i)\})\right)$$

tidak lebih kuat daripada tes WL 1-dimensi dalam membedakan graf yang tidak isomorfik.

Bukti.[Sketsa Bukti] Algoritma GNN berbasis agregasi dapat dipandang sebagai proses label refinement pada simpul: setiap langkah menggabungkan label node dengan multiset label tetangganya. Karena fungsi agregasi standar (mean, sum, max) tidak dapat membedakan multiset berbeda yang memiliki rata-rata atau maksimum sama, maka kapasitas diskriminasi GNN terbatas. Tes WL menggunakan fungsi hash injektif pada multiset, sehingga lebih kuat. Oleh karena itu, GNN berbasis agregasi \preceq WL dalam hal kekuatan representasi. ■

Teorema 2.4.18 (Hubungan dengan WL Test (Xu dkk., 2019)) *Jika fungsi agregasi AGGREGATE bersifat injektif terhadap multiset, maka GNN memiliki kekuatan representasi setara dengan tes WL satu dimensi.*

Bukti.[Sketsa Bukti] Anggap AGGREGATE injektif pada multiset label. Maka setiap langkah update GNN

$$h_i^{(l+1)} = \varphi\left(W^{(l)} \cdot \text{AGGREGATE}\left(\{h_i^{(l)}\} \cup \{h_j^{(l)} : j \in \mathcal{N}(i)\}\right)\right)$$

memberikan label unik yang ekuivalen dengan proses *hashing* dalam WL. Induksi pada l menunjukkan bahwa dua simpul yang dipisahkan WL juga dipisahkan GNN.

Sebaliknya, jika WL tidak dapat membedakan dua graf, maka GNN berbasis agregasi injektif juga gagal. Jadi kapasitas representasi $GNN = WL$. ■

Teorema 2.4.19 (Universal Approximation untuk Fungsi Graf (Xu dkk., 2019))

Misalkan $f : \mathcal{G} \rightarrow \mathbb{R}$ fungsi kontinu yang invarian terhadap isomorfisme graf. Maka terdapat GNN berbasis agregasi injektif yang dapat mengaproksimasi f dengan ketelitian arbitrer:

$$\forall \varepsilon > 0, \exists GNN g : \mathcal{G} \rightarrow \mathbb{R} \text{ sehingga } |f(\mathcal{G}) - g(\mathcal{G})| < \varepsilon.$$

2.5. Pembelajaran Representasi

Pembelajaran representasi (*representation learning*) adalah bidang dalam pembelajaran mesin yang berfokus pada pemetaan data mentah ke ruang fitur laten yang lebih sesuai untuk tugas prediksi maupun analisis. Alih-alih merancang fitur secara manual, pembelajaran representasi berupaya mempelajarinya secara otomatis dari data. Representasi ini dipandang sebagai transformasi $\phi : \mathcal{X} \rightarrow \mathcal{Z}$ dari ruang input \mathcal{X} ke ruang laten \mathcal{Z} , sehingga model prediksi $h : \mathcal{X} \rightarrow \mathcal{Y}$ dapat dituliskan sebagai $h = f \circ \phi$, dengan $f : \mathcal{Z} \rightarrow \mathcal{Y}$ relatif sederhana untuk dipelajari (Bengio, Courville, & Vincent, 2013; Goodfellow dkk., 2016).

Tujuan representasi adalah menghasilkan transformasi ϕ yang memiliki sifat-sifat:

1. *invarian* terhadap variasi tak relevan (misalnya translasi citra),
2. *diskriminatif* untuk membedakan kelas atau pola,
3. *kompak* dalam menyimpan informasi penting.

Dalam konteks data spasial-temporal, invariansi sering terkait dengan pergeseran lokasi atau waktu, sedangkan diskriminasi terkait dengan variasi antar lokasi maupun periode.

2.5.1. Konsep Dasar Representasi

Definisi 2.5.1 (Representasi Laten) Diberikan ruang input \mathcal{X} , ruang output \mathcal{Y} , dan ruang laten \mathcal{Z} , suatu representasi adalah pemetaan

$$\phi : \mathcal{X} \rightarrow \mathcal{Z},$$

sehingga model prediksi $h : \mathcal{X} \rightarrow \mathcal{Y}$ dapat ditulis sebagai $h = f \circ \phi$ dengan $f : \mathcal{Z} \rightarrow \mathcal{Y}$.

Contoh 2.5.2 Dalam klasifikasi citra, $x \in \mathcal{X}$ berupa piksel mentah dimensi 32×32 . Representasi ϕ berupa hasil konvolusi CNN, misalnya $\phi(x) \in \mathbb{R}^{128}$, yang berisi fitur tepi, tekstur, dan pola. Keluaran $\phi(x)$ ini kemudian dipakai oleh f berupa lapisan linear + softmax.

Lema 2.5.3 Jika ϕ adalah transformasi invertibel dan Lipschitz, maka mempelajari $f \circ \phi$ ekuivalen dengan mempelajari h langsung. Keuntungan diperoleh jika ϕ mendekati faktor penyebab data (underlying generative factors).

2.5.2. Representasi Klasik

Definisi 2.5.4 (Principal Component Analysis (PCA)) (Jolliffe, 2002)) Diberikan data $\mathbf{X} \in \mathbb{R}^{n \times p}$, PCA mencari proyeksi $\mathbf{Z} = \mathbf{XW}$ dengan $\mathbf{W} \in \mathbb{R}^{p \times k}$ yang memaksimalkan variansi:

$$\mathbf{W} = \arg \max_{\mathbf{W}^\top \mathbf{W} = \mathbf{I}_k} \text{tr}(\mathbf{W}^\top \mathbf{X}^\top \mathbf{X} \mathbf{W}).$$

Contoh 2.5.5 Untuk matriks kovarians $\mathbf{S} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, nilai eigen adalah $\{3, 1\}$. Komponen utama pertama adalah arah $(1, 1)^\top$, yang menjelaskan 75% variansi data.

Teorema 2.5.6 (Pythagorean Theorem pada PCA) Jika $\mathbf{X} = \mathbf{ZW}^\top + \mathbf{E}$, dengan

\mathbf{Z} proyeksi PCA dan \mathbf{E} residual, maka

$$\|\mathbf{X}\|_F^2 = \|\mathbf{Z}\mathbf{W}^\top\|_F^2 + \|\mathbf{E}\|_F^2.$$

Definisi 2.5.7 (Autoencoder Linear (Hinton & Zemel, 1994)) *Autoencoder linear meminimalkan loss rekonstruksi*

$$L(\mathbf{W}, \mathbf{V}) = \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{V}^\top\|_F^2,$$

dengan $\mathbf{W}, \mathbf{V} \in \mathbb{R}^{p \times k}$. Solusi optimal identik dengan PCA.

Contoh 2.5.8 Untuk $\mathbf{X} \in \mathbb{R}^{100 \times 50}$, autoencoder linear dengan $k = 10$ menghasilkan representasi laten $\mathbf{Z} \in \mathbb{R}^{100 \times 10}$ yang merekonstruksi \mathbf{X} dengan error minimum sama seperti PCA.

Selain PCA, terdapat representasi klasik non-linear:

- Isomap (Tenenbaum, de Silva, & Langford, 2000) membangun jarak geodesik di manifold.
- Locally Linear Embedding (LLE) (Roweis & Saul, 2000) memelihara hubungan linear lokal.

2.5.3. Representasi dalam Pembelajaran Mendalam

- CNN. Operasi konvolusi formalnya:

$$(\phi(x))_{ij} = \sum_{u,v} K_{uv} x_{i+u,j+v},$$

yang merepresentasikan fitur spasial.

- RNN. Representasi sekuensial diperbarui dengan

$$h_t = \varphi(Wx_t + Uh_{t-1} + b),$$

sehingga h_t merepresentasikan konteks hingga waktu t .

- **Transformer.** Representasi diperoleh dengan *self-attention*:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right) V,$$

yang menekankan ketergantungan jangka panjang.

- **GNN.** Representasi graf berbasis message passing:

$$h_i^{(l+1)} = \varphi\left(W^{(l)} \cdot \text{AGG}\{h_j^{(l)} : j \in \mathcal{N}(i)\}\right).$$

2.5.4. Landasan Teoretis Representasi

Definisi 2.5.9 (Invariansi) Representasi ϕ disebut *invariant terhadap grup transformasi* G jika

$$\phi(g \cdot x) = \phi(x), \quad \forall g \in G, x \in \mathcal{X}.$$

Contoh 2.5.10 Pada klasifikasi digit, translasi $g \cdot x$ (geser piksel) seharusnya tidak mengubah representasi digit tersebut.

Definisi 2.5.11 (Equivariansi) Representasi ϕ *equivariant* jika terdapat aksi T_g pada \mathcal{Z} sehingga

$$\phi(g \cdot x) = T_g \phi(x).$$

Contoh 2.5.12 Pada GNN, permutasi simpul graf menghasilkan permutasi representasi node yang ekuivalen.

Teorema 2.5.13 (Hipotesis Manifold (Fefferman, Mitter, & Narayanan, 2016)) *Jika data $x \in \mathbb{R}^D$ terletak pada manifold \mathcal{M} berdimensi rendah, maka terdapat representasi kontinu $\phi : \mathbb{R}^D \rightarrow \mathbb{R}^d$ dengan $d \ll D$ yang memparametrisasi koordinat intrinsik manifold.*

Teorema 2.5.14 (Universal Approximation untuk Representasi) *Jika ϕ_θ adalah jaringan saraf feedforward dengan fungsi aktivasi non-linear non-konstan, maka*

untuk setiap pemetaan kontinu $\phi^* : [0, 1]^p \rightarrow \mathbb{R}^d$ dan $\varepsilon > 0$, terdapat parameter θ sehingga

$$\sup_{x \in [0, 1]^p} \|\phi_\theta(x) - \phi^*(x)\| < \varepsilon.$$

2.5.5. Isu dan Permasalahan dalam Pembelajaran Representasi

1. **Generalisasi.** Representasi dapat overfit. Misalnya, autoencoder dengan dimensi laten sama dengan dimensi input akan mempelajari identitas, bukan struktur laten (Zhang, Bengio, Hardt, Recht, & Vinyals, 2017).
2. **Interpretabilitas.** Representasi sering sukar dimaknai. Contoh: node embedding pada GNN tidak selalu punya interpretasi geografis (Li, Chen, Wasserman, & Shi, 2020).
3. **Stabilitas.** Representasi bisa sensitif terhadap noise. Contoh: PCA sangat dipengaruhi oleh outlier.
4. **Non-stationaritas.** Pada data spasial-temporal, distribusi representasi dapat berubah menurut lokasi dan waktu, menuntut model adaptif.

Remark II.1 Meskipun teori menunjukkan kapasitas aproksimasi universal, isu-isu praktis di atas membuat pemilihan arsitektur dan regularisasi menjadi kunci sukses pembelajaran representasi.

2.6. Pembelajaran Semi-Terawasi

Pembelajaran semi-terawasi (semi-supervised learning, SSL) memanfaatkan data berlabel dan tidak berlabel sekaligus untuk meningkatkan performa pembelajaran dibandingkan pendekatan terawasi murni (*supervised*) ketika label langka atau mahal diperoleh (Chapelle, Schölkopf, & Zien, 2006; X. Zhu & Goldberg, 2009). Secara umum, SSL berusaha mengekstrak struktur pada distribusi tak-berlabel untuk memperkaya fungsi hipotesis, dengan asumsi tertentu mengenai kelanjutan, kluster, atau manifold yang mendasari data.

2.6.1. Konsep Dasar Pembelajaran Semi-Terawasi

Definisi 2.6.1 (Kerangka Data SSL (Chapelle dkk., 2006; X. Zhu & Goldberg, 2009)) Diberikan himpunan data berlabel dan tidak berlabel

$$\mathcal{D}_\ell = \{(\mathbf{x}_i, y_i)\}_{i=1}^\ell, \quad \mathcal{D}_u = \{\mathbf{x}_j\}_{j=\ell+1}^{\ell+u},$$

dengan $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$ dan $y \in \mathcal{Y}$ (diskret untuk klasifikasi, kontinu untuk regresi). Tujuan SSL adalah mempelajari hipotesis $h : \mathcal{X} \rightarrow \mathcal{Y}$ yang meminimalkan risiko terawasi sekaligus memenuhi regularisasi yang diekstrak dari \mathcal{D}_u . Formulasi umum:

$$\min_{h \in \mathcal{H}} \underbrace{\frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}_\ell(h(\mathbf{x}_i), y_i)}_{\text{kerugian terawasi}} + \lambda \underbrace{\mathcal{R}_u(h; \mathcal{D}_u)}_{\text{regularisasi tak berlabel}},$$

dengan $\lambda > 0$ parameter regularisasi dan \mathcal{R}_u menyandikan struktur tak berlabel (misal kehalusan pada graf).

Contoh 2.6.2 Untuk klasifikasi biner, pilih \mathcal{L}_ℓ sebagai kerugian logistik dan $\mathcal{R}_u(h) = \sum_{i,j} w_{ij} \|h(\mathbf{x}_i) - h(\mathbf{x}_j)\|^2$ dengan w_{ij} kemiripan berbasis kernel Gaussian pada seluruh $\mathcal{D}_\ell \cup \mathcal{D}_u$. Regularisasi ini mendorong prediksi serupa pada titik-titik yang saling mirip.

2.6.2. Kerangka Transduktif dan Induktif dalam Pembelajaran Semi-Terawasi

Definisi 2.6.3 (Inferensi Transduktif vs Induktif (Chapelle dkk., 2006; X. Zhu & Goldberg, 2009)) SSL transduktif mencari label untuk titik tak berlabel yang sudah tersedia pada saat pelatihan, tanpa menjamin generalisasi ke titik baru; formalnya, keluaran adalah $\{h(\mathbf{x}_j)\}_{j=\ell+1}^{\ell+u}$. SSL induktif mempelajari fungsi h yang dapat diaplikasikan pada data baru di luar himpunan pelatihan, sehingga h terdefinisi di seluruh \mathcal{X} .

Contoh 2.6.4 Propagasi label pada graf (Label Propagation) bersifat transduktif karena hanya memproduksi label untuk simpul tak berlabel yang ada pada graf pelatihan. Sebaliknya, model VAE semi-terawasi menghasilkan pemetaan $x \mapsto y$

yang dapat diaplikasikan pada sampel baru, sehingga bersifat induktif (Kingma, Mohamed, Rezende, & Welling, 2014).

2.6.3. Asumsi-Asumsi Fundamental pada Pembelajaran Semi-Terawasi

Definisi 2.6.5 (Asumsi Kelanjutan (*continuity*) (Chapelle dkk., 2006)) *Hipotesis h diasumsikan halus pada daerah berketataan tinggi; secara informal, titik-titik \mathbf{x} yang berdekatan dalam metrik yang relevan cenderung memiliki label yang sama atau mirip.*

Contoh 2.6.6 Jika kemiripan $w_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$ besar untuk pasangan tetangga, regularisasi graf $\sum_{i,j} w_{ij} \|h(\mathbf{x}_i) - h(\mathbf{x}_j)\|^2$ mengekspresikan asumsi kelanjutan.

Definisi 2.6.7 (Asumsi Klaster (*cluster assumption*) (Chapelle dkk., 2006)) *Data dari kelas yang sama membentuk klaster-klaster yang terpisah dalam ruang fitur, dan batas keputusan ideal melewati daerah berketataan rendah.*

Contoh 2.6.8 Jika dua klaster Gaussian berjarak jauh dan label diketahui hanya untuk beberapa titik, maka memaksa h konstan di masing-masing klaster (dan berubah tajam di area ketataan rendah) meningkatkan akurasi.

Definisi 2.6.9 (Asumsi Manifold (Belkin & Niyogi, 2004; Belkin, Niyogi, & Sindhwani, 2006)) *Data berbaring pada manifold berdimensi rendah $\mathcal{M} \subset \mathbb{R}^p$. Hipotesis yang baik harus halus sepanjang \mathcal{M} ; regularisasi yang sesuai adalah penalti energi Laplace–Beltrami yang didiskritisasi melalui Laplasian graf.*

Contoh 2.6.10 Poin-poin pada permukaan nonlinier (misal *Swiss roll*) dapat direpresentasikan pada graf k -tetangga; term kehalusan $h^\top Lh$ (dengan L Laplasian graf) menekan variasi h sepanjang manifold dan menghindari distorsi di luar struktur intrinsik.

2.6.4. Metode-Metode Klasik Pembelajaran Semi-Terawasi

Definisi 2.6.11 (Pelatihan Diri Sendiri (Self-Training)) *Mulai dari model terawasi awal $h^{(0)}$, iterasi: beri pseudo-label $\tilde{y}_j = \arg \max_y p(y \mid \mathbf{x}_j; h^{(t)})$ pada subset $\mathcal{S} \subset \mathcal{D}_u$ dengan kepercayaan tinggi, perbarui model $h^{(t+1)}$ dengan menambahkan $\{(\mathbf{x}_j, \tilde{y}_j)\}_{\mathbf{x}_j \in \mathcal{S}}$ ke data berlabel.*

Contoh 2.6.12 Pada klasifikasi sentimen, model awal yang dilatih pada 500 ulasan berlabel dapat memberikan pseudo-label untuk 5,000 ulasan tak berlabel paling yakin, lalu dilatih ulang untuk memperbaiki batas keputusan.

Definisi 2.6.13 (Co-Training (Blum & Mitchell, 1998)) *Diasumsikan dua pandangan (views) kondisional independen $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$. Latih dua pengklasifikasi $h^{(1)}, h^{(2)}$ pada view masing-masing. Pada setiap iterasi, masing-masing model melabeli sampel tak berlabel paling yakin untuk ditambahkan ke set latih model lain.*

Contoh 2.6.14 Klasifikasi halaman web memiliki view konten (bag-of-words) dan view tautan (anchor/graph). Co-training memanfaatkan keduanya saling melengkapi.

Definisi 2.6.15 (Propagasi Label Berbasis Graf (Zhou, Bousquet, Lal, Weston, & Schölkopf, 2004; X. Zhu, 2005)) *Bangun graf berbobot atas $\mathcal{D}_\ell \cup \mathcal{D}_u$ dengan matriks bobot $\mathbf{W} = [w_{ij}]$ dan normalisasi simetris $\mathbf{S} = \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$, $\mathbf{D} = \text{diag}(\sum_j w_{ij})$. Definisikan matriks label awal $\mathbf{Y} \in \mathbb{R}^{n \times C}$, baris- i adalah vektor satu-panas untuk titik berlabel dan nol untuk tak berlabel. Iterasi propagasi*

$$\mathbf{F}^{(t+1)} = \alpha \mathbf{S} \mathbf{F}^{(t)} + (1 - \alpha) \mathbf{Y}, \quad \alpha \in (0, 1),$$

konvergen ke solusi tertutup

$$\mathbf{F}^* = (1 - \alpha) (\mathbf{I} - \alpha \mathbf{S})^{-1} \mathbf{Y}.$$

Prediksi dilakukan dengan $\arg \max$ pada setiap baris \mathbf{F}^ .*

Contoh 2.6.16 Pada graf k -tetangga untuk data dua-klas, nilai $\alpha = 0.99$ dan kernel Gaussian σ sedang menghasilkan propagasi halus yang menghormati geometri data, memberi label konsisten dalam klaster.

Teorema 2.6.17 (Kekonvergenan Propagasi Label (Zhou dkk., 2004)) *Jika spektrum \mathbf{S} berada di $[-1, 1]$ dan $\alpha \in (0, 1)$, maka deret Neumann $(\mathbf{I} - \alpha\mathbf{S})^{-1} = \sum_{t=0}^{\infty} (\alpha\mathbf{S})^t$ konvergen sehingga iterasi di atas konvergen ke \mathbf{F}^* unik.*

Definisi 2.6.18 (Fungsi Harmonik pada Graf (X. Zhu, 2005)) *Partisi Laplasian graf $\mathbf{L} = \mathbf{D} - \mathbf{W}$ menjadi blok berlabel (L) dan tak berlabel (U):*

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{LL} & \mathbf{L}_{LU} \\ \mathbf{L}_{UL} & \mathbf{L}_{UU} \end{pmatrix}.$$

Carilah fungsi $f : \{1, \dots, n\} \rightarrow \mathbb{R}^C$ yang meminimalkan energi graf $E(f) = \frac{1}{2} \sum_{i,j} w_{ij} \|f_i - f_j\|^2$ dengan syarat batas $f_L = \mathbf{Y}_L$. Solusi unik pada simpul tak berlabel adalah

$$f_U^* = -\mathbf{L}_{UU}^{-1} \mathbf{L}_{UL} \mathbf{Y}_L.$$

Contoh 2.6.19 Pada graf garis dengan beberapa label jangkar di ujung-ujung, solusi harmonik menginterpolasi label secara halus di sepanjang graf, setara dengan solusi Dirichlet diskret.

2.6.5. Pendekatan Pembelajaran Semi-Terawasi pada Pembelajaran Mendalam

Definisi 2.6.20 (Regularisasi Konsistensi (Laine & Aila, 2017; Miyato, Maeda, Koyama, & Ishii, 2018; Tarvainen & Valpola, 2017)) *Hipotesis h_θ diharapkan konsisten terhadap augmentasi/noise \mathcal{T} pada data tak berlabel. Tambahkan penalti*

$$\mathcal{R}_u(\theta) = \frac{1}{u} \sum_{j=\ell+1}^{\ell+u} \mathbb{E}_{\tau \sim \mathcal{T}} [\|h_\theta(\tau(\mathbf{x}_j)) - \tilde{h}(\mathbf{x}_j)\|^2],$$

dengan target \tilde{h} adalah model sama (Φ -Model), ansambel temporal (Temporal Ensembling), atau guru EMA (Mean Teacher). VAT memilih gangguan adversarial kecil yang memaksimalkan perubahan prediksi (Miyato dkk., 2018).

Contoh 2.6.21 Pada klasifikasi gambar, τ berupa augmentasi kuat (flip, crop, jitter). Mean Teacher menggunakan \tilde{h} sebagai model guru dengan parameter rata-rata eksponensial $\theta_{\text{guru}} \leftarrow \alpha \theta_{\text{guru}} + (1 - \alpha)\theta$.

Definisi 2.6.22 (VAE Semi-Terawasi (Kingma dkk., 2014)) *Model generatif bersama $p_{\theta}(\mathbf{x}, \mathbf{z}, y) = p(y)p(\mathbf{z})p_{\theta}(\mathbf{x} \mid \mathbf{z}, y)$ dengan pengaproksimasi variational $q_{\phi}(\mathbf{z}, y \mid \mathbf{x}) = q_{\phi}(y \mid \mathbf{x})q_{\phi}(\mathbf{z} \mid \mathbf{x}, y)$. Tujuan memaksimalkan ELBO berlabel dan tak berlabel:*

$$\mathcal{L}_{\ell} = \mathbb{E}_{q_{\phi}(\mathbf{z} \mid \mathbf{x}, y)}[\log p_{\theta}(\mathbf{x} \mid \mathbf{z}, y)] - \text{KL}(q_{\phi}(\mathbf{z} \mid \mathbf{x}, y) \parallel p(\mathbf{z})),$$

$$\mathcal{L}_u = \sum_y q_{\phi}(y \mid \mathbf{x}) \left(\mathbb{E}_{q_{\phi}(\mathbf{z} \mid \mathbf{x}, y)}[\log p_{\theta}(\mathbf{x} \mid \mathbf{z}, y)] - \text{KL}(q_{\phi}(\mathbf{z} \mid \mathbf{x}, y) \parallel p(\mathbf{z})) \right) + \mathcal{H}(q_{\phi}(y \mid \mathbf{x})).$$

Contoh 2.6.23 Untuk MNIST, y adalah digit 0–9. VAE semi-terawasi memanfaatkan data gambar tak berlabel untuk menata ruang laten sehingga klasifikasi memerlukan label jauh lebih sedikit.

Definisi 2.6.24 (Klasifikasi Semi-Terawasi Berbasis GCN (Kipf & Welling, 2017)) *Berikan graf citra/teks dengan fitur simpul \mathbf{X} dan ketetanggaan $\tilde{\mathbf{A}}$ bertambah self-loop. GCN mempropagasi*

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}),$$

dengan kerugian hanya pada simpul berlabel: $\mathcal{L}_{\ell} = -\sum_{i \in \mathcal{V}_{\ell}} \sum_c Y_{ic} \log \hat{P}(y = c \mid i)$.

Contoh 2.6.25 Pada korpus sitiran (Cora/Citeseer), GCN menggunakan struktur graf sitiran sebagai *bias* kehalusan; node tak berlabel tetap mempengaruhi pembelajaran melalui operator propagasi.

2.6.6. Landasan Teoretis Pembelajaran Semi-Terawasi

Teorema 2.6.26 (Regularisasi Manifold (Belkin dkk., 2006)) *Dengan RKHS \mathcal{H}_K dan kernel K , fungsi $f \in \mathcal{H}_K$ dipelajari melalui*

$$\min_{f \in \mathcal{H}_K} \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}(f(\mathbf{x}_i), y_i) + \lambda_A \|f\|_{\mathcal{H}_K}^2 + \lambda_I \mathbf{f}^\top \mathbf{L} \mathbf{f},$$

di mana $\mathbf{f} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))^\top$ dan \mathbf{L} Laplasian graf atas seluruh data. Solusi optimal memiliki bentuk representer $f(\cdot) = \sum_{i=1}^n \alpha_i K(\cdot, \mathbf{x}_i)$.

Bukti. Ikuti Teorema Representer: penalti bergantung pada evaluasi f pada sampel $\{\mathbf{x}_i\}$ dan norma RKHS, sehingga minimizer berada pada span kernel $K(\cdot, \mathbf{x}_i)$. Substitusi ke objektif menghasilkan sistem linier dalam koefisien α (lihat Belkin dkk. 2006). ■

Teorema 2.6.27 (Solusi Harmonik Unik (X. Zhu, 2005)) *Dengan syarat batas $f_L = \mathbf{Y}_L$ dan graf terhubung, solusi energi minimum $E(f) = \frac{1}{2} \sum_{i,j} w_{ij} \|f_i - f_j\|^2$ adalah unik dan memenuhi persamaan Dirichlet diskret $\mathbf{L}f = 0$ pada simpul tak berlabel, sehingga $f_U^* = -\mathbf{L}_{UU}^{-1} \mathbf{L}_{UL} \mathbf{Y}_L$.*

Bukti. Energi graf adalah bentuk kuadrat positif semidefinit; mem-fiks f_L menyisakan masalah konveks dalam f_U . Menyetarakan turunan pertama dengan nol memberi $\mathbf{L}_{UU} f_U + \mathbf{L}_{UL} f_L = 0$. Karena \mathbf{L}_{UU} positif definit pada subgraf tak berlabel yang terhubung, solusi unik ada. ■

Cek this out.

BAB III

RENCANA PENELITIAN

3.1. Rencana Penelitian

Penelitian ini berfokus pada analisis spasial–temporal inflasi di Indonesia dengan unit observasi 38 provinsi. Pendekatan yang digunakan adalah regresi spasial–temporal (GTWR) dan pengembangannya dengan pembobotan berbasis *Graph Neural Network* (GTNNWR). Tujuannya untuk memahami disparitas inflasi antarprovinsi, faktor-faktor global maupun lokal yang memengaruhinya, serta potensi *spillover* antarwilayah.

3.2. Rencana Pengambilan Data

Data yang digunakan dalam penelitian ini berasal dari sumber resmi yang dapat diakses secara terbuka, yaitu:

1. **Inflasi Provinsi (bulanan)** dari BPS (IHK per provinsi, tahun dasar 2022=100).
2. **Suku Bunga Kebijakan (BI 7-Day Reverse Repo Rate)** dari Bank Indonesia.
3. **Nilai Tukar Rupiah/USD (JISDOR)** dari Bank Indonesia (harian, diagregasi bulanan).
4. **Uang Beredar (M2)** dari publikasi BI dan BPS.
5. **Harga Pangan Strategis** (beras, cabai, bawang, minyak goreng, daging ayam) dari PIHPS Nasional (harian, diagregasi bulanan).
6. **Variabel Struktural Provinsi** seperti PDRB, tingkat kemiskinan, dan UMP dari BPS.

7. **Statistik Ekonomi dan Keuangan Daerah (SEKDA)** dari BI, untuk kredit dan simpanan per provinsi.

3.3. Rencana Jadwal Penelitian

Jadwal penelitian direncanakan selama tiga bulan dengan tahapan sebagai berikut:

1. **Bulan 1: Persiapan dan Eksplorasi** Studi literatur, penyusunan proposal, pengunduhan dan pembersihan data, serta analisis deskriptif awal.
2. **Bulan 2: Estimasi Model** Estimasi model panel spasial dinamis sebagai baseline, dilanjutkan dengan implementasi GTWR dan GTNNWR.
3. **Bulan 3: Interpretasi dan Penulisan** Analisis hasil, interpretasi kebijakan, dan penyusunan laporan akhir penelitian.

BAB IV

METODOLOGI PENELITIAN

4.1. Literatur Optimasi Pembobotan pada Model Regresi Spasial

Kajian mengenai regresi berbobot spasial dan spasial-temporal telah berkembang dalam beberapa dekade terakhir. Evolusi model dapat dilacak dari regresi klasik berbasis kernel, regresi berbobot geografis, hingga integrasi dengan pembelajaran mesin dan jaringan saraf graf. Bagian ini memaparkan secara kronologis penelitian-penelitian terdahulu yang relevan dengan metodologi yang diusulkan.

4.1.1. *Geographically Weighted Artificial Neural Network (GWANN)*

Sebagai upaya mengintegrasikan *deep learning* ke dalam kerangka geografis, Hagenauer and Helbich (2022) mengusulkan *Geographically Weighted Artificial Neural Network (GWANN)*. Dalam GWANN, setiap lokasi memiliki model jaringan saraf tiruan (ANN) lokal:

$$y_i = f(\mathbf{x}_i; \theta(s_i)) + \varepsilon_i, \quad (4.1)$$

dengan parameter $\theta(s_i)$ dipengaruhi oleh bobot kernel spasial. ANN lokal di sini menggantikan model linear GWR, sehingga mampu menangkap hubungan non-linear antara kovariat dan respon. Kekurangannya adalah biaya komputasi yang tinggi karena ANN dilatih berulang pada tiap lokasi, serta ketidakstabilan parameter jika data lokal terbatas.

4.1.2. *Spatial and Attribute Neural Network Weighted Regression (SANNWR)*

Ni, Wu, and Zhang (2022) memperkenalkan *Spatial and Attribute Neural Network Weighted Regression (SANNWR)*, yang memperluas GWR dengan bobot berbasis gabungan jarak spasial dan atribut. Estimasi regresi tetap menggunakan

formulasi WLS, tetapi bobot w_{ij} dihitung melalui jaringan saraf yang mempelajari fungsi non-linear dari $(d_{ij}^{\text{spatial}}, d_{ij}^{\text{attribute}})$:

$$w_{ij} = \frac{\exp(\text{NN}(d_{ij}^{\text{spatial}}, d_{ij}^{\text{attribute}}))}{\sum_k \exp(\text{NN}(d_{ik}^{\text{spatial}}, d_{ik}^{\text{attribute}}))}. \quad (4.2)$$

Dengan demikian, bobot tidak lagi berupa fungsi kernel yang kaku, melainkan dipelajari adaptif. Inovasi ini menegaskan peran *deep learning* dalam merepresentasikan kedekatan spasial dan atribut secara simultan.

4.1.3. *Spatial Regression Graph Convolutional Neural Networks (SRGCNN)*

D. Zhu, Liu, Yao, and Fischer (2022) mengusulkan *Spatial Regression Graph Convolutional Neural Network* (SRGCNN), yang menggabungkan regresi spasial dengan konvolusi graf. Model ini merumuskan dependensi spasial dengan operator GCN:

$$H^{(\ell+1)} = \sigma\left(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(\ell)} W^{(\ell)}\right), \quad (4.3)$$

dan kemudian menghubungkannya dengan regresi linear

$$y = X\beta + \rho W y + \varepsilon, \quad (4.4)$$

sehingga efek lag spasial dan heterogenitas dapat ditangkap sekaligus. SRGCNN menandai transisi eksplisit penggunaan GNN dalam konteks regresi spasial.

4.1.4. *Spatio-Temporal Graph Convolutional Networks (ST-GCN)*

Dalam domain peramalan deret waktu spasio-temporal, Yu, Yin, and Zhu (2018) mengusulkan *Spatio-Temporal GCN* (ST-GCN). Model ini mengintegrasikan konvolusi graf spasial dengan konvolusi temporal berbasis deret:

$$H^{(\ell+1)} = \sigma(\text{ConvTemporal}(\tilde{A} H^{(\ell)} W^{(\ell)})). \quad (4.5)$$

ST-GCN dan variannya (ASTGCN, T-GCN) banyak digunakan untuk memprediksi lalu lintas, konsumsi energi, dan dinamika jaringan kompleks. Kontribusinya adalah

menunjukkan bahwa GNN dapat diperluas secara alami ke dimensi waktu.

4.1.5. *Geographically Neural Network Weighted Regression (GNNWR)*

Salah satu perkembangan mutakhir dalam literatur regresi spasial adalah pengenalan *Geographically Neural Network Weighted Regression* (GNNWR) oleh Yin, Huang, Du, Wu, dkk. (2024). Model ini berangkat dari keterbatasan *Geographically Weighted Regression* (GWR) yang menggunakan fungsi kernel parametrik (misalnya Gaussian atau bi-square) untuk menghitung bobot spasial. Dalam GWR, estimator koefisien lokal diberikan oleh

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^\top \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}(u_i, v_i) \mathbf{y}, \quad (4.6)$$

dengan $\mathbf{W}(u_i, v_i)$ matriks bobot kernel spasial berbasis jarak Euclidean antara titik observasi i dan titik lainnya. Namun, pemilihan kernel yang kaku sering gagal menangkap heterogenitas non-linear. Untuk itu, GNNWR mengganti $\mathbf{W}(u_i, v_i)$ dengan bobot yang dipelajari oleh *Spatial Weighted Neural Network* (SWNN). Secara umum, model GNNWR memandang koefisien lokal sebagai skala dari koefisien global,

$$\beta_k^{\text{Lokal}}(u_i, v_i) = w_k(u_i, v_i) \beta_k^{\text{Global}}, \quad \sum_{k=1}^p w_k(u_i, v_i) = 1, \quad w_k \geq 0, \quad (4.7)$$

dengan β_k^{Global} diperoleh dari model OLS baseline. Hubungan ini memastikan bahwa koefisien global dapat diinterpretasikan sebagai rata-rata tertimbang dari koefisien lokal.

Bobot $w_k(u_i, v_i)$ dipelajari menggunakan jaringan saraf tiruan yang menerima input berupa jarak spasial antar unit:

$$\mathbf{W}(u_i, v_i) = \text{SWNN}(d_{i1}^S, d_{i2}^S, \dots, d_{in}^S; \boldsymbol{\theta}), \quad (4.8)$$

dengan d_{ij}^S jarak spasial antara unit i dan j , serta $\boldsymbol{\theta}$ parameter jaringan saraf. Oleh

karena itu, prediksi \hat{y}_i dapat dituliskan sebagai

$$\hat{y}_i = \mathbf{x}_i^\top \mathbf{W}(u_i, v_i) (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}, \quad i = 1, \dots, n. \quad (4.9)$$

Persamaan (4.9) dapat dipadatkan menjadi bentuk matriks

$$\hat{\mathbf{y}} = \mathbf{S} \mathbf{y},$$

dengan \mathbf{S} bergantung pada bobot $\mathbf{W}(u_i, v_i)$ yang dilatih melalui neural network.

Yin dkk. (2024) kemudian memperluas kerangka ini menjadi *Geographically and Temporally Neural Network Weighted Regression* (GTNNWR), yang memasukkan dimensi waktu ke dalam pembobotan. Estimator GTNNWR ditulis sebagai

$$\hat{y}_i = \sum_{k=1}^p w_k(u_i, v_i, t_i) \beta_k^{\text{Global}} x_{ik} + \varepsilon_i, \quad (4.10)$$

dengan bobot spasio-temporal

$$\mathbf{W}(u_i, v_i, t_i) = \text{STWNN}(d_{i1}^{ST}, d_{i2}^{ST}, \dots, d_{in}^{ST}; \boldsymbol{\theta}_1), \quad (4.11)$$

dengan jarak spasio-temporal d_{ij}^{ST} diperoleh melalui *Spatiotemporal Proximity Neural Network* (STPNN), yaitu

$$d_{ij}^{ST} = \text{STPNN}(d_{ij}^S, d_{ij}^T; \boldsymbol{\theta}_2). \quad (4.12)$$

Dengan demikian, GTNNWR melibatkan dua tahap pembelajaran: (i) STPNN untuk menyintesis jarak spasio-temporal dari komponen spasial dan temporal, dan (ii) STWNN untuk mengubah jarak tersebut menjadi bobot regresi.

Sebagaimana pada GWR, prediksi $\hat{\mathbf{y}}$ dalam GNNWR dan GTNNWR tetap dapat dipandang melalui matriks $\mathbf{S}\mathbf{y}$. Akan tetapi, \mathbf{W} kini dihasilkan secara endogen melalui jaringan saraf. Fungsi kerugian yang umum digunakan adalah *mean*

squared error (MSE), yaitu

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

serta kriteria informasi seperti AIC dan AICc untuk menghindari *overfitting*, yaitu

$$\text{AIC}_c = 2k - 2 \ln(L) + \frac{2k(k+1)}{n-k-1},$$

dengan k jumlah parameter efektif dan L adalah *likelihood* model.

Metode GNNWR dan GTNNWR telah diaplikasikan pada berbagai domain, antara lain pemodelan kualitas udara (*air quality*) berbasis PM_{2.5} dan AOD, pemetaan nutrisi laut (*dissolved silicate*), serta estimasi emisi karbon perkotaan. Studi-studi tersebut menunjukkan bahwa dengan menggantikan kernel spasial konvensional dengan neural network, model mampu menangkap heterogenitas spasio-temporal yang lebih kompleks, sekaligus mempertahankan interpretabilitas koefisien lokal sebagaimana pada GWR.

4.2. Graph Neural Networks Geographically and Temporally Varying Coefficients (GNN-GTVC)

Keterbatasan utama pendekatan kernel pada GWR/GTWR (Fotheringham, Brunson, & Charlton, 2002b; Huang, Wu, & Barry, 2010) adalah penggunaan fungsi jarak parametrik yang kaku, sehingga sering gagal menangkap heterogenitas spasio-temporal yang non-linear. Literatur mutakhir seperti GWANN, SANNWR, dan SRGCNN menunjukkan manfaat *deep learning* untuk mempelajari kedekatan efektif yang adaptif. Namun, terdapat celah metodologis penting, yaitu GNNWR/GTNNWR (sebagaimana dipaparkan dalam Yin dkk. 2024) tidak melakukan estimasi koefisien lokal melalui WLS sebagaimana GWR/GTWR klasik, melainkan mengadopsi *skema multiplicative* yang pada hakikatnya adalah *varying-coefficient model* (VCM) Hastie and Tibshirani (1993). Di sisi lain, belum ada kerangka yang secara sistematis menggabungkan dua paradigma tersebut, yaitu bobot dipelajari secara multiplikatif dalam VCM dan bobot yang dipelajari dalam rumus estimator

WLS dalam satu metodologi yang konsisten, transparan, dan dapat dianalisis.

4.2.1. Formulasi Umum GNN-GTVC dan GNN-GTWR

Dimisalkan terdapat himpunan unit spasial–temporal $\mathcal{V} = \{(i, t)\}$ dengan fitur prediktor $x_{ik,t}$ dan lokasi–waktu $z_{i,t} = (u_i, v_i, t)$. VCM spasial–temporal adalah

$$y_{i,t} = \sum_{k=1}^p \beta_k(z_{i,t}) x_{ik,t} + \varepsilon_{i,t}, \quad \mathbb{E}[\varepsilon_{i,t} \mid X] = 0. \quad (4.13)$$

Pada model GTVC, koefisien lokal diparametrisasi dengan koefisien multiplikatif sebagaimana dilakukan oleh Yin dkk. 2024, yaitu

$$\beta_k(z_{i,t}) = w_k(z_{i,t}) \beta_k^{\text{global}}, \quad w_k(z_{i,t}) \geq 0, \quad \sum_{k=1}^p w_k(z_{i,t}) = 1, \quad (4.14)$$

sehingga

$$y_{i,t} = \sum_{k=1}^p w_k(z_{i,t}) \beta_k^{\text{global}} x_{ik,t} + \varepsilon_{i,t}. \quad (4.15)$$

Koefisien global β^{global} dapat diinisialisasi dengan OLS. Pembelajaran $w_k(\cdot)$ dilakukan oleh GNN.

Metodologi lainnya adalah menggunakan bobot yang dipelajari secara langsung ke dalam estimator WLS untuk menjaga sifat-sifat statistik dari GWR atau GTWR. Dengan kata lain,

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}. \quad (4.16)$$

Model ini dinamakan dengan GNN-GTWR. Kedua model layak dibandingkan secara performa.

4.2.2. Arsitektur GNN untuk Pembelajaran Bobot

Data yang dianalisis dapat dibangun berdasarkan graf spasio–temporal $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan k -NN spasial per waktu dan hubungan temporal $(i, t) \leftrightarrow (i, t \pm 1)$. Misal $\tilde{A} = A + I$ dan \tilde{D} adalah matriks derajatnya, maka tiga arsitektur yang

digunakan dalam penelitian ini adalah sebagai berikut.

(a) GCN yang dirumuskan sebagai

$$H^{(\ell+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(\ell)} W^{(\ell)}), \quad H^{(0)} = X, \quad (4.17)$$

dengan $H^{(\ell)}$ representasi (embedding) node pada layer- ℓ .

(b) GAT yang dirumuskan sebagai

$$\alpha_{ij}^{(\ell)} = \text{softmax}_j(\text{LeakyReLU}(a^\top [Wh_i^{(\ell)} \parallel Wh_j^{(\ell)}])), \quad (4.18)$$

$$h_i^{(\ell+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(\ell)} Wh_j^{(\ell)}\right). \quad (4.19)$$

(c) GraphSAGE yang dirumuskan sebagai

$$h_i^{(\ell+1)} = \phi\left(h_i^{(\ell)} \parallel \text{AGGREGATE}(\{h_j^{(\ell)} : j \in \mathcal{N}(i)\})\right), \quad (4.20)$$

dengan AGGREGATE (mean/pool/LSTM) dan ϕ MLP kecil. Ketiga backbone menghasilkan embedding $h_{i,t} \in \mathbb{R}^d$ yang menyandikan kedekatan efektif spasio-temporal.

4.2.3. Penentuan Bobot berdasarkan Hasil Representasi

Memprediksi langsung matriks bobot penuh, seperti $\mathbf{W} \in \mathbb{R}^{N \times N}$ atau vektor untuk setiap node-koefisien memberikan dimensionalitas dan biaya komputasi yang tinggi, yaitu $O(N^2)$. Selain itu, keterbatasan struktur seperti simetris, semidefinit positif, dan ternormalisasi baris sulit diterapkan dan rawan tidak stabil.

4.2.4. Penentuan Bobot berdasarkan Hasil Representasi

Arsitektur GNN menghasilkan representasi laten atau *embedding* node $h_i \in \mathbb{R}^d$ untuk setiap unit spasio-temporal (i, t) . Namun, dalam kerangka GTVC maupun GTWR, yang dibutuhkan bukanlah embedding h_i secara langsung, melainkan bobot w yang menentukan kontribusi relatif dari masing-masing koefisien atau tetangga

dalam estimasi regresi. Oleh karena itu, diperlukan tahap proyeksi dari embedding ke bobot, yaitu $h \mapsto W$.

Secara umum, konversi ini memastikan dua sifat utama: (i) bobot non-negatif dan ter-normalisasi, sehingga dapat ditafsirkan secara probabilistik; (ii) dimensi bobot konsisten dengan skema regresi yang digunakan, baik berbasis koefisien (GTVC) maupun berbasis tetangga (GTWR). Berikut dipaparkan beberapa pendekatan yang umum digunakan.

Dalam GTVC, setiap node (i, t) dipetakan ke vektor bobot $\mathbf{w}_{i,t} \in \Delta^{p-1}$ pada p koefisien regresi. Pemetaan ini dapat diformulasikan sebagai berikut:

$$\text{Dot-product: } w_{i,t}^{(k)} = \frac{\exp(h_{i,t}^\top v_k / \tau)}{\sum_{j=1}^p \exp(h_{i,t}^\top v_j / \tau)}, \quad (4.21)$$

$$\text{Cosine similarity: } w_{i,t}^{(k)} = \frac{\exp\left(\frac{h_{i,t}^\top v_k}{\|h_{i,t}\| \|v_k\| \tau}\right)}{\sum_{j=1}^p \exp\left(\frac{h_{i,t}^\top v_j}{\|h_{i,t}\| \|v_j\| \tau}\right)}, \quad (4.22)$$

$$\text{Gaussian kernel: } w_{i,t}^{(k)} = \frac{\exp\left(-\frac{\|h_{i,t} - v_k\|^2}{2\sigma^2}\right)}{\sum_{j=1}^p \exp\left(-\frac{\|h_{i,t} - v_j\|^2}{2\sigma^2}\right)}, \quad (4.23)$$

$$\text{Learned attention: } w_{i,t}^{(k)} = \frac{\exp(a_k^\top h_{i,t})}{\sum_{j=1}^p \exp(a_j^\top h_{i,t})}, \quad (4.24)$$

$$\text{MLP mapping: } \mathbf{w}_{i,t} = \text{softmax}(\text{MLP}(h_{i,t})). \quad (4.25)$$

di mana $v_k \in \mathbb{R}^d$ adalah vektor representasi koefisien ke- k , $\tau > 0$ parameter temperatur, dan a_k vektor bobot perhatian yang dipelajari. Dengan proyeksi ini, model GTVC dituliskan sebagai

$$y_{i,t} = \sum_{k=1}^p w_{i,t}^{(k)} \beta_k^{\text{global}} x_{ik,t} + \varepsilon_{i,t}. \quad (4.26)$$

Dalam GTWR, estimasi koefisien lokal diperoleh melalui *weighted least squares* (WLS). Untuk setiap target (i, t) , dibentuk bobot terhadap tetangga $(j, s) \in$

$\mathcal{N}(i, t)$. Representasi $h_{i,t}$ dipetakan ke bobot tetangga sebagai berikut:

$$\text{Dot-product: } \tilde{w}_{(i,t) \rightarrow (j,s)} = \frac{\exp(h_{i,t}^\top h_{j,s} / \tau)}{\sum_{(k,r) \in \mathcal{N}(i,t)} \exp(h_{i,t}^\top h_{k,r} / \tau)}, \quad (4.27)$$

$$\text{Cosine / RBF / Attention / MLP: } \text{serupa dengan (4.22)–(4.25),} \quad (4.28)$$

$$\text{dibatasi pada himpunan tetangga } \mathcal{N}(i, t). \quad (4.29)$$

Bobot ini disusun ke dalam matriks diagonal $W_{i,t} = \text{diag}(\tilde{w}_{(i,t) \rightarrow (j,s)})$ sehingga estimator lokal diperoleh sebagai

$$\hat{\beta}_{i,t} = \arg \min_{\beta} \sum_{(j,s) \in \mathcal{N}(i,t)} \tilde{w}_{(i,t) \rightarrow (j,s)} (y_{j,s} - x_{j,s}^\top \beta)^2. \quad (4.30)$$

Solusi tertutup dari (4.30) adalah

$$\hat{\beta}_{i,t} = (X^\top W_{i,t} X)^{-1} X^\top W_{i,t} y, \quad (4.31)$$

dengan X matriks kovariat global dan y vektor respon global. Formulasi ini menegaskan perbedaan utama GTWR terhadap GTVC: bobot digunakan untuk menentukan kontribusi observasi lain dalam regresi lokal, bukan distribusi antar-koefisien.

4.2.5. Fungsi Objektif dan Proses Pelatihan

Setelah embedding dipetakan menjadi bobot W , tahap berikutnya adalah pelatihan model melalui optimisasi fungsi objektif. Fungsi ini dirancang untuk mengukur perbedaan antara nilai aktual y dan prediksi \hat{y} yang dihasilkan oleh model GTVC maupun GTWR.

Fungsi objektif utama yang digunakan adalah *mean squared error* (MSE), didefinisikan sebagai

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4.32)$$

dengan n jumlah observasi, y_i nilai aktual, dan \hat{y}_i prediksi dari model. Formulasi (4.32) bersifat umum dan berlaku baik untuk GTVC (dengan prediksi berbasis

(4.26)) maupun GTWR (dengan prediksi berbasis estimasi WLS (4.30)).

Sebagai alternatif, beberapa literatur menggunakan kriteria informasi seperti *Akaike Information Criterion* terkoreksi (AICc) untuk mengendalikan kompleksitas model:

$$\text{AIC}_c = 2k - 2 \ln(L) + \frac{2k(k+1)}{n-k-1}, \quad (4.33)$$

dengan k jumlah parameter efektif dan L nilai likelihood model. Kriteria ini relevan terutama untuk mencegah *overfitting* pada model dengan banyak parameter bobot yang dipelajari.

Untuk meningkatkan stabilitas, fungsi objektif dapat dilengkapi dengan komponen regularisasi:

1. Entropi bobot:

$$\mathcal{R}_{\text{entropy}} = - \sum_i \sum_j w_{ij} \log w_{ij},$$

yang mencegah distribusi bobot menjadi terlalu tajam (over-confident).

2. Smoothness spasio-temporal:

$$\mathcal{R}_{\text{smooth}} = \sum_{(i,t) \sim (j,s)} w_{(i,t),(j,s)} \|\beta_{i,t} - \beta_{j,s}\|^2,$$

yang mendorong koefisien di lokasi dan waktu berdekatan agar tidak berbeda terlalu jauh, sesuai prinsip kontinuitas spasio-temporal.

Dengan demikian, fungsi objektif total dapat dituliskan sebagai

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \lambda_1 \mathcal{R}_{\text{entropy}} + \lambda_2 \mathcal{R}_{\text{smooth}}, \quad (4.34)$$

dengan $\lambda_1, \lambda_2 \geq 0$ parameter regularisasi.

Pelatihan dilakukan dengan algoritma gradien stokastik. Dua algoritma populer yang digunakan adalah:

1. **Stochastic Gradient Descent (SGD):**

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_{\theta} \mathcal{L}(\theta^{(t)}),$$

dengan η laju belajar (*learning rate*) dan θ parameter model.

2. **Adam optimizer:** varian adaptif dari SGD yang memperbarui parameter dengan estimasi momentum gradien pertama dan kedua:

$$m^{(t+1)} = \beta_1 m^{(t)} + (1 - \beta_1) \nabla_{\theta} \mathcal{L}(\theta^{(t)}), \quad (4.35)$$

$$v^{(t+1)} = \beta_2 v^{(t)} + (1 - \beta_2) (\nabla_{\theta} \mathcal{L}(\theta^{(t)}))^2, \quad (4.36)$$

$$\theta^{(t+1)} = \theta^{(t)} - \eta \frac{m^{(t+1)} / (1 - \beta_1^{t+1})}{\sqrt{v^{(t+1)} / (1 - \beta_2^{t+1})} + \epsilon}. \quad (4.37)$$

Adam banyak digunakan karena stabil dalam pelatihan jaringan saraf dalam, termasuk GNN.

DAFTAR PUSTAKA

- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers. (Foundational text on spatial econometric methods)
- Axler, S. (2024). *Linear Algebra Done Right* (4th ed.). Cham: Springer. doi: 10.1007/978-3-031-41026-0
- Belkin, M., & Niyogi, P. (2004). Semi-supervised learning on riemannian manifolds. *Machine Learning*, 56(1-3), 209–239. doi: 10.1023/B:MACH.0000033121.75173.49
- Belkin, M., Niyogi, P., & Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7, 2399–2434. Diambil dari <http://jmlr.org/papers/v7/belkin06a.html>
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. doi: 10.1109/TPAMI.2013.50
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.
- Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th annual conference on computational learning theory (colt)* (pp. 92–100). doi: 10.1145/279943.279962
- Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18–42. doi: 10.1109/MSP.2017.2693418
- Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-supervised learning*. Cambridge, MA: MIT Press.
- Chung, F. R. K. (1997). *Spectral graph theory*. Providence, RI: American Mathematical Society.

- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303–314. doi: 10.1007/BF02551274
- Dhrymes, P. J. (2013). *Mathematics for Econometrics* (4th ed.). New York: Springer. doi: 10.1007/978-1-4614-8145-4
- Fan, J., & Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1(1), 179–195. (A comprehensive review of varying coefficient models)
- Fefferman, C., Mitter, S., & Narayanan, H. (2016). Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4), 983–1049. doi: 10.1090/jams/852
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002a). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: John Wiley & Sons. (Definitive guide to geographically weighted regression methods)
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002b). . Chichester: Wiley.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *34th international conference on machine learning (icml)*. Diambil dari <https://arxiv.org/abs/1704.01212>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. Diambil dari <http://www.deeplearningbook.org>
- Greene, W. H. (2002). *Econometric Analysis* (5th ed.). Upper Saddle River, New Jersey: Pearson Education.
- Greub, W. (1978). (2nd ed.). New York: Springer-Verlag.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. New York: McGraw-Hill/Irwin.
- Hackbusch, W. (2012). *Tensor Spaces and Numerical Tensor Calculus*. Berlin Heidelberg: Springer. doi: 10.1007/978-3-642-28027-6
- Hagenauer, J., & Helbich, M. (2022). Geographically weighted artificial neural

- networks: A framework for non-linear local modelling. *International Journal of Geographical Information Science*, 36(2), 231–252.
- Hamilton, W., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *31st conference on neural information processing systems (neurips)*. Diambil dari <https://arxiv.org/abs/1706.02216>
- Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757–796.
- Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). New York: Prentice Hall.
- Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*. Diambil dari <https://arxiv.org/abs/1606.08415>
- Hinton, G. E., & Zemel, R. S. (1994). Autoencoders, minimum description length, and helmholtz free energy. In *Advances in neural information processing systems* (Vol. 6, pp. 3–10). Morgan Kaufmann.
- Horn, R. A., & Johnson, C. R. (2012). *Matrix Analysis* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139020411
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251–257. doi: 10.1016/0893-6080(91)90009-T
- Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383–401.
- Jiang, T.-X., Huang, T.-Z., Zhao, X.-L., & Deng, L.-J. (2020, July). Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm. *Journal of Computational and Applied Mathematics*, 372, 112680. Diambil dari <http://dx.doi.org/10.1016/j.cam.2019.112680> doi: 10.1016/j.cam.2019.112680
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer. doi: 10.1007/b98835
- Kingma, D. P., Mohamed, S., Rezende, D. J., & Welling, M. (2014). Semi-

- supervised learning with deep generative models. In *Advances in neural information processing systems (neurips)* (Vol. 27, pp. 3581–3589). Diambil dari <https://arxiv.org/abs/1406.5298>
- Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *5th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1609.02907>
- Laine, S., & Aila, T. (2017). Temporal ensembling for semi-supervised learning. In *5th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1610.02242>
- Lang, S. (1987). *Linear Algebra* (3rd ed.). New York: Springer. doi: 10.1007/978-1-4757-1949-9
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: CRC Press. (Comprehensive introduction to spatial econometric models)
- Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6), 861–867. doi: 10.1016/S0893-6080(05)80131-5
- Li, Y., Chen, C., Wasserman, E. A., & Shi, Z. (2020). Deep neural network interpretability: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9), 3121–3138. doi: 10.1109/TNNLS.2019.2955144
- Maas, A. L. (2013). *Rectifier nonlinearities improve neural network acoustic models* (Unpublished doctoral dissertation). Stanford University.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133. doi: 10.1007/BF02478259
- Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.
- Miyato, T., Maeda, S.-i., Koyama, M., & Ishii, S. (2018). Virtual adversarial training: A regularization method for supervised and semi-supervised le-

- arning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8), 1979–1993. doi: 10.1109/TPAMI.2018.2858821
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). (5th ed.). Hoboken, NJ: Wiley.
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (icml)* (pp. 807–814).
- Ni, X., Wu, H., & Zhang, X. (2022). Sannwr: Spatial and attribute neural network weighted regression for modeling complex spatial heterogeneity. *International Journal of Geographical Information Science*, 36(7), 1401–1426.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. doi: 10.1037/h0042519
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323–2326. doi: 10.1126/science.290.5500.2323
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. doi: 10.1038/323533a0
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61–80. doi: 10.1109/TNN.2008.2005605
- Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3), 83–98. doi: 10.1109/MSP.2012.2235192
- Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems (neurips)* (Vol. 30, pp. 1195–1204). Diambil dari <https://arxiv.org/abs/1703.01780>
- Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric fra-

- mework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319–2323. doi: 10.1126/science.290.5500.2319
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In *6th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1710.10903>
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. doi: 10.1109/TNNLS.2020.2978386
- Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? In *7th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1810.00826>
- Yin, P., Huang, B., Du, Z., Wu, H., et al. (2024). Gnnwr: An r package for geographically neural network weighted regression and its spatiotemporal extension. *Journal of Geographical Systems*.
- Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the twenty-seventh international joint conference on artificial intelligence (ijcai-18)* (pp. 3634–3640).
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. In *5th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1611.03530>
- Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. In *Advances in neural information processing systems (neurips)* (Vol. 16, pp. 321–328).
- Zhu, D., Liu, Y., Yao, X., & Fischer, M. M. (2022). Spatial regression graph convolutional neural networks (srgcnns). *GeoInformatica*, 26(3), 557–582.
- Zhu, X. (2005). Semi-supervised learning literature survey. *Computer Sciences Technical Report*, 1530. Diambil dari <http://pages.cs.wisc.edu/>

`~jerryzhu/pub/ssl_survey.pdf`

Zhu, X., & Goldberg, A. B. (2009). *Introduction to semi-supervised learning*. San Rafael, CA: Morgan & Claypool. doi: 10.2200/S00196ED1V01Y200906AIM006