Laporan Kuis
Data Science

Listra Imelda Sidabutar

11422050

Sarjana Terapan Teknologi Rekayasa Perangkat Lunak

Fakultas Vokasi Institut Teknologi Del

Tahun Ajaran 2024/2025

Linear Algebra In Data Science

Aljabar Linier adalah cabang matematika yang fokus pada studi vektor, matriks, dan transformasi linier. Keberadaan aljabar linier sangat krusial dalam bidang data science dan machine learning, mengingat banyak algoritme yang dibangun berdasarkan konsep-konsepnya. Dalam praktiknya, banyak dataset serta model machine learning yang sering kali disajikan dan divisualisasikan dalam format matriks, menjadikannya sebagai alat yang esensial dalam analisis data.

Aljabar linier sering dianggap sebagai fondasi dari machine learning, yang merupakan salah satu komponen utama dalam data science. Penyampaian data dan representasi berpikir matematis di dalam aljabar linier memungkinkan para profesional untuk memvisualisasikan model-model machine learning dengan lebih jelas. Selain itu, aljabar linier juga memberikan keuntungan dalam mengoptimalkan proses komputasi saat menangani dataset yang besar dan kompleks.

Menurut pengajaran Luis Serrano, matematika yang digunakan dalam machine learning dan aljabar linier merupakan bagian dari program online dasar yang disusun oleh DeepLearning.ai. Program ini dirancang untuk membantu peserta memahami fundamental matematika yang penting bagi penguasaan machine learning. Dengan menguasai aljabar linier, para praktisi data science akan lebih siap dalam mengembangkan dan menerapkan algoritme machine learning yang efektif.

Secara keseluruhan, aljabar linier merupakan alat yang sangat berharga dalam dunia data science dan machine learning, karena kemampuannya untuk merepresentasikan data secara matematis dan mendasari proses analisis serta pengambilan keputusan berbasis data. Terdapat beberapa konsep dasar dari Linear algebra yaitu :

1. Vektor
   Vektor merupakan objek matematis yang memiliki magnitude dan arah yang direpresentasikan menjadi array satu dimensi serta ruang vektor yang merupakan kumpulan dari vektor yang dapat ditambahkan satu sama lain dan dapat dikalikan dengan skalar. Vektor bisa direpresentasikan sebagai daftar angka yang menggambarkan titik-titik dalam ruang.
2. Matriks
   Matriks merupakan array dua dimensi yang digunakan untuk menyajikan atau merepresentasikan sistem linear. Matriks sering digunakan dalam pemodelan data dan operasi dalam data science.
3. Transformasi Linear
   Yang berfungsi untuk memetakan vektor ke vektor lain dengan cara mempertahankan operasi penjumlahan dan perkalian skalar.

Contoh dari Implementasi aljabar linear yang didalamnya ada vektor dan Matriks

```python
import numpy as np

# Membuat vektor
v1 = np.array([1, 2, 3])
v2 = np.array([4, 5, 6])

# Penjumlahan vektor
v_sum = v1 + v2
print("Penjumlahan Vektor:", v_sum)

# Membuat matriks
A = np.array([[1, 2], [3, 4]])
B = np.array([[5, 6], [7, 8]])

# Perkalian matriks
C = np.dot(A, B)
print("Perkalian Matriks:\n", C)

# Invers matriks
A_inv = np.linalg.inv(A)
print("Invers Matriks A:\n", A_inv)
```

```
Penjumlahan Vektor: [5 7 9]
Perkalian Matriks:
 [[19 22]
 [43 50]]
Invers Matriks A:
 [[-2.   1. ]
 [ 1.5 -0.5]]
```

Vectors

Vektor adalah cara untuk merepresentasikan data dalam berbagai dimensi. Setiap titik data memiliki beberapa komponen yang membentuk vektor, di mana setiap komponen mewakili satu dimensi tertentu. Vektor juga bisa diartikan sebagai perpindahan dari titik A ke titik B, dengan panjang garis antara kedua titik disebut sebagai besar vektor, dan arah perpindahannya disebut sebagai arah vektor AB.

Dalam Data Science, vektor sering digunakan untuk menggambarkan fitur, titik data, bobot dalam model, dan gradien dalam proses optimasi. Misalnya, dalam klasifikasi gambar, sebuah gambar dapat direpresentasikan sebagai kumpulan angka dalam bentuk vektor. Dalam pemrosesan bahasa alami, sebuah kalimat juga dapat dikonversi menjadi vektor menggunakan teknik tertentu untuk memahami maknanya.

Rumus dari Vector adalah:

Magnitude of a Vector Formula

$$|A| = \sqrt{x_1^2 + y_1^2 + z_1^2}$$

$$|\vec{v}| = \sqrt{x^2 + y^2}$$

$$|\vec{v}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Berikut merupakan Penggunaan vector pada dataset IrisDataset.csv :

```
import numpy as np

# Mengambil fitur numerik dari dataset
features = df.drop(columns=['species']).values

# --- Operasi pada Vektor ---
vector1 = features[0]
vector2 = features[1]

# Penjumlahan vektor
vector_sum = vector1 + vector2

# Perkalian skalar
scalar = 2
times_vector = scalar * vector1

# Panjang (norma) vektor
vector_norm = np.linalg.norm(vector1)

# Menampilkan hasil operasi vektor
print("Penjumlahan Vektor:", vector_sum)
print("Perkalian Skalar:", times_vector)
print("Norma Vektor:", vector_norm)
```

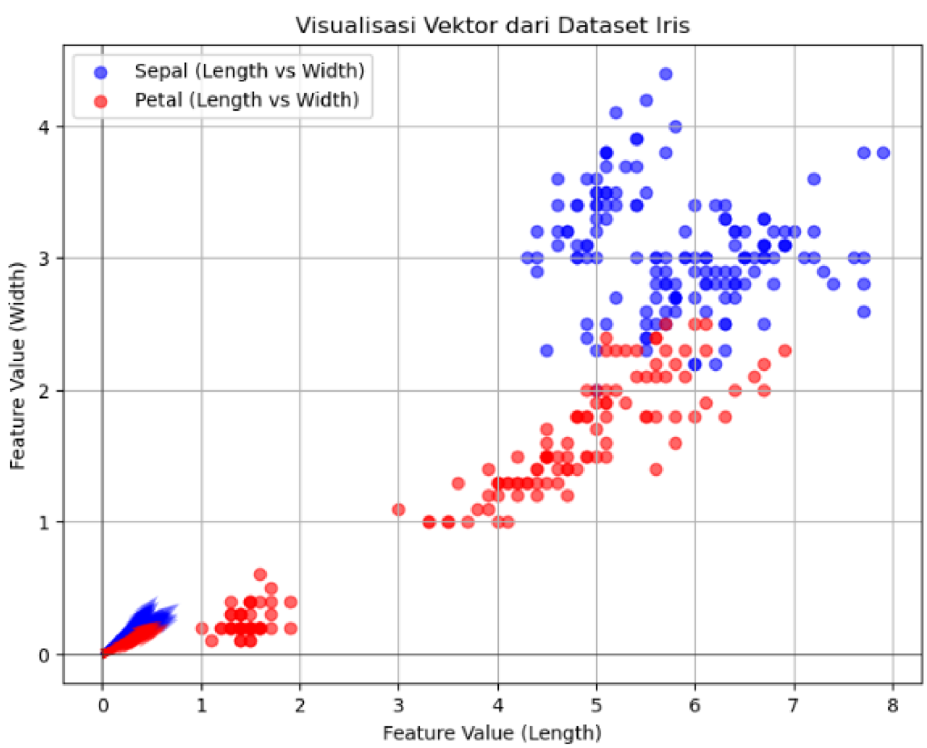
Maka hasilnya adalah :

```
Penjumlahan Vektor: [10.   6.5  2.8  0.4]
Perkalian Skalar: [10.2  7.   2.8  0.4]
Norma Vektor: 6.345076831686122
```

Memvisualisasikan data numerik dari fitur-fitur bunga iris dalam bentuk vektor dan scatter plot:

```
# Mengonversi fitur numerik menjadi matriks (numpy array)
X = df.iloc[:, :-1].values  # Mengambil semua kolom kecuali 'species'

# Scatter plot untuk representasi vektor
plt.figure(figsize=(8, 6))
plt.scatter(X[:, 0], X[:, 1], alpha=0.6, label="Sepal (Length vs Width)", color='blue')
plt.scatter(X[:, 2], X[:, 3], alpha=0.6, label="Petal (Length vs Width)", color='red')

origin = np.zeros((X.shape[0], 2))
plt.quiver(origin[:, 0], origin[:, 1], X[:, 0], X[:, 1], angles='xy', scale_units='xy', scale=10, color='blue', alpha=0.3)
plt.quiver(origin[:, 0], origin[:, 1], X[:, 2], X[:, 3], angles='xy', scale_units='xy', scale=10, color='red', alpha=0.3)
plt.xlabel("Feature Value (Length)")
plt.ylabel("Feature Value (Width)")
plt.title("Visualisasi Vektor dari Dataset Iris")
plt.axhline(0, color='black', linewidth=0.5)
plt.axvline(0, color='black', linewidth=0.5)
plt.legend()
plt.grid(True)
plt.show()
```

Hasilnya adalah :



Matrics

Matriks adalah struktur data yang berbentuk array dua dimensi, terdiri dari baris dan kolom, yang digunakan untuk menyimpan dan merepresentasikan data dalam bentuk angka. Dengan adanya matriks, kita dapat mengorganisir informasi secara sistematis dan melakukan berbagai operasi matematis yang kompleks dengan cara yang lebih sederhana dan efisien. Matriks juga menjadi alat penting dalam banyak bidang, termasuk matematika, fisika, statistik, dan terutama dalam data science serta machine learning, di mana ia digunakan untuk memodelkan dan menganalisis data.

Salah satu aplikasi utama matriks adalah dalam memodelkan transformasi linier. Transformasi linier merupakan konsep yang digunakan untuk merubah posisi, ukuran, dan orientasi objek dalam ruang multidimensi tanpa mengubah bentuk dasarnya. Matriks memungkinkan kita untuk melakukan transformasi tersebut dengan merepresentasikan fungsi matematika dalam bentuk yang lebih mudah dikelola dan dioperasikan.

Beberapa operasi dasar matriks meliputi :

1. Penjumlahan matriks yaitu penjumlahan elemen-elemen yang bersesuaian dari dua matriks
   Rumus :

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} + \begin{bmatrix} j & k & l \\ m & n & o \\ p & q & r \end{bmatrix} = \begin{bmatrix} a+j & b+k & c+l \\ d+m & e+n & f+o \\ g+p & h+q & i+r \end{bmatrix}$$

Atau

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \qquad B = \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$$

$$A + B = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} + \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 & a_2 + b_2 \\ a_3 + b_3 & a_4 + b_4 \end{bmatrix}$$

2. Perkalian matriks yaitu mengalikan baris dari matriks pertama dengan kolom dari matriks kedua.
   Rumus :

$$c_{ij} = \sum_{k-1}^{n} a_{ik} b_{kj}$$

3. Invers matriks yaitu matriks yang dikalikan dengan matriks asli dan menghasilkan matriks identitas.
   Rumus :
   Jika $A \times A^{-1} = I$,

Yang dimana $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Maka invers matriks dapat dihitung dengan :

$$A^{-1} = \frac{1}{\det(A)} \mathrm{adj}(A)$$

Secara keseluruhan, matriks adalah alat yang sangat fleksibel dan kuat dalam analisis matematis. Dengan memanfaatkan operasi dasar pada matriks, para profesional dapat melakukan analisis data yang mendalam, merancang algoritme yang kompleks, serta mengembangkan model yang dapat memprediksi dan memberikan wawasan berharga dari data yang ada.

Berikut merupakan Penggunaan matrics pada dataset IrisDataset.csv :

```python
# --- Operasi pada Matriks ---
matrix = features[:5]

# Transpos matriks
matrix_transpose = matrix.T

# Perkalian matriks
matrix_product = np.dot(matrix, matrix_transpose)

# Invers matriks (hanya jika matriks persegi dan tidak singular)
if matrix.shape[0] == matrix.shape[1]:
    matrix_inverse = np.linalg.inv(matrix)
else:
    matrix_inverse = "Matriks tidak persegi, tidak dapat dihitung inversenya."

# Determinan matriks (jika persegi)
if matrix.shape[0] == matrix.shape[1]:
    matrix_determinant = np.linalg.det(matrix)
else:
    matrix_determinant = "Matriks tidak persegi, determinan tidak dapat dihitung."

# Menampilkan hasil operasi matriks
print("Transpos Matriks:\n", matrix_transpose)
print("Perkalian Matriks:\n", matrix_product)
print("Invers Matriks:", matrix_inverse)
print("Determinan Matriks:", matrix_determinant)
```

Maka hasil nya adalah:

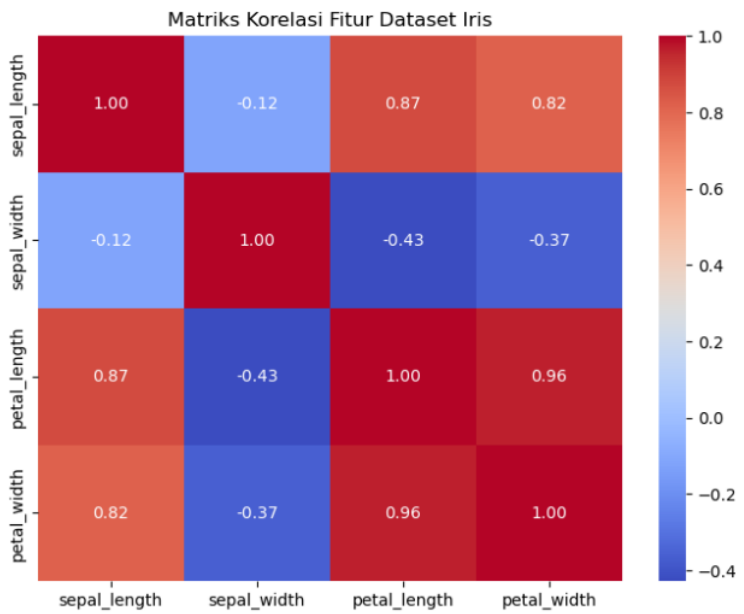
```
Transpos Matriks:
 [[5.1 4.9 4.7 4.6 5. ]
 [3.5 3.  3.2 3.1 3.6]
 [1.4 1.4 1.3 1.5 1.4]
 [0.2 0.2 0.2 0.2 0.2]]
Perkalian Matriks:
 [[40.26 37.49 37.03 36.45 40.1 ]
 [37.49 35.01 34.49 33.98 37.3 ]
 [37.03 34.49 34.06 33.53 36.88]
 [36.45 33.98 33.53 33.06 36.3 ]
 [40.1  37.3  36.88 36.3  39.96]]
Invers Matriks: Matriks tidak persegi, tidak dapat dihitung inversenya.
Determinan Matriks: Matriks tidak persegi, determinan tidak dapat dihitung.
```

Menampilkan satu sampel bunga.

```
# Mengambil fitur numerik dan mengonversinya menjadi matriks
X = df.iloc[:, :-1].values

# Menampilkan sebagian data sebagai matriks
print("Matriks Data (5 Sampel Pertama):")
print(X[:5, :])

# Menghitung matriks korelasi
correlation_matrix = df.iloc[:, :-1].corr()

# Visualisasi matriks korelasi dengan heatmap
plt.figure(figsize=(8,6))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Matriks Korelasi Fitur Dataset Iris")
plt.show()
```

Hasilnya adalah :

```
Matriks Data (5 Sampel Pertama):
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]]
```



Pada gambar dapat kita ketahui bahwa visualisasinya menampilkan bahwa setiap baris merepresentasikan satu sampel bunga. Pada visualisasi headmap ini dapat diketahui bahwa Terdapat korelasi positif yang kuat antara Sepal Length dan Petal Length (0.87), yang berarti semakin panjang sepal, semakin panjang pula petalnya, sedangkan Sepal Width dan Petal Length memiliki korelasi negatif sedang (-0.43), menunjukkan bahwa semakin lebar sepal, panjang petal cenderung berkurang; hubungan antara Petal Length dan Petal Width sangat kuat (0.96), menandakan bahwa keduanya cenderung bertambah bersama, sementara Sepal Width dan Sepal Length memiliki korelasi sangat lemah (-0.12), yang berarti hampir tidak ada hubungan antara keduanya.

Referensi :

https://arsipdosen.amikom.ac.id/Ferry-Wahyu-Wibowo-SSi-MCs-Dr_20240223231443_Handbook_Aljabar_Linier.pdf

https://www.researchgate.net/profile/Swasti-Maharani/publication/348177064_ALJABAR_LINEAR/links/5ff2a63845851553a0199329/ALJABAR-LINEAR.pdf

https://repository.ung.ac.id/get/kms/14308/Resmawan-Aljabar-Linear-Elementer-Vektor.pdf

https://365datascience.com/tutorials/python-tutorials/linear-algebra-data-science/

https://medium.com/@adebayoaliyyah2020/how-linear-algebra-and-statistics-are-related-to-data-science-69c963a12a14

https://blog.myskill.id/istilah-dan-tutorial/dasar-dan-penerapan-linear-algebra-dalam-data-analysis/