

# Numpy\_With\_Dataset\_Iris

June 7, 2025

## 0.1 NumPy with Datasets Iris Dataset by Isha

### 0.2 1. Import Libraries

```
[6]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

### 0.3 2. Load Dataset using Pandas and Convert to NumPy

```
[7]: # Load dataset
df = pd.read_csv('iris_dataset.csv')
print("Pandas DataFrame:")
print(df.head())
# Convert to NumPy array
data_np = df.to_numpy()
print("\nNumPy Array:")
print(data_np[:5])
```

Pandas DataFrame:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	\
0	5.1	3.5	1.4	0.2	
1	4.9	3.0	1.4	0.2	
2	4.7	3.2	1.3	0.2	
3	4.6	3.1	1.5	0.2	
4	5.0	3.6	1.4	0.2	

	target
0	0
1	0
2	0
3	0
4	0

NumPy Array:

```
[[5.1 3.5 1.4 0.2 0. ]
 [4.9 3.  1.4 0.2 0. ]
 [4.7 3.2 1.3 0.2 0. ]
```

```
[4.6 3.1 1.5 0.2 0. ]
[5.  3.6 1.4 0.2 0. ]]
```

## 0.4 3. Basic NumPy Operations

```
[8]: # Separate features and target
X = data_np[:, :-1] # all columns except target
y = data_np[:, -1]  # target
# Mean and Std of each feature
print("Mean of each feature:", np.mean(X, axis=0))
print("Std of each feature:", np.std(X, axis=0))
```

```
Mean of each feature: [5.84333333 3.05733333 3.758      1.19933333]
Std of each feature: [0.82530129 0.43441097 1.75940407 0.75969263]
```

## 0.5 4. Filtering: Select Setosa Species Only

```
[9]: # Setosa has target label 0
setosa = data_np[data_np[:, -1] == 0]
print("Setosa samples:", setosa.shape[0])
print(setosa[:5])
```

```
Setosa samples: 50
[[5.1 3.5 1.4 0.2 0. ]
 [4.9 3.  1.4 0.2 0. ]
 [4.7 3.2 1.3 0.2 0. ]
 [4.6 3.1 1.5 0.2 0. ]
 [5.  3.6 1.4 0.2 0. ]]
```

## 0.6 5. Feature Normalization

```
[12]: # Normalize features
X_norm = (X - np.mean(X, axis=0)) / np.std(X, axis=0)
print("Normalized feature matrix (first 5 rows):\n", X_norm[:5])
```

```
Normalized feature matrix (first 5 rows):
[[-0.90068117  1.01900435 -1.34022653 -1.3154443 ]
 [-1.14301691 -0.13197948 -1.34022653 -1.3154443 ]
 [-1.38535265  0.32841405 -1.39706395 -1.3154443 ]
 [-1.50652052  0.09821729 -1.2833891  -1.3154443 ]
 [-1.02184904  1.24920112 -1.34022653 -1.3154443 ]]
```

## 0.7 6. Visualize with Matplotlib

```
[13]: # Scatter plot of normalized sepal length vs width
plt.scatter(X_norm[:, 0], X_norm[:, 1], c=y, cmap='viridis')
plt.xlabel('Sepal Length (Normalized)')
plt.ylabel('Sepal Width (Normalized)')
plt.title('Iris Sepal Length vs Width')
```

```
plt.colorbar(label='Species')  
plt.show()
```

