🧩 Exercise List

Step 1 – Data Generation

Generate two datasets corresponding to two hypothetical classes, following independent bivariate normal distributions:

- **Class 1:**
  $\mu_x = 25$, $\mu_y = 10$, $\sigma_x^2 = 100$, $\sigma_y^2 = 25$
  (150 samples)

- **Class 2:**
  $\mu_x = 60$, $\mu_y = 30$, $\sigma_x^2 = 36$, $\sigma_y^2 = 144$
  (150 samples)

Combine both classes into a single dataset containing **300 points**.

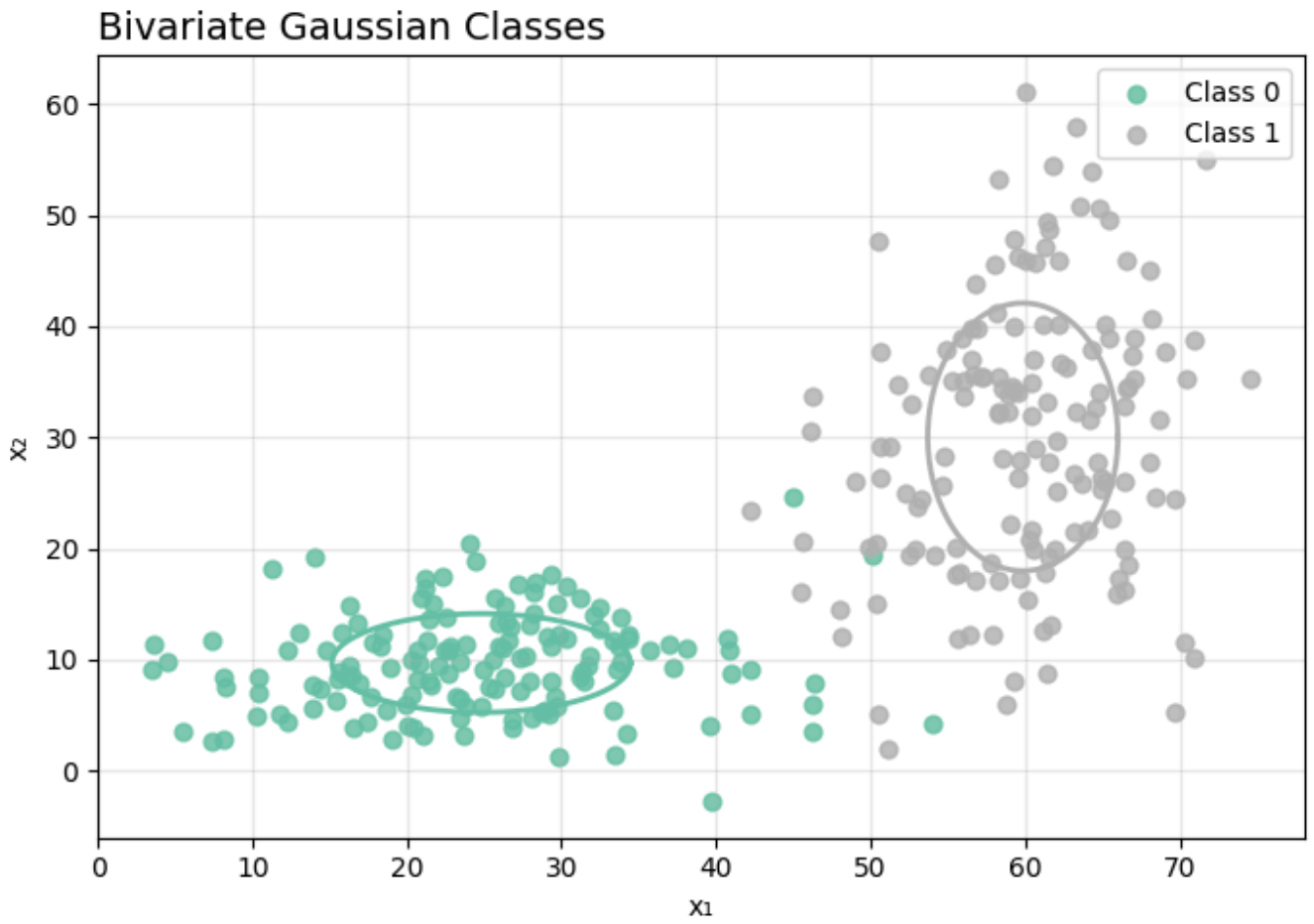
```
%run 00-setup.py
```

```
import yaml
import numpy as np

from pathlib import Path

from ml.viz import scatter_data
from ml.data import generate_data, save_dataset, export_csv
```

```
# Load YAML configuration
CFG_PATH = Path("../configs/default.yaml")
cfg = yaml.safe_load(open(CFG_PATH))

# Generate dataset based on YAML parameters
X, y = generate_data(cfg)

print(f"Dataset generated: {X.shape[0]} samples, {X.shape[1]} features")
print(f"Classes: {set(y)}")
```

```
Dataset generated: 300 samples, 2 features
Classes: {np.int64(0), np.int64(1)}
```

```
# Visualize the two Gaussian classes

scatter_data(X, y,title="Bivariate Gaussian Classes")
```




```
(<Figure size 700x500 with 1 Axes>,
 <Axes: title={'left': 'Bivariate Gaussian Classes'}, xlabel='x₁', ylabel='x₂'>)
```

```
for cls in np.unique(y):
    Xc = X[y == cls]

    print(f"\nClass {cls}")
    print("Mean:", np.round(Xc.mean(axis=0), 2))
    print("Var :", np.round(Xc.var(axis=0), 2))
```

```
Class 0
Mean: [24.81  9.68]
Var : [93.24 19.79]

Class 1
Mean: [59.85 30.04]
Var : [ 37.84 145.36]
```

```
# Save generated dataset

save_dataset(X, y, "../data/data_bivariate_gaussian.npz")
```

```
WindowsPath('../data/data_bivariate_gaussian.npz')
```

```
# Export generated dataset

export_csv(X, y, "../data/data_bivariate_gaussian.csv")
```

```
WindowsPath('../data/data_bivariate_gaussian.csv')
```