### **Task 8**

**Machine Learning**

Upload .py or Ipynb extension file on GitHub public repo “100DaysofBytewise" and share the link in the submission form by 3 July 2024.

**Exercise: Calculate the mean, median, and mode of the sepal lengths in the Iris dataset.**

**CODE:** import pandas as pd

from sklearn.datasets import load\_iris

iris = load\_iris()

iris\_df = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

sepal\_length = iris\_df['sepal length (cm)']

mean = sepal\_length.mean()

median = sepal\_length.median()

mode = sepal\_length.mode()[0]

print(f'Mode: {mode}')

print(f'Mean: {mean}')

print(f'Median: {median}')

**Exercise: Calculate the variance and standard deviation of the petal widths in the Iris dataset.**

**Code:** import pandas as pd

from sklearn.datasets import load\_iris

iris = load\_iris()

iris\_df = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

petal\_width = iris\_df['petal width (cm)']

std\_dev = petal\_width.std()

variance = petal\_width.var()

print(f'Standard deviation: {std\_dev}')

print(f'Variance: {variance}')

**Exercise: Create a summary table that includes the mean, median, variance, and standard deviation for all numerical features in the dataset.**

**Code:** import pandas as pd

from sklearn.datasets import load\_iris

iris = load\_iris()

iris\_df = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

summary = pd.DataFrame({

    'mean': iris\_df.mean(),

    'median': iris\_df.median(),

    'variance': iris\_df.var(),

    'standard deviation': iris\_df.std()

})

print(summary)

**Exercise: Define a random variable for the sepal length and calculate the probability distribution of sepal lengths.**

**Code:** import pandas as  pd

import numpy as np

from sklearn.datasets import load\_iris

import matplotlib.pyplot as plt

iris=load\_iris()

iris\_df=pd.DataFrame(data=iris.data,columns=iris.feature\_names)

sepal\_length=iris\_df['sepal length (cm)']

hist,bin\_edges=np.histogram(sepal\_length,bins='auto',density=True)

distribution = pd.DataFrame({

    'sepal\_length': bin\_edges[:-1],

    'probability': hist

})

Print(distribution)

**Exercise: Plot the probability distribution of sepal lengths using a histogram.**

**Code:** import pandas as  pd

import numpy as np

from sklearn.datasets import load\_iris

import matplotlib.pyplot as plt

iris=load\_iris()

iris\_df=pd.DataFrame(data=iris.data,columns=iris.feature\_names)

sepal\_length=iris\_df['sepal length (cm)']

hist,bin\_edges=np.histogram(sepal\_length,bins='auto',density=True)

distribution = pd.DataFrame({

    'sepal\_length': bin\_edges[:-1],

    'probability': hist

})

plt.figure(figsize=(10,4))

plt.hist(sepal\_length,bins='auto',color='black',alpha=0.7,density=True)

plt.title('Probability Distribution ')

plt.grid(True)

plt.show()

**Exercise: Calculate the cumulative distribution function (CDF) for the petal lengths and plot it.**

**Code:** import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.datasets import load\_iris

iris = load\_iris()

i\_df = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

petal\_length = i\_df['petal length (cm)']

x = np.sort(petal\_length)

y = np.arange(1, len(x) + 1) / len(x)  # formula for CDF

plt.figure(figsize=(10, 4))

plt.plot(x, y, marker='o', linestyle='--', color='purple')

plt.grid(True)

plt.show()

**Exercise: Calculate and plot the probability density function (PDF) for sepal width.**

**Code:** import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.datasets import load\_iris

iris = load\_iris()

i\_df = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

sepal\_width = i\_df['sepal width (cm)']

plt.figure(figsize=(10, 4))

sns.kdeplot(sepal\_width, bw\_adjust=0.1, fill=True, color='black')

plt.grid(True)

plt.show()

**Exercise: Determine the probability of a randomly selected iris flower having a petal length greater than a given value.**

**Code:** import pandas as pd

from scipy.stats import norm

from sklearn.datasets import load\_iris

iris = load\_iris()

iris\_df = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

petal\_length = iris\_df.iloc[:, 2]

mu, std = norm.fit(petal\_length)

value = float(input('Enter your desired value: '))

x = 1 - norm.cdf(value, loc=mu, scale=std)

print(f"The probability of {value} is : {x}")

**Exercise: Perform a hypothesis test to determine if there is a significant difference in the mean petal length between two species of iris flowers.**

**Code: in progress**

**Exercise: Calculate and interpret the covariance and correlation between sepal length and sepal width.**

**Code:** import pandas as pd

from sklearn.datasets import load\_iris

iris = load\_iris()

df = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

sepal\_length = df['sepal length (cm)']

sepal\_width = df['sepal width (cm)']

covariance = sepal\_length.cov(sepal\_width)

correlation = sepal\_length.corr(sepal\_width)

print(f"Covariance between sepal length and sepal width: {covariance:.4f}")

print(f"Correlation between sepal length and sepal width: {correlation:.4f}")