**Övning 2**

# Import necessary libraries

from sklearn.datasets import load\_iris, fetch\_california\_housing

import pandas as pd

# Load the datasets

iris = load\_iris()

california = fetch\_california\_housing()

# Convert the datasets into DataFrames for easier exploration

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

iris\_df['target'] = iris.target

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

california\_df['target'] = california.target

# Display the first few rows of each dataset to understand the data structure

print("Iris Dataset (Classification Task):")

print(iris\_df.head())

print("\nCalifornia Housing Dataset (Regression Task):")

print(california\_df.head())

# Brief descriptions of the datasets to guide discussion

print("\nIris target variable (classification):", set(iris.target))

print("California Housing target variable (regression): range =", (california\_df['target'].min(), california\_df['target'].max()))

**Exercise Questions**

1. **Dataset Exploration**: After running the code, examine the output for each dataset.
   * What are the features in each dataset?
   * What is the target variable in each dataset?
2. **Problem Type Identification**:
   * Which dataset is suitable for classification? Why?
   * Which dataset is suitable for regression? Why?
3. **Explain Your Reasoning**:
   * Explain in a few sentences why the Iris dataset is suited for classification and the California Housing dataset for regression, based on the target variable type.