General Exercises for SK-Learn Regression datasets

Scikit-learn (sklearn) provides several built-in datasets suitable for regression tasks, which are useful for learning and experimenting with various regression algorithms.

Accessing Built-in Regression Datasets:

These datasets are part of the sklearn.datasets module. They are typically loaded using specific functions that return a Bunch object, which behaves like a dictionary containing the data, target variable, feature names, and a description.

Examples of Built-in Regression Datasets:

- load_diabetes(): This dataset contains 442 samples, each with 10 physiological features and a quantitative measure of disease progression one year after baseline. It's a classic dataset for regression.
  Python

```python
from sklearn.datasets import load_diabetes
diabetes = load_diabetes()
X = diabetes.data  # Features
y = diabetes.target # Target variable (disease progression)
```

- load_boston(): This dataset contains information about housing in Boston, including features like crime rate, proportion of residential land, and average number of rooms per dwelling, with the target being the median value of owner-occupied homes. Note: While historically popular, this dataset is sometimes avoided due to ethical concerns related to its origins and potential for misuse.
  Python

```python
# from sklearn.datasets import load_boston # (Consider alternatives if ethical concerns are present)
# boston = load_boston()
# X = boston.data
# y = boston.target
```

- make_regression(): This function generates a synthetic regression dataset with a specified number of samples, features, and noise. This is particularly useful for creating custom datasets for testing and demonstration purposes.
  Python

```python
from sklearn.datasets import make_regression
X, y = make_regression(n_samples=100, n_features=1, noise=10)
```

Using the Datasets:

Once loaded, these datasets can be used to train and evaluate various regression models available in scikit-learn, such as LinearRegression, DecisionTreeRegressor, RandomForestRegressor, etc.

Part 0 - "All Regression SK Learn" algo To DO:

Take a dataset above and apply "all Regression SK Learn" algo(s) studied and follow – normal Machine learning workflow

Put code in your personal Public code GitHub repository.

Thanks.