

Numpy Exercise

The following is the set of numpy functions which are used while working with a dataset.

Numpy Documentation:

https://numpy.org/doc/stable/user/absolute_beginners.html

(Please refer the above documentation for further knowledge and practice)

This PDF uses the Wine Recognition Dataset provided by Scikit Learn Library.

Dataset Details:

https://scikit-learn.org/stable/datasets/toy_dataset.html#wine-recognition-dataset

(If you are using Jupyter notebook, then install the library first by using the following code in a code block)

```
!pip install scikit-learn
!pip install numpy
```

(scikit learn is pre-installed in google colab)

→ Encourage to use Jupyter notebooks for this exercise

STEPS TO IMPORT DATASET FROM SCIKIT LEARN:

```
# 1. Load the wine dataset from scikit-learn and print the names of
the features.

from sklearn.datasets import load_wine

wine = load_wine()
feature_names = wine.feature_names
print("Feature names:", feature_names)

# 2. Load data into respective variables
X = wine.data      # Independent variables (feature matrix)
y = wine.target    # Dependent variables (target column)
```

EXECUTE THE FOLLOWING NUMPY FUNCTIONS AND GIVE ONE LINE EXPLANATION FOR THE OUTPUT:

1. Using the shape method, find the shape of Independent and dependent variables separately.
 - a. `arr.shape`
2. Print the feature names
 - a. `wine.feature_names`
3. Extract the first 5 samples from the feature matrix (X) and print them.
 - a. Use indexing method
4. Extract the first 5 target labels (y) and print them.
 - a. Same as above. Use indexing method
5. Find the minimum and maximum values of each feature in the dataset.
 - a. `arr.min()` → Choose the correct parameters within the function i.e. input and axis
6. Calculate the mean value of the alcohol content feature.
 - a. `np.mean(independent_variable[:, feature_names.index(column_name)])`
 - b. Input the correct 'independent_variable' and 'column_name'
7. Print the unique class labels in the target vector (y).
 - a. Use the unique function → check documentation for execution
8. Find the wine(s) with the highest magnesium measurement
 - a. Refer the magnesium column using indexing and store in a variable → `X[:,4]`
 - b. Find max of this column using max function → refer documentation for the same
9. Calculate the mean, median, and standard deviation of 'proanthocyanins', 'color_intensity', and 'hue' column separately
 - a. Refer the columns using indexing and store in variable as above
 - b. Use the mean, median, std functions of numpy to display the output. → refer documentation
 - c. Also print the column names while printing.
eg. 'Mean of proanthocyanins is ____'