Iris Classifier problem statement

Supervised learning problem

Specifications used are :-

- length and width of petals and sepals

Outcome is to create a machine capable of identifying iris species from specified measurements .

This type of problem is a classification problem which will have classes as outputs . Each Iris would belong to one of the classes in the dataset .

Using train_test_split :-

-using a portion of the dataset to train the model
-using separate portion to test the model

Features and Targets in machine learning :-

-Features are basically input variables to the model
-Targets are classifications made on the basis of features

**default test size if not specified is 0.25 .
**(.shape) returns number of rows and columns

Inspecting your data :-

Using scatter plots to visualise data (only 2d or 3d)
A work-around for this limitation is to use pair plots that utilise all possible combinations of features

Steps to create a scatter plot are :-

1) Create a scatter matrix (panda dataFrame)
We can observe that the diagonals contain histograms of the features

2) Observe the datapoints of the 3 classes and if they are well separated , this means the machine model is likely to distinguish between the classes.

K - Nearest neighbour :-

eg:- Take 3 nearest points to the current locus
Check the class of the points
Use majority of the 3 classes to predict the new locus

Next steps are to call the fit method on the knn object for the parameters x_train and y_train

Using predict method to then use the y-pred
