

Python for Machine Learning

Q1. (Install)

- Install anaconda.
- Create an environment called dmml.
- Activate the environment.
- Install jupyter
- Install pandas
- Install scikit-learn
- Start jupyter notebook.
- Test (1+1)

Q2. (arrays)

- Create a python notebook for arrays.
- Run the example in the notes.
- Create 1, 2 and 3D arrays.
- Obtain slices of arrays.
- Reshape a 1D array to a 2D array.
- Introduce appropriate cells.

Q3. (Dataframe)

- Read the iris dataset from a csv file stored in a folder called data, into a dataframe.
- Get the type, head, columns, shape, and index of the dataframe
- Get the value of a row of the df
- Get the value of an entry in the df
- Take a copy of the dataframe (df2) and set the value of an entry in df2. Print the value of the entry in df and df2.
- Define dataframes X and y where X contains the feature values in the dataframe and y contains the target values.
- Convert X and y to ndarrays.
- Define dataframes X and y where X contains the feature values in the dataframe and y contains the target values. Use df.drop() to define X and df.species to define y.

Q4. Confusion matrix

For the values `actual = [0, 1, 0, 1]` and `predicted = [1, 1, 1, 0]`

- Find the confusion matrix.
- Print it out
- Find the values of TN, FP, FN, TP

Do the same for `actual = ['B', 'M', 'B', 'M']` and `predicted = ['M', 'M', 'M', 'B']`

Do the same for `actual = ['W', 'M', 'W', 'M']` and `predicted = ['M', 'M', 'M', 'W']` but make W the negative and M the positive.

Do the same for `actual = ["cat", "ant", "cat", "cat", "ant", "bird"]` and `predicted = ["ant", "ant", "cat", "cat", "ant", "cat"]`.

Q5. Confusion matrix

The following is actual and predicted values for a cyber-security model.

```
actual = ['attack', 'normal', 'normal', 'attack', 'normal', 'normal',  
'attack', 'normal', 'attack']  
predicted = ['attack', 'normal', 'attack', 'attack', 'normal', 'normal',  
'normal', 'normal', 'normal']
```

Print out the confusion matrix and the values of tn, fp, fn, tp.

Q6. (Builtin Datasets)

- Use `load_iris` from `sklearn.datasets`.
- Print out the type, `feature_names` and `target_names` for the iris dataset.
- Set X to be the iris data and y to be the iris target
- Print out the description (DESCR) of the dataset and read it.

Q7. (Onehot Encoding)

Load the mushroom dataset from a csv file.

Set X to be all columns except type.

Set y to be the type. (classification problem.)

Use `X = pd.get_dummies(X)` to one hot encode X.