



Hands on Introduction to Python and its Scientific Library for Data Science/Mining

Instructor:
Heru Praptono

Some based on scikit learn: <http://scikit-learn.org/stable/index.html>



UNIVERSITAS
INDONESIA
Veritas, Probitas, Justitia

Agenda



- Setup your tools (5 minutes: 14:45-14.50)
(NECESSARY)
- Familiarise with Python, numpy, scipy, (or pandas, or anything else you wish) [14.50-15.50]
(IMPORTANT)
- Try to run something in Scikit-Learn (feel free to explore, it will not cost you :)) [15.50-16.15]
(YOUR CREATIVITY)



UNIVERSITAS
INDONESIA
Veritas, Probatum, Tutelis

Tools



- Setup your tools/environment (5 minutes)
 - Command prompt
 - Jupyter Notebook (recommended)
 - PyCharm (recommended)

Get your hands dirty with python

- Find the attached document on Moddle

Try Scikit Learn: Load Example Dataset

- **Load Example Dataset**

```
$ python
```

```
>>> from sklearn import datasets
```

```
>>> iris = datasets.load_iris()
```

```
>>> digits = datasets.load_digits()
```



UNIVERSITAS
INDONESIA
Veritas, Probitas, Justitia



Try Scikit Learn: Load Example Dataset

- A dataset is a dictionary-like object that holds all the data and some metadata about the data
`{'key' : 'value'}`
- See what keys: `[dataset].keys()`

Try Scikit Learn: Load Example Dataset

- In the case of the digits dataset, **digits.data** gives access to the features that can be used to classify the digits samples

```
>>> print(digits.data)
```

- **digits.target** gives the ground truth for the digit dataset, that is the number corresponding to each digit image that we are trying to learn:

```
>>> digits.target
```



UNIVERSITAS
INDONESIA

Veritas, Probatum, Tutis



Try Scikit Learn: Load Example Dataset

- Shape of the data arrays

The data is always a 2D array, shape (**n_samples**, **n_features**), although the original data may have had a different shape. In the case of the digits, each original sample is an image of shape (8, 8) and can be accessed using:

```
>>> digits.images[0]
```

(see the explanation on board)

Try Scikit Learn: example - Learning and Inference

- An example of an estimator is the class `sklearn.svm.SVC` that implements support vector classification. The constructor of an estimator takes as arguments the parameters of the model, but for the time being, we will consider the estimator as a black box:

```
>>> from sklearn import svm
```

```
>>> clf = svm.SVC(gamma=0.001, C=100.)
```

- [for the parameter, see the documentations]



UNIVERSITAS
INDONESIA
Veritas, Prodesse, Tutare



Try Scikit Learn: example - Learning and Inference

- Suppose: we want to predict an instance. It means, that one instance should be excluded from training set
- We select this training set with the `[:-1]` Python syntax, which produces a new array that contains all but the last entry of `digits.data`



UNIVERSITAS
INDONESIA



pusilkom ui

Try Scikit Learn: example - Learning and Inference

- Learning

```
>>> clf.fit(digits.data[:-1], digits.target[:-1])
```

- Inference

```
>>> clf.predict(digits.data[-1:])
```



UNIVERSITAS
INDONESIA
Veritas, Probatum, Tutelis

Now...

- Explore on your own..
- See:

<http://scikit-learn.org/stable/tutorial/index.html>