SC201 Lecture 6

Pandas

- 1D ➡ _____
- 2D ➡ _____(More than 1 series)

Series

import panda as pd

s = _____(_____)

< Attributes >

_____ # _____

_____ # _____

_____ # _____

< Methods >

_____

_____ # _____

_____ # _____

Show the first/last n

_____ / _____

_____ / _____

_____ / _____

_____ / _____

Counts of uniques

_____

Fill missing data

_____

DataFrame d = _____(_____)

d

0

1

2

< Attributes >

_____ # _____ } Get _____

_____ } Get _____

_____

_____ # _____

_____ #

_____ #

< Add new column >

_____ = _____

= _____

d

0

1

2

< Add new row >

_____ = { _____ }

_____ = _____(_____)

< Count non-NA cells >

_____

scikit-learn

- _____
- _____ ➡ Easy to use!

< from sklearn import preprocessing >

Normalization

_____ = _____

X = _____

_____

X = _____

Standardization

_____ = _____

X = _____

= _____

_____

X = _____

< from sklearn import linear_model >

Logistic regression

h = _____

_____ = _____

_____ = _____

Linear regression

h = _____

_____ = _____

Dtrain → ☐ → ☐ ← **Dtest**

↓

predictions

< from sklearn import Polynomial Features >