ScikitLearn操作記錄單(ScikitLearn pre-Processing)

組別: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 學號: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 姓名: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. 請根據以下教學資源操作: <http://www.cse.msu.edu/~ptan/dmbook/tutorials/tutorial4/tutorial4.html>

若對python 及 pandas 不熟悉，請從<http://www.cse.msu.edu/~ptan/dmbook/software/>中學習更前面的教學內容。

1. 請自行查詢了解下列scikit-learn 模組的功能作用<https://scikit-learn.org/stable/>

|  |  |  |  |
| --- | --- | --- | --- |
| Pre-processing | Module | Function | 試寫程式，實驗該函式所提供功能及主要參數設定效果並簡要敘述 |
| Missing data | sklearn.preprocessing | Imputer() |  |
| Sampling | sklearn.model\_selection | train\_test\_split() |  |
| sklearn.utils.random | sample\_without\_replacement() |
| Binarize | sklearn.preprocessing | Binarizer() |  |
| sklearn.preprocessing | OneHotEncoder() |
| Discretization | sklearn.preprocessing | K-bins discretization() |  |
| Standardize | sklearn.preprocessing | StandardScaler()  Scale() |  |
| Normalize | sklearn.preprocessing | MinMaxScaler() |  |
| Dimension reduction | sklearn.decomposition | PCA() |  |
| Feature selection | sklearn.feature\_selection | VarianceThreshold() |  |
| sklearn.feature\_selection | SelectKBest()  chi2 |  |
| sklearn.feature\_selection | SelectFromModel() |  |
|  | sklearn.feature\_selection | RFE() |  |

其他參考資源:

* machine learning 參考書: "[Introduction to Machine Learning with Python](https://www.amazon.com/Introduction-Machine-Learning-Python-Scientists/dp/1449369413)"

之github code <https://github.com/amueller/introduction_to_ml_with_python>

* Scikit Learn 官方documentation: <https://scikit-learn.org/stable/>
* 自尋搜尋其他可信網路資源