Lecture : PlaY Data

Detail Outline:

- [Talk] Asking Questions & Get the Solutions
    a. Questions in Real World
    b. Questions & Solutions in Machine Learning & Data Mining
    c. Solutions in Real World
    d. Standard Working Flow of Problem-Solving & Data-Playing
- [Hands-on Orange] Typical Problem-Solving Skills in Machine Learning & Data Mining
    a. Understanding the Working Flow of Data Analysis
    b. Hands-on PlaY Data
    c. Supervised Problem
    d. Unsupervised Problem
- [Talk] A Lighting Talk in Data Sciience
    a. What is Data ?
    b. What is Science ? To see is to believe ?
    c. How to hear the voice from data ?
- [Hands-on Pandas] Preprocessing before Machine Learning & Data Mining:
    a. What is a DataFrame ?
    b. Basic Manipulation of DataFrame
    c. Basic Statistics in pandas
        i. cut, qcut, hist, ….
    d. A Lighting Talk in Probability
    e. Aggregation Skills & Conditional Probability
    f. More Useful Skills in pandas
        i. pivot_table & crosstable
        ii. merge & concat
- [Hands-on Matplotlib] Visualization Skills:
    a. Plot with DataFrame Directly
    b. Basic Objects in Matplotlib
    c. More Skills in Matplotlib
    d. Exploratory Data Analysis with Matplotlib
- [Talk] A Lighting Talk in Visualization:
    a. Visualization of Data
    b. Visualization of Model
    c. Visualization of Problem
    d. Visualization of Concept
    e. Teach Yourserlf & Teach Others
- [Hands-on sklearn] PlaY Data with Scikit-Learn:
    a. Preprocessers
    b. Analysis Tools

- [Hands-on sklearn&scipy] PlaY Natural Language Processing (with PTT corpus) :
    a. Basic Working Flow
        i. Text To Vector
        ii. Feature Selection & Vector Quantization
    b. Text Classification with Scikit-Learn
    c. Text Similarity & Clustering with Scikit-Learn
    d. Solving Large Scale Problem with scipy
    e. A Lighting Talk in Language Modeling

Timeline :

- [Talk 10 min] Asking Questions & Get the Solutions
- [Hands-on Orange 40 min] Typical Problem-Solving Skills in MLDM
- [Break 10 min]
- [Talk 15 min] A Lighting Talk in Data Sciience
- [Hands-on Pandas 80 min] Preprocessing before Machine Learning & Data Mining
- [Break 10 min]
- [Hands-on Matplotlib 50 min] Visualization Skills
- [Talk 10 min] A Lighting Talk in Visualization
- [Break 10 min]
- [Hands-on sklearn 40 min] PlaY Data with Scikit-Learn
- [Break 5 min]
- [Hands-on sklearn&scipy 80 min] PlaY Natural Language Processing (with PTT corpus)

Environment Settings:

- Required Python Packages:
    - PyMongo
    - Orange
    - pandas
    - matplotlib
    - scikit-learn
    - numpy & scipy
    - jieba
    - nltk
    - gensim
    - cv2 (OpenCV)
- Other Requirements:
    - mongodb
    - OpenCV

Required Skills:

- Linear Algebra
- Multivariate Calculus
- Basic Probability
- Basic Python Manipulation
- (optional) some basic concept in Convex optimization & Discrete Optimization
- (optional) some basic concept in Machine Learning or Data Mining
- (optional) some basic concept in Image Processing & Digital Signal Processing
- (optional) some basic concept in Natural Language Processing