**JAX – ML Course Plan**

Below are the dates and times we prefer to teach the course:

Thu Feb 13, 2pm-5pm

Fri Feb 14, 10am-3pm

Thu Feb 20, 10am-3pm

Feb Feb 21, 10am-3pm

Asa and I discussed about the course, and came up with the draft lesson plan below. We consider the course in eight two-hour blocks: 90 mins lesson/hands-on practice and 30 mins exercise. We want to cover a specific topic in each block.

We would like to build the lesson material around a dataset relevant to the general interests of the participants, e.g to  gene expression dataset with some clinical features.

1st block – 2 hour 1pm-2:45pm: Overview of the relevant Python library modules – getting familiar with the dataset, generating plots to examine feature/class distributions.

2nd block –2 hour 3pm – 5pm: Introduction to ML – k-Nearest Neighbor Classification – train-validation-test procedure

3rd block: Pre-processing – data type conversion, normalization, feature selection…

4th block: Performance evaluation – ROC, Precision-recall… overfitting

5th and 6th blocks: 2-3 classification algorithms – e.g. Naïve Bayes, Random Forest, Neural Networks – classification as a whole including pre-processing, feature selection and performance evaluation

7th block: Regression; Unsupervised learning - clustering (k-means, PCA, TSNE)

8th block: Deep Learning – convolutional neural networks

**1st Block – 1:45 hours:**

**Sue – Introduction – 10 mins**

Asa – Introduction to the workshop – 5 mins

Asa - **Python Review - 15 mins:**

***Live Coding basic python functions (Improvisational):***

* Conditionals (if, elif, else) (also explain indenting for code blocks)
* Lists: initializing, appending, adding, extending
* for/while loops
* writing a function ( def myFunction(): )
* Importing (show numpy and pandas)
* Using methods from a library using the dot (.) operator, also explain “tabbing” through methods
* Googling a method to understand what it does
* **parameters/arguments** and **return values**

Asa - **Numpy Arrays – 20 mins**

Asa – **Exercise 20 mins –** **load the MiceProtein\_2f2c.csv**

Asa **– Plotting 15 mins – Scatter plot, line plot, bar plot**

--------

**2nd Block – 2 hours:**

Asli – **ML overview slide – 10 mins**

Asli – **kNN overview slide 5 mins**

Asli **– kNN implementation – 30 mins**

Asli **– kNN from Scikit package overview and application – 15 mins**

Asli **– Exercise – 30 mins**

Test different k values using the package (in a for loop), put the results (probability) in an array and generate the line plot.

**3rd block – 2 hours:**

Asli – **Normalization 10 mins** – problem and solution using the exercise

Asli – **Pre-processing – slides 30 mins**: data type conversion (one hot encoding), normalization (standard normalization, min-max normalization), handling missing values – Scikit functions to use the functions – a separate jupyter notebook to show Scikit functions on small random datasets, and exercises on the actual dataset

Asli – **Exercise - 30 mins: load the actual dataset and apply normalization and handling missing values-**

Asli – **Train-test calculate accuracy** - **30** **mins**, using pre-defined k value – repeat the exercise from 2nd block with accuracy instead of probability

**4th block – 2 hours:**