IE 345 - K "Introduction to Deep Learning: Fundamentals Concepts"

Prof. Yuzo

K Nearest Neighbor

pg. 56 - 63

```python
In [1]:  import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
```

```python
In [2]:  dataset = pd.read_csv('diabetes.csv')
         dataset.head(5)
```

Out[2]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcor
0	6	148	72	35	0	33.6	0.627	50	
1	1	85	66	29	0	26.6	0.351	31	
2	8	183	64	0	0	23.3	0.672	32	
3	1	89	66	23	94	28.1	0.167	21	
4	0	137	40	35	168	43.1	2.288	33	

```python
In [3]:  dataset.shape
```

Out[3]: (768, 9)

```python
In [4]:  dataset.describe()
```

Out[4]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunc
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420

```python
In [5]:  # Data frame correlation function
         corr = dataset.corr()

         fig, ax = plt.subplots(figsize = (13, 13))
         ax.matshow(corr) #color code the rectangles by correlation value
         plt.xticks(range(len(corr.columns)), corr.columns) #draw x tick marks
         plt.yticks(range(len(corr.columns)), corr.columns) #draw y tick marks
```

```
Out[5]:  ([<matplotlib.axis.YTick at 0x1f6ba5de198>,
           <matplotlib.axis.YTick at 0x1f6ba5d6a90>,
           <matplotlib.axis.YTick at 0x1f6ba5c0978>,
           <matplotlib.axis.YTick at 0x1f6ba8658d0>,
           <matplotlib.axis.YTick at 0x1f6ba865da0>,
           <matplotlib.axis.YTick at 0x1f6ba865e80>,
           <matplotlib.axis.YTick at 0x1f6ba85d710>,
           <matplotlib.axis.YTick at 0x1f6ba86e278>,
           <matplotlib.axis.YTick at 0x1f6ba86e780>],
          <a list of 9 Text yticklabel objects>)
```
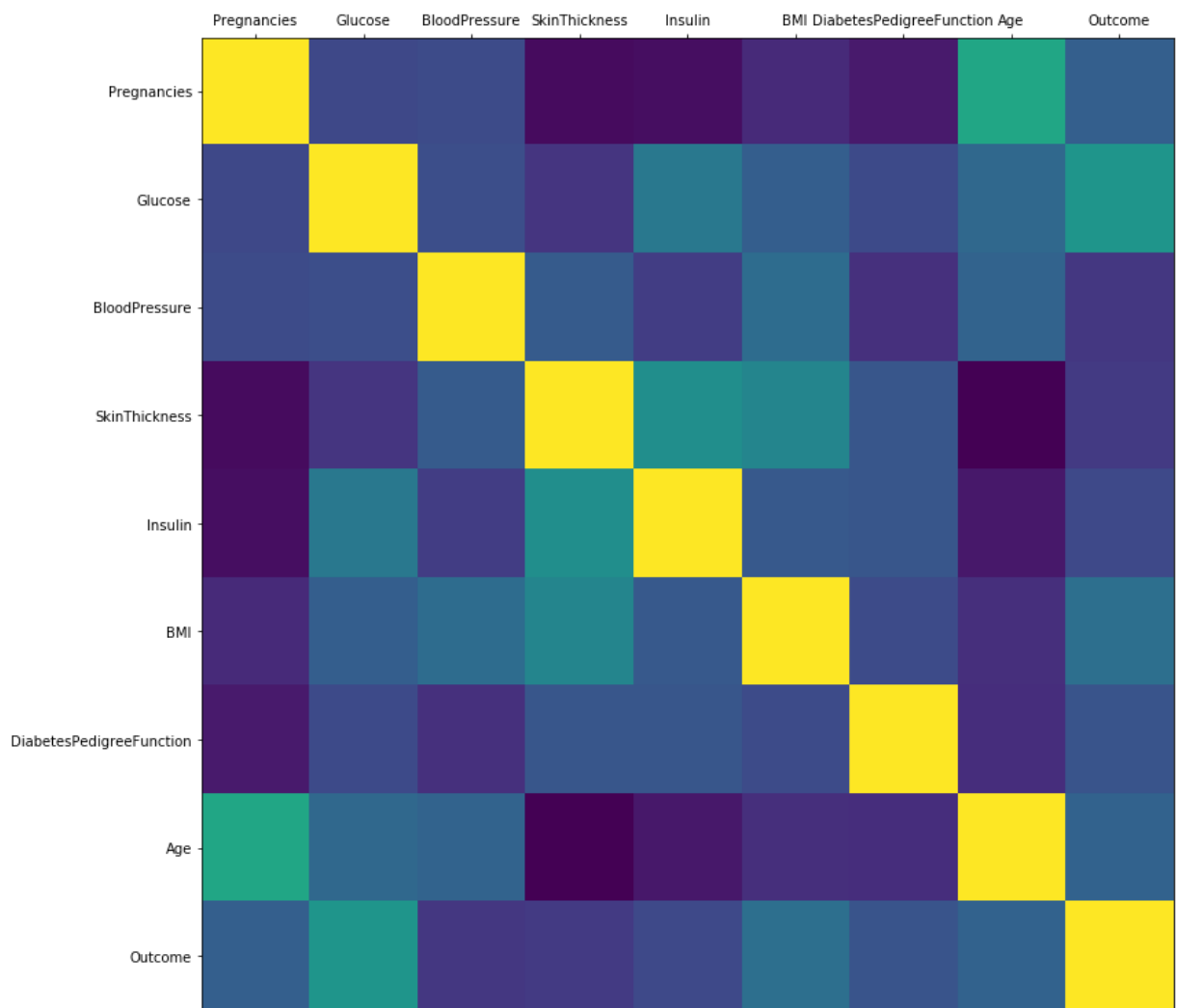


```python
In [6]:  features = dataset.drop(['Outcome'], axis=1)
         labels = dataset['Outcome']
```

```python
In [7]:  from sklearn.model_selection import train_test_split

         features_train, features_test, labels_train, labels_test = train_test_split(features, lab
         els, test_size=0.25)
```

```
In [8]:  # Importing the model
         from sklearn.neighbors import KNeighborsClassifier

         classifier = KNeighborsClassifier()
         classifier.fit(features_train, labels_train)

         pred = classifier.predict(features_test)
```

```
In [9]:  from sklearn.metrics import accuracy_score

         accuracy = accuracy_score(labels_test, pred)
         print('Accuracy: {}'.format(accuracy))
```

Accuracy: 0.7083333333333334

Pablo David Minango Negrete

pablodavid218@gmail.com

Lisber Arana Hinostroza

lisberarana@gmail.com