# Untitled

October 8, 2022

[1]:
```
import pandas as pd
from sklearn.neural_network import MLPClassifier
from sklearn.svm import SVC

from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.feature_extraction import DictVectorizer

from sklearn.pipeline import Pipeline
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV, ParameterGrid

import numpy as np
```

[2]:
```
df_train = pd.read_csv('train.csv')
df_test = pd.read_csv('test.csv')
```

[3]:
```
df_train.head()
```

[3]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th… | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | | | | | |
| 3 | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | | | | | |
| 4 | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | | | | | |

```
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

[4]:
```
df_test.head()
```

[4]:
```
   PassengerId  Pclass                                          Name     Sex  \
0          892       3                              Kelly, Mr. James    male
1          893       3              Wilkes, Mrs. James (Ellen Needs)  female
2          894       2                     Myles, Mr. Thomas Francis    male
3          895       3                              Wirz, Mr. Albert    male
4          896       3  Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female

    Age  SibSp  Parch   Ticket     Fare Cabin Embarked
0  34.5      0      0   330911   7.8292   NaN        Q
1  47.0      1      0   363272   7.0000   NaN        S
2  62.0      0      0   240276   9.6875   NaN        Q
3  27.0      0      0   315154   8.6625   NaN        S
4  22.0      1      1  3101298  12.2875   NaN        S
```

[5]:
```
df_train.describe(include="all")
```

[5]:
```
        PassengerId    Survived      Pclass                     Name   Sex  \
count    891.000000  891.000000  891.000000                      891   891
unique          NaN         NaN         NaN                      891     2
top             NaN         NaN         NaN  Braund, Mr. Owen Harris  male
freq            NaN         NaN         NaN                        1   577
mean     446.000000    0.383838    2.308642                      NaN   NaN
std      257.353842    0.486592    0.836071                      NaN   NaN
min        1.000000    0.000000    1.000000                      NaN   NaN
25%      223.500000    0.000000    2.000000                      NaN   NaN
50%      446.000000    0.000000    3.000000                      NaN   NaN
75%      668.500000    1.000000    3.000000                      NaN   NaN
max      891.000000    1.000000    3.000000                      NaN   NaN

                Age       SibSp       Parch  Ticket        Fare    Cabin  \
count   714.000000  891.000000  891.000000     891  891.000000      204
unique         NaN         NaN         NaN     681         NaN      147
top            NaN         NaN         NaN  347082         NaN  B96 B98
freq           NaN         NaN         NaN       7         NaN        4
mean     29.699118    0.523008    0.381594     NaN   32.204208      NaN
std      14.526497    1.102743    0.806057     NaN   49.693429      NaN
min       0.420000    0.000000    0.000000     NaN    0.000000      NaN
25%      20.125000    0.000000    0.000000     NaN    7.910400      NaN
50%      28.000000    0.000000    0.000000     NaN   14.454200      NaN
75%      38.000000    1.000000    0.000000     NaN   31.000000      NaN
max      80.000000    8.000000    6.000000     NaN  512.329200      NaN
```

```
       Embarked
count       889
unique        3
top           S
freq        644
mean        NaN
std         NaN
min         NaN
25%         NaN
50%         NaN
75%         NaN
max         NaN
```

[6]:
```
df_train = df_train.drop(["Name", "Age", "Ticket", "Cabin"], axis=1)
df_test = df_test.drop(["Name", "Age", "Ticket", "Cabin"], axis=1)
```

[7]:
```
df_test.head()
```

[7]:

| | PassengerId | Pclass | Sex | SibSp | Parch | Fare | Embarked |
|---|---|---|---|---|---|---|---|
| 0 | 892 | 3 | male | 0 | 0 | 7.8292 | Q |
| 1 | 893 | 3 | female | 1 | 0 | 7.0000 | S |
| 2 | 894 | 2 | male | 0 | 0 | 9.6875 | Q |
| 3 | 895 | 3 | male | 0 | 0 | 8.6625 | S |
| 4 | 896 | 3 | female | 1 | 1 | 12.2875 | S |

[8]:
```
df_train = df_train.dropna(subset=["Embarked"])
```

[9]:
```
df_train["Sex"] = np.where(df_train["Sex"] == "female", 1, 0)
df_test["Sex"] = np.where(df_test["Sex"] == "female", 1, 0)
```

[10]:
```
emb_hot = pd.get_dummies(df_train["Embarked"], prefix="embarked")
sex_hot = pd.get_dummies(df_train["Sex"], prefix="sex")
plcass_hot = pd.get_dummies(df_train["Pclass"], prefix="pclass")

df_train = pd.concat([df_train, emb_hot, sex_hot, plcass_hot], axis=1)
df_train = df_train.drop(["Pclass", "Sex", "Embarked"], axis=1)
```

[11]:
```
df_train.head()
```

[11]:

| | PassengerId | Survived | SibSp | Parch | Fare | embarked_C | embarked_Q | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 7.2500 | 0 | 0 | |
| 1 | 2 | 1 | 1 | 0 | 71.2833 | 1 | 0 | |
| 2 | 3 | 1 | 0 | 0 | 7.9250 | 0 | 0 | |
| 3 | 4 | 1 | 1 | 0 | 53.1000 | 0 | 0 | |
| 4 | 5 | 0 | 0 | 0 | 8.0500 | 0 | 0 | |

```
   embarked_S  sex_0  sex_1  pclass_1  pclass_2  pclass_3
0           1      1      0         0         0         1
1           0      0      1         1         0         0
2           1      0      1         0         0         1
3           1      0      1         1         0         0
4           1      1      0         0         0         1
```

[13]:
```
emb_hot = pd.get_dummies(df_test["Embarked"], prefix="embarked")
sex_hot = pd.get_dummies(df_test["Sex"], prefix="sex")
plcass_hot = pd.get_dummies(df_test["Pclass"], prefix="pclass")

df_test = pd.concat([df_test, emb_hot, sex_hot, plcass_hot], axis=1)
df_test = df_test.drop(["Pclass", "Sex", "Embarked"], axis=1)
```

[14]:
```
df_test.head()
```

[14]:
```
   PassengerId  SibSp  Parch     Fare  embarked_C  embarked_Q  embarked_S  \
0          892      0      0   7.8292           0           1           0
1          893      1      0   7.0000           0           0           1
2          894      0      0   9.6875           0           1           0
3          895      0      0   8.6625           0           0           1
4          896      1      1  12.2875           0           0           1

   sex_0  sex_1  pclass_1  pclass_2  pclass_3  embarked_C  embarked_Q  \
0      1      0         0         0         1           0           1
1      0      1         0         0         1           0           0
2      1      0         0         1         0           0           1
3      1      0         0         0         1           0           0
4      0      1         0         0         1           0           0

   embarked_S  sex_0  sex_1  pclass_1  pclass_2  pclass_3
0           0      1      0         0         0         1
1           1      0      1         0         0         1
2           0      1      0         0         1         0
3           1      1      0         0         0         1
4           1      0      1         0         0         1
```

[ ]: