09.03.2017 task2

```
In [73]: import numpy as np
          from sklearn import tree
          from sklearn.tree import DecisionTreeClassifier
          import pandas as pd
          from sklearn.model selection import cross val score
          import graphviz
          import matplotlib.pyplot as plt
          from IPython.display import Image
          %matplotlib inline
  In [7]: X = pd.read csv('german.data', sep=' ', header=None)
          cat features = [0, 2, 3, 5, 6, 8, 9, 11, 13, 14, 16, 18, 19]
          new cat features = pd.get_dummies(X[cat_features])
          X.drop(cat features, axis=1, inplace=True)
          X = pd.concat((X, new cat features), axis=1)
          y = X[20]
          X.drop(20, axis=1, inplace=True)
In [110]: classifier = DecisionTreeClassifier(max depth=3, criterion='gini')
          classifier.fit(X, y)
          classifier.score(X, y)
Out[110]: 0.751
```

09.03.2017 task2

```
In [111]: tree.export graphviz(classifier, 'tree.dot')
               !dot -Tpng tree.dot > output.png
               Image(filename = "output.png", width=1000, height=100)
Out[111]:
                                                                                        X[10] \le 0.5
                                                                                         gini = 0.42
                                                                                       samples = 1000
                                                                                     value = [700, 300]
                                                                                                        False
                                                                               True
                                                                     X[0] \le 22.5
                                                                                                           X[49] \le 0.5
                                                                     gini = 0.4869
                                                                                                           gini = 0.2062
                                                                    samples = 606
                                                                                                           samples = 394
                                                                   value = [352, 254]
                                                                                                          value = [348, 46]
                                                                     X[26] \le 0.5
                                                                                                           X[31] \le 0.5
                                       X[12] \le 0.5
                                                                                                                                        X[4] \le 30.5
                                       gini = 0.4438
                                                                     gini = 0.4973
                                                                                                           gini = 0.4175
                                                                                                                                        gini = 0.1502
                                      samples = 349
                                                                     samples = 257
                                                                                                           samples = 64
                                                                                                                                        samples = 330
                                     value = [233, 116]
                                                                   value = [119, 138]
                                                                                                          value = [45, 19]
                                                                                                                                       value = [303, 27]
                  gini = 0.4271
                                       gini = 0.3878
                                                          gini = 0.4722
                                                                              gini = 0.4717
                                                                                                   gini = 0.3911
                                                                                                                      gini = 0.375
                                                                                                                                        gini = 0.2622
                                                                                                                                                            gini = 0.0806
                                                                                                                      samples = 4
                  samples = 330
                                       samples = 19
                                                          samples = 89
                                                                              samples = 168
                                                                                                   samples = 60
                                                                                                                                        samples = 116
                                                                                                                                                           samples = 214
                                                         value = [55, 34]
                 value = [228, 102]
                                      value = [5, 14]
                                                                             value = [64, 104]
                                                                                                  value = [44, 16]
                                                                                                                      value = [1, 3]
                                                                                                                                       value = [98, 18]
                                                                                                                                                           value = [205, 9]
```

09.03.2017 task2

```
In [114]: plt.plot(depths, cv_score, label='Cross validation score')
    plt.plot(depths, test_score, label='Test score')
    plt.xlabel('max_depth')
    plt.ylabel('accuracy')
    plt.legend(loc='upper left')
    plt.show()
```

