

# KNearestNeighbors-Loans

April 27, 2017

## 1 K-Nearest Neighbor - KD-Trees

We are now going to apply k-nearest neighbor to the loans dataset with and without kd-trees.  
First we load all the needed libraries.

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model, datasets, neighbors
from sklearn import model_selection
from sklearn.model_selection import StratifiedKFold
from matplotlib.colors import ListedColormap
import time
%matplotlib inline
```

```
/home/xcs/anaconda3/lib/python3.6/site-packages/matplotlib/font_manager.py:280: Use
'Matplotlib is building the font cache using fc-list. '
```

Next, we load the dataset that is included in the Scikit-Learn dataset module.

```
In [2]: loans = pd.read_csv('LoansNumerical.csv')
target = 'safe_loans'
features = loans.columns[loans.columns!=target]

x = loans[features]
y = loans[target]
```

Now we apply plain k-nearest neighbor with a k of 15 and evaluate it using 10 fold crossvalidation

```
In [3]: k = 15
knn = neighbors.KNeighborsClassifier(n_neighbors=k, algorithm='brute')
knn_kdtree = neighbors.KNeighborsClassifier(n_neighbors=k, algorithm='kd_tree')

In [4]: %%time
start_kdtree = time.process_time()
knn_eval_kd = model_selection.cross_val_score(knn_kdtree, x, y, cv=StratifiedKFold(10))
end_kdtree = time.process_time()
```

CPU times: user 11.6 s, sys: 308 ms, total: 11.9 s  
Wall time: 11.9 s

```
In [5]: time_taken_kdtree = end_kdtree-start_kdtree
        print("%d-nearest-neighbor\tt=%.3f\tAccuracy=%.3f\tStd=%.3f"%(k,time_taken_
```

15-nearest-neighbor	t=11.932	Accuracy=0.806	Std=0.001
---------------------	----------	----------------	-----------

```
In [6]: %%time
        start_vanilla = time.process_time()
        knn_eval = model_selection.cross_val_score(knn, x, y, cv=StratifiedKFold(n_
        end_vanilla = time.process_time()
```

CPU times: user 8min 23s, sys: 1min 33s, total: 9min 57s  
Wall time: 7min 5s

```
In [7]: time_taken_vanilla = end_vanilla-start_vanilla
        print("%d-nearest-neighbor\tt=%.3f\tAccuracy=%.3f\tStd=%.3f"%(k,time_taken_
```

15-nearest-neighbor	t=597.443	Accuracy=0.806	Std=0.001
---------------------	-----------	----------------	-----------