Final Course Assignment

Mark Niehues, Stefaan Hessmann
Mathematical Aspects in Machine Learning

13. Juli 2017

1 Introduction

In the past course we dealt with the broad mathematical foundations of machine learning. To get an idea of what the consequences of those mathematical theorems and approaches are and to get in touch with the standard Python tools, we have evaluated an comparatively easy data science example found on `kaggle.com`. Since this was our first machine learning project, we decided to deal with an rather simple problem. The example dataset [1] consists of the historic passenger records of the disastrous Titanic maiden voyage in 1912. The goal in this Challenge was to predict if a passenger survived the accident based on informations like for example age, sex and payed ticket fare. Therefore it is in terms of machine learning a **classification problem** with two classes: survived and not survived.

Inspired by sample solutions from the website, we first took a deeper look on the dataset and tried to select the most significant influences by reviewing the statistical properties of the dataset. In the following we implemented an naive Sequential Minimal Optimization (SMO) algorithm and ran a few tests with them in order to finally compare the results with other machine learning algorithms.

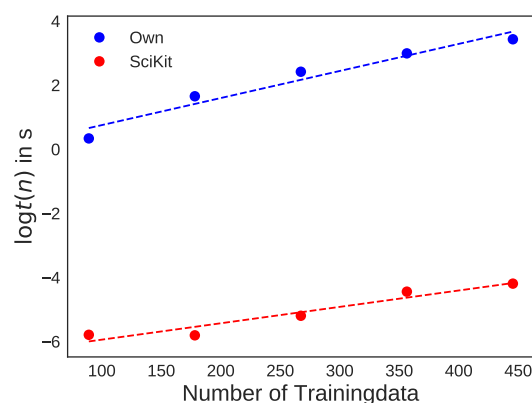## 2 Applying Machine Learning Methods on the Titanic Disaster

blalba

Abbildung 1: Benchmark

3 Implementation of an easy SMO Algorithm

Listing 1: Hello World

```python
0  #    Copyright (C) 2017 Mark Niehues, Stefaan Hessmann
1  #
2  #    This program is free software: you can redistribute it and/or modify
3  #    it under the terms of the GNU General Public License as published by
4  #    the Free Software Foundation, either version 3 of the License, or
5  #    (at your option) any later version.
6  #
7  #    This program is distributed in the hope that it will be useful,
8  #    but WITHOUT ANY WARRANTY; without even the implied warranty of
9  #    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
10 #    GNU General Public License for more details.
11 #
12 #    You should have received a copy of the GNU General Public License
13 #

16 import numpy as np


19 class Kernels:
20     """
21     Class that holds different Kernels
22     """
23     def __init__(self, gamma):
24         self.gamma = gamma
25         self.kernels = {
26             "rbf" : self.kernel_rbf,
27             "linear": self.kernel_lin}

29     def get_kernel(self, kernel_name):
30         return self.kernels[kernel_name]

32     def kernel_lin(self, x, y):
33         """
34         Linear kernel
35         """
36         return x.dot(y)

38     def kernel_rbf(self, x, y):
39         """
40         RBF Kernel
41         """
42         d = x - y
43         return np.exp(-np.dot(d, d) * self.gamma)
```

Literatur

[1] Kaggle. *Titanic: Machine Learning from Disaster*. 13. Juli 2017. URL: https://www.kaggle.com/c/titanic.