Neural networks for text classification

Marko Sahan
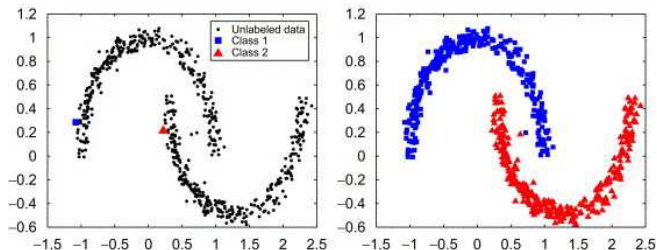
Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering

21.06.2018

Content

- Description of The Semi-supervised Learning Problem
- Text Representation
- Classification Methods
- Binary Classification
- Multiclass Classification
- Bayessian Neural Networks

Description of the Semi-supervised Learning Problem

* from Wang, Zhang, Robust self-tuning semi-supervised learning, Neurocomputing, 70, 2007.

Semi-supervised Learning of Text Documents

Sources of Text Information

- ► Social Media
- ► Emails
- ► Text Messages

Text Representation

Types of text representation

- One-hot Bag-of-Words
- Frequency Bag-of-Words
- Bigram Bag-of-Words
- Embeddings vector

Text Representation

One-hot Bag-of-Words text representation

#	Words			
1	1	0	0	1 ...
2	0	1	0	1 ...
3	0	0	0	0 ...
4	1	1	0	1 ...
5	0	0	1	0 ...
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Main problem while processing is high dimensionality

Dimensionality reduction

For One-hot Bag-of-Words representation, to reduce noise and dimensionality are used such methods as

- Stop-words
- Lemmatization

Dataset Description

UCI machine learning library. Dataset that consists of 20000 messages taken from 20 newsgroups. Four newsgroups are taken for the classification

- atheism
- science (electronics)
- science (medicine)
- politics (guns)

For binary classification are used "atheism" and "science (electronics)" categories. Each category has 1000 documents.

Dimensionality reduction

Amount of columns in Bag-of-Words with respect to the
dimensionality reduction methods

Usual Bag-of-words	Bag-of-words with stop-words	Bag-of-words with lemmatization	Bag-of-words with stop-words and lemmatization
20177	20039	15714	15590

Supervised Classification Methods

In this work are used following classification methods

- ▶ Neural Networks
- ▶ Support Vector Machines
- ▶ Naive Bayes

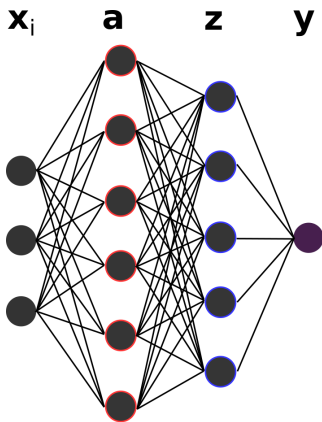# Classification Methods: Feed Forward Neural Network

Definition
Two Layers Feed Forward Neural Network is defined with following equations

$$a_k = \mathbf{w}_{1,k}^T \mathbf{x}_i + b_k,$$
$$z_r = \mathbf{w}_{2,r}^T f(\mathbf{a}) + b_r,$$
$$y_i = \mathbf{w}_{3,i}^T f(\mathbf{z}) + b_i,$$

where $\mathbf{x_i}$ is an input vector, $\mathbf{w}_{lm}$ are weights, $\mathbf{b}$ is a vector of bias values, $\mathbf{a}$ and $\mathbf{z}$ are outputs of the hidden layers and $\mathbf{y}$ is a vector from an output layer. Function f is an activation function.

Classification Methods: Feed Forward Neural Network
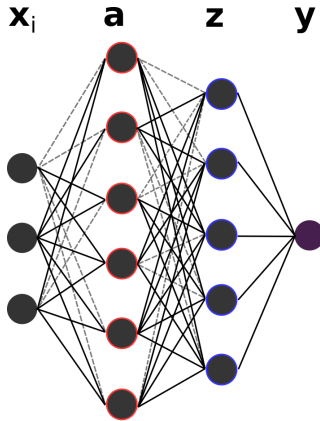
Feed Forward Neural Network Illustration

Neural Network is a universal approximator. Main disadvantage that it is easy to overfit.
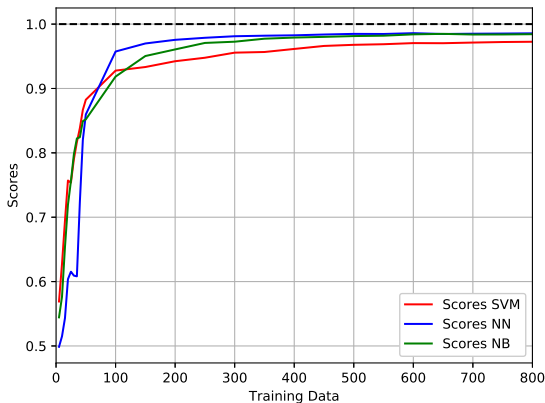
Classification Methods: Feed Forward Neural Network
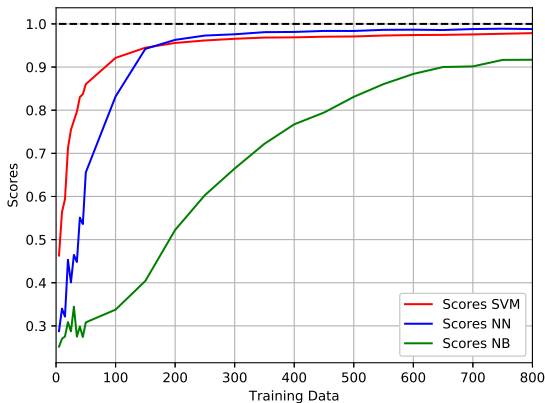
Illustration of the Dropout method

Supervised Binary Classification

Binary classification scores

# Supervised Multiclass Classification

Multiclass classification scores

Active semi-supervised learning

- ▶ Aim is to teach Neural Network to ask for label on low number of samples
- ▶ The task is to choose documents where we expect high information gain
- ▶ Decision theory requires integration over the space of parameters
  - ▶ Bayesian estimation of parameters

Bayesian Neural Network
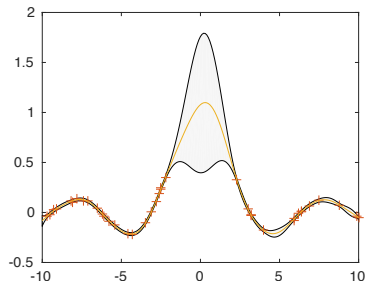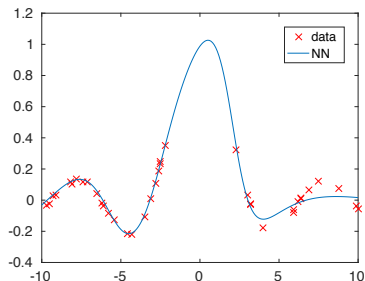
Possibilities:

- Bayesian stochastic gradient descent
- Hamiltonian Monte Carlo

Thank you for attention