Identifying Depression with Machine Learning

Ivan Lovrenčić

University of Zagreb, Faculty of Electrical Engineering and Computing

ivan.lovrencic@fer.hr

Identifying Depression on Twitter(2016)
Authors: M. Nadeem, M. Horn, G. Coppersmith, et al.

Overview

Motivation

- More people than ever being diagnosed with a mental disorder
- Increase of young people with MDD (major depressive disorder)
- Mental disorders (specifically, depression) detection has not changed for nearly 50 years
- Early depression detection through user's posts on Twitter

Background

- Rich bodies of work on depression have been performed within the psychiatry, psychology, medicine, and sociolinguistic fields
- **Shallow ML algorithms + text features**
- More recent approaches involve deep learning and topic modeling
- Goal: To detect and predict Major Depressive Disorder (MDD) and other mental illnesses

Pros

1. Clarity and interpretable nature
2. Don't demand a lot of data preparation
3. Not computationally costly

Cons

1. Easy to overfit
2. Don't perform adequately on more complicated tasks

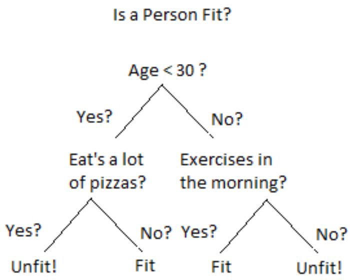

Figure: An example of simple decision tree

Logistic regression vs Linear regression

Logistic regression

- Generalized linear model
- Sigmoid function
- Probabilistic output
- Logistic loss
- Iterative optimization

Linear regression

- Generalized linear model
- Identity function
- No probabilistic output
- Quadratic loss
- Closed-form solution

Logistic regression vs Linear regression

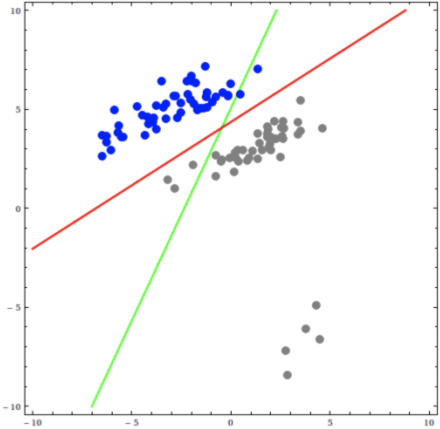

Figure: Logistic regression (red line) vs Linear regression (green line)

Logistic regression vs Linear regression

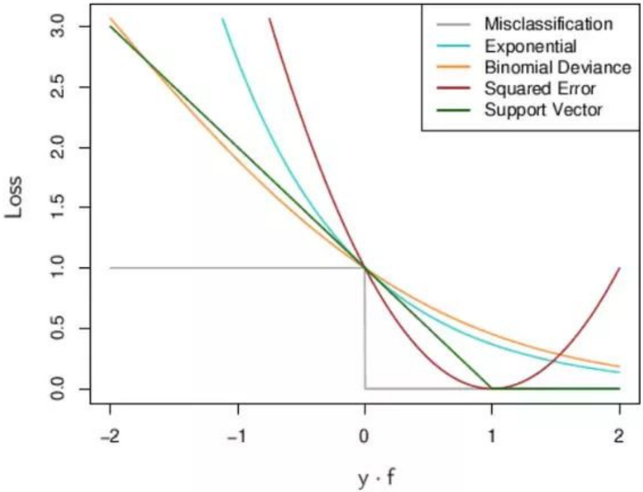

Figure: Comparison of different loss functions

Support vector machine

- Maximum margin hyperplane
- Generalization power
- Dual vs primary formulation
- Hard margin vs soft margin
- Hinge loss function

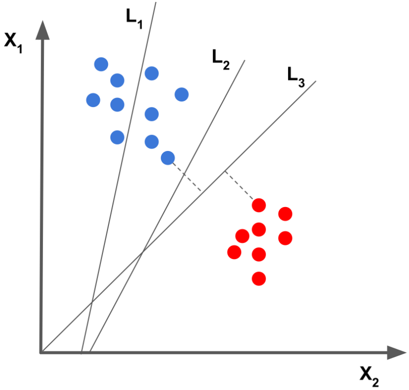

Figure: L3 model has maximum margin hyperplane

Support vector machine

Dual vs primary formulation

$h(x) = w^T x + w_0$ = primary formulation
$h(x) = \sum_{i=1}^{N} \alpha_i y^i \mathbf{x}^T \mathbf{x}^i + w_0$ = dual formulation

Support vector machine

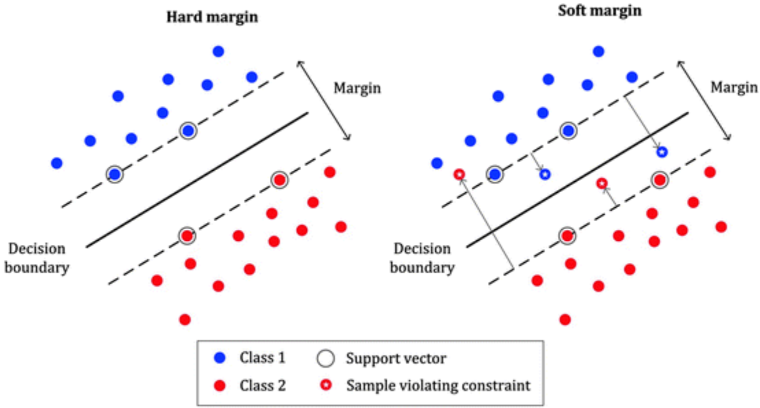

Figure: Difference between hard-margin (left) and soft-margin (right) model.

Support vector machine

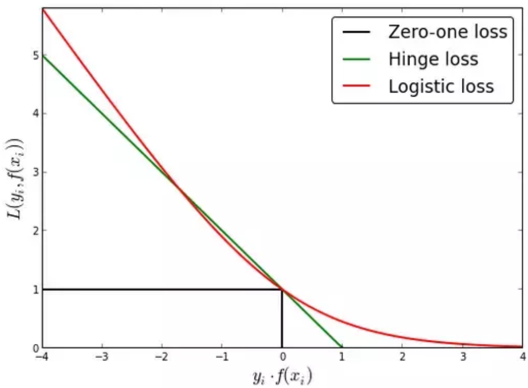

Figure: Hinge loss vs logistic loss function

Naive Bayes

- Generative probabilistic model
- Bayes theorem
- Probabilistic output
- Parameter estimation - MLE or MAP

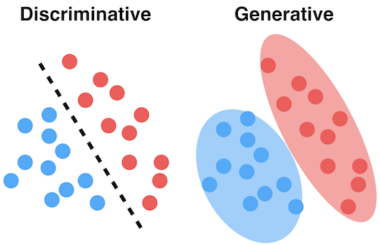

Figure: Discriminative vs Generative model

Naive Bayes

Theorem (Bayes theorem)

$P(y|x) = \frac{p(x|y)*P(y)}{P(x)}$

- $P(y|x)$ - the probability of label y for an example x (a posteriori)
- $P(y)$ - the probability of label y (a priori)
- $P(x)$ - the probability distribution of the examples
- $P(x|y)$ - the probability of example x if there is a label y (likelihood)

Naive Bayes

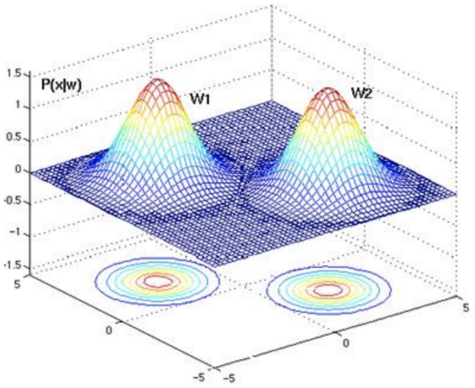

Figure: Binary classification with normal distributions for class likelihood

Dataset

CLPsych 2015 Twitter dataset		
Mental condition	Number of users	Number of tweets
No condition	574	1,253,594
Depression	426	742,560
Total	**1000**	**2,000,000**

Table: Distribution of the CLPsych 2015 dataset

Dataset

Reddit Self-Reported Depression Diagnosis (RSDD) dataset		
Data split	Number of users	Number of posts
Train	486	295,509
Validation	206	118,937
Test	200	117,899
Total	**892**	**532,345**

Table: Distribution of the RSDD dataset based on data split

Reddit Self-Reported Depression Diagnosis (RSDD) dataset		
Mental condition	Number of users	Number of posts
No condition	755	450864
Depression	137	81761
Total	**892**	**532,345**

Table: Distribution of the RSDD dataset based on mental condition

Dataset

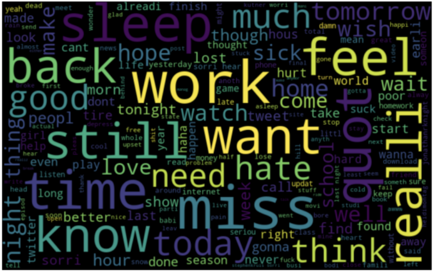

Figure: Most common words among depressive users in Twitter dataset

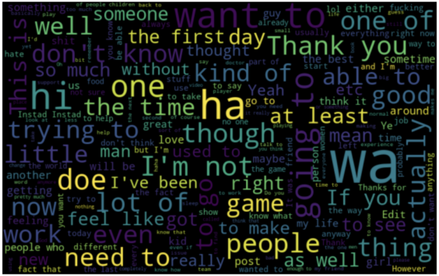

Figure: Most common words among depressive users in Reddit dataset

Model

- Five binary classifiers (shallow ML algorithm + bag-of-words/tf-idf)
- Sckit-learn libraries + Python3
- RSDD dataset

Results

The Johns Hopkins paper results				
Algorithm	Precision	Recall	F1-score	Accuracy
Decision Trees	0.67	0.68	0.75	0.67
LinearSVC	0.83	**0.83**	0.83	0.82
Naive Bayes	0.81	0.82	0.81	**0.86**
Logistic	**0.86**	0.82	**0.84**	0.82
Ridge Classifier	0.81	0.79	0.78	0.79

Table: The evaluation of the Johns Hopkins models

Results

The reimplementation results				
Algorithm	Precision	Recall	F1-score	Accuracy
Decision Trees	0.62	0.62	0.62	0.62
LinearSVC	0.60	0.59	0.58	0.59
Naive Bayes	0.69	0.68	0.68	0.68
Logistic	**0.71**	**0.70**	**0.70**	**0.70**
Ridge Classifier	0.67	0.67	0.67	0.67

Table: The evaluation of the reimplemented models

Conclusion

- Bag of words + shallow ML performs worse but still adequately on different text format and domain
- Add more task specific features
- Explore with deep learning

The End