Assignment 8,9 + Text Classification Basics

(SNLP Tutorial 8)

Vilém Zouhar

22nd, 24th June

Overview

- Decision Trees
- Naïve Bayes
- kNN
- Perceptron
- SVM
- Homework

Classification

TODO

Decision Trees

TODO

Naïve Bayes

TODO

kNN

TODO

SVM

- Find a boundary that maximizes the distance to closest vectors
- If not possible, find one that minimizes the error
- Add the kernel trick

Perceptron

- Binary classification
- Linear boundary in feature space
- $\hat{y} = \text{sign}(wx + b)$

Perceptron

- Binary classification
- Linear boundary in feature space
- $\hat{y} = \text{sign}(wx + b)$

Algorithm:

- $w_0 = \overrightarrow{0}$
- For every data point x_i
  - $\hat{y}_i = \text{sign}(w_k x_i + b)$
  - if $\hat{y}_i \neq y_i$:
    - $w_{k+1} = w_k - \hat{y}_i \cdot x$
  - else:
    - $w_{k+1} = w_k$

Perceptron

- Binary classification
- Linear boundary in feature space
- $\hat{y} = \text{sign}(wx + b)$

Algorithm:

- $w_0 = \overrightarrow{0}$
- For every data point x_i
- ▸ $\hat{y}_i = \text{sign}(w_k x_i + b)$
- ▸ if $\hat{y}_i \neq y_i$:
- ▸ ⋆ $w_{k+1} = w_k - \hat{y}_i \cdot x$
- ▸ else:
- ▸ ⋆ $w_{k+1} = w_k$
- TODO: illustration
- TODO: advantages/disadvantages

Resources

1. UdS SNLP Class, WSD: https://teaching.lsv.uni-saarland.de/snlp/