Project-II by Group LasVegas

Igor Kulev
EPFL

Andrii Maksai
EPFL

Marjan Shahpaski
EPFL

{igor.kulev, andrii.maksai, marjan.shahpaski}@epfl.ch

Abstract

This report provides a summary of our work done for the second project of the PCML course. The project consists of solving two problems. In the first problem we need to build a system that can recognize whether a person is present in an image or not. In the second problem we need to build a music recommendation system that can predict the number of times a particular user will listen to a particular song.

1 Introduction

The second project of the Pattern Classification and Machine Learning course focuses on applying machine learning techniques to real-world data. Two different datasets were therefore provided, one for each of the problems. The ultimate goal is to train a model which will be able to produce accurate response predictions to unseen input data. We have applied different techniques for both problems in order to produce relavant predictions. The tasks involved in solving these problems are data transformation and applying different models to fit our data. For the person detection problem we have used two different dimensionality reduction techniques: PCA and t-SNE, and few different classifiers: sparse logistic regression, neural networks and Support Vector Machines (SVM). For the music recommendation problem we have built a baseline model and we have substracted the baseline predictions from the true values. After that we have used two different types of methods for generating predictions: neighbour-based methods and matrix factorization methods.

2 Music recommendation problem

We can put few sentences here what is the main challenge for this problem.

2.1 Data Description

We can plot the log of the counts (FIGURE). We can give some statistics about the data set.

2.2 Models

ewfweffwe

2.2.1 Baseline

We can present few different baselines. We can plot the learning curve (FIGURE).

2.2.2 Matrix factorization techniques

We can explain the different models that can be used here. We can explain that if we didn't substract the baseline, we get very worse results. We can plot the test error as a function of the number of features for fixed value of lambda (FIGURE).

2.2.3 Neighbourhood-based methods

We can explain the different models that can be used here. We can explain that if we didn't substract the baseline, we get very worse results. We can plot the test error as a function of the number of features for fixed value of lambda (FIGURE).

2.3 Results

efewf

3 Person detection problem

Regression models try to establish the relationships between the input and the output variables(s). It is therefore used for predicting future outcomes for new (unobserved) data, or for interpreting the underlying connection(s) between the input and the output variables. In this project we will use linear regression models, which assume a linear relation between the inputs and the outputs. We will also use feature transforms which allow non-linear predictions.

3.1 Data Description

The training data for regression consists of $N_{tr} = 1400$ input and output data s

3.2 Data transformation

wefwe

3.2.1 PCA

fwefwe

3.2.2 t-SNE

fwefwe

3.3 Classification models

ewfweg

3.4 Results

efewf

4 Conclusion

fgfdgdfgdgdfgdfg