Task 1.1. Supervised Learning: Standard Classifier

Garoe Dorta-Perez
CM50246: Machine Learning and AI

November 11, 2014

1 Introduction

Given pictures from the world and been asked to classify them in several groups, we are faced with a problem of multi-class classification. One of the options would be to create N one-against-all binary classifiers. Using x to denote our data, ω to for the world state, and *lambda* for the probability of observing the given class.

$$Pr(\omega|\mathbf{x}) = Bern_w[\lambda] \tag{1}$$

However a better one involves using a categorical distribution to model our world. Where $\boldsymbol{\lambda}$ is a vector that contains a λ for each class.

$$Pr(\omega|\mathbf{x}) = Cat_w[\boldsymbol{\lambda}[\mathbf{x}]] \tag{2}$$

2 Mathematical derivation

As stated in the introduction, we are going to fit a Categorical probability model into our data. Denoting I as the total number of data points that we are given. Then, using Bayes' rule we have:

$$Pr(\theta|x_{1\cdots I}) = \frac{\prod_{i=1}^{I} Pr(\omega = k_n|x, \theta)Pr(\theta)}{Pr(x_{1\cdots I})} \tag{3}$$

We need N activation functions to enforce the constrains. Since we are solving for multi-class classification a logistic sigmoid function as activation will not be valid. Therefore a softmax function is used instead for each activation a_n.

$$a_n = \phi_n^T x \tag{4}$$

$$\lambda_n = softmax_n[a_1, a_2 \cdots a_N] = \frac{exp[a_n]}{\sum_{m=1}^{N} exp[a_m]} \tag{5}$$

In order to simplify the calculations we are going to minimise the log of the probability:

$$Ł = -log \sum_{i=1}^{I} P(\omega = k|\mathbf{x}, \theta) = -log \sum_{i=1}^{I} softmax[a_n] \quad (6)$$

3 Implementation

4 Results

Digits dataset, with prior 100, initial phi ones Elapsed time is 414.619122 seconds. Hits: 71.87

5 Conclusion