

知能システム学特論レポート（第hdp2班）

2016年7月14日（中間発表）

1 出席者

16344203 井上 聖也

16344216 田中 良道

16344217 津上 祐典

16344229 沈 歩偉

2 概要

2.1 Hadoopとは

2.2 Sparkとは

2.3 機械学習のテーマ

学習テーマはスパムメールの分類とした。データセットとして Spambase Data Set を用いた。このデータセットは 1813 通のスパムメールと 2788 通の非スパムメールから構成されており、すでに 57 次元のベクトルとして特徴量が抽出済みである。学習アルゴリズムとして、ロジスティック回帰、ナイーブベイズを使用した。ロジスティック回帰とは、識別関数としてシグモイド関数を用いた回帰モデルである。パラメータを決定する際には確率的勾配法や最急降下法、準ニュートン法などが挙げられる。ナイーブベイズとは、ベイズの定理を用いた分類アルゴリズムである。パラメータ推定には最尤法が用いられている。

2.4 結果と考察

3 今後の展望