The Experiment Report of Machine Learning



December 14, 2017

Grade:

Undergraduate

Student ID：

201530611890

Supervisor:

Mingkui Tan

Author:

Weizhao Li

**SUBJECT:**SOFTWARE ENGINEERING

**SCHOOL:** SCHOOL OF SOFTWARE ENGINEERING

[[1]](#footnote-2)Linear Regression, Linear Classiﬁcation andGradient Descent

Abstract—in this experiment, we compare the difference between NAG, RMSProp, AdaDelta and Adam, these four different optimization methods, under the implementations of logistic regression and linear SVM. The experiment result shows that these methods have different rates to decent, but they will reach similar point to convergence in the end.

# INTRODUCTION

Gradient descent method is an optimization algorithm, usually called the steepest descent method. The steepest descent method is one of the simplest and oldest methods for solving unconstrained optimization problems. Although it is not practical now, many effective algorithms are based on it and are improved and corrected. The steepest descent method is to use the negative gradient direction for the search direction, steepest descent method closer to the target value, the smaller the step, the slower the progress. We aim to compare NAG, RMSProp, AdaDelta and Adam. We use these four methods to update the models of logistic regression and SVM linear classification.

# METHODS AND THEORY

A. Logistic regression

In statistics, logistic regression is a regression model where the dependent variable (DV) is categorical. This article covers the case of a binary dependent variable—that is, where the output can take only two values, "0" and "1", which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick. Cases where the dependent variable has more than two outcome categories may be analyzed in multinomial logistic regression, or, if the multiple categories are ordered, in ordinal logistic regression. In the terminology of economics, logistic regression is an example of a qualitative response/discrete choice model.

In this experiment, the labels are binary. And for all sample, the log-likelihood loss function is

# Experiment

# conclusion

In this experiment, through implementing these four different optimization algorithms on logistic regression and linear SVM classification, we can summarize that different optimization algorithms have different performance on rate .RMSProp and Adam decent with a slower rate than the other two. NAG and AdaDelta have similar decent rate. About the accuracy, the four algorithms performs similarly, the deviation is less than 1%,which results in these algorithms converge to the same point. What’s more, we find that the parameters values in the algorithms are very important. In the beginning, we set the epsilon in AdaDelta as 1e-9, the loss is a straight line. Then we change the epsilon to 1e-5. The loss becomes converge.

1. [↑](#footnote-ref-2)