South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Rachel cao

Supervisor:
Qingyao Wu

Student ID：
201330610031

Grade:
Undergraduate or Graduate

December 15, 2017

I. Logistic Regression, Linear Classification and Stochastic Gradient Descent

Abstract—
I introduce the experiment that I have done. In the experiment, first I implement the logistic regression&linear SVM and four SGD algorithms(NAG,RMSProp,AdaDelta,Adam).
Secondly,I run them corroborate the theoretical application about them.
At last I do control experiment to adjust the parameter, for inding the best paraments for this two model with the givind dataset.

II. INTRODUCTION

Linear SVM and logistic regression are two basic algorithm for solving the binary classification problem.Yet ,NAG,RMSProp, AdaDelta,Adam are the well-known SGDalgorithms.I learn them this term but just studying through books is not enough.

So now I do such a experimentt including implementing them accordint to the theories, running them and ajust the parameter,for understanding them more thoroughly.

III. METHODS AND THEORY

The main formula that use in the experiment:

1.LinearSVM:

Loss:

$$\min_{\mathbf{w},b} \frac{\|\mathbf{w}\|^2}{2} + \frac{C}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(\mathbf{w}^\top$$

Gradient:

$$\nabla_{\mathbf{w}} L(\mathbf{w}, b) = \mathbf{w} + \frac{C}{n}\sum_{i=1}^{n} g_{\mathbf{w}}(\mathbf{x}_i)$$

$$\nabla_{b} L(\mathbf{w}, b) = \frac{C}{n}\sum_{i=1}^{n} g_{b}(\mathbf{x}_i)$$

$$g_{\mathbf{w}}(\mathbf{x}_i) = \begin{cases} -y_i\mathbf{x}_i & 1 - y_i(\mathbf{w}^\top\mathbf{x}_i + b) >= 0 \\ 0 & 1 - y_i(\mathbf{w}^\top\mathbf{x}_i + b) < 0 \end{cases}$$

$$g_{b}(\mathbf{x}_i) = \begin{cases} -y_i & 1 - y_i(\mathbf{w}^\top\mathbf{x}_i + b) >= 0 \\ 0 & 1 - y_i(\mathbf{w}^\top\mathbf{x}_i + b) < 0 \end{cases}$$

2.Logistic regression:

Loss(with regularization):

$$J(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + e^{-y_i \cdot \mathbf{w}^\top\mathbf{x}_i}) + \frac{\lambda}{2}||\mathbf{w}||_2^2$$

Gradient:

$$-\frac{1}{n}\sum_{i=1}^{n} \frac{y_i\mathbf{x}_i}{1 + e^{y_i \cdot \mathbf{w}^\top\mathbf{x}_i}} + \lambda\,\mathbf{w}$$

3.NAG,RMSProp,AdaDelta,Adam:

NAG:

$$\begin{aligned} \mathbf{g}_t &\leftarrow \nabla J(\boldsymbol{\theta}_{t-1} - \gamma\mathbf{v}_{t-1}) \\ \mathbf{v}_t &\leftarrow \gamma\mathbf{v}_{t-1} + \eta\mathbf{g}_t \\ \boldsymbol{\theta}_t &\leftarrow \boldsymbol{\theta}_{t-1} - \mathbf{v}_t \end{aligned}$$

RMSProp:

$$\begin{aligned} \mathbf{g}_t &\leftarrow \nabla J(\boldsymbol{\theta}_{t-1}) \\ G_t &\leftarrow \gamma G_t + (1-\gamma)\mathbf{g}_t \odot \mathbf{g}_t \\ \boldsymbol{\theta}_t &\leftarrow \boldsymbol{\theta}_{t-1} - \frac{\eta}{\sqrt{G_t + \epsilon}} \odot \mathbf{g}_t \end{aligned}$$

AdaDelta:

$$\begin{aligned} \mathbf{g}_t &\leftarrow \nabla J(\boldsymbol{\theta}_{t-1}) \\ G_t &\leftarrow \gamma G_t + (1-\gamma)\mathbf{g}_t \odot \mathbf{g}_t \\ \Delta\boldsymbol{\theta}_t &\leftarrow -\frac{\sqrt{\Delta_{t-1} + \epsilon}}{\sqrt{G_t + \epsilon}} \odot \mathbf{g}_t \\ \boldsymbol{\theta}_t &\leftarrow \boldsymbol{\theta}_{t-1} + \Delta\boldsymbol{\theta}_t \\ \Delta_t &\leftarrow \gamma\Delta_{t-1} + (1-\gamma)\Delta\boldsymbol{\theta}_t \odot \Delta\boldsymbol{\theta}_t \end{aligned}$$

Adam:

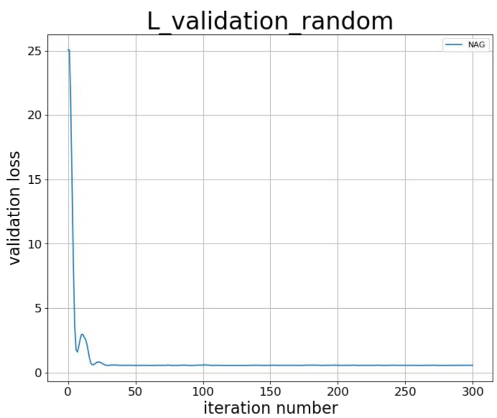
$$\mathbf{g}_t \leftarrow \nabla J(\boldsymbol{\theta}_{t-1})$$
$$\mathbf{m}_t \leftarrow \beta_1 \mathbf{m}_{t-1} + (1 -$$
$$G_t \leftarrow \gamma G_t + (1-\gamma)\mathbf{g}$$
$$\alpha \leftarrow \eta \frac{\sqrt{1-\gamma^t}}{1-\beta^t}$$
$$\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t-1} - \alpha \frac{\mathbf{m}_t}{\sqrt{G_t +}}$$

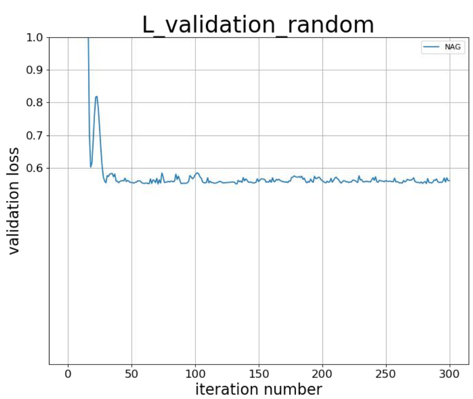
4. I use python to implement it and test. Because numpy and sklearn etc are very convenient.

IV. EXPERIMENT

1. Try to learn and compare the linear SVM and logistic regression

1.a the linear SVM

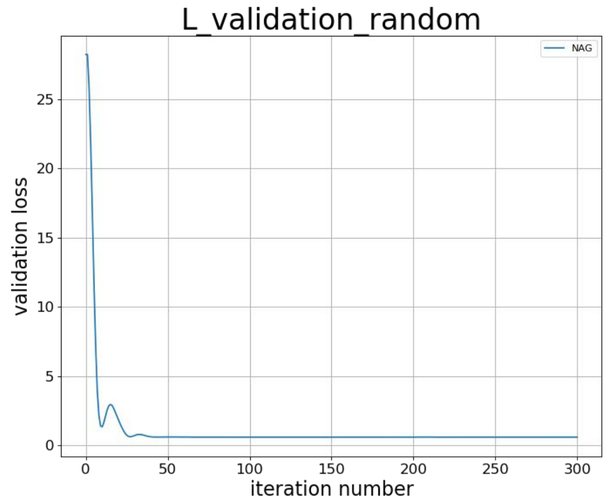


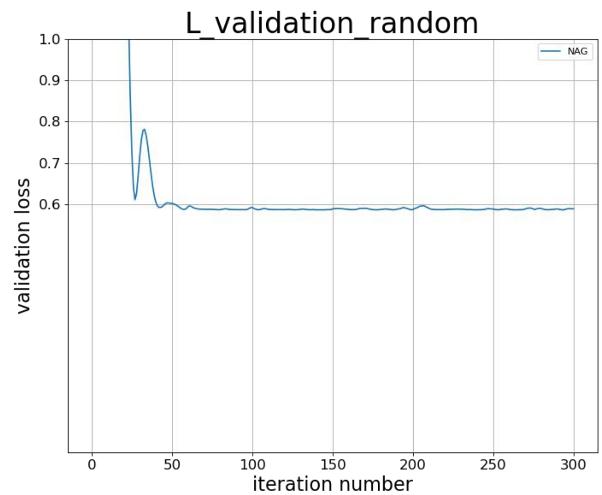


1.b logistic regression

First I try it with no regulation and the final loss is about 0.33

Then I try it with regulation:





It is seen that their final loss is 0.6. So,as I think ,the margin of the linear SVM makes some effect as the regulation.

2.Find the best parameter

C = 1,

The paramenter of regulation = 1

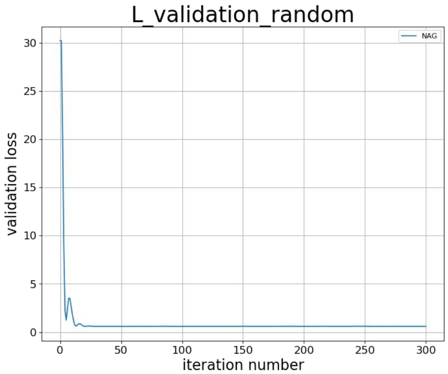
And I use the random to init

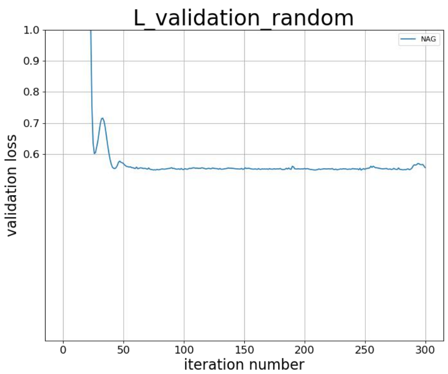
I think loss <= 0.6 is good enough

2.1 Logistic regression

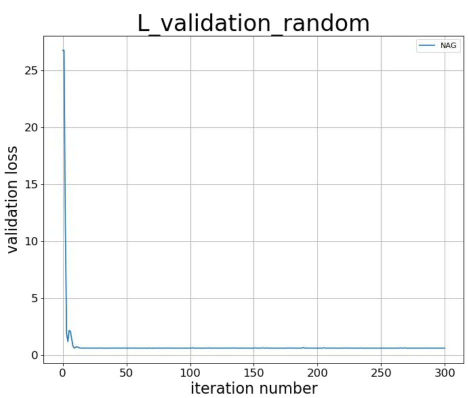
2.1.a NAG:

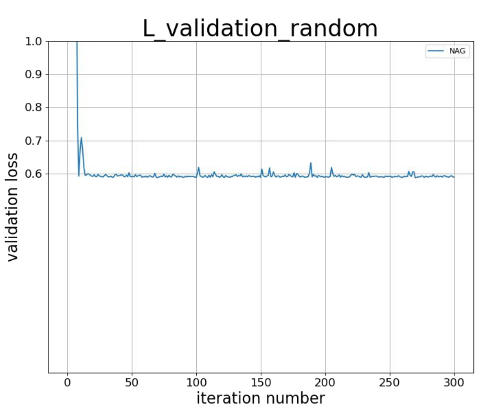
lr=0.16:



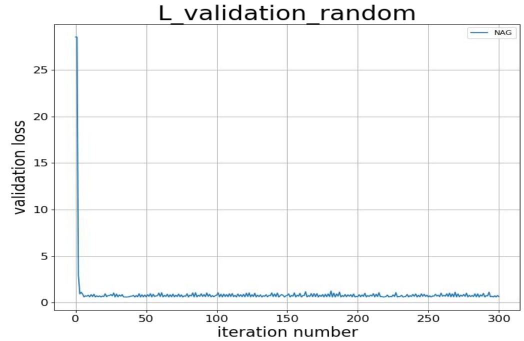


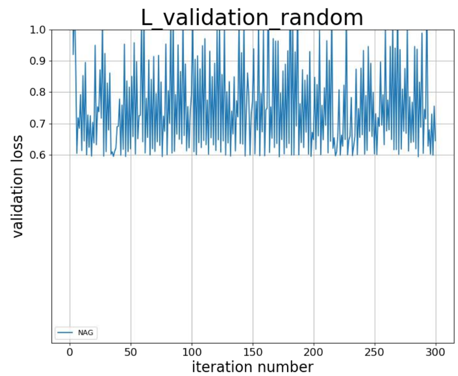
lr=0.32:





lr=0.64:

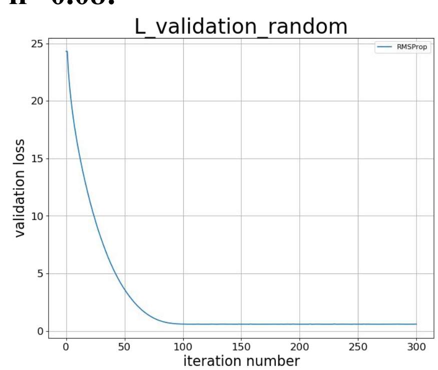


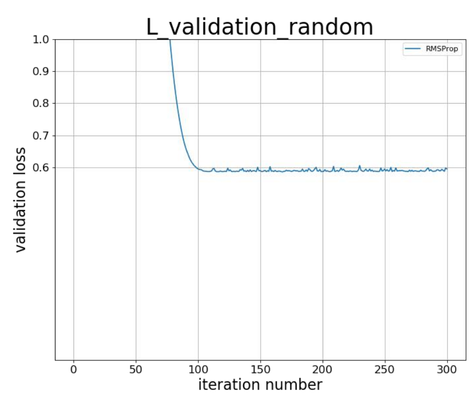


I think lr = 0.16 is good for NAG

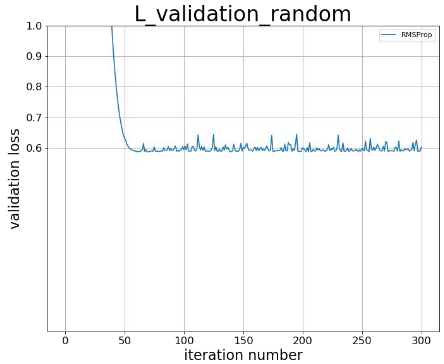
2.1.b RMSProp:

lr=0.08:





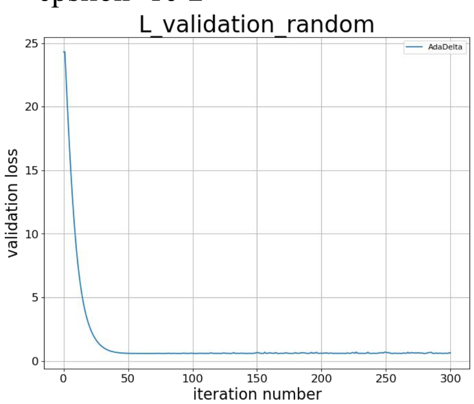
lr=0.16:

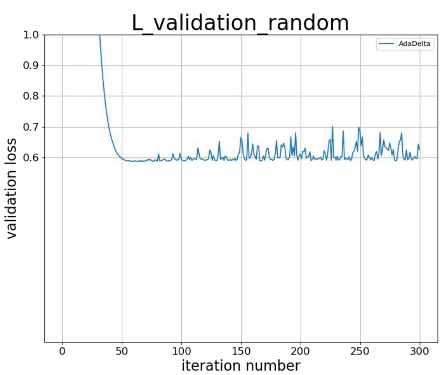


I think lr = 0.08 is good for NAG

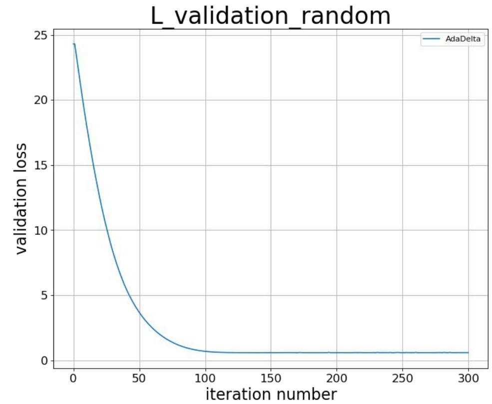
2.1.c AdaDelta:

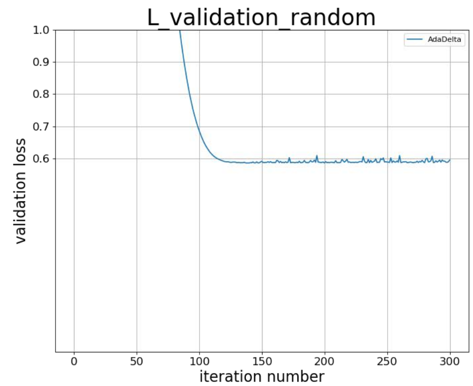
epsilon=1e-2





epsilon=1e-3:

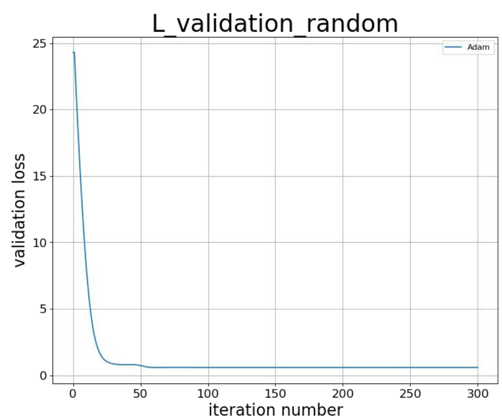


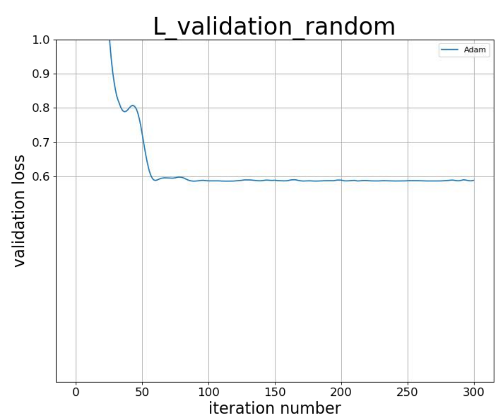


I think epsilon = 1e-3 is good for AdaDelta

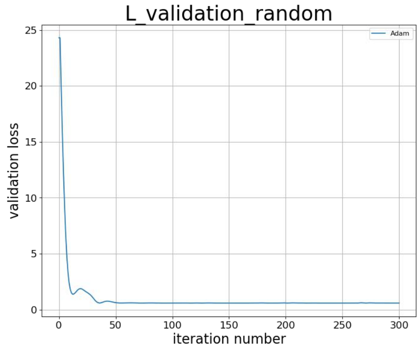
2.1.d Adam:

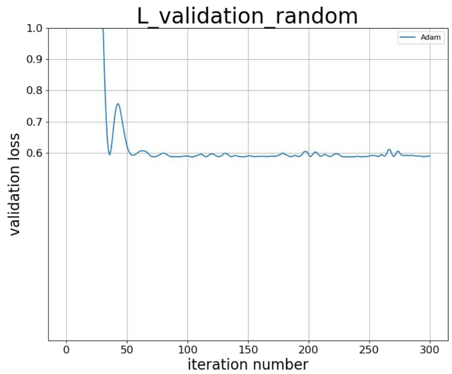
lr=0.32



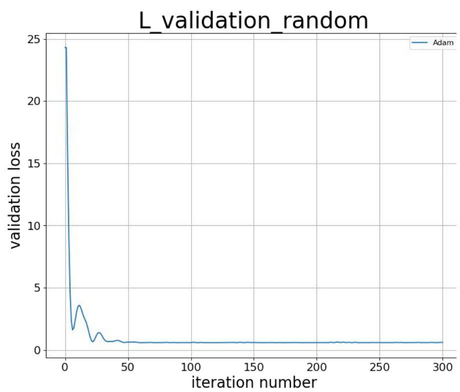


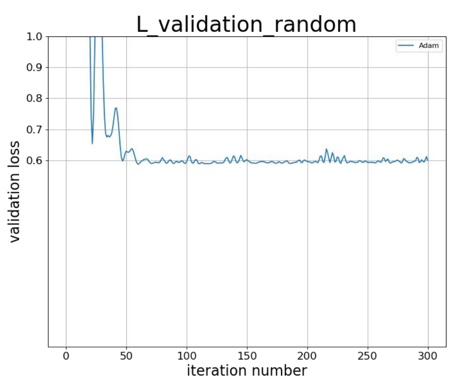
lr=0.64:





lr=1.28:

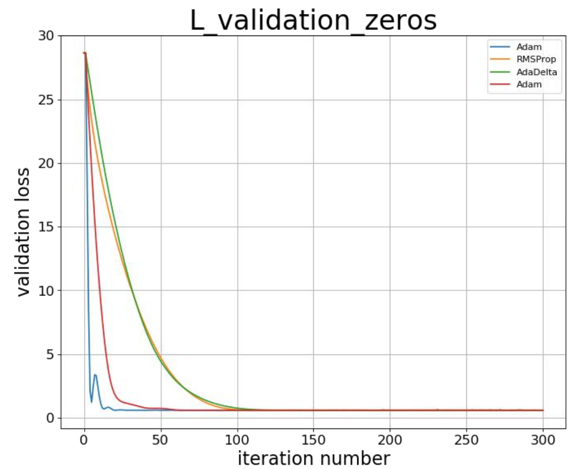


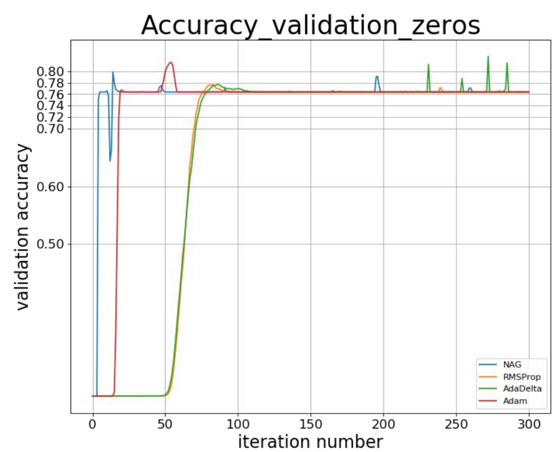


I think lr = 0.32 or 0.64 is good for Adam, I pick the 0.32

2.1.e their comparison after ajusting:

the good parameter
for linear SVM:
NAG: lr = 0.16
RMSProp: lr = 0.08
AdaDelta: epsilon = 1e-3
Adam : lr = 0.32

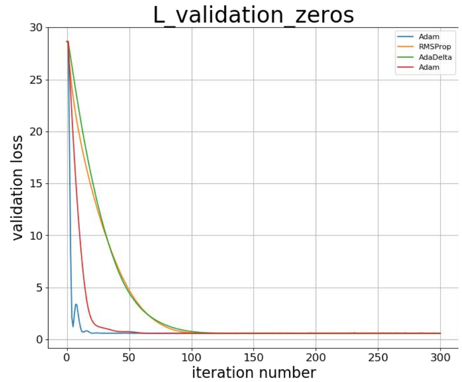


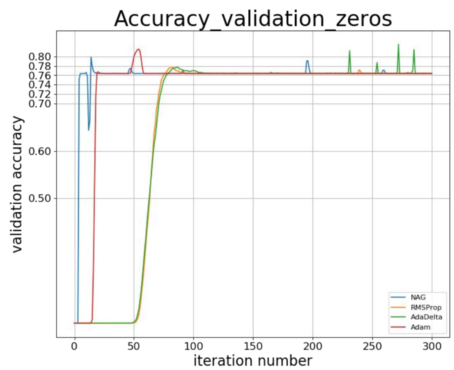


2.2Linear SVM:

In a similar way that I do on the Logistic regression:

I get the good parameter
for linear SVM:
NAG: lr = 0.08
RMSProp: lr = 0.08
AdaDelta: epsilon = 1e-3
Adam : lr = 0.64





V. CONCLUSION

Through the experiment I have very harvest. Via implementing the algorithm by myself, I have a deep understanding of them really.

And I have a experient ofadjusting the parameter. I understand the importance of adjusting too.