[CMPUT 466/566, Winter 2020] Introduction to Machine learning

Coding Assignment 2

Due: March 23 (11:59pm Edmonton time)

Prepared by the instructor: Lili Mou
LMOU@ualberta.ca
https://lili-mou.github.io/
with TA: Zehra

Submission Information

The student should submit a zip file containing a pdf report and all the code which should replicate the results.

Problem 1 [70%]

Download: Codebase

In this coding problem, we will implement the softmax regression for multi-class classification using the MNIST dataset.

Dataset

First, download the datasets from the link above. You need to unzip the .gz file by either double clicking or some command like `gunzip -k file.gz`

The dataset contains 60K training samples, and 10K test samples. Again, we split 10K from the training samples for validation. In other words, we have 50K training samples, 10K validation samples, and 10K test samples. The target label is among {0, 1, …, 9}.

Algorithm

We will implement stochastic gradient descent (SGD) for cross-entroy loss of softmax as the learning algorithm. The measure of success will be the accuracy (i.e., the fraction of correct predictions).

The general framework for this coding assignment is the same as SGD for linear regression, so you may re-use most of the code. However, you shall change the computation of output, the loss function, the measure of success, and the gradient whenever needed.

Implementation trick

For softmax classification, you may encounter numerical overflow if you just follow the equation mentioned in the lecture.

$$z_i = \exp\{\boldsymbol{w}_i^\top \boldsymbol{x} + b_i\}$$

$$y_i = \frac{\exp\{z_i\}}{\sum_j \exp\{z_j\}} = \frac{\exp\{\boldsymbol{w}_i^\top \boldsymbol{x} + b_i\}}{\sum_j \exp\{\boldsymbol{w}_j^\top \boldsymbol{x} + b_j\}}$$

The observation is that the exp function increases very fast with its input, and very soon exp(z) will give NAN (not a number).

The trick is to subtract every z_i by the maximum value $z_1, \cdots, z_K$.

In other words, we compute

$\tilde{z}_i = z_i - z_{\max}$, where $z_{\max} = \max_{k=1}^{K} z_k$, and then we have

$$y_i = \frac{\exp\{\tilde{z}_i\}}{\sum_j \exp\{\tilde{z}_j\}}$$

Note that the gradient is computed with y, and since subtracting a constant before softmax doesn't affect y, it doesn't affect the gradient either.

Problem 2 [30%]

Ask one meaningful scientific question yourself, design your experimental protocol, present results, and draw a conclusion.

Note:

A *scientific question* means that we can give a verifiable hypothesis that can be either confirmed or declined.

> Example of a scientific question: Is the learned classifier for this dataset better than majority guess? Your hypothesis could be either yes or no, and it can be verified by experiments.
> Example of a non-scientific question: Does the learned classifier become better if I have super-power? My hypothesis could be either yes or no, but cannot be verified it by any experiment. I don't know what superpower is, and I can say yes, or I can also say no. Neither is wrong, nor even correct.

A *meaningful* scientific question means that you'll learn something from the experiment. Of course, what is meaningful itself is subjective. In terms of this coding assignment, the scientific question is considered meaningful as long as the student would learn something, or verify some results we mentioned in lectures.

> Example of a meaningful scientific question: How does linear regression perform for classification for this dataset? Doing this experiment will give us first-hand experience on why we shall not use regression models to do classification. But you **cannot** ask this question as the solution. You need to ask your own scientific question that interests you and/or inspires others.
>
> Example of a not-so-meaningful scientific question: Is the learned classifier for this dataset better than majority guess? Ok, this question is considered scientific, but too trivial for us, although it is not necessarily trivial for those who don't know machine learning at all. [Again this shows the subjectivity of evaluating the significance of science.]