Machine Learning 2021

Homework 2

Due date : 2021/12/3 23:59:59

1 Classification Problem

You are given a dataset of handwritten character digits (EMNIST.zip) derived from the EMNIST dataset. This dataset contains 8 classes with 128 different images in each class. Supervised learning is performed for training data $\{\mathbf{x}_n, \mathbf{y}_n = \{y_{nk}\}\}$. In this exercise, you need to implement the following classifiers

(1) least squares for classification

(2) logistic regression model for classification

Note: You need to normalize the data samples before training and randomly select 32 images as test data for each class and the remaining images as training data.

1. Implement the least squares for classification. You should use a 1-of-K binary coding scheme for the target vector $\mathbf{t}$. **Show** the classification accuracy and loss value of training and test data.

2. Implement the logistic regression model using batch GD (batch gradient descent), SGD (stochastic gradient descent) and mini-batch SGD with softmax activation. Set the initial weight vector $\mathbf{w}_k = [w_{k1}, \ldots, w_{kF}]$ to be a zero vector where F is the number of features and K is the number of classes.

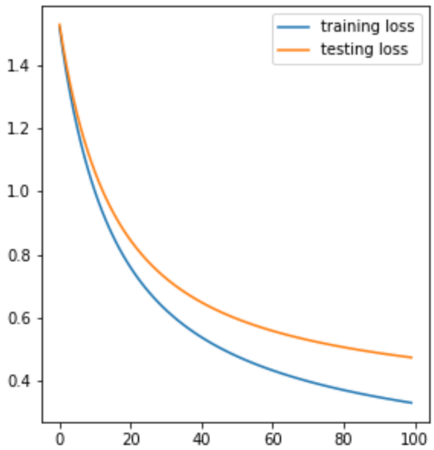
Algorithms	Batch size	No. of iterations in each epoch
batch GD	N	1
SGD	1	N
mini-batch SGD	B	N/B

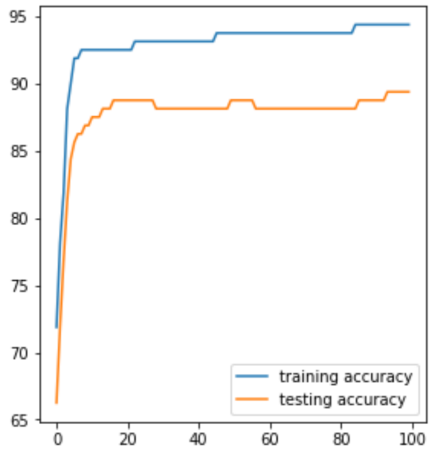
N is the number of training data. B is the batch size.

The error function is defined as

$$E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \log y_{nk}$$

(a) **Plot** the learning curves of the loss function and the accuracy of classification versus the number of epochs until convergence for training data as well as test data, e.g.





(b) **Show** the final classification accuracy and loss value of training and testing data.

(c) Based on your observation about different algorithms (batch GD, SGD and mini-batch SGD), please **make some discussion**.

3. **Make some discussion** about the difference between the results of 1 and 2. From these results, should we use the least squares model for classification problem? why or why not?

2 Gaussian Process for Regression

In this exercise, please implement Gaussian process (GP) for regression. The file x.csv and t.csv have the input data $\mathbf{x} : \{x_1, x_2, \ldots, x_{300}\}, 0 < x_i < 1$ and the corresponding target data $\mathbf{t} : \{t_1, t_2, \ldots, t_{300}\}$. Please take the first 150 points as the training set and the rest as the test set. A regression function $y(\cdot)$ is used to express the target value by

$$t_n = y(x_n) + \epsilon_n$$

where the noisy signal ϵ_n is Gaussian distributed, $\epsilon_n \sim \mathcal{N}(0, \beta^{-1})$ with $\beta^{-1} = 1$.

1. Please construct a kernel function using the basis functions in a form of polynomial model

$$\phi(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^{D} w_i x_i + \sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij} x_i x_j \quad (M = 2)$$

   and implement the Gaussian process for regression.

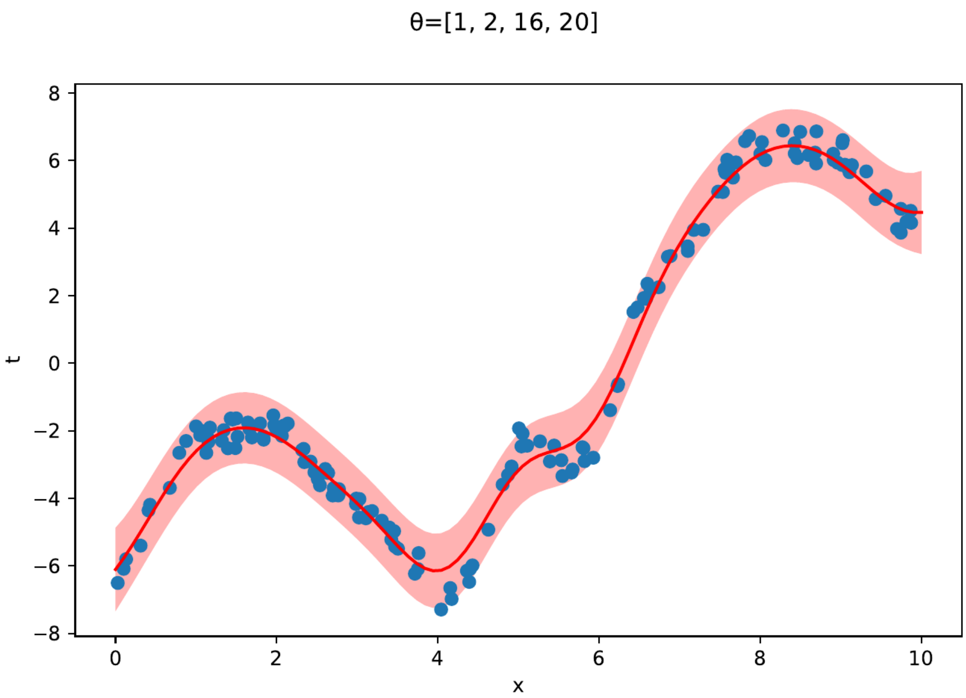
2. Repeat 1 by using the widely used exponential-quadratic kernel function given by

$$k(\mathbf{x}_n, \mathbf{x}_m) = \theta_0 \exp\left\{-\frac{\theta_1}{2} \|\mathbf{x}_n - \mathbf{x}_m\|^2\right\} + \theta_2 + \theta_3 \mathbf{x}_n^\top \mathbf{x}_m$$

   where the hyperparameters $\boldsymbol{\theta} = \{\theta_0, \theta_1, \theta_2, \theta_3\}$ are fixed. Please use the training set with four different combinations:

   - linear kernel $\boldsymbol{\theta} = \{0, 0, 0, 1\}$
   - squared exponential kernel $\boldsymbol{\theta} = \{1, 8, 0, 0\}$
   - exponential-quadratic kernel $\boldsymbol{\theta} = \{1, 1, 0, 16\}$
   - exponential-quadratic kernel $\boldsymbol{\theta} = \{1, 1, 32, 0\}$

3. Please **plot** the prediction results in 1 and 2 like Figure 6.8 of textbook for training set but one standard deviation instead of two is shown and without the need of showing green curve. The **title of the figure** in 2 should be the value of the hyperparameters used in the models. The red line shows the mean $m(\cdot)$ of the Gaussian process predictive distribution. The pink region corresponds to plus and minus one standard deviation. Training data points are shown in blue. An example is provided in below.



4. **Show** the corresponding root-mean-square errors

$$E_{\text{RMS}} = \sqrt{\frac{1}{N}\left(m\left(x_n\right) - t_n\right)^2}$$

for both training and test sets with respect to different kernels in 1 and 2.

5. Try to **tune the hyperparameters** $\boldsymbol{\theta}$ in 2 by yourself to find the best combination for the dataset. You can tune the hyperparameters by trial and error or use Automatic relevance determination (ARD) in Section of 6.4.4 of textbook.

6. Explain your findings and **make some discussion**.

3 Rules

- Please name the assignment as hw2_StudentID.zip (e.g. hw1_0123456.zip).
- In your submission, it needs to contain three files.
  - **.ipynb** file which contains all the results and codes for this homework.
  - **.py** file which is downloaded from the .ipynb file
  - **.pdf** file which is the report that contains your description for this homework.
- Implementation will be graded by
  - Completeness
  - Algorithm Correctness
  - Model description
  - Discussion
- Only Python implementation is acceptable.
- Only the packages we provided are acceptable.
- DO NOT PLAGIARIZE. (We will check program similarity score.)