OPTIMIZATION. HOMEWORK 5

OSCAR DALMAU

(1) Maximize the function $h(\boldsymbol{\beta}, \beta_0)$ defined below with respect to β, β_0 using the steepest descent algorithm

$$
\begin{aligned}
h(\boldsymbol{\beta}, \beta_0) &= \sum_{i=1}^{n} y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i) \\
\pi_i := \pi_i(\boldsymbol{\beta}, \beta_0) &= \frac{1}{1 + \exp\left(-\boldsymbol{x}_i^T \boldsymbol{\beta} - \beta_0\right)}
\end{aligned}
$$

where $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$ is obtained from the train_set, see details below, and $\boldsymbol{x}_i \in \mathbb{R}^{784}$ and and $y_i \in \{0, 1\}$.

(2) Using $\hat{\boldsymbol{\beta}}, \hat{\beta}_0$, computed in the previous item, compute the error

$$
error = \frac{1}{n} \sum_{i=1}^{n} |\mathbf{1}_{\pi_i(\hat{\boldsymbol{\beta}}, \hat{\beta}_0) > 0.5}(\boldsymbol{x}_i) - y_i|
$$

where $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$ is obtained from the test_set, see details below, and $\boldsymbol{x}_i \in \mathbb{R}^{784}$ and and $y_i \in \{0, 1\}$.

(3) The train_set and test_set, i.e. the sets $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, can be obtained from the dataset mnist.pkl.gz, which is provided in the homework. The dataset mnist.pkl.gz can be read using

```
import cPickle, gzip, numpy
# Load the dataset
f = gzip.open('mnist.pkl.gz', 'rb')
train_set, valid_set, test_set = cPickle.load(f)
f.close()
```

- train_set[0] is a matrix of size (50000, 784) where $n = 50000$ is the number of observations and each row represents an observation $\boldsymbol{x}_i \in \mathbb{R}^{784}$
- train_set[1] is a vector of size (50000) where each entry $y_i \in \{0, 1, \cdots, 9\}$

- test_set[0] is a matrix of size (10000, 784) where $n = 10000$ is the number of observations and each row represents an observation $\boldsymbol{x}_i \in \mathbb{R}^{784}$
- test_set[1] is a vector of size (10000) where each entry $y_i \in \{0, 1, \cdots, 9\}$

- Select from train_set[0] and train_set[1] the set of observations $\{(\boldsymbol{x}_i, y_i)\}$ with $\boldsymbol{x}_i \in \mathbb{R}^{784}$ and $y_i \in \{0, 1\}$ and estimate the parameters $\hat{\boldsymbol{\beta}}, \hat{\beta}_0$

(4) **Note:**

- Each row $\boldsymbol{x}_i$ of train_set[0] and/or test_set[0] can be shown as an image as follows:

```
import matplotlib.pyplot as plt
idx = 1 # index of the image
im = train_set[0][idx].reshape(28, -1)
plt.imshow(im, cmap=plt.cm.gray)
print('Label: ', train_set[1][idx])
```

- The equations in item (1) correspond to the log-likelihood of the logistic regression model for two-class classification.
- You can find more information available on line at: http://yann.lecun.com/exdb/mnist/