Neural Networks: Assignment 1

(Due: 2025.4.29 11:59 pm)

Formatting: please include both a **.ipynb** and **.pdf** file in your homework submission, named **studentID_homework1.ipynb** and **studentID_homework1.pdf**. Please don't change the filename.

1. Logistic Regression for Binary Classification

In multiclass classification, we typically use the exponential model

$$p(y|\mathbf{o}) = \mathrm{softmax}(\mathbf{o})_y = \frac{\exp(o_y)}{\sum_{y'} \exp(o_{y'})}.$$

1.1 For binary classification, i.e. whenever we have only two classes $\{-1, 1\}$, we can arbitrarily set $o_{-1} = 0$. Using the shorthand $o = o_1$, show that this is equivalent to

$$p(y = 1|o) = \frac{1}{1 + \exp(-o)}.$$

1.2 Show that the log-likelihood loss (often called logistic loss) for labels $y \in \{-1, 1\}$ is thus given by

$$-\log p(y|o) = \log(1 + \exp(-y \cdot o)).$$

Now, we will model covariate shift and attempt to fix it using logistic regression. This is a fairly realistic scenario for data scientists. To keep things well under control and understandable we will use Fashion-MNIST as the data to experiment on.

2. Logistic Regression

2.1 Implement the logistic function $l(y, f) = -\log(1 + \exp(-yf))$.

2.2 Plot its values and its derivative for $y = 1$ and $f \in [-5, 5]$, using automatic differentiation.

2.3 Generate training and test datasets for a binary classification problem using Fashion-MNIST with class 1 being a combination of 'sneaker' and 'pullover' and class -1 being the combination of 'sandal' and 'shirt' categories.

2.4 Train a binary classifier of your choice (it can be linear or a simple MLP such as from a previous lecture) using half the data (i.e. $12,000$ observations mixed as above) and one using the full dataset (i.e. $24,000$ observations as arising from the 4 categories) and report its accuracy.

Hint - you should encapsulate the training and reporting code in a callable function since you'll need it quite a bit in the following.

3. Covariate Shift

Your goal is to introduce covariate shit in the data and observe the accuracy. For this, compose a dataset of 12,000 observations, given by a mixture of sneaker and pullover and of sandal and shirt respectively, where you use a fraction $\lambda \in \{0.05, 0.1, 0.2, \ldots 0.8, 0.9, 0.95\}$ of one and a fraction of $1 - \lambda$ of the other datasets respectively. For instance, you might pick for $\lambda = 0.1$ a total of 600 sneaker and 5,400 pullover images and likewise 600 sandal and 5,400 shirt photos, yielding a total of 12,000 images for training. Note that the test set remains unbiased, composed of 2,000 photos for the sneaker + pullover category and of the sandal + shirt category each.

Generate training sets that are appropriately biased. You should have 11 datasets. Train a binary classifier using this and report the test set accuracy on the unbiased test set.

4. Covariate Shift Correction

Having observed that covariate shift can be harmful, let's try fixing it. For this we first need to compute the appropriate propensity scores $\frac{dp(x)}{dq(x)}$. For this purpose pick a biased dataset, let's say with $\lambda = 0.1$ and try to fix the covariate shift.

4.1 When training a logistic regression binary classifier to fix covariate shift, we assumed so far that both sets are of equal size. Show that re-weighting data in training and test set appropriately can help address the issue when both datasets have different size. What is the weighting?

4.2 Train a binary classifier (using logistic regression) distinguishing between the biased training set and the unbiased test set. Note - you need to weigh the data.

4.3 Use the scores to compute weights on the training set. Do they match the weight arising from the biasing distribution λ?

4.4 Train a binary classifier of the covariate shifted problem using the weights obtained previously and report the accuracy. Note - you will need to modify the training loop slightly such that you can compute the gradient of a weighted sum of losses.