05-04-2022

# Assignment 3
2019051

Q5. (a) Activation function : $\sigma(x) = \frac{1}{1+e^{-x}}$

$\because \quad 0 < \sigma(x) < 1$

If $\sigma(x) > 0.5$ : belongs to +1
$\sigma(x) < 0.5$ : " " -1

Taking the loss as the threshold function, if $x$ is misclassified then the loss is $L = -y(\beta^T x + \beta_0)$, where $y$ is the true label, $\beta$ and $\beta_0$ are the parameters.

Applying gradient descent

$$\frac{\partial L}{\partial \beta} = -yx \quad ; \quad \frac{\partial L}{\partial \beta_0} = -1$$

$$\Rightarrow \beta_{new} = \beta_{old} - \eta(-xy)$$

$$\Rightarrow \beta_{0\,new} = \beta_{0\,old} - \eta(-1)$$

$\eta$ : the learning rate.

Since, the loss is same, we get the same update rule.

(b) $\phi(\beta, \beta_0) = -\sum_{i=1}^{N} y_i(\beta^T x_i + \beta_0)$

given that $\beta^T\beta = 1$

From lagrange multiplication for $\beta$

$$\frac{\partial}{\partial \beta}\left[\phi(\beta,\beta_0) - \lambda_1(\beta^T\beta - 1)\right] = 0$$

$$\Rightarrow -\sum_{i=1}^{N} y_i x_i - 2\lambda_1\beta = 0$$

$$\cancel{\Rightarrow 2\lambda_1 = \sum_{i=1}^{N} y_i x_i / \beta}$$

$$\Rightarrow \beta = \sum_{i=1}^{N} y_i x_i \Big/ -2\lambda_1$$

$\because \beta^T\beta = 1$

$$\Rightarrow \sum_{i=1}^{N}\left(\frac{y_i x_i}{-2\lambda_1}\right)^T \cdot \left(\sum_{i=1}^{N}\frac{y_i x_i}{-2\lambda_1}\right) = 1$$

$$\Rightarrow \sum_{i=1}^{N} y_i^2 x_i^T x_i = 4\lambda_1^2$$

$$\Rightarrow \lambda_1 = \frac{\sqrt{\sum_{i=1}^{N} y_i^2 \, x_i^T x}}{2}$$

$$\Rightarrow \beta = \frac{\sum_{i=1}^{N} y_i x_i}{-2\lambda}$$

Ans.

where ; $\lambda = \sqrt{\sum_{i=1}^{N} (y_i x_i)^T (y_i x_i)}$

$\phi'$ : lagrange function, as defined above

$$\frac{\partial \phi'}{\partial \beta_0} = -\sum_{i=1}^{N} y_i - 0$$

Thus,

$$\frac{\partial \phi'}{\partial \beta_1} = -\sum_{i=1}^{N} (y_i x_i) - 2\lambda\beta$$

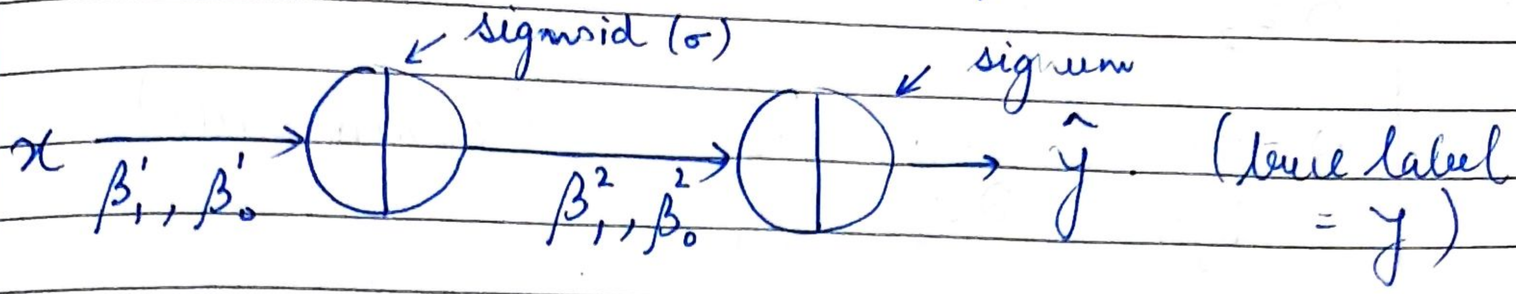
and $\frac{\partial \phi'}{\partial \beta_0} = -\sum_{i=1}^{N} y_i$

Given update rule:

$$\beta_{1_{new}} = \beta_{1_{old}} - \eta \frac{\partial \phi'}{\partial \beta_{1_{old}}}$$

$$\beta_{0_{new}} = \beta_{0_{old}} - \eta \frac{\partial \phi'}{\partial \beta_0}$$

86. The neural network that forms here:




Output from first perceptron = $\sigma(\beta_1^1 x + \beta_0^1)$

Let $z_1 = \beta_1^1 x + \beta_0^1$ ; $\cancel{z_2}$

$a_1 = \sigma(z_1)$ : activation from 1st perceptron

Now, $z_2 = \beta_1^2 x + \beta_0^2$

$\Rightarrow a_2 = \sigma(z_2)$ : activation from 2nd perceptron

Final loss = $-y z_2$

Now, we need to find all the parameters to train the model

$$\frac{\partial L}{\partial z_2} = -y$$

$$\Rightarrow \frac{\partial L}{\partial \beta_1^2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial \beta_1^2} \quad \text{(Chain rule)}$$

$$\because \boxed{\frac{\partial L}{\partial \beta_1^2} = -y a_1 \text{ and } \frac{\partial L}{\partial \beta_0^2} = -y}$$

Now, $$\frac{\partial L}{\partial \beta_1^1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial \beta_1^1} \quad \text{(Chain rule)}$$

$$\frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \quad \text{(Back propagation)}$$

$$\Rightarrow \frac{\partial L}{\partial a_1} = \frac{\partial L}{\partial z_2} \frac{\partial z_2}{\partial a_1} = -y\beta_1^2$$

$$\Rightarrow \frac{\partial L}{\partial z_1} = -y\beta_1^2 \sigma(z_1)(1-\sigma(z_1))$$

$$\frac{\partial L}{\partial z_1} = -y\beta_1^2 (1-a_1) a_1$$

Hence, $$\boxed{\frac{\partial L}{\partial \beta_1^1} = -y\beta_1^2 a_1 (1-a_1) x}$$

$$\beta^1_{1\,\cancel{old}\,new} = \beta^1_{1\,old} - \eta \frac{\partial L}{\partial \beta_1}$$

$$\boxed{\frac{\partial L}{\partial \beta_0^1} = -y\beta_1^2\, a_1(1-a_1)}$$

$$\beta^1_{0\,new} = \beta^1_{0\,old} - \frac{\partial L}{\partial \beta_0^1}\cdot\eta$$

For the second perceptions the parameters are as follows:

$$\beta^2_{1\,new} = \beta^2_{1\,\cancel{new}\,old} - \frac{\partial L}{\partial \beta_1^2}\,\eta$$

$$\beta^2_{0\,new} = \beta^2_{0\,old} - \frac{\partial L}{\partial \beta_0^2}\,\eta$$

# Assignment 3

## Kesar Shrivastava

## 2019051

### Question 1

- Unpickle the data files.
- Since the data is of images, reshape the data to fit into the image category.
- Then visualize the images using the cv2 library.
- I have visualized five images of each class from each batch of the training set.
- To apply LDA, I extend the dataset using all the training batches into one training dataset, and similarly for the testing dataset.
- I use the LDA from sklearn library.
- After the model is trained, I use it to predict the test set and the accuracy comes out to be 0.3713

### Question 2

- There are two datasets: the training dataset and the testing dataset.
- I create two readers, one for the image dataset and one for the test dataset which is the label dataset.
- After that, I flatten the dataset to feed into PCA.
- PCA returns the data with less and desired dimensions.
- Then we apply LDA.
- We see that as the number of components increases the accuracy increases because the loss of information decreases as we increase the number of components.
- Hence, **n_components = 15 gives best accuracy**.

### Question 3

- Since here the data is already in CSV format, I read it directly into data frames.
- To apply FDA, I group the data points according to their classes such that $X_i$ : d X $c_i$; $c_i$ is the number of data points in class i.
- Now, I make the final X = [$X_1$ $X_2$ … $X_c$]

- The Scatter function gives the scatter matrix of any matrix that is fed in it.
- For FDA, I calculate the within scatter and between scatter. W is the eigenvalues of $S_w^{-1}.S_b$
- Sort the eigenvectors in descending order of eigenvalues and take the first c-1 vectors to modify the data.
- Then project the data as $w^T x$
- After feeding the data into the LDA, the accuracy is reduced a little due to loss of information.

**Question 4**

- After reading the file as done before, I apply PCA to bring it down to 15 dimensions.
- Then apply the same steps as described in question 3.
- Apply LDA on the dataset.
- The accuracy is 0.7959