Deep Learning Homework 1

Rohit Sah

1 Data Preprocessing

1.a

train_valid_split splits dataset into 2 sets out of which 1 will be used for training the model and the other for validation of the model. Validation dataset is used to provide an unbiased evaluation of the model fit on the training dataset while tuning model hyperparameters.

1.b

It is correct to retrain the model with whole training set.After training it is always better to train with more data.

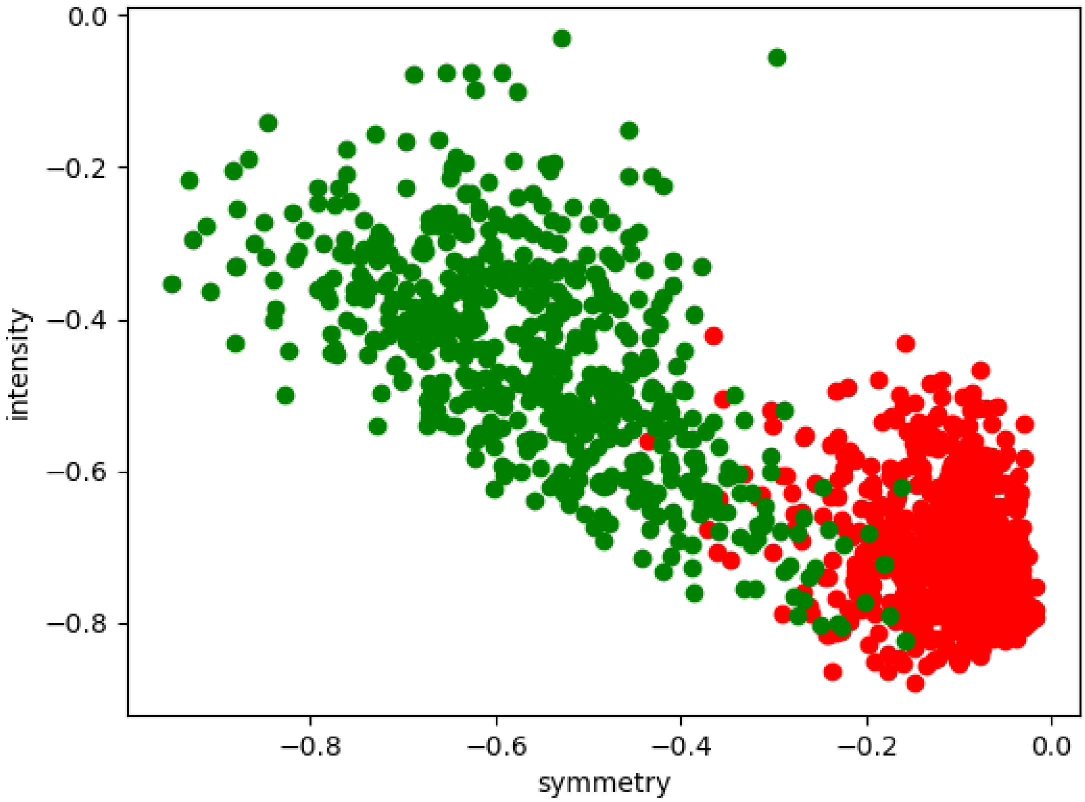
1.d

This 1 denotes w_0 in weights vector which corresponds to the bias constant which is added to w_i^Tx.

1.f

Test results on test data:

Below is the scatterplot graph where x-axis corresponds to the *symmetry* and y-axis corresponds to the *intensity*. The dots in *Green* color represents class -1 and *Red* color represents class 1



2 Cross-entropy loss

2.a

Loss Function:

$$E(w) = \frac{1}{N}\sum_{n=1}^{N} ln(1 + e^{-y_n \cdot w^T x})$$

For one training sample (x_n,y_n), $\quad E(w) = \frac{1}{N}\sum_{n=1}^{N} e(w, x_n, y_n)$

2.b

Gradient:

$$
\begin{aligned}
\nabla E(w) &= \frac{\delta E}{\delta w} \\
&= \frac{\delta E}{\delta u} \cdot \frac{\delta u}{\delta w} && u = 1 + e^{-y_n \cdot w^T x} \\
&= \frac{1}{N * (1 + e^{-y_n \cdot w^T x})} * \sum_{n=1}^{N} \frac{\delta u}{\delta \theta} \cdot \frac{\delta \theta}{\delta w} && \theta = -y_n w^T x \\
&= \frac{1}{N * (1 + e^{-y_n \cdot w^T x})} * \sum_{n=1}^{N} -yX * e^{-y_n w^T x} \\
&= \frac{-yX * e^{-y_n w^T x}}{(1 + e^{-y_n \cdot w^T x})}
\end{aligned} \tag{1}
$$

2.c

we use the Sigmoid function when we need to map predicted values to probabilities. The result can be improved if there are not too many data points near the decision boundary. Picking a decision boundary means the amount of certainty for the prediction. For example, if $\theta(\mathrm{w}^T x)$=0.6 then we can say that there is an 60% probability that y_i=1 and a 40% probability that y_i=0.

2.d

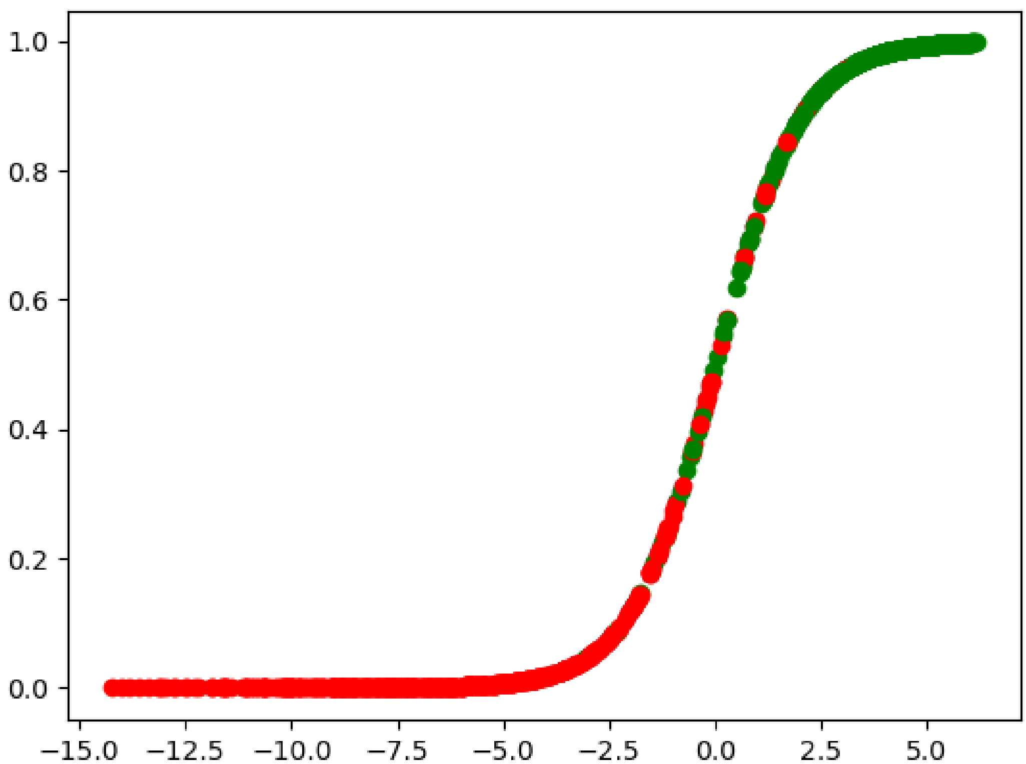
A decision boundary is linear as long as there exists a function H(x) = 0 + T x such that h(x) = I(H(x) 0)$.The further a point is from the decision boundary, the more certain we are about the decision.$

2.e

essential property: A decision boundary is linear as long as there exists a function H(x) = 0 + T x such that h(x) = I(H(x) 0).

3 Sigmoid logistic regression

3.d



3.e

weights: [2.7209728 19.35966376 -4.77828619]

accuracy: 0.9437229437229437

4 Softmax logistic regression

4.d

4.e

weights: [[8.62652205 0.96767364 12.6229412]
[-0.76716833 18.96164337 -8.84047454]
[-5.67709009 -17.78058276 -1.99515445]]

accuracy: 0.8635809987819733

5 Softmax logistic vs Sigmoid logistic

5.a

5.b

References