Assignment-2

1) Given binary label as $y_i \in \{-1, +1\}$. To prove that the empirical loss function on a data sample takes the form of:-

$$l(w; x_i, y_i) = \log(1 + \exp(-y_i w^T x_i))$$

→ Representing logistic regression probability model:-

Logistic regression models the probability of having input x_i and belonging to class $y_i \in \{-1, +1\}$ as:-

$$P(y_i|x_i) = \frac{1}{1 + \exp(-y_i w^T x_i)}$$, here w is the weight vector.

→ Using log-likelihood function:-

The likelihood of observing the entire dataset $\{(x_i, y_i)\}$ where $i = 1$ to n under the given model is: $L(w) = \prod_{i=1}^{n} P(y_i|x_i)$

logarithm of log-likelihood is given as:- $\log(L(w))$

$$\Rightarrow \log(L(w)) = \sum_{i=1}^{n} \log(P(y_i|x_i))$$

Substituting $P(y_i|x_i)$:

$$\log(L(w)) = \sum_{i=1}^{n} \log\left(\frac{1}{1 + \exp(-y_i w^T x_i)}\right)$$

$$= \sum_{i=1}^{n} -\log(1 + \exp(-y_i w^T x_i))$$

→ Negative log-likelihood as loss function:-

From the above step if we define empirical loss function; we minimize the negative log-likelihood by the below statement:-

$$l(w; x_i, y_i) = \log(1 + \exp(-y_i w^T x_i))$$

Thus, we have shown that the empirical loss function for logistic regression would be in the form as:-

$$l(w; x_i, y_i) = \log(1 + \exp(-y_i w^T x_i))$$

2) The objective function of logistic regression with L2-regularization is also called as Ridge regression.

Given sensitive dataset $\{x_i \mid i \in [1,n]\}$, where binary label is represented as $y_i \in \{-1,+1\}$, here empirical loss for single data point is given by:-

$$l(w; x_i, y_i) = \log(1+\exp(-y_i w^T x_i))$$

→ Consider Average Empirical loss:-

The total empirical loss over n samples is deduced as:-

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} \log(1+\exp(-y_i w^T x_i))$$

→ Introducing Regularization term:-

To avoid overfitting, we introduce an L2-norm regularization term, which is given by :- $\frac{\lambda}{2} \|w\|^2 = \frac{\lambda}{2} \sum_{j=1}^{d} w_j^2$; here $\lambda > 0$ which is a regularization parameter, this controls the overfitting of data and maintains models complexity low.

After introducing regularization term, the objective function is given by

$$J(w) = \frac{1}{n} \sum_{i=1}^{n} \log(1+\exp(-y_i w^T x_i)) + \frac{\lambda}{2} \sum_{j=1}^{d} w_j^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \log(1+\exp(-y_i w^T x_i)) + \frac{\lambda}{2} \|w\|^2$$

Therefore, the objective function of logistic regression, considering average loss of data samples and to use L2-norm to avoid overfitting is given by:-

$$J(w) = \frac{1}{n} \sum_{i=1}^{n} \log(1+\exp(-y_i w^T x_i)) + \frac{\lambda}{2} \|w\|^2$$