ECE-363 Assignment 2

Lavanya Verma (lavanya18155@iiitd.ac.in)

April 25, 2021

1. Theory Problem 1

Let P(z) be True distribution and Q(z) be the estimated distribution.
Here Z are the outputs, z(s) are elements of Z.

$$
\begin{aligned}
H(P) &= -\sum_{z\in Z} P(z)\log(P(z)) \text{ \textit{(Discrete Entropy)}} \\
H(P,Q) &= -\sum_{z\in Z} P(z)\log(Q(z)) \text{ \textit{(Discrete Entropy)}} \\
D_{KL}(P||Q) &= \sum_{z\in Z} P(z)\log(\frac{P(z)}{Q(z)}) \text{ \textit{(Discrete KL divergence)}} \\
D_{KL}(P||Q) &= \sum_{z\in Z} P(z)\log(\frac{P(z)}{Q(z)}) \\
D_{KL}(P||Q) &= -\sum_{z\in Z} P(z)\log(Q(z)) + \sum_{z\in Z} P(z)\log(P(z)) \\
D_{KL}(P||Q) &= H(P,Q) - H(P)
\end{aligned}
$$

If P is represented as one hot vector(multi-class classification) for every Z, then Entropy term H(P) would be zero, since P(z) would have no uncertainty either 1 or 0.
In that case,

$$
\begin{aligned}
H(P) &= -\sum_{z\in\{0,1\}} P(z)\log(P(z)) \\
&= 0 \\
D_{KL}(P||Q) &= H(P,Q)
\end{aligned}
$$

However, if the labels are zero/ones not necessary one hot vector(multi-label classification), entropy would still be zero. As entropy considers the input to be a probability distribution, therefore it has to be normalized.

2. Theory Problem 2

Here, for brevity the activation functions of neurons N_1, N_2 and N_3 are denoted by their identified symbol i.e., $N_1()$, $N_2()$ and $N_3()$ respectively.
Output

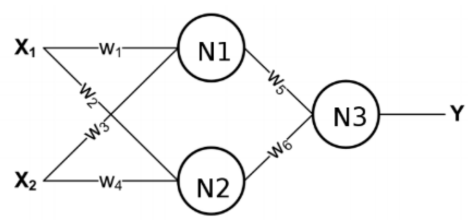
$$Y = N_3(W_5 N_1(X_1W_1 + X_2W_3) + W_6 N_2(X_1W_2 + X_2W_4))$$

Since, the output of NN is output of binary logistic linear classifier, that forces the $N_1($ and $N_2()$ to be linear function of their inputs. As per the choices available, $N_1(), N_2()$ would be A_2

$$A_2 : L(x) = Cx, \text{C is some arbitrary constant}$$

Following the same argument, $N_3()$ should be A_1.

$$A_1 : S(x) = sign\left[\sigma(x) - 0.5\right] = sign\left[\frac{1}{1+\exp(-x)} - 0.5\right]$$



$$\begin{aligned} Z &= W_5C(X_1W_1 + X_2W_3) + W_6C(X_1W_2 + X_2W_4) \\ &= X_1C(W_1W_5 + W_2W_6) + X_2C(W_3W_5 + W_4W_6) \\ Z' &= \sigma(Z);\ (\mathrm{P(Y = 1\ |X)}) \end{aligned}$$

Upon comparison with logistic classifier expression

$$\begin{aligned} \beta_1 &= C(W_1W_5 + W_2W_6) \\ \beta_2 &= C(W_3W_5 + W_4W_6) \end{aligned}$$

References

- PP 33 Lecture 5(Click Here)

3. Theory Problem 3

A single perceptron can for linearly separable data. It can represent **AND**, **OR**, **NOT** gate. A D-Dimensional boolean can be represented as canonical sum of products(C.S.O.P.). In cases where C.S.O.P. contains term more than 2^{d-1}, their expression be minimized to fewer terms.
So, The first layer(**INPUT**) of multi-layer perceptron network will contain the inputs(d features), where $X_i \in \{0,1\} \forall 1 \leq i \leq d$. Second layer(**HIDDEN 1**) will contain the computation (done by the connections and activation(threshold)) i.e., minterms. Third Layer(**OUTPUT**) will output the linear combination of minterms, hence the boolean function.

References

- Lecture Slides of DL Course(Click Here)