CIE6032 and MDS6232: Homework #1

Due on Sunday, September 30, 2018, 5:30pm (DYB 224)

Problem 1

[25 points]

Cross entropy is often used as the objective function when training neural networks in classification problems. Suppose the training set includes N training pairs $\mathcal{D} = \{(\mathbf{x}_i^{(\text{train})}, y_i^{(\text{train})})\}_{i=1}^N$, where $\mathbf{x}_i^{(\text{train})}$ is a training sample and $y_i^{(\text{train})} \in \{1, \ldots, c\}$ is its class label. $\mathbf{z}_i$ is the output of the network given input $\mathbf{x}_i^{(\text{train})}$ and the nonlinearity of the output layer is softmax. $\mathbf{z}_i$ is a c dimensional vector, $z_{i,k} \in [0, 1]$ and $\sum_{k=1}^{c} z_{i,k} = 1$. Please (1) write the objective function of cross entropy with softmax activation function, and the gradient of hidden-to-output weights (Hits: two conditions of softmax) (20 points) (2) it is equivalent to the negative log-likelihood on the training set, assuming the training samples are independent. (5 points)

Problem 2

[**25 points**]

x_1 and x_2 are two input variables, and y is the target variable to be predicted. The network structure is shown in Figure 1(a). $h_{11} = f_{11}(x_1)$, $h_{12} = f_{12}(x_2)$, and $y = g(h_{11}, h_{12})$.

- Assuming $x_1 \in [0, 1]$ and $x_2 \in [0, 1]$, in order to obtain the decision regions in Figure 1(b), decide functions f_{11}, f_{12}, and g. [**5 points**]
- Now we extend the range of x_1 and x_2 to $[0, 2]$. Please add one more layer to Figure 1(a) in order to obtain the decision regions in Figure 1(c). [**5 points**]
- Although the decision boundaries in Figure 1(c) look complicated, there exist regularity and global structure. Please explain such regularity and global structure. Based on your observation, draw the decision boundaries when the range of x_1 and x_2 are extended to [0, 4]. [**5 points**]

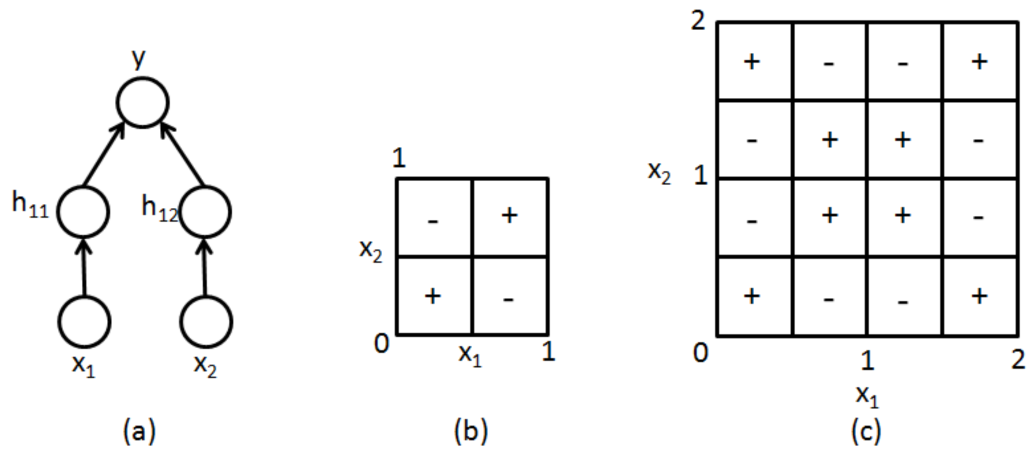

Figure 1:

- Following the question above and assuming the range of x_1 and x_2 is extended to $[0, 2^n]$, draw the network structure and the transform function in each layer, in order to obtain the decision regions with the same regularity and global structure in Figure 1 (b) and (c). The complexity of computation units should be $O(n)$. [**5 points**]
- Assuming the range of x_1 and x_2 is $[0, 2^n]$ and only one hidden layer is allowed, specify the network structure and transform functions. [**5 points**]

Problem 3

[50 points]

- Finish the Planar_MLP.md according to requirements, i.e., fill in code in required lines [**30 ponits**]
- According to the tutorial MLP.md introduced on course, please rewrite the MLP model defined from scratch in Planar_MLP.md using mxnet and powerful autograd function. [**20 ponits**]