526 UU1180 Lecture Notes Chapter 3 Supplementary Material Setting the Weights in Multilayer Perceptron

Lectruer Cheng-Yuan Liou

Department of Computer Science and Information Engineering, National Taiwan University, Republic of China

Abstract

Instead of training the weights, this note shows that the weights in Chapter 3 can be preset and used in the multilayer perceptron. It has a perfect performance for any training datasets. Its performance is comparable to that of SVM for the testing datasets. I illustrate an example in this material.

1 Designing the weights

Let the set of all patterns be $X = {\mathbf{x}^p, p = 1, ..., P}$. Each pattern \mathbf{x}^p is a *D*-dimensional column vector. The label function, $C : \mathbb{R}^D \to N$, maps each pattern, \mathbf{x}^p , to its class identity number, c_p . Suppose there are 3 hidden layers in the network, $\{m = 1, 2, ..., L\}$, refer Figure 1(c). Let n_m be the total number of neurons in the *m*th layer.

The method in Chapter 3 provids a design for the initial weights for the MLP. It is a civide-and-conquer design. I will show that a general-position twoclass classification problem can be solved perfectly with three hidden layers. This design is very different from all BP algorithms that solve the complexity, $\sum_{k=0}^{n_{m-1}} \binom{n_m}{k}$ [6], in the succeeding MLP layers. It is also a divide-and-conquer design. Figure 1(a) illustrates the network for a two-class problem, $c_1 = 1$ and $c_2 = 2$, in a two dimensional space, $D = n_0 = 2$.

In this D space, a center line of a strip, $\overline{\mathbf{x}^p \mathbf{x}^q}$, is allocated for two near patterns, \mathbf{x}^p and \mathbf{x}^q , that are in a same class c_1 , $\mathbf{x}^p \in c_1$ and $\mathbf{x}^q \in c_1$. We assume that c_1 contains fewer number of patterns than that of c_2 . Then, this center line is split into two parallel lines, line \overline{a} and line \overline{b} . They are in the two opposite sides of the center line and parallel to the center line, $\overline{a} ||\overline{\mathbf{x}^p \mathbf{x}^q}||\overline{b}$.

For \overline{a} , pick a pattern \mathbf{x}^r , $\mathbf{x}^r \in c_2$, where \mathbf{x}^r is closest to $\overline{\mathbf{x}^p \mathbf{x}^q}$. \mathbf{x}^r and \overline{a} are in the same side of $\overline{\mathbf{x}^p \mathbf{x}^q}$. (Note that this \mathbf{x}^r may not exist. The line $\overline{\mathbf{x}^p \mathbf{x}^q}$ is suffice for the strip.border.) Plot a parallel line \overline{a}^r , $\overline{a}^r || \overline{\mathbf{x}^p \mathbf{x}^q}$, that passes the pattern \mathbf{x}^r . Pick a pattern \mathbf{x}^s , $\mathbf{x}^s \in c_1$, that is in between the two lines, \overline{a}^r and $\overline{\mathbf{x}^p \mathbf{x}^q}$, and is the closest pattern to the line \overline{a}^r . Plot a parallel line, \overline{a}^s , $\overline{a}^s || \overline{\mathbf{x}^p \mathbf{x}^q}$, that passes the pattern \mathbf{x}^s . (Note that this \mathbf{x}^s may not exist. We will use the $\overline{\mathbf{x}^p \mathbf{x}^q}$ as the \overline{a}^s line.) Plot a decision border line, \overline{a}^{rs} , right in between the two parallel lines, \overline{a}^r and \overline{a}^s . \overline{a}^{rs} has wide margin between the pair ($\mathbf{x}^r, \mathbf{x}^s$). The two patterns, \mathbf{x}^r and \mathbf{x}^s , serve as the margin-limiting stops of the region in between the two lines, \overline{a}^r and \overline{a}^s .

The decision border line \overline{b}^{uv} for \overline{b} can be accomplished in a similar way on the other side of $\overline{\mathbf{x}^p \mathbf{x}^q}$. These two decision lines are used as the two neurons in the first hidden layer. All patterns in between the two lines \overline{a}^{rs} and \overline{b}^{uv} belong to the same class c_1 . These patterns will be substracted from the pattern set c_1 and will not be used for the determination of all other strips. This is a divideand-conquer scheme.

Note that the width between the two decision lines, \overline{a}^{rs} and \overline{b}^{uv} , is useful in the determination of the significance of the two neurons. Those neurons with large width are preferable and will be preserved with high priority in certain training operations. Small width neurons will be eliminated occasionally.

The patterns in between the two lines, \overline{a}^{rs} and \overline{b}^{uv} , are well isolated from the patterns in the other class c_2 . The stops \mathbf{x}^r and \mathbf{x}^s are different from the support vectors in SVM. The space in between the two parallel lines, \overline{a}^{rs} and \overline{b}^{uv} , is a sector of the D space. An example of the typical decision regions is illustrated in the Figure 1(b). The decision regions contain four bar-like strips. There exists physiological evidences on receptive fields, D = 2, for the bar-like strips, [2, 3]. Note that there are many techniques to pick the center patterns \mathbf{x}^p and \mathbf{x}^q to build a strip. One way to do this is to select all patterns, $\{\mathbf{x}^p,$ $\mathbf{x}^q, \mathbf{x}^r, \mathbf{x}^s, \mathbf{x}^u$ and $\mathbf{x}^v\}$, in a predefined neighborhood region. The size of the neighborhood region can be tuned during the training process. One may include a penality cost to set the borders \overline{a}^{rs} and \overline{b}^{uv} in a way similar to that for SVM.

As for the general-position two-class classification problem, a single 'AND'

function is used for a neuron in the second hidden layer, n_2 , to represent the patterns in one individual strip, see Figure 1(c). A global 'OR' function is used for the output neuron, n_3 , to represent all patterns in class c_1 that are in all strips. To our knowlede, this is the simplest MLP architecture in many aspects.

Alternatively, one can fix the two hidden layers, n_2 and n_3 , with 'OR' and 'AND' functions and use the BP to train the n_1 layer only. In the n_1 layer we set D + 1 neurons connected to each n_2 neuron, or each 'AND' neuron. There are $n_1 = n_2(D + 1)$ neurons in the n_1 layer. These D + 1 neurons can enclose an isolated polyhedron region (cell) and represent all patterns in that cell. All patterns in a single cell must belong to the same class after the training process. In Figure 1(c) there are $n_1 = 2n_2$ in the first hidden layer and each strip is enclosed by two parallel hyperplanes.

This article shows a divide-and-conquer design. Both the number of neurons and the number of layers in the tiling algorithm [5] are highly sensitive to the setting of the origin, the absolute coordinates, of the patterns. The relative distances between patterns are used in Chapter 3 and this material. This relative distance gives the classification quality. The network in Fig. 1(c) will give a perfect performance for any training datasets. To our knowledge, the performance of this network is comparable to that of SVM [1] for any testing datasets. Note that this network can be extended to an isolated polyhedral region that is encompassed by (D + 1) neorons. It is easy to extend the design to the multiple-class problem. It is easy to obtain all positive weights as those obtained by the NMF method.



Figure 1: The concept of the weight design method in [4].

References

- Boser, B.E., Guyon, I.M., and Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory (1992) 144–152
- [2] Daugman, J.: Two-dimensional spectral analysis of cortical receptive field profiles. Vision Research 20 (1980) 847–856
- [3] Dobbins, A., Zucker, S.W., and Cynader, M.S.: Endstopped neurons in the visual cortex as a substrate for calculating curvature. Nature 329 (1987) 438–441
- [4] Liou, C.-Y. and Yu, W.-J.: Initializing the weights in multilayer network with quadratic sigmoid function. In: Proceedings of the International Conference on Neural Information Processing (1994) 1387–1392
- [5] Mézard, M. and Nadal, J.P.: Learning in feedforward layered networks: the tiling algorithm. Journal of Physics A22 (1989) 2191–2203
- [6] Mirchandini, G. and Cao, W.: On hidden nodes in neural nets. IEEE Transaction Circuits and Systems 36 (1989) 661–664