TR-I-0116

ON INTERPRETATIONS OF A FEED-FORWARD NEURAL NETWORK

フィードフォワードニューラルネットワークの解釈

Shin'ichi Tamura

田村　震一

1989 . 10.23

Abstract

In this technical report, interpretations of a feed-forward neural network are described. For the interpretations, a network is decomposed into a successive combination of unit transformations. A unit transformation stands for a transformation from one layer output to the next higher layer output. This transformation is further divided into an affine part and a sigmoid non-linear part. The affine part is interpreted using the singular value decomposition technique and the sigmoid non-linear part is interpreted based on a classification of its input space into direction-invariant sub-space. Also described is an application of the interpretations, which is applied to an exclusive OR feed-forward neural network.

ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所

CONTENTS

ON INTERPRETATIONS OF A FEED-FORWARD NEURAL NETWORK

Shin'ichi TAMURA

ATR Interpreting Telephony Research Laboratories
Seika-cho, Souraku-gun, Kyoto, 619-02, Japan

ABSTRACT

In this paper, interpretations of a feed-forward neural network are described. For the interpretations, a network is decomposed into a successive combination of unit transformations. A unit transformation stands for a transformation from one layer output to the next higher layer output. This transformation is further divided into an affine part and a sigmoid non-linear part. The affine part is interpreted using the singular value decomposition technique and the sigmoid non-linear part is interpreted based on a classification of its input space into direction-invariant subspaces. Also described is an application of the interpretations, which is applied to an exclusive OR feed-forward neural network.

1. INTRODUCTION

Neural networks have been applied to many areas such as speech recognition, image coding, expert systems, and so forth, because of their potential capabilities[1][2]. In such applications, one of the major problems is the inability to explain the neural network internal functions. At present there are few papers which describe interpretations of neural network internal functions compared with the number of papers which describe specific applications of neural networks.

Although there have been theoretical studies concerning the capabilities of feed-forward neural networks which state that a three- or four- layered feed-forward neural network can, in principle, realize any continuous non-linear mapping[3][4][5][6][7][8][9], the proofs are based on the Kolmogorov or Stone-Weierstrass theorems, or multi-dimensional Fourier transform, and thus assume special structures on neural networks, or are existence theorems. They are difficult to use for interpretations of a feed-forward neural network with a specified number of hidden units. They are, of course, important in neural network theory, but in order to interpret a feed-forward neural network, a different approach seems to be required.

In this paper, based on the singular value decomposition technique of linear algebra and a classification of an input space of a non-linear mapping, a feed-forward neural network is interpreted and the interpretations are applied to an exclusive OR neural network.

In Section 2, the interpretations of a feed-forward neural network are described. In Section 3, an application of the interpretations to an exclusive OR network is described.

2. INTERPRETATIONS

The feed-forward neural network considered here is a network which is allowed all connections between layers. The transformation from a network input to a network output is made by successive applications of a unit transformation, which stands for a transformation from one layer output to the next higher layer output. Thus, interpreting a unit transformation is essential.

2-1. UNIT TRANSFORMATION

A unit transformation is a transformation from one layer output to the next higher layer output of a feed-forward neural network. We assume here that all the units at the next higher layer, or the output layer of a unit transformation, have a slightly modified sigmoid function. Fig. 1 shows the plot of this sigmoid function which takes values between *-1* and *1*. The gradient at $x = 0$ is normalized to *1* for simplicity. Fig. 2

This paper was presented at IJCNN89 (International Joint Conference on Neural Networks 89 sponsored by IEEE and INNS, at WASHINGTON D.C., June 18-22, 1989)

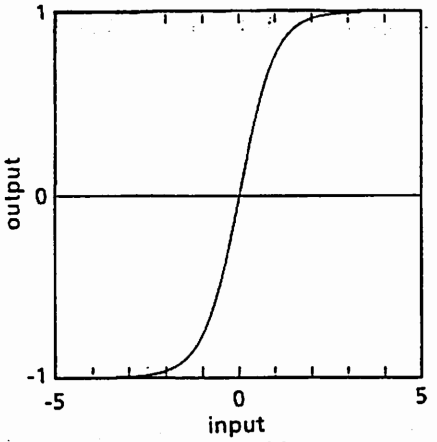

Fig. 1Sigmoid function

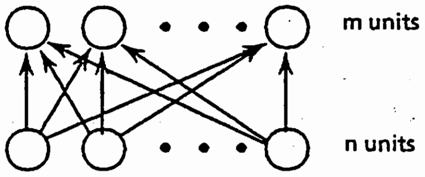

Fig. 2 Unit transformation

illustrates the unit transformation. The numbers of units at the input and output layers of the unit transformation are assumed to be n and m, respectively. Let x be an n-dimensional input vector of the unit transformation, and z an m-dimensional output vector of the unit transformation. Using matrix notation, x and z can be related as follows:

$$\mathrm{y} = A\mathrm{x} + \mathrm{b} \text{ ----- (1)} \qquad \mathrm{z} = \mathrm{f}(\mathrm{y}) \text{ ----- (2)}$$

where A is an m×n matrix of the connection weights between layers, b an m-dimensional bias vector, y an m-dimensional vector of the input to f, and f an m-dimensional sigmoid function, i.e. $\mathrm{f}_i = (1 - exp(2\mathrm{y}_i))/(1 + exp(-2\mathrm{y}_i))$ $(i = 1,2,3,\ldots,m)$.

Formula (1) expresses the affine part of the transformation and Formula (2) expresses the purely non-linear part of the transformation. An input vector of the transformation, x, is first transformed by Formula (1) and then deformed by Formula (2). Here, the case where $n \geqq m$ is considered, and a standard Euclidean distance and a standard inner product are assumed. The discussion below also holds for the case where n < m.

2-2. The affine transformation

Formula (1) stands for the affine transformation. A vector x is linearly transformed by the matrix A and next is shifted by the vector b. To see the structure of the linear transformation, A: R^n --> R^m, the singular value decomposition technique[6] can be used.

It is known that an optimal lower dimensional approximation of a finite dimensional linear mapping, T : R^s --> R^t, can be given by singular value decomposition of the mapping. H.Bourlard & Y.Kamp pointed out that the identity mapping network, or the auto-association network performs data compression by singular value decomposition[7]. In this paper, however, the singular value decomposition technique is adopted to structure the linear transformation part, A : R^n --> R^m.

Let the rank of A be m. By singular value decomposition, A is decomposed into a product of matrices and a diagonal matrix, i.e.

$$A = UDV^t \qquad (t{:}transpose)$$

where V is an $n \times m$ matrix and its column vectors $\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_m$ are orthonormal vectors of the space R^n. U is an $m \times m$ matrix and its column vectors $\mathrm{u}_1, \mathrm{u}_2, \ldots, \mathrm{u}_m$ are orthonormal vectors of the space R^m. D is an $m \times m$ diagonal matrix, $diag(d_1, d_2, \ldots, d_m)$ $(d_1 \geqq d_2 \geqq \ldots \geqq d_m \geqq 0)$. d_i is called a singular value. In the process of calculating Ax, x is first projected onto each row vector of V^t to give new orthonormal coordinate values in the space R^m. Second, each new orthonormal coordinate value is re-scaled by being multiplied by a corresponding singular value, d_i. And last, an exchange of orthonormal coordinates in the space R^m is carried out. From these considerations, the following can be stated:

(1) The first transformation V^t can be regarded as the linear characteristic measuring operator. These linear characteristics are determined by the orthonormal column vectors of V. Thus, the generalization of the unit transformation is realized in the space which is orthogonal to the linear subspace spanned by the column vectors of V. The unit transformation outputs the same

vector **z** if the input vectors all have the same projection onto the linear subspace spanned by the column vectors of V. This transformation is related to the input space $\mathbf{R}^n$.

(2) The second transformation D re-scales each of these linear characteristics of the input vector. The generalization of the unit transformation is realized mainly in the directions of the orthonormal coordinate axes which correspond to the smaller singular values.

(3) The last transformation U carries out an exchange of orthonormal coordinates in the space $\mathbf{R}^m$. Because this last transformation performs only an exchange of orthonormal coordinates, this transformation doesn't change any geometric feature in the space. This transformation alone seems to have no meaning and is related to the sigmoid non-linear transformation, whose output depends on the input vector position. It can also be seen that no generalization occurs at this stage of the unit transformation.

After the linear transformation A, the position shift by the vector **b** is carried out. Of course, this shift operation doesn't change any geometric features at all and thus this transformation is also related to the sigmoid non-linear transformation.

2-3. The sigmoid non-linear transformation

Formula (2) stands for the sigmoid non-linear transformation. By Formula (2), the input vector lengths and directions are changed and this effect is dependent on the input vector position.

Let us consider a classification of the input space of the transformation in terms of vector directions. Let the input vector of the non-linear transformation $\mathbf{y} = (\, y_1, y_2, \ldots, y_m \,)$ and the output vector $\mathbf{z} = (\, z_1, z_2, \ldots, z_m \,)$. If $/y_1/ = /y_2/ = \ldots = /y_m/$, it can be seen that these input vector directions are not changed by this transformation because each element of the input vector is compressed equally, and the sigmoid function doesn't change the input vector element sign. And if $/y_{i_1}/ = /y_{i_2}/ = ,\ldots, = /y_{i_a}/ = 0$ and $/y_{j_1}/ = /y_{j_2}/ = ,\ldots, = /y_{j_b}/ \doteqdot 0$ where $a + b = m$, $1 \leqq a,b$, it can also be seen that these input vector directions are not changed by the transformation because the sigmoid function maps 0 to 0. Thus, there are 2^m (corresponding to $/y_1/ = /y_2/ = \ldots = /y_m/$) + { ${}_mC_1 * 2^{m-1} + {}_mC_2 * 2^{m-2} + {}_mC_{m-1} * 2$ } (corresponding to $/y_{i_1}/ = /y_{i_2}/ = ,\ldots, = |y_{i_a}/ = 0$ and $/y_{j_1}/ = /y_{j_2}/ = ,\ldots, = /y_{j_b}/ \doteqdot 0$ where $a + b = m$, $1 \leqq a,b$) one-dimensional subspaces where vector directions are preserved after the transformation. They are determined by, for example, vectors $(\, 1,1,\ldots,1 \,)$, $(\, -1,1,\ldots,1 \,)$, $(\, 1,-1,1,\ldots,1 \,)$, ... $(\, 0,1,\ldots,1 \,)$, $(\, 1,0,1,\ldots,1 \,)$, ... and so forth.

Let us consider the region R: $(\, y_1, y_2, \ldots, y_m \,)$ of the input space, where $0 \leqq y_1,\ 0 \leqq y_2, \ldots, 0 \leqq y_m$. Because the sigmoid function maps positive values to positive values, first, it can be seen that the transformation maps this region into the same region R′ : $(\, z_1, z_2, \ldots, z_m \,)$ where $0 \leqq z_1,\ 0 \leqq z_2, \ldots, 0 \leqq z_m$. Let the angle between a vector **a** and **b** be Angle(**a**,**b**).

Angle(**a**,**b**) = acos((**a**,**b**)/ ‖ **a** ‖ * ‖ **b** ‖)
where(**a**,**b**) is an inner product of **a** and **b**, and ‖ . ‖ stands for a norm of a vector, i.e. square root ((**v**,**v**)).

The direction-invariant subspace of this region is, for example, determined by the vector $(\, 1,1,\ldots,1 \,)$. Now let **s** be $(\, 1,1,\ldots,1 \,)$ and let us consider Angle(**s**,**z**) **z** ∊ R′ .
Then,

$$(\mathbf{s},\mathbf{z}) = z_1 + z_2 + \ldots + z_m$$
$$\|\,\mathbf{z}\,\| = \textit{square root}(\, z_1^2 + z_2^2 + \ldots + z_m^2 \,)$$
$$\|\,\mathbf{s}\,\| = \textit{square root}(\, m \,)$$

Both the inner product and norm are unchanged by a permutation of the vector **z** elements, and the element-permuted output vector **z** obviously corresponds to the element-permuted input vector **y**. Thus, in terms of the vector angle with respect to the vector **s**, considering only a subregion,

$$\text{subR} = \{(\, y_1, y_2, \ldots, y_m \,);\, 0 < y_1 < y_2 <, \ldots, < y_m\}$$

is necessary and sufficient. Other subregions, $\{(\, y_{s(1)}, y_{s(2)}, \ldots, y_{s(m)} \,)\,;\, 0 < y_1 < y_2 <, \ldots, < y_m\}$ ($s()$: *permutation*) are mapped onto this region, subR, using the inverse permutation operation, s^{-1}: $s(i) \,\text{->}\, i$ ($i = 1.2.\ldots.m.$). The sigmoid non-linear transformation maps this region into subR' , where

$$\mathrm{subR}' = \{(z_1,z_2,\ldots,z_m);0 < z_1 < z_2 <,\ldots,< z_m\}$$

In this subregion, consider the angles between s and y,z. Let us take a projection of the subregion onto a two-dimensional linear subspace, which is determined by a combination of two natural bases i.e.

$e_i = (\,0,\ldots,0,\underset{i\text{-th element}}{1},0,\ldots,0\,)$ and $e_j = (\,0,\ldots,0,\underset{j\text{-th element}}{1},0,\ldots,0\,)$

$(\, i < j \,)$.Let a projection of an input vector y be y' and let a projection of the corresponding output vector be z' . Of course, $y' = (\, y_i,y_j \,)$, $z' = (\, z_i,z_j \,)$ and s' , a projection of s, is $(\,1,1\,)$. Using two constants c_i, c_j, the ratio z_j/z_i is written as follows:

$$z_j/z_i = (\, c_jy_j \,)/(\, c_iy_i \,) \qquad \text{---}(\,11\,)$$

c_i, c_j is the compression ratio of the sigmoid. Because the sigmoid is a strictly monotonic increasing function and because y_j,y_i $(\, i < j \,)$ are positive, the following holds.

$$0 < c_iy_i < c_jy_j \qquad \text{---}(\,12\,)$$

Because the sigmoid function is increasingly saturate as its input grows larger, also holds the following inequality:

$$0 < c_j < c_i < 1 \qquad \text{---}(\,13\,)$$

Using $(\,13\,)$, the right side of the equation $(\,11\,)$ can be written as follows:

$$(\, c_jy_j \,)/(\, c_iy_i \,) = (\, c_j/c_i \,)(\, y_j/y_i \,) < y_j/y_i \qquad \text{---}(\,14\,)$$

From the equation (12), $(\, c_jy_j \,)/(\, c_iy_i \,)$ is larger than 1. Thus, from the inequality $(\,14\,)$,

$$1 < (\, c_jy_j \,)/(\, c_iy_i \,) < y_j/y_i \qquad \text{---}(\,15\,)$$

holds. Taking an arc-tangent of each term of the inequality $(\,15\,)$, we obtain the following inequality.

$$45^\circ < \theta' < \theta \qquad \text{---}(\,16\,)$$

$\theta - 45^\circ$, $\theta' - 45^\circ$ are the angles between y' and s' and between z' and s' , respectively. Thus it is proved that, in terms of the angle, y' approaches s' by the sigmoid non-linear transformation. Fig. 3 shows a sample projection of input and output vectors of the transformation.

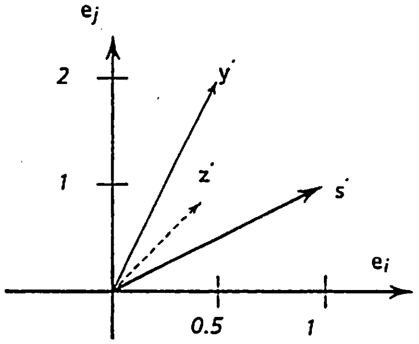

Fig. 3 Projection of input and output vectors of the transformation

The discussion above is obviously independent of the specific choice of e_i,e_j, and thus, in any projection onto any e_i-e_j linear subspace, y' approaches s' by the transformation. Hence, it can be concluded that the sigmoid non-linear operation makes an input vector in the region R closer to the direction-invariant subspace of the region. The same thing applies to the other regions, i.e. $\{(\, y_1,y_2,\ldots, y_m \,);\, 0 \leqq y_1 \,\&\, 0 \leqq y_2, \ldots, 0 \leqq y_{m-1} \,\&\, y_m \leqq 0 \,\}$, $\{(\, y_1,y_2,\ldots, y_m \,);\, 0 \leqq y_1 \,\&\, 0 \leqq y_2, \ldots, 0 \leqq y_{m-2} \,\&\, y_{m-1} \leqq 0 \,\&\, 0 \leqq y_m \,\}$, $\{(\, y_1,y_2,\ldots, y_m \,);\, 0 \leqq y_1 \,\&\, 0 \leqq y_2, \ldots, 0 \leqq y_{m-3} \,\&\, y_{m-2} \leqq 0 \,\&\, 0 \leqq y_{m-1} \,\&\, 0 \leqq y_m \,\},\ldots, \{(\, y_1,y_2,\ldots, y_m \,);\, y_1 \leqq 0 \,\&\, y_2 \leqq 0,\ldots, y_m \leqq 0 \,\}$. The proof is almost the same as the proof above.

Fig. 4 shows 2-dimensional input and output vectors of the transformation. It can be observed that vectors in each region of $\{\, (\, y_1,y_2 \,)\,;\, y_1 \geqq 0 \,\&\, y_2 \geqq 0 \,\}$, $\{\, (\, y_1,y_2 \,)\,;\, y_1 \leqq 0 \,\&\, y_2 \leqq 0 \,\}$, $\{\, (\, y_1,y_2 \,)\,;\, y_1 \leqq 0 \,\&\, y_2 \geqq 0 \,\}$, $\{\, (\, y_1,y_2 \,)\,;\, y_1 \geqq 0 \,\&\, y_2 \leqq 0 \,\}$ are made closer to the direction-invariant subspace of the region.

To summarize, the sigmoid non-linear transformation makes all vectors in each region closer to the direction-invariant subspace of the region. This effect is, of course, dependent on the input vector length. The longer the input vector is, the larger this effect becomes.

Each region can be further divided into subregions, $\{\, (\, y_{s(1)},y_{s(2)},\ldots,y_{s(m)} \,) : |y_1| \leqq |y_2| \leqq \ldots$

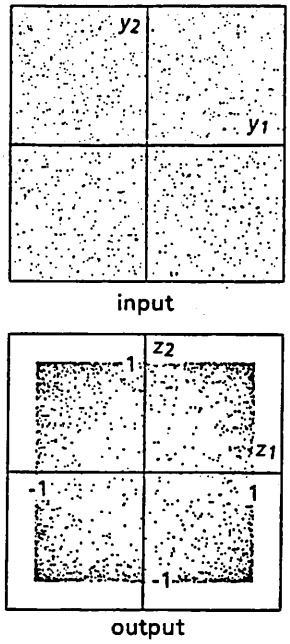

Fig. 4 Sample input and output of the transformation

$\leq |y_m| \}$ (*s() : permutation*) and these regions are equivalent in terms of the deformation of the sigmoid non-linear transformation.

2-4. A structured unit transformation

From these considerations, the unit transformation from one layer output to the next higher layer output can be structured. Fig. 5 illustrates this structured unit transformation.

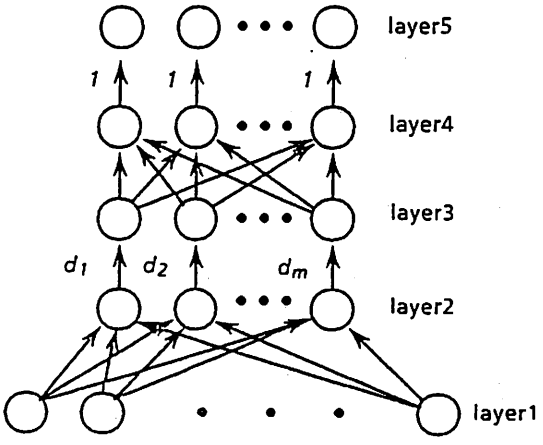

Fig. 5 Structured unit transformation

All the units at layer 1, layer 2 and layer 3 are linear units and have no bias values. All the units at layer 4 are also linear units and have bias values. Layer 5 corresponds to the next higher layer of the original unit transformation and thus all the units at this layer have sigmoids.

The transformation from the layer 1 output to the layer 2 output corresponds to matrix V^t operation in Section 2-1 and thus is a characteristic measuring liner transformation. The next transformation, from layer 2 to layer 3, corresponds to the matrix D, $diag(d_1,d_2,...,d_m)$ in Section 2-1. Each unit has only one link to a layer 3 unit. This transformation deforms layer 2 output vectors using singular values. The transformation from the layer 3 output to the layer 4 output corresponds to the matrix U operation and adds the bias vector **b** in Section 2-1. This transformation adjusts layer 3 output vector positions properly in the space $\mathbf{R}^m$ for the next non-linear transformation. The last non-linear transformation at layer 5 is the sigmoid non-linear transformation of Section 2-2. This transformation non-linearly deforms vector position relationships in the layer 4 space.

3. AN APPLICATION TO AN EXCLUSIVE OR NETWORK

A feed forward neural network realizing exclusive OR[12] is shown in Fig. 6. In this case, the output unit and all the input units have no sigmoid. The numbers within circles indicate a unit number. Fig. 7 shows the input vectors to the network. The network maps solid circles to -*1* and hollow circles to *1*. Table 1 shows the parameters of the exclusive OR neural network obtained by the back-propagation learning algorithm[12]. $w_{i,j}$ is a link weight from unit j to unit i. b_1 and b_2 correspond to bias values of

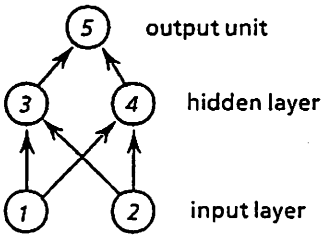

Fig. 6 Exclusive OR network

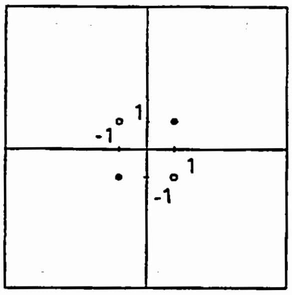

Fig. 7 Input vectors to the Exclusive OR network

units *3* and *4*, respectively, and b'_{1} is a bias of unit *5*.

connection weights

$w_{3,1}$	-1.27863
$w_{3,2}$	1.27863
$w_{4,1}$	-1.42848
$w_{4,2}$	1.42848
$w_{5,3}$	-1.90863
$w_{5,4}$	1.89593

biases

b_1	1.09786
b_2	-1.29396
b'_{1}	1.58892

Table 1. Parameters of the Exclusive OR net

This network can be divided into two unit transformations. One is the transformation from the input layer output to the hidden layer output, and the other is the transformation from the hidden layer output to the output unit.

Fig. 8 illustrates an equivalent structured unit transformation. This corresponds to the transformation from the input layer output of the network to the hidden layer output of the network. Let us consider this unit transformation.

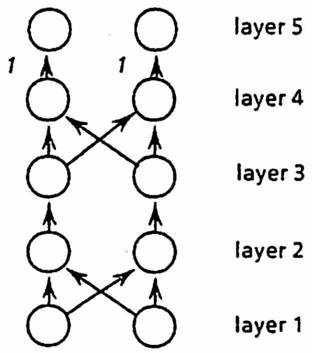

Fig. 8 Structured unit transformation

Coefficients A, b of the affine part are as follows:

$$A = \begin{pmatrix} -1.27863, 1.27863 \\ -1.42848, 1.42848 \end{pmatrix} \quad \mathrm{b} = \begin{pmatrix} 1.09786 \\ -1.29396 \end{pmatrix}$$

A can be decomposed by singular value decomposition as follows:

$$A = UDV^t \qquad (t\text{:transpose})$$

where

$$U = \begin{pmatrix} -0.666944, -0.745107 \\ -0.745107, 0.666944 \end{pmatrix} D = \begin{pmatrix} 2.71125, 0 \\ 0, 0 \end{pmatrix}$$

$$V^t = \begin{pmatrix} 0.707107, -0.707107 \\ -0.707107, -0.707107 \end{pmatrix}$$

Layer 1 to layer 2: As described in Section 2-3, operator V^t measures input vector characteristics. These characteristics are determined by the row vectors of V^t. In this case these characteristics are determined by the vectors,

$$\mathrm{v}_1 = \begin{pmatrix} 0.707107 \\ -0.707107 \end{pmatrix} \quad \mathrm{v}_2 = \begin{pmatrix} -0.707107 \\ -0.707107 \end{pmatrix}$$

Interestingly, the direction of v_1 is equal to the first principal axis of the data of category "1" and the direction of v_2 is equal to the first principal axis of the data of category "-1". At the first stage of a structured unit transformation between layers, the network must take as much category "1" as well as category "-1" information as possible. Thus, this result seems reasonable. Fig. 9 shows the plot of the transformation output.

Layer 2 to layer 3: This transformation is determined by the matrix D. The first element of the input vector is multiplied by the constant *2.71125*, and the second element by zero. The dimension of the input space is reduced to one. From this result, it can be seen that the generalization has been realized in the direction of the characteristic measuring vector v_2. Fig. 10 is the plot of the transformation output.

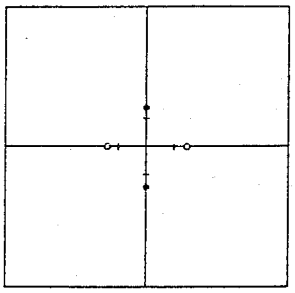

Fig. 9 Output of the transformation from layer 1 to layer 2

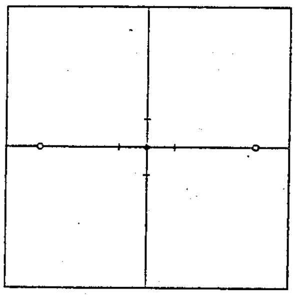

Fig. 10 Output of the transformation from layer 2 to layer 3

Layer 3 to layer 4: Here, first, an exchange of coordinates which is determined by the matrix U is carried out. Because the determinant of the matrix U is negative, an inversion of the coordinates has been carried out. The inverted coordinates, $(1,0)$, $(0,-1)$, are rotated 131.832° clockwise. Fig. 11 shows this transformation output.The output vectors are shifted by the vector b. Fig. 12 is the plot of the shifted output.

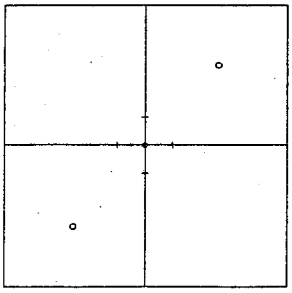

Fig. 11 Output of the transformation from layer 3 to .layer 4

At layer 5: As Fig. 12 illustrates, the outputs of the affine transformation part are not linearly separable. After the sigmoid non-linear operation at layer 5, only an affine transformation, which is the transformation between the hidden layer and the output unit, is carried out. This affine transformation measures a signed distance of a vector from a specified line. In this case, the distance of vectors in category "1" should be 1 and the distance of vectors in category "-1" should be -1. Thus, the sigmoid operation must deform the input vectors so that the vectors in category "1" and the vectors in category "-1" are linearly separable. As can be seen in Fig. 12, the input vectors in category "1" belong to regions $\{(y_1,y_2);y_1 \leqq 0 \ \& \ y_2 \geqq 0\}$, $\{(y_1,y_2);y_1 \geqq 0 \ \& \ y_2 \leqq 0\}$ and the input vectors in category "-1" belong to a region $\{(y_1,y_2);y_1 \geqq 0 \ \& \ y_2 \geqq 0\}$. It can also be seen in Fig. 12 that the input vector lengths of category "1" are much longer than those of the input vectors of category "-1". Thus, the sigmoid non-linear transformation makes the input vectors of category "1" much closer to the direction-invariant subspaces of the regions. This causes the vectors in category "1" and the vectors in category "-1" to become linearly separable. Fig. 13 shows the output of the sigmoid non-linear transformation and a distance-measuring line (a dotted line) determined by the link weights $w_{5,4}$, $w_{5,4}$, and the bias b'_1. As can be seen in Fig. 13, the sigmoid non-linear transformation has successfully deformed the input vectors. The transformations between layer 2 and layer 3 and between layer 3 and layer 4 have contributed to properly arrange the vectors for the sigmoid non-linear deformation. The sigmoid non-linearity is essential in this task.

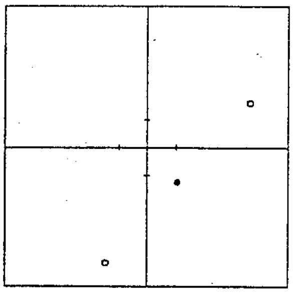

Fig. 12 Output of the transformation from layer 3 to layer 4

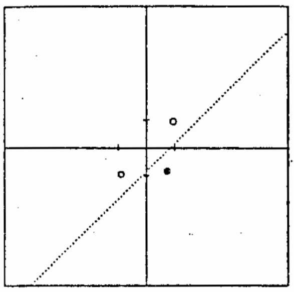

Fig. 13 Output of the transformation at layer 5

4. CONCLUSION

In this paper, the interpretations of a feed-forward neural network have been described. Based on singular value decomposition and a classification of an input space of the sigmoid non-linear transformation, the following have been clarified.

(1) The first transformation V^t can be regarded as the linear characteristic measuring operator. These linear characteristics are determined by the orthonormal column vectors of V. The generalization of the unit transformation is realized in the space which is orthogonal to the linear subspace spanned by the column vectors of V.

(2) The second transformation D re-scales each of these linear characteristics of the input vector. The generalization of the unit transformation is realized mainly in the directions of the orthonormal coordinate axes which correspond to the smaller singular values.

(3) The last transformation U carries out an exchange of orthonormal coordinates in the space $\mathbf{R}^m$. This transformation is related to the sigmoid non-linear transformation. No generalization occurs at this stage of the unit transformation.

(4) The sigmoid non-linear transformation makes all vectors in the each region of the input space closer to the direction-invariant subspace of the region.

Acknowledgements

The author is grateful to Dr. Kuremastu, the president of ATR Interpreting Telephony Research Laboratories, Dr. Shikano, the head of the Speech Processing Department, and Dr. Kuwabara, the supervisor of the Speech Processing Department, for their constant help and encouragement. The author is also grateful to Mr. Tanaka, an engineer of the Speech Processing Department, for his cooperation in computer programming.

REFERENCES

[1]Proc. ICNN 87, 1987.

[2]Proc. ICNN 88, 1988.

[3]K.Funahashi, "ON THE APPROXIMATION OF CONTINUOUS MAPPINGS BY NEURAL NETWORKS", MBE 88-9 1988, in Japanese.

[4]K.Funahashi, "ON THE CAPABILITIES OF NEURAL NETWORKS", MBE 88-52 1988, in Japanese.

[5]G. Cybenko, "Continuous Valued Neural Networks with Two Hidden Layers are Sufficient", preprint, March, 1988.

[6]G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function", preprint, October 24, 1988.

[7]K. Hornik, M. Stinchcombe and H. White, "MULTI-LAYER FEEDFORWARD NETWORKS ARE UNIVERSAL APPROXIMATORS", preprint, June, 1988.

[8]B. Irie and S. Miyake, "Capabilities of Three-layered Perceptrons", IEEE ICNN 88, 1988.

[9]R. Hecht-Nielsen, "Kolmogorov's mapping neural network existence theorem", IEEE ICNN 87, Proc. III-11 - III-14, 1987.

[10]C.R.Rao, *Linear Statistical Inference and Its Applications*, pp. 39 - 40, Tokyotosho co., 1983, in Japanese, translated from English.

[11]H.Bour, Y.Kamp, "Auto-Association by Multilayer Perceptrons and Singular Value Decomposition", NEURAL NETWORKS FOR COMPUTING, April 6-9, 1988, Snowbird Utah.

[12]D.E.Rumelhart, J.L.McClelland and the PDP Research Group, *Parallel Distributed Processing*, Vol.1, MIT Press, 1986.