Proceedings of the 7th WSEAS International Conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006 (pp102-107)

# On Implementation of Nested Rectangular Decision Regions by Multi-Layer Perceptrons II: Properties and Feasibility

CHE-CHERN LIN
Department of Industrial Technology Education
National Kaohsiung Normal University
116 Ho Ping First Road,Kaohsiung, 802
TAIWAN, ROC

*Abstract:* - In this paper, the author theoretically discusses the partitioning capabilities of the up-down algorithm and proves the implementation feasibility of the algorithm. Four propositions are presented in the paper. The first three propositions explain the nature of the up-down algorithm and the last one proves the implementation feasibility of the algorithm. The author also discusses the implementation feasibility of the partially nested rectangular decision regions using the proposed algorithm.

*Key-Words:* - Classification; Multi-layer perceptron; Nested decision region; Partitioning capability.

## 1 Properties

At the beginning of this section, the author first presents a well-known necessary and sufficient condition for two-class classification problems, and then discusses some properties necessary for proving the feasibility of the implementation of the nested rectangular decision regions using the up-down algorithm.

**Criterion 1:** The necessary and sufficient condition for implementing two-class classification problems in a decision region is the minimum $\theta$ value of sub-regions belonging to class A must be greater than the maximum $\theta$ value of sub-regions belonging to class B.

The trick of the up-down algorithm is to sequentially add an outermost ring to a nested decision region starting from the 1A-1B case and to guarantee these additions of rings are successfully implemented.

To do this, Criterion 1 must be satisfied. Figure 1 shows the case of adding an outermost ring of class B to the 2A-1B decision region, resulting in a 2A-2B one. In this figure the original 2A-1B decision region consists of sub-regions A1 to A17 and B1 to B8, and the added outmost ring consists of sub-regions B9 to B32. To implement this, one adds four partitioning lines to the original decision region (the 2A-1B decision region): $w_{left}$, $w_{right}$, $w_{bottom}$, and $w_{top}$. All of sub-regions in the original decision region are on the '0' side of $w_{right}$. This implies $w_{right}$ does not contribute anything to the $\theta$ values of the original sub-regions during the addition of the outermost ring. Similarly $w_{top}$ doesn't affect the $\theta$ values of the original sub-regions. On the other hand, all of original sub-regions are on the '1' side of $w_{left}$. $w_{left}$ (=3) is therefore added to the $\theta$ value of each sub-region in the original decision region during the addition of the outermost ring. Similarly, $w_{bottom}$ (=3) is also added to the $\theta$ value of each sub-region during the addition. The classification condition is still guaranteed because all sub-regions in the original decision region are offset with a same value (=6), the sum of $w_{left}$ and $w_{bottom}$. Figure 1 shows the changes on the $\theta$ values before and after adding the outermost ring to the 2A-1B decision region. In the figure, the number in each sub-region indicates its $\theta$ value after adding the outermost ring while the number in parentheses indicates the $\theta$ value before adding the outermost ring. The sum of $w_{left}$ and $w_{bottom}$ (=6) is added to the $\theta$ value of each sub-region in the original decision during the addition of the outermost ring.

To implement the classification, Criterion 1 must be guaranteed, i.e., the maximum $\theta$ value of the added ring (sub-regions B9 to B32) is less than the minimum $\theta$ value of the original sub-regions belonging to class A (sub-regions A1 to A17). The maximum $\theta$ values of the added ring occur in the sub-corner sub-regions of the added ring including sub-regions B10, B14, B16, B22, B20, B26, B27, and B31 with the same $\theta$ value (=3). The author will prove it later in Proposition 2. The nested decision region (after the addition of the outermost ring) is successfully implemented if one selects a proper threshold $\theta_h$ (=3.5).

It is important to note that the $\theta$ values of sub-regions in a nested decision region are vertically and horizontally symmetrical if the weights are selected by the up-down algorithm. The $\theta$ values are therefore symmetrical to the innermost ring due to the vertical and horizontal symmetries. Figure 1 shows these symmetrical properties of the $\theta$values in the 2A-2B nested decision region. For example, sub-region B9 is horizontally symmetrical to sub-region B15 and vertically symmetrical to sub-region B21. One gets $\theta_{B9} = \theta_{B15}$ and $\theta_{B9} = \theta_{B21}$. In addition, since sub-region 21 is also horizontally symmetrical to sub-region 32, one obtains $\theta_{B21} = \theta_{B32}$. From the above description, one concludes$\theta_{B9} = \theta_{B32}$ (symmetrical to the innermost ring). The following proposition proves these symmetrical properties.

**Proposition 1**: For a nested decision region implemented by the up-down algorithm, the$\theta$values in the decision region are

(a) horizontally symmetrical,
(b) vertically symmetrical, and
(c) symmetrical to the innermost ring of the decision region.

**Proof**

(a) The author discusses the case of the nested decision region whose outermost ring belongs to class B, and then generalizes to the case of the nested decision whose outermost ring belongs to class A. Consider a nested decision region whose outermost ring belongs to class B with *p* partitioning lines, as shown in Figure 2. The decision region consists of ($p$/2 + 1) vertically layered portions from its bottom to top. The author proves that the $\theta$ values in the lowest portion of the decision region are horizontally symmetrical and then generalizes the symmetrical properties to the rest of the layered portions. Let $q$ be $p$/4. For convenience, the author numbers the sub-regions in the lowest portion of the decision region from 0 to 2$q$ (SR 0 to SR 2$q$) from the left hand side to the right hand side. The author first proves the leftmost sub-region of the lowest portion of the decision region (sub-region 0) is horizontally symmetrical to the rightmost sub-region (sub-region 2$q$), and then generalizes the horizontally symmetrical property to the rest of sub-regions in the lowest portion of the decision region. The $\theta$ value of sub-region 0 is equal to 0 because it is on the '0' sides of all partitioning lines. However, sub-region 2$q$ is on the '1' sides of all vertical partitioning lines and on the '0' sides of all horizontal partitioning lines. The $\theta$value of sub-region 2$q$ is given by

$$\begin{aligned}\theta_{2q} &= \sum_{k=1}^{2q} w_k \\ &= \underbrace{(q)}_{w_1} + \underbrace{(-q+1)}_{w_2} + \underbrace{(q-2)}_{w_3} + \Lambda + \underbrace{(-q+2)}_{w_{2q-2}} + \underbrace{(q-1)}_{w_{2q-1}} + \underbrace{(-q)}_{w_{2q}} = 0\end{aligned} \quad (1)$$

From the above discussion, one gets $\theta_0 = \theta_{2q}$. Now consider the second leftmost sub-region of the lowest portion of the decision region (sub-region 1) and its horizontally symmetrical sub-region (sub-region 2$q$-1). The only difference between sub-region 1 and sub-region 0 is that sub-region 0 is on the "0" side of $w_1$, while sub-region 1 is on the "1" side of $w_1$. $\theta_1$ is given by

$$\theta_1 = \theta_0 + w_1 = 0 + q = q \quad (2)$$

Similarly, the only difference between sub-region 2$q$-1 and sub-region 2$q$ is that they are on different sides of partitioning line $w_{2q}$. $\theta_{2q-1}$ is given by

$$\theta_{2q-1} = \theta_{2q} - w_{2q} = 0 - (-q) = q \quad (3)$$

As discussed earlier, $w_1$ and $w_{2q}$ are horizontally anti-symmetrical. This anti-symmetrical property guarantees$\theta_1 = \theta_{2q-1}$. By the same procedure one concludes the $\theta$values are horizontally symmetrical for the rest of sub-regions of lowest portion of the nested decision region. One can easily generalize the horizontally symmetrical property to the rest of the layered portions of the decision region.

Using the same procedure as above, one can obtain the horizontally symmetrical property for a nested decision region whose outermost ring belongs to class A.

(b) Using a similar procedure as part (a), one can conclude the$\theta$values in the nested decision region are vertically symmetrical.

(c) By parts (a) and (b), one concludes the$\theta$values in the nested decision region are symmetry to the innermost ring. QED.

By the symmetrical properties discussed above and a similar proving procedure as Proposition 1, one can conclude the $\theta$ values of the sub-corner sub-regions of the outermost ring of a nested decision region are the same.

**Proposition 2**: Consider a nested decision region with *p* partitioning lines (again, *p* is dividable by 4). If one uses the up-down algorithm to add an outermost ring to the decision region, the following statements are true.

(a) If the added ring belongs to class B, the maximum $\theta$values of the added ring occur in the sub-corner sub-regions of the added ring with the same $\theta$value (= $p$/4).
(b) If the added ring belongs to class A, the minimum$\theta$values of the added ring occur in the sub-corner sub-regions of the added ring with the same $\theta$value (= – $p$/4).

**Proof:**

(a) Consider a nested decision region with *p* partitioning lines, as shown in Figure 2. Due to the symmetrical properties of $\theta$values, one only needs to discuss the sub-regions in the left part of the bottom

Proceedings of the 7th WSEAS International Conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006 (pp102-107)

portion of the outermost ring (SR 0 to SR $q$) . Let $q = p/4$. It is important to note that $q$ is an odd number in this case. The vertical partitioning lines are selected as follows: $w_1 = q$, $w_2 = -(q-1)$, $w_3 = q-2$, $w_4 = -(q-3)$, …, $w_q = 1$. Again, one numbers these sub-regions from 0 to $q$ from the left hand side to the right hand side. $\theta_0$ is equal to 0 because sub-region 0 is on the '0' side of all partitioning lines. The $\theta$ values of the rest of the sub-regions are:

$$\theta_{2q} = \sum_{k=1}^{2q} w_k = \underbrace{(q)}_{w_1} + \underbrace{(-q+1)}_{w_2} + \underbrace{(q-2)}_{w_3} + \underbrace{(-q+3)}_{w_4} + \ldots + \underbrace{(-1)^{q+r}(q-r+1)}_{w_r} \quad \text{for } r = 1,2,\ldots q \tag{4}$$

Next the author will prove the maximum $\theta$ value occurs in sub-region 1 with a value of $q$.

To prove this, the author divides the discussion into three cases:

Case 1: sub-region 0: as just discussed above, $\theta_0 = 0$.

Case 2: the sub-region number is even (not including sub-region 0). The stopping point of Eq. (4) is at an even term. Rearranging Eq. (4), one gets

$$\theta_r = \sum_{k=1}^{r} w'_k = \underbrace{(q-q+1)}_{w_1+w_2=1} + \underbrace{(q-2-q+3)}_{w_3+w_4=1} + \ldots + \underbrace{(q-(r-1))-(q-r+1)}_{w_{r-1}+w_r=1} \quad \text{for } r = 2,4,6,\ldots,(q-1) \tag{5}$$

Observing Eq. (5), one get $\theta_2 = 1$, $\theta_4 = 2$, $\theta_6 = 3$,…,$\theta_{q-1} = (q-1)/2$. Eq. (5) therefore forms a strictly increasing sequence with a relationship of $\theta_2 < \theta_4 < \theta_6 < \cdots < \theta_{q-1}$. The maximum value of the sequence occurs at $\theta_{q-1}$ with a value of $(q-1)/2$.

Case 3: the sub-region number is odd. The stopping point of Eq. (4) is at an odd term. Rearranging Eq. (4), one gets

$$\theta_r = \sum_{k=1}^{r} w'_k = \underbrace{(q)}_{w_1} + \underbrace{(-q+q+1-2)}_{w_2+w_3=-1} + (-q+q+3-4) + \ldots + \underbrace{\{-[q-(r-1)+1] + (q-r+1)\}}_{w_{r-1}+w_r=-1} \quad \text{for } r = 1,3,\ldots,q \tag{6}$$

Again, observing Eq. (6), one get $\theta_1 = q$, $\theta_3 = q-1$, $\theta_5 = q-2$,…,$\theta_q = q-(q-1)/2$. Eq. (6) therefore forms a strictly decreasing sequence with a relationship of $\theta_1 > \theta_3 > \theta_5 > \cdots > \theta_q$. The maximum value of the sequence occurs at $\theta_1 = 1$ with a value of $q$.

To sum up the discussion, the maximum $\theta$ values for Case 1, Case 2, and Case 3 are:

$\theta_0 = 0$ occurring in sub-region 0 for Case1,

$\theta_q = (q-1)/2$ occurring in sub-region $q$ for Case 2, and

$\theta_1 = q$ occurring in sub-region 1 for Case 3.

It is easy to prove that $q > (q-1)/2$ since $q$, as mentioned earlier, is a positive odd number. The global maximum of $\theta$ value of the above three cases therefore occurs in sub-region 1, a sub-corner sub-region of the added ring, with a value of $q$ ($= p/4$). By the symmetrical properties of $\theta$ values, one concludes that the maximum $\theta$ values of the added ring occur in the sub-corner sub-regions of the added ring with the same value ($= p/4$).

(b) Using a similar procedure as part (a), one can prove the minimum $\theta$ values of the added ring occur in the sub-corner sub-regions of the added ring with the same value ($= -p/4$). QED.

**Proposition 3**: For a nested decision region implemented by the up-down algorithm, the following statements are true.

(a) The maximum $\theta$ value of each ring belonging to class B is the same.

(b) The minimum $\theta$ value of each ring belonging to class A is the same.

**Proof**

(a) Consider a nested decision region whose outermost ring belongs to class B with $p$ partitioning lines, as indicated in Figure 2. By Proposition 2, the maximum $\theta$ value of the outermost ring occurs in its sub-corner sub-regions (for example, sub-region $s$ adjacent to the left-top corner sub-region of the outermost ring) with a value of $p/4$. Letting $q = p/4$, one gets $\theta_s = q$. The maximum $\theta$ value of the second outermost ring of class B also occurs in its sub-corner sub-regions (for example, sub-region $t$ adjacent to the left-top corner sub-region of the second outermost ring). Observing Figure 2, one finds that the differences between sub-region $s$ and sub-region $t$ are that they are on the different side of $w_2$, $w_3$, $w_{4q-1}$, and $w_{4q}$. The sum of $w_2$ and $w_3$ is equal to –1 since they are adjacent lines determined by the up-down algorithm. Similarly, the sum of $w_{4q-1}$ and $w_{4q}$ is also equal to –1. The $\theta$ value of sub-region $t$ is given by

$$\begin{aligned}\theta_t &= \theta_s + w_2 + w_3 - (w_{4q-1} + w_{4q}) \\ &= q + \underbrace{(-q+1)+(q-2)}_{w_2+w_3=-1} - [\underbrace{(-q)+(q-1)}_{w_{4q-1}+w_{4q}=-1}] = q\end{aligned} \tag{7}$$

Using the same procedure, one concludes that the maximum $\theta$ value of the third outermost ring belonging to class B is also equal to $q$ because of the horizontally and vertically anti-symmetrical properties of the weights. Using the same procedure, one can conclude the maximum $\theta$ value of each ring belonging to class B is the same.

One can apply the same procedure to a nested decision region whose outermost ring belongs to class A and obtain the same conclusion.

(b) Using the same procedure as part (a), one concludes the minimum $\theta$ value of each ring belonging to class A is the same. QED.

## 2 Feasibility

To successfully add an outermost ring to a nested decision region implemented by the up-down algorithm, Criterion 1 must be guaranteed. The following proposition will prove it.

**Proposition 4**: Consider a nested decision region implemented by the up-down algorithm. If one uses the up-down algorithm to add a ring to the decision region, the following statements are true.

(a) If the added ring belongs to class B, the maximum $\theta$ value of the added ring is less than the minimum $\theta$ value of the sub-regions of class A of the original nested decision region.

(b) If the added ring belongs to class A, the minimum $\theta$ value of the added ring is greater than the maximum $\theta$ value of the sub-regions of class B of the original nested decision region.

**Proof**

(a) Consider an original decision region with $p$ partitioning lines. Let $q = p/4$. By Propositions 2 and 3, the minimum $\theta$ value of sub-regions belonging to class A in the original decision region is $-q$. To add an outermost ring of class B to the original decision, one needs to add 4 more partitioning lines on the left, right, top, and bottom of the decision region. The number of total partitioning lines is therefore p + 4. Let $q' = (p+4) / 4 = q + 1$. By the up-down algorithm, the four weights associated with these partitioning lines are selected as follows: $w_{left} = q'$, $w_{right} = -q'$, $w_{bottom} = q'$, and $w_{top} = -q'$. As discussed earlier, the original decision region is on the '1' sides of $w_{left}$ and $w_{bottom}$, and the '0' sides of $w_{right}$ and $w_{top}$. This implies $w_{left}$ and $w_{bottom}$ are added to the $\theta$ values of all sub-regions in the original decision region during adding the four partitioning lines while $w_{right}$ and $w_{top}$ make no contribution. After adding the outermost ring, the minimum $\theta$ value of the sub-regions of class A of the original nested region becomes $-q + w_{left} + w_{bottom} = -q + 2q' = q+2$. By Proposition 2, the maximum $\theta$ value of the added ring (class B) is $q' = q + 1$. Since $q+2 > q+1$, one concludes the maximum $\theta$ value of the added ring is less than the minimum $\theta$ value of the sub-regions of class A of the original nested region.

(b) Using the same procedure as part (a), one concludes the minimum $\theta$ value of the added ring is greater than the maximum $\theta$ value of the sub-regions of class B of the original nested region, if the added ring belongs to class A. QED.

Proposition 4 guarantees the feasibility of implementation of the nested decision regions using the up-down algorithm. A nested decision can be successfully implemented by sequentially adding the rings from the innermost ring to the outermost ring and by selecting a proper threshold $\theta_h$.

Consider the partially nested rectangular decision regions mentioned in the first paper of the series of the studies (Part I: Algorithm). Using a similar proving procedure as Proposition 4, one can easily prove that the implementation feasibilities of these partially nested decision regions are still guaranteed because Criterion 1 is always guaranteed when one adds any combination of the four partitioning lines to the original nested decision region.

## 3 Discussion

The author theoretically discussed the partitioning capabilities of the up-down algorithm and proved the implementation feasibility of the algorithm. Four propositions were presented in the paper. The first three propositions explained the nature of the up-down algorithm and the last one proved the implementation feasibility of the up-down algorithm. The author also discussed the implementation feasibility of the partially nested rectangular decision regions using the proposed algorithm.

*References:*


[1] C. Lin, A. El-Jaroudi, An Algorithm to Determine the Feasibilities and Weights of Two-Layer Perceptrons for Partitioning and Classification, *Pattern Recognition*, Vol. 31, No. 11, 1998, pp. 1613-1625.

[2] P. J. Zwietering, E. H. L. Arts, J. Wessels, Exact Classification with Two-Layered Perceptrons, *Int. Journal of Neural Systems*, Vol. 3, No. 2, 1992, pp. 143-156.

[3] J. Makhoul, A. El-Jaroudi, R. Schwartz, Partitioning Capabilities of Two-Layer Neural Networks, *IEEE Trans. on Signal Processing*, Vol. 39, No. 6, 1991, pp. 1436-1440.

[4] R. P. Lippmann, An Introduction to Computing with Neural Nets, *IEEE ASSP Mag.*, Vol. 4, 1987, pp. 4-22.

[5] G. Cybenko, Approximation by Superpositions of a Sigmoidal Function, *Math. Contr., Signals, Syst.*, 1989, pp. 303-314.

[6] R. Shonkwiler, Separating the Vertices of N-cubes by Hyperplanes and its Application to Artificial Neural Networks, *IEEE Trans. on Neural Networks*, Vol. 4, No. 2, 1993, pp. 343-347.

[7] C. Lin, A. El-Jaroudi, A Study on the Partitioning Capabilities of Two-Layer Neural Networks, *IEEE Int. Conf. on Neural Network*, Orlando, Florida, July 1994, pp. I-360-365.

[8] G. J. Gibson, C. F. Cowan, On the Decision Regions of Multilayer Perceptrons, *Proceeding of the IEEE*, Vol. 78, No. 10, 1990, pp.

Proceedings of the 7th WSEAS International Conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006 (pp102-107)

1590-1594.

[9] C. Cabrelli, U. Molter, R. Shonkwiler, A Constructive Algorithm to Solve "Convex Recursive Deletion" (CoRD) Classification Problems via Two-Layer Perceptron Networks, *IEEE Trans. On Neural Networks*, Vol. 11, No. 3, 2000, pp. 811-816.

[10] W. Fan, L Zhang, Applying SP-MLP to Complex Classification Problems, *Pattern Recognition Letters*, Vol. 21, 2000, pp. 9-19.

[11] S. Draghici, The Constraint Based Decomposition (CBD) Training Architecture, *Neural Networks*, Vol. 14, 2001, pp. 527-550.

[12] V. Deolalikar, A two-layer paradigm capable of forming arbitrary decision regions in input space, IEEE Trans. On Neural Networks, Vol. 13, No. 1, 2002, pp. 15-21.

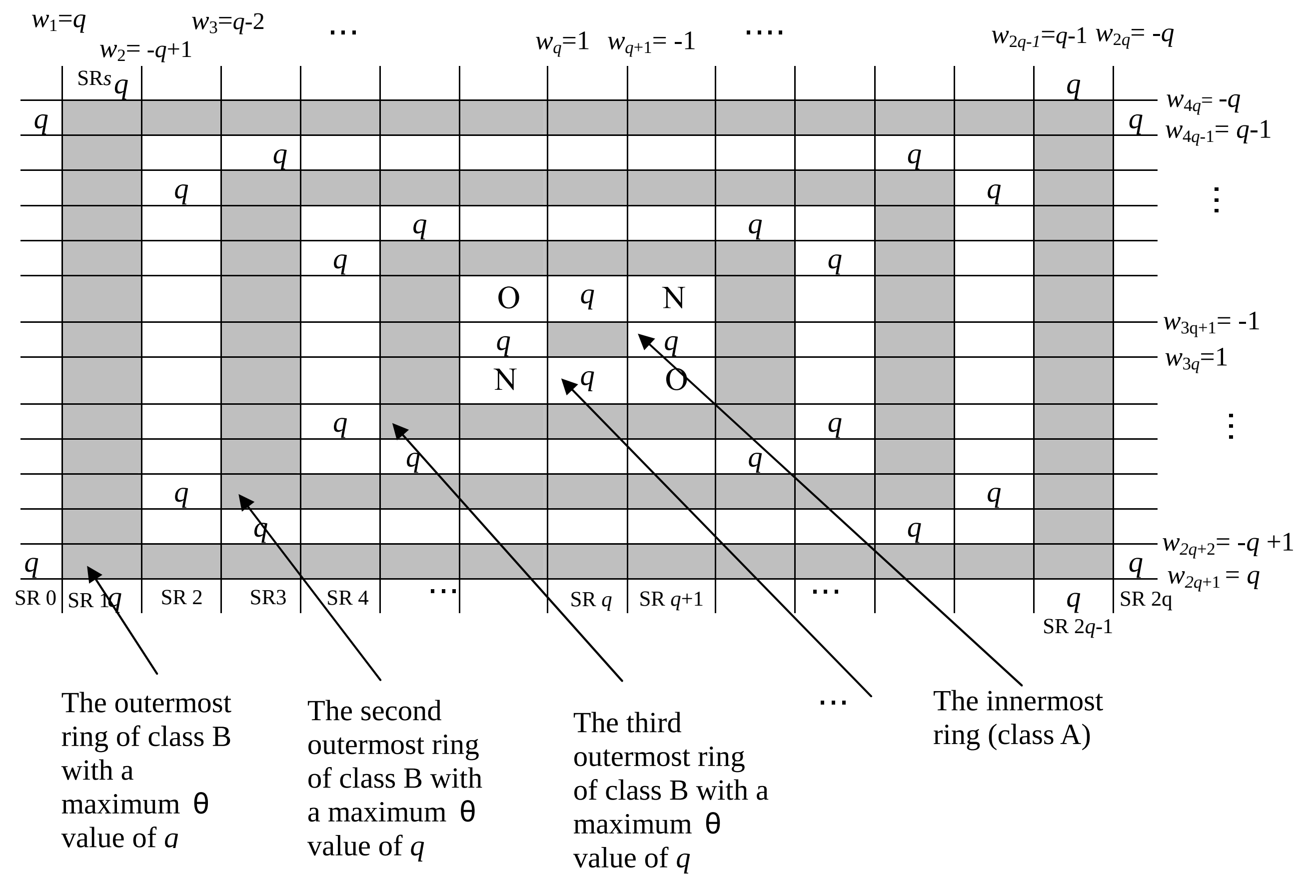
[13] G. Huang, Y Chen, H. A. Babri, Classification ability of single hidden layer feedforward neural networks, IEEE Trans. On Neural Networks, Vol. 11, No. 3, 2000, pp. 799-801.

.

| $\mathbf{w_{left}}$ **= 3** | $w_1$ = -2 | $w_2$ = 1 | $w_3$ = -1 | $w_4$ = 2 | $\mathbf{w_{right}}$ **= -3** | | |
|---|---|---|---|---|---|---|---|
| 0<br>B9 | 3<br>B10 | 1<br>B11 | 2<br>B12 | 1<br>B13 | 3<br>B14 | 0<br>B15 | $\mathbf{w_{top}}$ **= -3** |
| 3<br>B16 | 6 (0)<br>A2 | 4 (-2)<br>A3 | 5 (-1)<br>A4 | 4 (-2)<br>A5 | 6 (0)<br>A6 | 3<br>B22 | $w_8$ = 2 |
| 1<br>B17 | 4 (-2)<br>A7 | 2 (-4)<br>B1 | 3 (-3)<br>B2 | 2 (-4)<br>B3 | 4 (-2)<br>A8 | 1<br>B23 | $w_7$ = -1 |
| 2<br>B18 | 5 (-1)<br>A9 | 3 (-3)<br>B4 | 4 (-2)<br>A1 | 3 (-3)<br>B5 | 5 (-1)<br>A10 | 2<br>B24 | $w_6$ = 1 |
| 1<br>B19 | 4 (-2)<br>A11 | 2 (-4)<br>B6 | 3 (-3)<br>B7 | 2 (-4)<br>B8 | 4 (-2)<br>A12 | 1<br>B25 | $w_5$ = -2 |
| 3<br>B20 | 6 (0)<br>A13 | 4 (-2)<br>A14 | 5 (-1)<br>A15 | 4 (-2)<br>A16 | 6 (0)<br>A17 | 3<br>B26 | $\mathbf{w_{bottom}}$ |
| 0<br>B21 | 3<br>B27 | 1<br>B28 | 2<br>B29 | 1<br>B30 | 3<br>B31 | 0<br>B32 | |

Remark:
The number in each sub-region indicates its θvalue after the addition of the outmost ring while the number in parentheses indicates theθvalue before the addition .

Figure 1: The example of adding an outermost ring to a 2A-1B decision region, resulting in a 2A-2B decision region.




Remarks:

1. The symbol $q$ in the sub-regions indicates the maximum of the rings belonging to class B.
2. The maximum θvalue of the rings belonging to class B occurs in the sub-corner sub-regions of these rings.
3. SR represents sub-region; e.g., SR 0 = sub-region 0; SR $q$ = sub-region $q$

Figure 2: Explanations of the outermost and innermost rings in a nested decision region with $p$ (= $4q$) partitioning lines and the conceptual picture for Propositions 1, 2 and 3.