Application of analysis of overgeneralization in semantic comprehension

Tomohiro Washino
washino@libe.nara-k.ac.jp
Department of Liberal Studies
National Institute of Technology
Nara College
22 Yata, Yamatokoriyama
Japan

Tadashi Takahashi
ttakahashi@hagoromo.ac.jp
Faculty of Social Sciences
HAGOROMO University
of International Studies
1-89-1 Hamadera Minami, Nishi, Sakai
Japan

Abstract

When two concepts contain a common concept, overgeneralization (the phenomenon of overgeneralizing specific rules or semantic features) may occur in the process of learners gaining an understanding of the two concepts in relation to each other. In the present research, we use a neural network to analyze the overlap singularity phenomenon and the elimination singularity phenomenon in singular regions and perform simulations on the loss surface reported in previous research [4], [5]. As an extension of [6], we trained a neural network by inputting test data for a technical college and analyzed the "symbolic comprehension" of three of that college's classes in previous research [7]. In the present paper, the analysis technique used in "symbolic comprehension" is applied to "semantic comprehension". Additionally, we analyzed whether correct answers are misconception and calculated the semi-correct factor.

1 Introduction

Interrelationship problems can be thought of as problems that require consideration of the common concepts of permutations and combinations to derive an answer. For example, consider the following question: "How many ways are there to choose 3 people from the ten members of a committee?" The correct answer is ${}_{10}P_3$, whereas the answer ${}_{10}C_3$ is considered a semi-correct answer. In this study, we separate the conceptual understanding of "permutation" and "combination" into two parts: "symbolic comprehension" and "semantic comprehension". If we classify mathematics problems into categories of "semantic comprehension", the test in this case is to correctly judge the meanings of "arrange" and "choose" in the question sentence. We attempted to apply the results of analyzing the overlap singularity phenomenon and the elimination singularity phenomenon in singular regions, as a method for visualizing the state of "semantic comprehension" [4], [5] and "symbolic comprehension" [6]. As an extension of [6], we trained a neural network by inputting test data for a technical college and analyzed the "symbolic comprehension" of three of that college's classes [7].

2 Preparation for analysis

For more details on the basic definitions of learning theory, see [1], [2].

To test semantic comprehension in the field of "permutations" and "combinations" in the unit "The Number of Cases and Probability" in high school mathematics, we performed two examinations (1st and 2nd) and created a set of three questions as follows: (a) "How many ways can 3 people are taken from 9 people and arranged in a row?" (b) "How many ways are there to choose 2 cards out of 7 ?" (c) "How many ways can a group of 4 students arrange themselves in a low?" Questions (a) and (c) are permutation problems; question (b) is a combinatorial problem. The total score for permutations was set to 0.5, as was the total score for combinations. A student was credited with a semi-correct answer if that student mistakenly treated question (a) as a combination problem, question (b) as a permutation problem, or answered question (c) as 4 instead of 4!. The scores for correct answers were (a) 0.4 points, (b) 0.5 points, and (c) 0.1 points; and the scores for semi-correct answers were (a) 0.2 points, (b) 0.25 points, and (c) 0.05 points. We excluded students whose scores in both the permutation and combination areas were less than 0.5 and used the remaining data for analysis.

A schema is a conceptual structure created by knowledge from past experiences. The concept of "arranging" a permutation can be considered to be a transformation of a schema after learning combination "choosing".

We defined two learning stages: learning stage (1), in which correct answers are obtained by considering and progressing in the understanding of both permutations and combinations; and learning stage (2), in which semi-correct answers are influenced by the study of combinations and are obtained despite little progress in understanding combinations.

Three student groups were defined based on questions (a) through (c): student group (i), students with full points in the permutation problems and partial points in the combination problem; student group (ii), students with partial points in the permutation problems and full points in the combination problem; and student group (iii), students with full points in both the permutation and combination problems. We compared student groups (i) and (iii) in learning stage (1), and student groups (i) and (ii) in learning stage (2).

Definition 1 *Given inputs x, θ_0, the output Y of the two-layer neural network is defined as follows: $Y := f(x, \theta_0) = w_3 \tanh(w_1 x) + w_4 \tanh(w_2 x)$.*

The coordinate transformation from the parameters $\theta = (w_1, w_2, w_3, w_4)$ to the new parameters $\xi = (a, b, v, w)$ can be defined as follows: $a = w_2 - w_1$, $b = \frac{w_3 - w_4}{w_3 + w_4}$, $v = \frac{w_3 w_1 + w_4 w_2}{w_3 + w_4}$, $w = w_3 + w_4$.

For the two student groups, let c be the average score for the combinatorial problem, d be the average score for the permutation problems, and e and f be the numbers of students in the two groups. The weights of the neural network can then be determined as follows: $w_1' = c$, $w_2' = d$, $w_3 = \frac{e}{e+f}$, $w_4 = \frac{f}{e+f}$. For more details on correction, see [4], [5].

Parameter a is the difference between the average score for the permutation problems and the average score for the combinatorial problem. Parameter b is the difference in the relative proportions of the two student groups being considered.

We investigated these proportions by selecting two groups from student groups (i), (ii), and (iii). For the selected groups, we can visualize the understanding of permutations and combinations on the learning loss surface. For more details on the construction of neural networks and learning loss surfaces, see [3], [4].

Using Mathematica, a NetGraph is constructed with the loss function as the log likelihood ratio function to define *trainingNet*. We input the following:

```
trainingNet[a_, b_, v_, w_] := NetGraph[<|"params" -> parameterNet[a, b, v, w], "lhood" ->
ThreadingLayer[gaussianLikelihood],"neglog" -> ElementwiseLayer[-Log[#]&]|>,NetPort["Output"],
NetPort["params", "Output1"] ->"lhood", "lhood" -> "neglog" -> NetPort["Loss"]].
```

ParameterNet (two-layer neural network) and *trainingNet* are shown on the left and right sides of Figure 1, respectively.

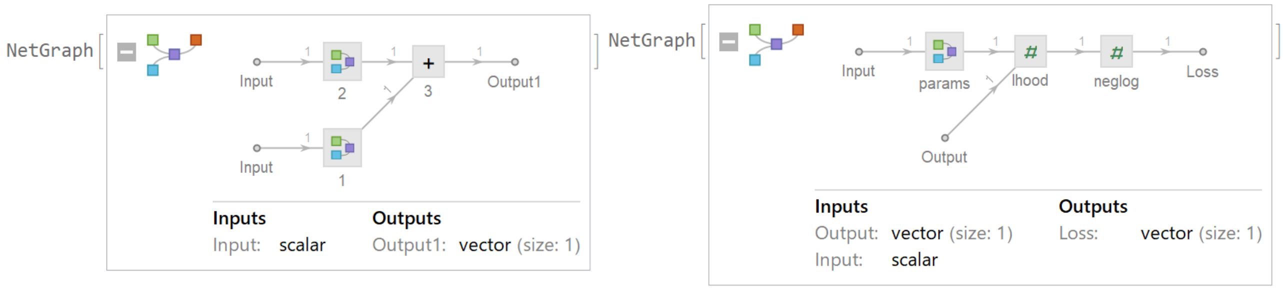

Figure 1: `parameterNet, trainingNet`

Near the singular region where the number of intermediate units changes from 2 to 1, a plateau phenomenon occurs where learning stagnates. We are interested in the plateau phenomenon in which learning is affected by singular regions and stagnates in learning. For more details on overlap singularity and cross-overlap singularity (where the learning process crosses the overlap and reaches the global optimum after training) in the context of information science, see [4]. For more details on elimination singularity, near-elimination singularity, and fast convergence, see [5]. In the following sections, the results obtained from real data are used to perform simulations in which the initial values are changed in 0.05 increments from -0.6 to 0.6 for a and from -1.1 to 1.1 for b in agreement with [4], [5]. In the figures, the initial values for the 1st examination is indicated by ◎, and the true distribution is indicated by $\times$. For the dynamics from the 1st to the 2nd examination, the initial value (simulation) is shown in blue, and the value after the simulation is shown in red.

3 Application to semantics comprehension

As an example of the application of a learning system to education at a technical college, the academic areas of permutations and combinations are taught at the end of the first year, and the academic area of probability is not taught until the fourth year at the technical college where I work. In this section, we consider a neural network trained using the 1st exam results for fourth-year technical college students (121 people) as initial values and the 2nd exam results for a high school as true data. We visualized all data from the technical college and analyzed the comprehension process for three classes (A, B, and C). Table 2 shows the mean scores, number of examinees, and correction coefficients for the correct (left-most two columns, center two columns) and semi-correct (right-most two columns) answers to the interrelationship question in the 2nd high school examination (again correcting for weights) and the technical college examination (overall, three classes).

2nd examination	(i) (full, partial) points	(iii) (full, full) points	(ii) (partial, full) points	(iii) (full, full) points	(i) (full, partial) points	(ii) (partial, full) points
Average c, d	0.25	0.5	0.283333333	0.5	0.25	0.276666667
Number of people e, f	3	23	6	23	8	60
Overall average h_2	0.471153846		0.455172414		0.273529412	
Total number of people	26		29		68	
Correction factor $\frac{g}{h_2}$	0.830007981		1.008761277		0.983827258	
Correction w_1, w_2	0.207501995	0.41500399	0.285815695	0.504380638	0.245956814	0.272192208
Overall technical college	**(i) (full, partial) points**	**(iii) (full, full) points**	**(ii) (partial, full) points**	**(iii) (full, full) points**	**(i) (full, partial) points**	**(ii) (partial, full) points**
Average c, d	0.25	0.5	0.3625	0.5	0.166666667	0.3
Number of people e, f	2	50	4	50	3	12
Overall average h_1	0.490384615		0.489814815		0.273333333	
Total number of people	52		54		15	
Correction factor $\frac{g}{h_1}$	0.797458648		0.937416124		0.984533016	
Correction w_1, w_2	0.199364662	0.398729324	0.339813345	0.468708062	0.164088836	0.295359905
Class A	**(i) (full, partial) points**	**(iii) (full, full) points**	**(ii) (partial, full) points**	**(iii) (full, full) points**	**(i) (full, partial) points**	**(ii) (partial, full) points**
Average c, d	0.25	0.5	0.35	0.5	0.25	0.3
Number of people e, f	2	15	2	15	1	4
Overall average h_1	0.470588235		0.482352941		0.29	
Total number of people	17		17		5	
Correction factor $\frac{g}{h_1}$	0.831005587		0.951917706		0.927950659	
Correction w_1, w_2	0.207751397	0.415502793	0.333171197	0.475958853	0.231987665	0.278385198
Class B	**(i) (full, partial) points**	**(iii) (full, full) points**	**(ii) (partial, full) points**	**(iii) (full, full) points**	**(i) (full, partial) points**	**(ii) (partial, full) points**
Average c, d	0	0.5	0.3	0.5	0.125	0.275
Number of people e, f	0	17	1	17	2	4
Overall average h_2	0.5		0.488888889		0.225	
Total number of people	17		18		6	
Correction factor $\frac{g}{h_2}$	0.782122905		0.939191534		1.196025294	
Correction w_1, w_2	0	0.391061453	0.28175746	0.469595767	0.149503162	0.328906956
Class C	**(i) (full, partial) points**	**(iii) (full, full) points**	**(ii) (partial, full) points**	**(iii) (full, full) points**	**(i) (full, partial) points**	**(ii) (partial, full) points**
Average c, d	0	0.5	0.45	0.5	0	0.325
Number of people e, f	0	18	1	18	0	4
Overall average h_3	0.5		0.497368421		0.325	
Total number of people	18		19		4	
Correction factor $\frac{g}{h_3}$	0.782122905		0.92317945		0.828017511	
Correction w_1, w_2	0	0.391061453	0.415430752	0.461589725	0	0.269105691

Figure 2: Examination results

3.1 Analysis of students: learning stage (1)

3.1.1 Simulation of changing a

The dynamics of the simulation with a varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 3.

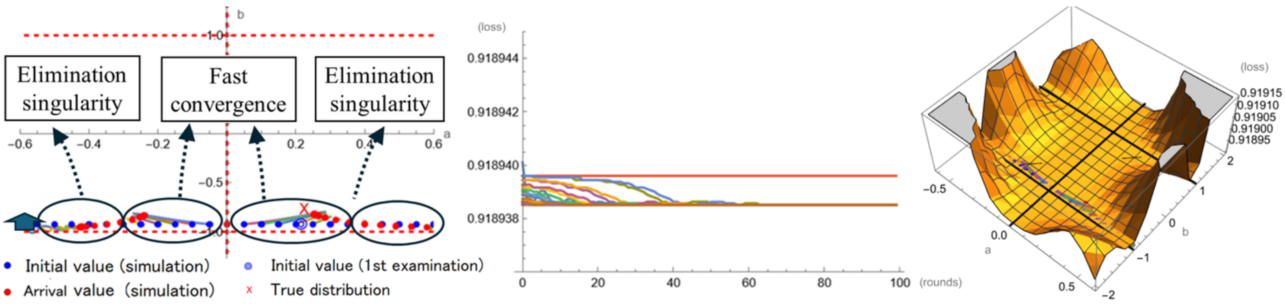

Figure 3: a varied: correct answer (technical college overall)

Learning begins from the state $b = -0.92$, where the proportion of student group (iii) is much greater than that of student group (i). As the overgeneralization continues to change within $|a| < 0.28$, fast convergence occurs at $a = \pm 0.28$ without the influence of the critical line $a = 0$. Overgeneralization also increases for $|a| > 0.28$, and the critical line $b = -1$ is affected and an elimination singularity occurs, namely all students are in student group (iii). There is little change in the proportion of student group (i), meaning the learning of combinations does not progress and a transformation of the permutational schema does not take place.

3.1.2 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 4, comparing the class A (left) and classes B and C (right).

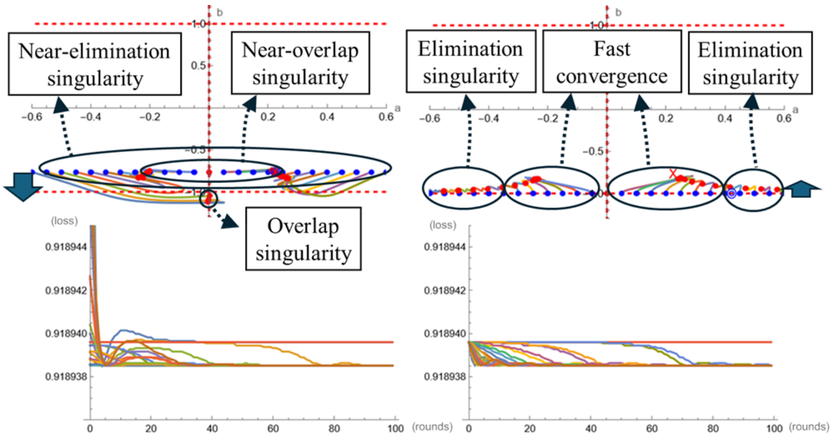

Figure 4: a varied: correct answer (three classes)

In class A, learning begins from the state $b = -0.76$, where the proportion of student group (iii) is greater than that of student group (i). As overgeneralization changes within $|a| < 0.22$, a near-overlap singularity occurs due to the influence of the critical line $a = 0$, with learning stagnating at $a = \pm 0.22$. When $a < -0.22$, an overlap singularity occurs due to the influence of the critical line $a = 0$, namely all students are in student group (iii). Since the proportion of student group (i) decreased from $b = -0.73$ to $b = -1.0$, the learning of combinations progresses and a transformation of the permutational schema takes place. When $a > 0.22$, a near-elimination singularity occurs due to the influence of the critical line $b = -1$ and a transformation of the permutational schema does not take place. Analysis of classes B and C gives the same results as analysis of "overall" for the technical college: learning begins from the state $b = -1.0$.

3.1.3 Simulation of changing b

The dynamics of the simulation with b varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 5.

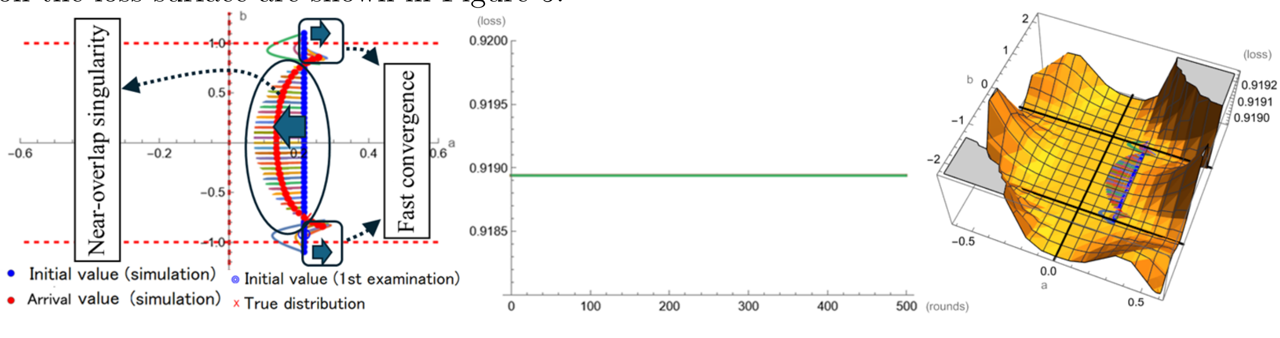

Figure 5: b varied: correct answer (technical college overall)

Learning begins from the state $a = 0.22$, where the overgeneralization of permutations as combinations is large. When the proportion of student group (i) changes within $-0.79 \leq b \leq 0.79$, a becomes small and a near-overlap singularity occurs. A positive transition occurs as the difference from the overgeneralization becomes smaller. As the proportion of student group (iii) changes when $b < -0.79$ or $0.79 < b$, a becomes big and converges (fast convergence) without the influence of the critical line $a = 0$. The difference from the overgeneralization of permutations as combinations becomes even bigger, leading to a negative transition.

3.1.4 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 6, comparing class A (left) and classes B and C (right).

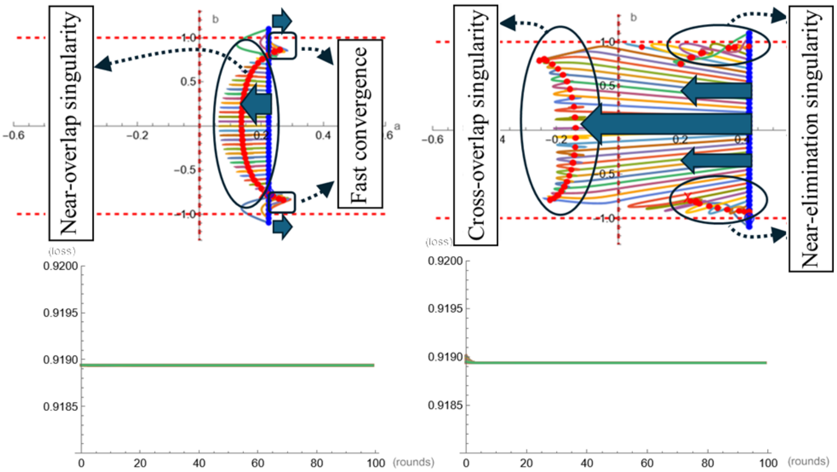

Figure 6: b varied: correct answer (three classes)

Analysis of class A gives the same results as analysis of "overall" for the technical college. In classes B and C, learning starts from the state $a = 0.42$, where the overgeneralization of permutations as combinations is large. First, as the proportion of student group (i) becomes larger (b increases), a near-elimination singularity occurs due to approaching the critical line $b = -1$ and returning to the original state. A positive transition occurs because a large overgeneralization of permutations as combinations is suppressed. Additionally as the proportion of student group (i) continues to become larger (b increases), a cross-overlap singularity occurs beyond the critical line $a = 0$. When the overgeneralization of permutations as combinations is large, weights w_1 and w_2 both equal 0.42, since $v = 0.42$, and there is no difference from the overgeneralization of combinations as permutations. Finally, the overgeneralization of combinations as permutations continues further to become large, leading to a negative transition.

3.2 Analysis of students: learning stage (2)

3.2.1 Simulation of changing a

The dynamics of the simulation with a varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 7.

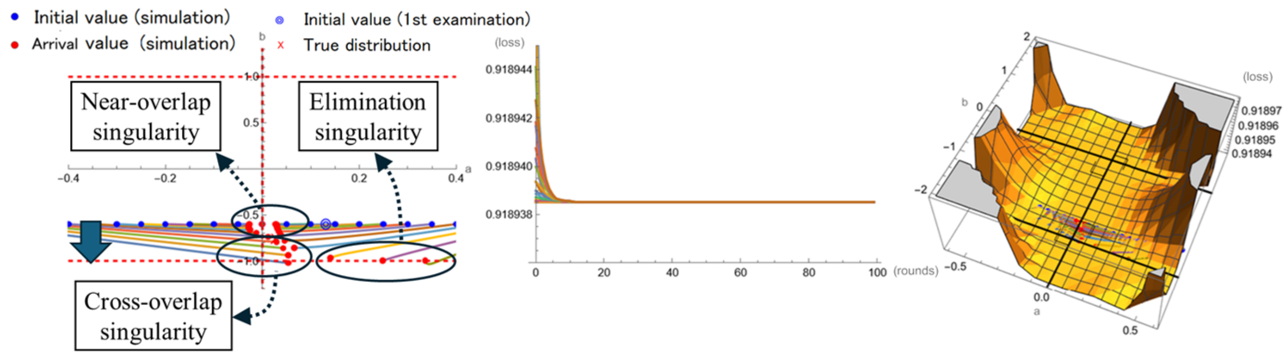

Figure 7: a varied: semi-correct answer (technical college overall)

Learning begins from the state $b = -0.60$, where the proportion of student group (ii) is greater than the proportion of student group (i). When overgeneralization changes within $-0.50 < a < 0.45$, a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = \pm 0.03$. On the other hand, as the overgeneralization of permutations as combinations increases (a increases) when $a > 0.45$, an elimination singularity occurs due to the influence of the critical line $b = -1$, resulting in only student group (ii). Additionally, as the overgeneralization of combinations as permutations increases (a decreases) from $a < -0.50$, a cross-overlap singularity occurs, stagnating at $a = 0.06$ due to crossing the critical line $a = 0$. Since the proportion of student group (i) has decreased from $b = -0.60$ to $b = -1.0$, the learning of combinations progress and a transformation of the permutational schema takes place.

3.2.2 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 8, comparing classes A (left), B (center), and C (right).

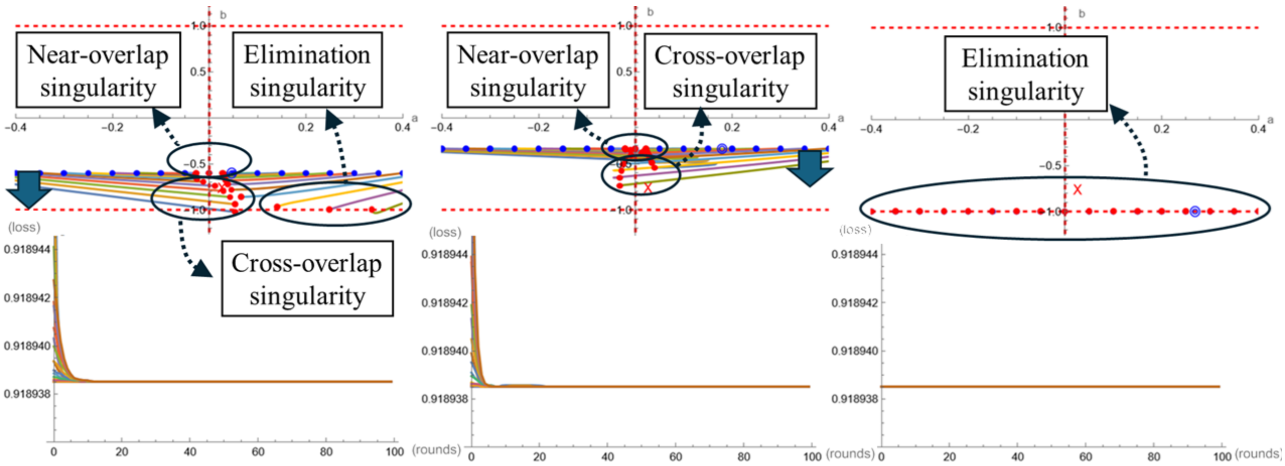

Figure 8: a varied: semi-correct answer (three classes)

Analysis of class A gives the same results as analysis of "overall" for the technical college. In class B, learning begins from the state $b = -0.33$, where the proportion of student group (ii) is slightly greater than the proportion of student group (i). When overgeneralization changes within $-0.25 < a < 0.35$, a near-overlap singularity occurs due to the influence of the critical line, stagnating at $a = \pm 0.02$. On the other hand, as the overgeneralization changes from $a < -0.25$ or $0.35 < a$, a cross-overlap singularity occurs, stagnating at $a = 0.04, -0.03$ due to crossing the critical line $a = 0$. Since the proportion of student group (i) decreases from

$b = -0.33$ to $b = -0.73$, learning of combinations progresses and a transformation of the permutational schema takes place. In class C, learning begins from the state $b = -1.0$, which means learning is only in student group (ii). As b changes, an elimination singularity occurs, which is only in student group (ii), due to the critical line $b = -1$. The learning of combinations does not progress and a transformation of the permutational schema does not take place.

3.2.3 Simulation of changing b

The dynamics of the simulation with b varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 9.

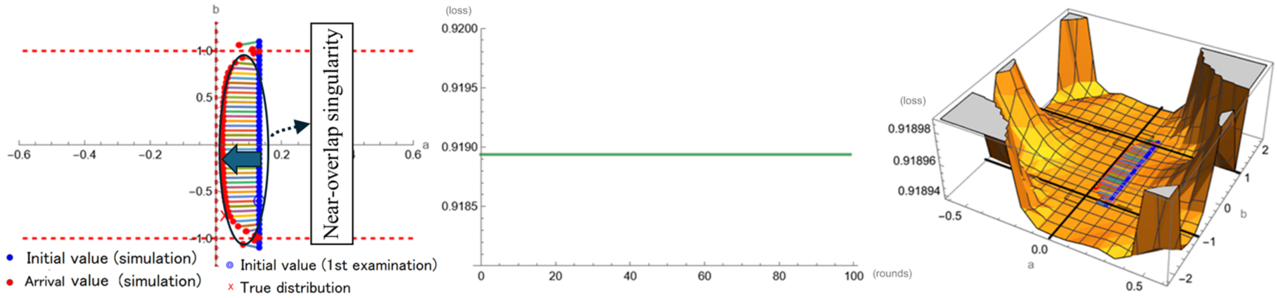

Figure 9: b varied: semi-correct answer (technical college overall)

Learning begins from the state $a = 0.13$, where the overgeneralization of permutations as combinations is large. As b changes, a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = 0.02$. The overgeneralization of permutations as combinations becomes smaller, and therefore a positive transition occurs.

3.2.4 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 10, comparing classes A (left), B (center), and C (right).

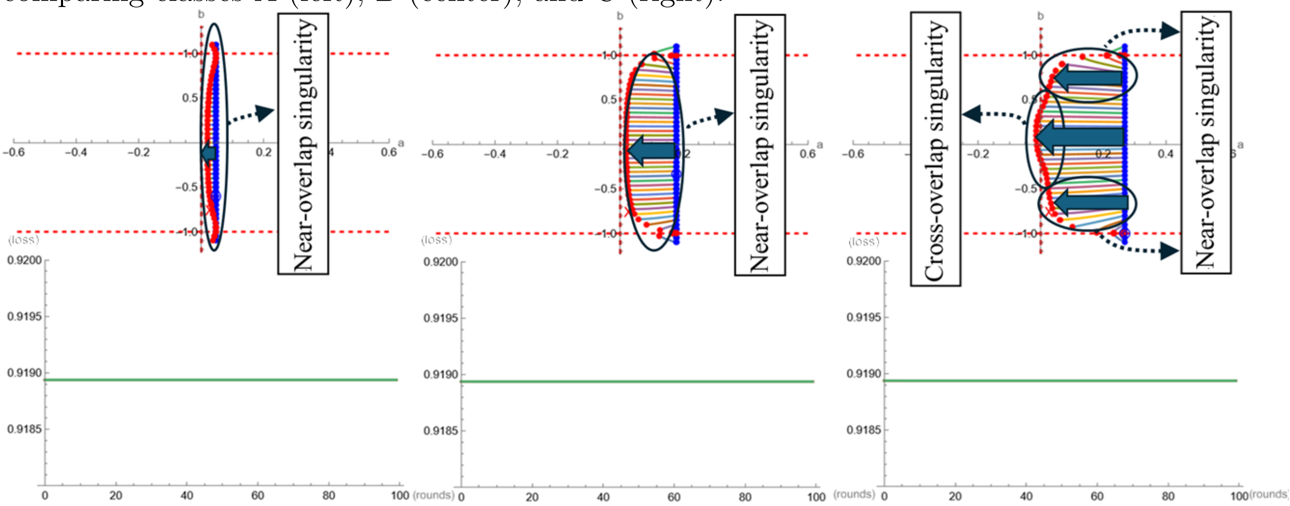

Figure 10: b varied: semi-correct answer (three classes)

In class A, learning begins from the state $a = 0.04$. In class B, learning begins from the state $a = 0.18$. Analysis of classes A and B gives the same results as analysis of "overall" for the technical college. In class C, learning begins from the state $a = 0.27$, where the overgeneralization of permutations as combinations is larger than the overgeneralization in classes A and B. As b changes within either $b < -0.17$ or $0.37 < b$, a near-overlap singularity occurs due to the influence of the critical line $a = 0$. When $b = -0.17$ or 0.37, weights w_1 and w_2 both equal 0.27, since $v = 0.27$, and the overgeneralization of combinations as permutations and the overgeneralization of permutations as combinations are equal (overlap singularity). The overgeneralization of permutations as combinations becomes smaller, so positive transitions occur. When $-0.17 < b < 0.37$, a cross-overlap singularity occurs, stagnating at $a = -0.016$. In addition, the overgeneralization of combinations as permutations becomes large, which leads to a negative transition.

3.3 Overall analysis of technical college and comparison of three classes

3.3.1 Correct answer

For the technical college overall, the learning of both permutations and combinations progresses is preserved by changing the overgeneralization. Deterioration of the understanding of permutations is caused by the overgeneralization of permutations as combinations. If the proportion of student group (i) increases, the overgeneralization of permutations as combinations becomes small, and positive transitions occur. Therefore, misconceptions are unlikely to occur.

For class A, no matter how large the overgeneralization of permutations as combinations is, not everyone will be in student group (iii). Deterioration of the understanding of permutations is affected by the overgeneralization of permutations as combinations. The proportion of student group (i) increases, and the overgeneralization of permutations as combinations becomes small and positive transitions occur. Therefore misconceptions are unlikely to occur. For classes B and C, the proportion of student group (iii) decreases such that the overgeneralization of combinations as permutations becomes big, and negative transitions occur, and therefore misconceptions are likely to occur. Therefore, understanding of combinations deteriorates.

Based on the above dynamics, we propose a strategy for teachers to improve learning guidance. In class A, learning should be promoted so that understanding of permutations does not deteriorate, with both attention to the overgeneralization of permutations as combinations and consideration of their interrelationships. In classes B and C, there is little change in the proportion of any student group during the learning process, but learning should be promoted so that understanding of combinations does not deteriorate, with attention to the overgeneralization of combinations as permutations (misconceptions).

3.3.2 Semi-correct answer

For the technical college overall, learning begins with some students having semi-correct comprehension. The overgeneralization of permutations as combinations is suppressed as the understanding of combinations progresses. Therefore, it is necessary to consider this as a possible error in the developmental process. Classes A and B include students who are semi-correct. The overgeneralization of permutations as combinations becomes small as the understanding of combinations progresses. Therefore, it is necessary to consider this as a possible error in the developmental process. For class A, if the proportion of student group (ii) increases, students for whom the semi-correct factor is semantic comprehension may be included, due to the overgeneralization of permutations as combinations becoming big. However, for class B, if the proportion of student group (ii) increases, semantic comprehension is not considered to be a factor for semi-correct answers, due to the overgeneralization of combinations as permutations becoming big. In class C, learning begins where no students have a semi-correct answer in semantic comprehension. If the proportion of student group (i) is close to the proportion of student group (ii), then the deterioration of the understanding of combinations is affected by the overgeneralization of combinations as permutations.

Based on the above dynamics, we propose a strategy for teachers to improve learning guidance. In class A, attention should be paid to the overgeneralization of combinations as permutations to prevent the understanding of combinations from deteriorating and reduce the number of students for whom the semi-correct factor is semantic comprehension. The learning of combinations should be promoted by interrelating them with permutations, with attention to the overgeneralization of combinations as permutations. In class B, learning can be promoted to reduce the number of students for whom the semi-correct factor is semantic comprehension by increasing the overgeneralization of permutations as combinations. Since the overgeneralization of permutations as combinations has little effect, the learning of combinations should be promoted. In class C, learning should be promoted to prevent a deterioration in understanding of combinations. Attention should be paid to the overgeneralization of combinations as permutations when the proportion of student group (i) is close to that of student group (ii).

Acknowledgment

This work was supported by the Japan Society for the Promotion of Science, Grant Numbers 25K17021.

References

[1] S. Watanabe (2006), Algebraic geometry and statistical learning theory, Morikita Publishing Co., Ltd.

[2] S. Amari (2014), New developments in information geometry, Saiensu-sya Co., Ltd.

[3] T. Washino and T. Takahashi: On the analysis of singularity structure in learning, Proceedings of the 26th Asian Technology Conference in Mathematics (ATCM 2021), pp. 297-307 (2021)

[4] T. Washino and S. Ohashi: Learning guidance based on the overlap singularity phenomenon, Sci. Math. Japonicae, e-2023, No. 12, pp. 1-20 (2023)

[5] T. Washino and T. Takahashi: Learning guidance based on the elimination singularity phenomenon, Proceedings of 28th Asian Technology Conference in Mathematics (ATCM 2023), pp. 352-361 (2023)

[6] T. Washino and T. Takahashi: Learning guidance based on analysis of symbolic comprehension, Proceedings of 29th Asian Technology Conference in Mathematics (ATCM 2024), pp. 200-209 (2024)

[7] T. Washino and T. Takahashi: Analysis of overgeneralization in symbolic comprehension and its application, The Electronic Journal of Mathematics and Technology, Vol. 19, No. 2, pp. 142-161 (2025)