The Electronic Journal of Mathematics and Technology, Volume 19, Number 2, ISSN 1933-2823

Analysis of overgeneralization in symbolic comprehension and its application

Tomohiro Washino
washino@libe.nara-k.ac.jp
Department of Liberal Studies
National Institute of Technology
Nara College
Yamatokoriyama, Nara 639-1080
Japan

Tadashi Takahashi
ttakahashi@hagoromo.ac.jp
Faculty of Social Sciences
Hagoromo University
of International Studies
Sakai, Osaka 592-8344
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 order to analyze the overlap singularity and elimination singularity phenomena in singular regions, we constructed a learning system that performs simulations on loss surfaces using neural networks. We first defined four learning stages in the process of symbolic comprehension of the concepts of "permutations" and "combinations" in high school and analyzed the change in training loss and the dynamics on the loss surface. From this analysis, we propose learning guidance for mathematics teachers. As an application of the learning system, we trained a neural network by inputting test data for a technical college and analyzed the understanding of three of that college's classes.

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 10 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 understanding" and "semantic understanding". If we classify mathematics problems into categories of "semantic understanding", 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]). Considering how to answer an interrelationship problem requires understanding the semantics and then symbols. To consider a transformation of the concept of “arranging” in permutations, we need to analyze the state of “symbolic comprehension” in learning mathematics.

2 Preparation for analysis

For more details on the basic definitions of learning theory, see [1], [2]. If we classify mathematics problems into categories of “symbolic comprehension”, the test here is to correctly compute permutations and combinations of symbols. To test symbolic comprehension in the academic area of “permutations” and “combinations” in the unit “The Number of Cases and Probability” in high school mathematics, we performed three examinations (1st, 2nd, and 3rd) and created a set of four questions: (a) ${}_5\mathrm{P}_3$ (1st and 2nd), ${}_8\mathrm{P}_4$ (3rd); (b) 6! (1st and 2nd), ${}_7\mathrm{C}_3$ (3rd); (c) ${}_6\mathrm{C}_4$ (1st), ${}_8\mathrm{C}_3$ (2nd); and (d) ${}_{20}\mathrm{C}_{18}$ (1st), ${}_{16}\mathrm{C}_{14}$ (2nd). Questions (a) (1st to 3rd) and (b) (1st and 2nd) are permutation problems; questions (b) (3rd), (c) (1st to 3rd), and (d) (1st to 3rd) are combinatorial problems. The total scores for permutations and combinations were each set to 0.5. A student was credited with a semi-correct answer if that student mistakenly treated question (a) as a combination problem, question (c) or (d) as a permutation problem. For the 1st and 2nd examinations, the scores for correct answers were (a) 0.4 points, (b) 0.1 points, (c) 0.3 points, and (d) 0.2 points; and the scores for semi-correct answers were (a) 0.2 points, (c) 0.15 points, and (d) 0.1 points. For the 3rd examination, the scores for correct answers were (a) 0.5 points and (b) 0.5 points; and the scores for semi-correct answers were (a) 0.25 points and (b) 0.25 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.

We defined four learning stages: learning stage (1), in which correct answers are obtained by considering and progressing in understanding both permutations and combinations between the 1st and 2nd examinations; learning stage (2), in which semi-correct answers are influenced by the study of combinations and are obtained despite little progress in understanding combinations between the 1st and 2nd examinations; learning stage (3), in which semi-correct answers are influenced by the study of combinations between the 2nd and 3rd examinations; learning stage (4), in which correct answers are obtained by considering the interrelationship with combinations between the 2nd and 3rd examinations.

Three student groups were defined based on questions (a) through (d): (i) students with full points in the permutation problems and partial points in the combination problem; (ii) students with partial points in the permutation problems and full points in the combination problem; (iii) students with full points in both the permutation and combination problems. We compared student groups (i) and (iii) in learning stage (1), student groups (i) and (ii) in learning stages (2) and (3), and student groups (ii) and (iii) in learning stage (4).

Definition 1 *For an input X following a probability distribution $q(x)$ and a function $f(x,\theta)$ from $\mathbb{R}$ to $\mathbb{R}$ with parameter θ, consider a random variable Z on $\mathbb{R}$ with mean 0 and variance 1. Then a random variable Y on $\mathbb{R}$ is called a function approximation model if it takes the following form: $Y := f(x,\theta) + Z$.*

Definition 2 *Given inputs x, θ_0, the output Y of the two-layer neural network is defined as follows:*

$$Y := f(x,\theta_0) = w_3 \tanh(w_1 x) + w_4 \tanh(w_2 x).$$

Definition 3 *The conditional probability $p(y|x,\theta)$ that the function approximation model Y follows is defined as a statistical model, and if this statistical model realizes the true distribution, then the conditional probability $q(y|x)$ that the output Y follows is determined to be the true distribution:*

$$p(y|x,\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{|y-f(x,\theta)|^2}{2\sigma^2}\right),\ q(y|x) = p(y|x,\theta_0) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{|y-f(x,\theta_0)|^2}{2\sigma^2}\right).$$

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: $w_1' = c$, $w_2' = d$, $w_3 = \frac{e}{e+f}$, $w_4 = \frac{f}{e+f}$. Then, the score obtained by subtracting 0.5 from the average of the sum of the permutation and combination results from the first to the third examinations is denoted as $g = \frac{ce+df}{e+f} = w_1' w_3 + w_2' w_4$, whereas the corresponding first to third examination scores are denoted as h_1, h_2, and h_3, respectively. Then, a correction for weights is determined as $w_1 = w_1' \times \frac{g}{h_i}$, $w_2 = w_2' \times \frac{g}{h_i}$.

Figure1 shows the average score, number of students, and correction coefficients for students who gave correct answers (two left-most columns, two center columns) and semi-correct answers (two right-most columns) to the interrelationship question (d) in the three examinations.

1st to 3rd examinations	(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	0.282352941	0.5	0.127777778	0.5	0.291071429	0.172727273
Number of people e,f	34	121	9	121	28	11
Overall average g	0.452258065		0.474230769		0.257692308	
Total number of people	155		130		39	
1st 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	0.282142857	0.5	0.15	0.5	0.3	0.425
Number of people e,f	28	50	6	50	7	2
Overall average h_1	0.421794872		0.4625		0.327777778	
Total number of people	78		56		9	
Correction factor $\frac{g}{h_1}$	1.072222767		1.025363825		0.786179922	
Correction w_1, w_2	0.302519995	0.536111383	0.153804574	0.512681913	0.235853977	0.334126467
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	0.283333333	0.5	0	0.5	0.305263158	0.4
Number of people e,f	6	28	0	28	19	2
Overall average h_2	0.461764706		0.5		0.314285714	
Total number of people	34		28		21	
Correction factor $\frac{g}{h_2}$	0.979412369		0.948461538		0.81993007	
Correction w_1, w_2	0.277500171	0.489706185	0	0.474230769	0.250294442	0.327972028
3rd 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	0.5	0.083333333	0.5	0.125	0.035714286
Number of people e,f	0	43	3	43	2	7
Overall average h_3	0.5		0.472826087		0.055555556	
Total number of people	43		46		9	
Correction factor $\frac{g}{h_3}$	0.904516129		1.002970822		4.638461538	
Correction w_1, w_2	0	0.452258065	0.083580902	0.501485411	0.579807692	0.165659342

Figure 1: Examination results

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 this difference for selections of two groups from the three student groups (i), (ii), and (iii). For a pair of 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]. Here, the state of understanding is displayed as points for the 1st examination (blue), 2nd examination (red), 3rd examination (green), and 1st to 3rd examinations (yellow). Student groups (i) and (iii) are shown on the left side of Figure 2; student groups (i) and (ii) are shown in the center; and student groups (ii) and (iii) are shown on the right side.

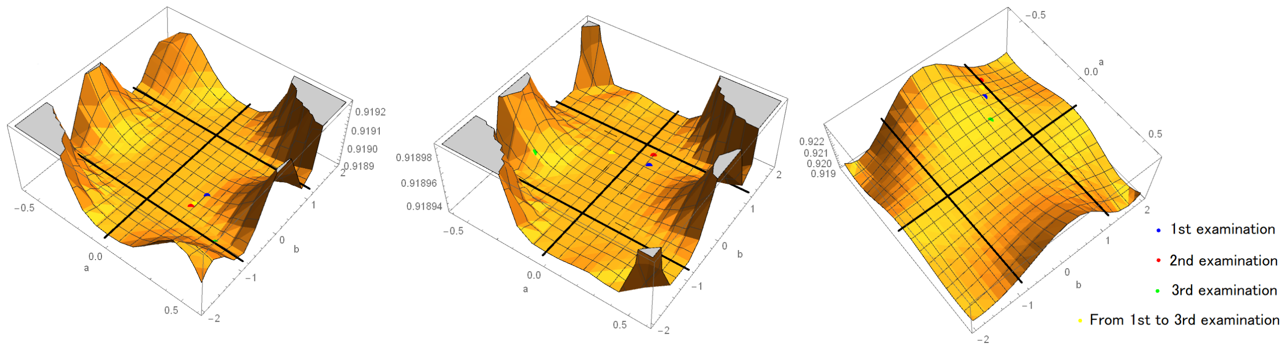

Figure 2: Learning loss surface of symbolic comprehension

We next define the overlap singularity and the elimination singularity as singular regions.

Definition 4 *An overlap singularity is defined as the region in the parameter space where θ satisfies*

$$R_0 := \{\theta \in \mathbb{R}^4 | w_1 = w_2\}.$$

An elimination singularity is defined as the region in the parameter space where θ satisfies

$$R_1 := \{\theta \in \mathbb{R}^4 | w_3 = 0\} \cup \{\theta \in \mathbb{R}^4 | w_4 = 0\}.$$

For the details of specifically a cross-overlap singularity in information science, see [4], and for the details of specifically a near-elimination singularity and fast convergence, see [5].

3 Simulation

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 ([4], [5]). In the figures, the initial values for the 1st (blue), 2nd (red), and 3rd (green) examinations are indicated by $\circledcirc$, 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 arrival value (simulation) is shown in red. For the dynamics from the 2nd to the 3rd examination, the initial value (simulation) is shown in red, and the arrival value (simulation) is shown in green.

3.1 Analysis of students: learning stage (1)

We analyzed the change from the 1st to the 2nd examination to target the students who gave correct answers. We can now consider the process of understanding the two concepts by comparing student groups (i) and (iii).

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.

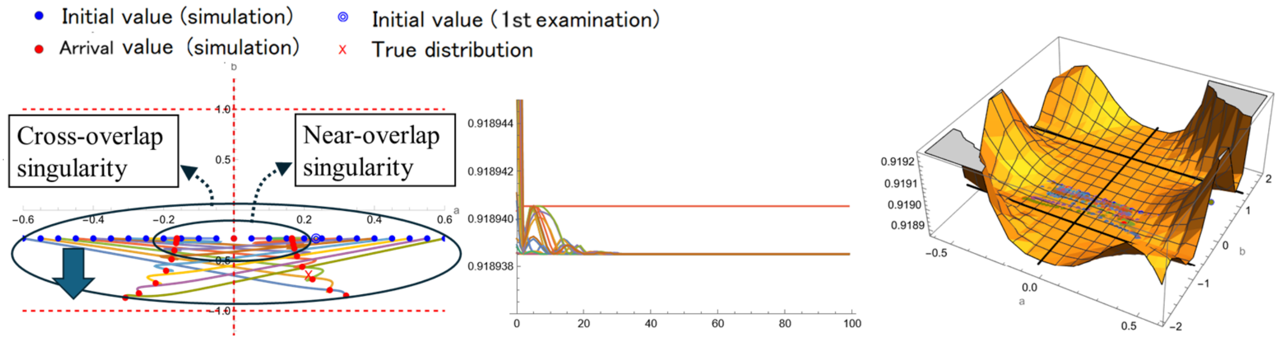

Figure 3: a varied: correct answer

Learning begins from the state $b = -0.28$, where the proportion of student group (iii) is slightly greater than that of student group (i). As the overgeneralization of combinations as permutations increases (a decreases), a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = 0.16$. Furthermore, an overlap singularity occurs under the influence of the critical line $a = w_2 - w_1 = 0$. The state of the overgeneralization of combinations as permutations is large, weights w_1 and w_2 both equal 0.45, since $v = \frac{w_3 w_1 + w_4 w_2}{w_3 + w_4} = 0.45$, and there is no difference from the overgeneralization of permutations as combinations. Finally a cross-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.16$. Since the proportion of student group (iii) decreased from $b = -0.28$ to $b = -0.83$, the learning of combinations progresses and a transformation of the permutational schema takes place. This can be considered a natural way of understanding in the developmental process.

The dynamics of the simulation with b varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 4.

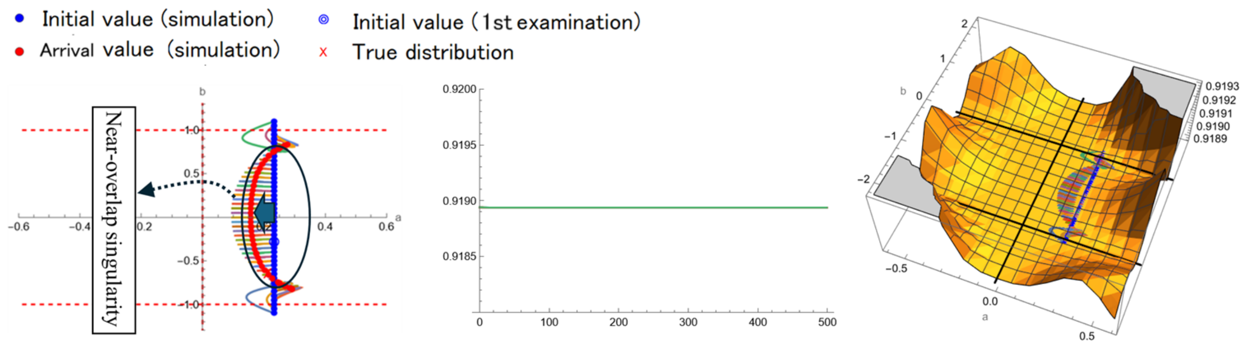

Figure 4: b varied: correct answer

In this case, learning starts from $a = 0.233$, a state in which the overgeneralization of permutations as combinations is highly large. As the proportion of student group (iii) increases (b decreases), a becomes small when $b < -0.74, b > 0.74$ and converges (fast convergence) without the influence of the critical line $a = 0$. From $b = -0.28$ to $b = -0.74$, the difference from the overgeneralization of permutations as combinations becomes smaller, so that a positive transition occurs. From $b = -0.74$ to $b = -0.83$, the difference from the overgeneralization of permutations as combinations becomes even bigger, leading to a negative transition. When the proportion of student group (i) increases (b increases), a becomes small within $-0.74 \leq b \leq 0.74$ and a near-overlap singularity occurs. A positive transition occurs as the difference from the overgeneralization becomes smaller.

3.2 Analysis of students: learning stage (2)

We analyzed the change from the 1st to the 2nd examination to target the students who gave semi-correct answers. We can now consider the process of understanding the two concepts by comparing student groups (i) and (ii).

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

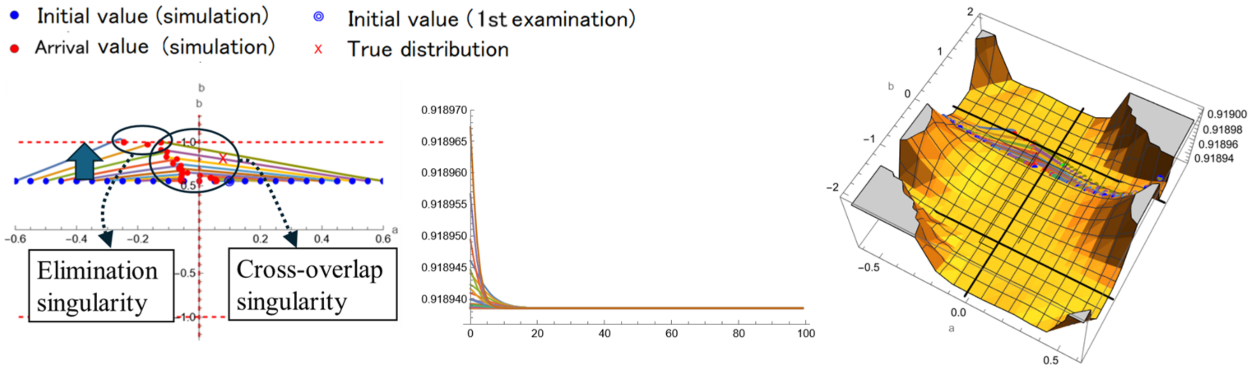

Figure 5: a varied: semi-correct answer

Learning begins from the state ($b = 0.55$), where the proportion of student group (i) is larger than the proportion of student group (ii). When the overgeneralization of combinations as permutations increases (a decreases), the critical line $b = 1$ is affected and an elimination singularity occurs. Only student group (i) is included; there is no student group (ii). If the overgeneralization of combinations as permutations increases (a is smaller than -0.5), the learning of combinations does not progress and a transformation of the permutation schema takes place. As the overgeneralization of permutations as combinations increases (a increases), a near-overlap singularity occurs due to the influence of the critical line $a = 0$, the difference in overgeneralization disappears from $v = 0.25$ (overlap singularity), and a cross-overlap singularity occurs beyond the critical line $a = 0$. Since the proportion of student group (i) increases slightly, the learning of combinations progresses and a transformation of the permutational schema takes place.

The dynamics of the simulation with b varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 6.

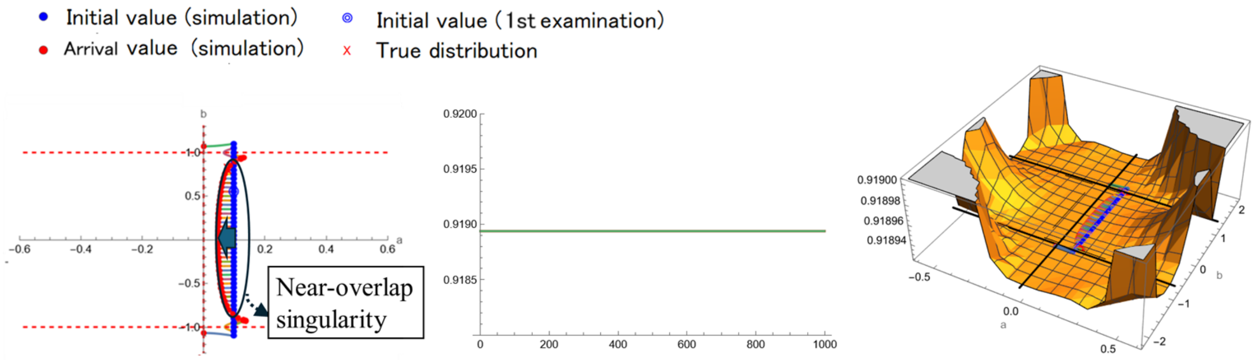

Figure 6: b varied: semi-correct answer

Learning starts from the state $a = 0.09$, where the overgeneralization of permutations as combinations is large. First, as the proportion of student group (i) becomes larger (b increases), an overlap singularity occurs under the influence of the critical line $a = 0$. When the overgeneralization of permutations as combinations is large, weights w_1 and w_2 both equal 0.25, since $v = 0.25$, and there is no difference from the overgeneralization of combinations as permutations. Therefore, a positive transition occurs. When the proportion of student group (ii) becomes larger (b decreases), the same transition as when the proportion of student group (i) becomes larger (b increases) occurs.

3.3 Analysis of students: learning stage (3)

In order to target students who are able to answer semi-correctly, we analyzed changes between the 2nd and 3rd examinations. We can now consider the process of understanding the two concepts by comparing student groups (i) and (ii).

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.

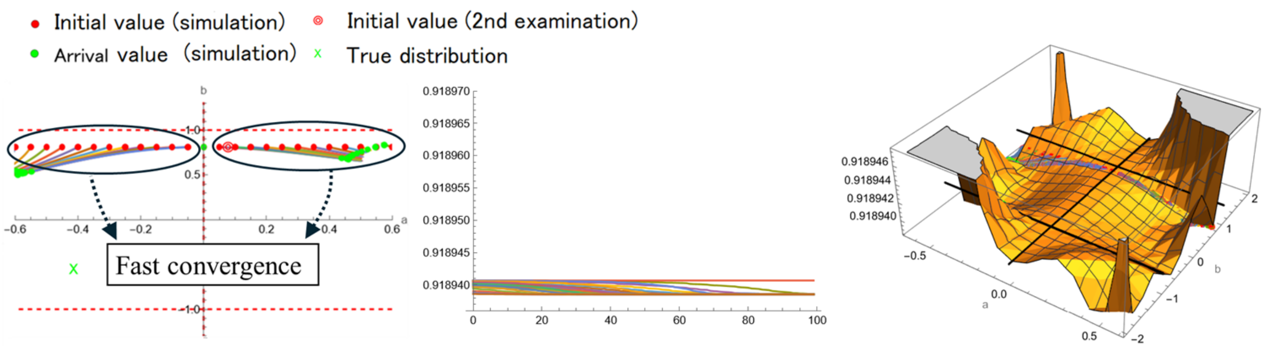

Figure 7: a varied: correct answer

Learning begins from the state ($b = 0.80$), where the proportion of student group (i) is larger than the proportion of student group (ii). The overgeneralization of combinations as permutations increases (a decreases) and converges at $a = -0.57$ (fast convergence) without the influence of the critical line $a = 0$. Since the proportion of student group (i) decreased from $b = 0.80$ to $b = 0.49$, a transformation of the combinational schema takes place. As the overgeneralization of permutations as combinations increases (a increases), the same transformation as when the overgeneralization of combinations as permutations increases (a decreases) and fast convergence occur. The proportion of student group (ii) increased slightly.

The dynamics of the simulation with b varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 8.

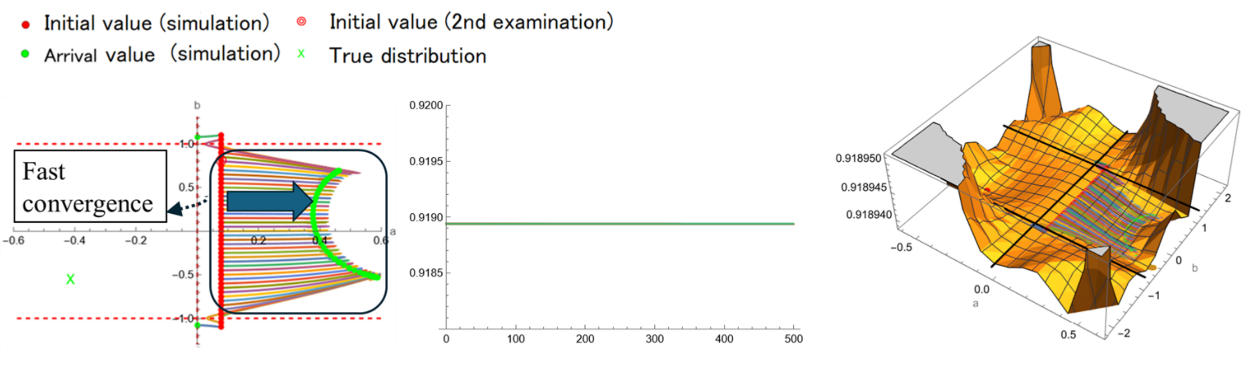

Figure 8: b varied: correct answer

In this case, learning starts from $a = 0.03$, a state in which the overgeneralization of permutations as combinations is small. As the proportion of student group (ii) increases (b decreases), fast convergence occurs without the influence of the critical line $a = 0$. The overgeneralization of permutations as combinations is even greater, so that a negative transition occurs. As the proportion of student group (ii) increases (b increases), fast convergence occurs without the influence of the critical line $a = 0$. A large overgeneralization of permutations as combinations is suppressed, so that a negative transition occurs. By learning combinations (relearning permutations), the overgeneralization of permutations as combinations can be suppressed from $a = 0.46(0.58)$ to $a = 0.37$.

3.4 Analysis of students: learning stage (4)

In order to target students who are able to answer correctly, we analyzed changes between the 2nd and 3rd examinations. We can now consider the process of understanding the two concepts by comparing student groups (ii) and (iii).

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

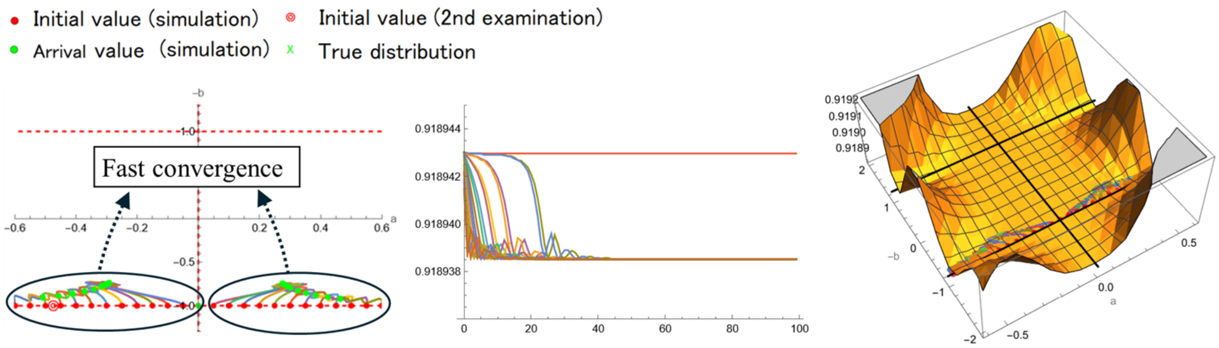

Figure 9: a varied: semi-correct answer

Learning starts from $b = 1$ when only student group (iii) is included. As the overgeneralization of combinations as permutations increases ($a > -0.3$ decreases), fast convergence occurs at $a = -0.29$ without the influence of the critical line $a = 0$. Moreover, the overgeneralization of combinations as permutations increases ($a < -0.3$ decreases) and a near-overlap singularity occurs due to the influence of the critical line $a = 0$. Since the proportion of student group (iii) decreased from $b = 1.0$ to $b = 0.73$, a transformation of the combinational schema takes place. As the overgeneralization of permutations as combinations increases (a increases), the same transformation as when the overgeneralization of combinations as permutations increases (a decreases) and fast convergence occurs at $a = -0.29$.

The dynamics of the simulation with b varied, the change in training loss, and the dynamics on the loss surface are shown in Figure 10.

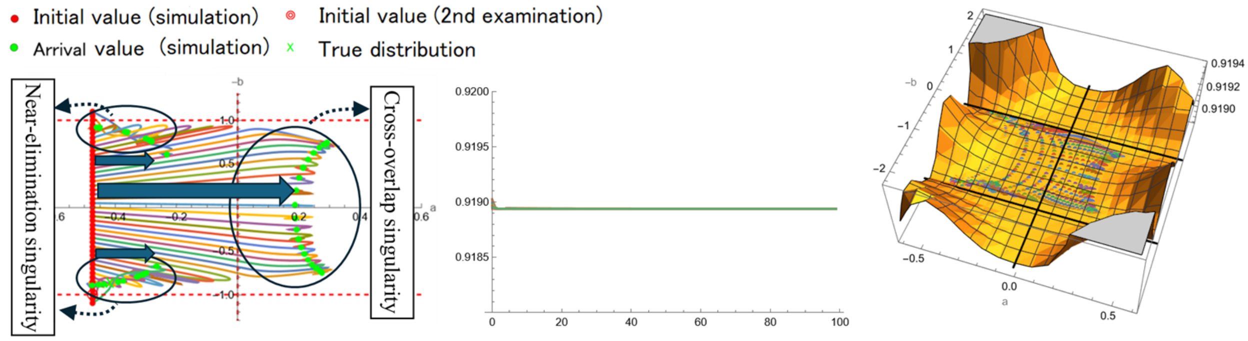

Figure 10: b varied: semi-correct answer

Learning starts from the state $a = -0.47$, where the overgeneralization of combinations as permutations is large. At first, as the proportion of student group (ii) becomes larger (b decreases), 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 combinations as permutations is suppressed. As the proportion of student group (iii) becomes 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 0.47, since $v = 0.47$, and there is no difference from the overgeneralization of combinations as permutations. Furthermore, the overgeneralization of permutations as combinations becoming large leads to a negative transition.

3.5 Direction of teacher guidance

By simulating changes in a and b, the transformation and transition of the schema can be assumed, allowing the teacher to devise a teaching strategy.

3.5.1 Direction of teacher guidance: learning stage (1)

From varying a, by increasing the overgeneralization, a transformation of the schema of permutations is caused to occur, and student understanding of both permutations and combinations progresses. Students answered correctly and were able to identify the problem correctly as a permutation. It appears that understanding is achieved by considering the interrelationship with permutations in trying to understand combinations. From varying b, if we increase the proportion of student group (iii), the overgeneralization of permutations as combinations increases and negative transitions occur. If the proportion of student group (i) is increased, the overgeneralization of combinations as permutations is reduced, and positive transitions occur. Although the problem is answered correctly, the above relationship is affected by the overgeneralization of permutations as combinations. Therefore, even after learning combinations, it is necessary to proceed with the relearning of permutations while still paying attention to the understanding of permutations.

3.5.2 Direction of teacher guidance: learning stage (2)

From varying a, by increasing the overgeneralization, a transformation of the schema of permutations is caused to occur, but student understanding of combinations progresses little. From varying b, if the proportion of student group (ii) is increased, the overgeneralization of permutations as combinations is reduced, and positive transitions occur when student understanding of combinations progresses. Although the problem is answered partially correctly, the above relationship is not affected by the overgeneralization of permutations as combinations. In symbolic comprehension, the proportion of student group (ii) is not increased in the process of understanding the concept of "combinations" when the students are given the opportunity to learn about combinations. Although the students answered partially correctly to the interrelationship question, it can be assumed that the reason for these partially correct answers is not symbolic comprehension. It is necessary to provide guidance for deepening the understanding of combinations by considering the interrelationships between permutations and combinations. Students need to be taught to make correct judgments for problems that require an understanding of these interrelationships.

3.5.3 Direction of teacher guidance: learning stage (3)

By increasing the overgeneralization of permutations as combinations by varying a, the schema of the combinations is changed. However, the proportion of student group (ii) is not increased, so the learning of combinations does not progress. From varying b, increasing the proportion of student group (i) increases the overgeneralization of permutations as combinations, so that a negative transition occurs even as the learning of combinations progresses, but the overgeneralization of permutations as combinations can be slightly suppressed by learning combinations (relearning permutations). Although the problem is answered partially correctly, the above relationship is greatly affected by the overgeneralization of permutations as combinations. This can be considered as result of a superficial understanding of the developmental process. In symbolic comprehension, however, learning combinations does not affect learning permutations, and thus it is considered necessary to learn combinations, paying attention to the understanding of permutations even after learning combinations.

3.5.4 Direction of teacher guidance: learning stage (4)

By increasing the overgeneralization of combinations as permutations by varying a, the schema of combinations is changed, but the learning of both permutations and combinations is improved. From varying b, increasing the proportion of student group (ii) increases the overgeneralization of combinations as permutations, so that a positive transition occurs even as the relearning of combinations progresses. (The levels of understanding of both permutations and combinations are stable.) Additionally, decreasing the proportion of student group (iii) increases the overgeneralization of permutations as combinations, so that a negative transition occurs despite the learning of permutations progressing little. Although the problem is answered correctly, the above relationship may be affected by the overgeneralization of combinations as permutations due to an insufficient understanding of permutations. Students need to be relearning permutations continuously to make correct judgments in problems that require understanding the interrelationships between permutations and combinations while still paying attention to the reduction in understanding permutations due to the influence of the overgeneralization of permutations as combinations.

4 Application to symbolic comprehension in technical colleges

4.1 Preparation for analysis

As an example of the application of the learning system to education at a college of technology (technical college), the academic area of permutations and combinations is 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 chapter, the neural network is trained using the 1st exam results for the fourth-year technical college students (121 people) as initial values and the 2nd exam results for the 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 11 shows the mean scores, number of examinees, and correction coefficients for the correct (two left-most columns, two center columns) and semi-correct (two right-most columns) answers to the interrelationship question (d) 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	0.283333333	0.5	0	0.5	0.305263158	0.4
Number of people e, f	6	28	0	28	19	2
Overall average h_2	0.461764706		0.5		0.314285714	
Total number of people	34		28		21	
Correction factor $\frac{g}{h_2}$	0.984365953		0.960795455		0.825284091	
Correction w_1, w_2	0.278903686	0.492182976	0	0.480397727	0.251928828	0.330113636
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	0.266666667	0.5	0.4	0.5	0.2875	0.1
Number of people e, f	9	45	1	45	8	1
Overall average h_1	0.461111111		0.497826087		0.266666667	
Total number of people	54		46		9	
Correction factor $\frac{g}{h_1}$	0.985761227		0.964991068		0.97265625	
Correction w_1, w_2	0.262869661	0.492880613	0.385996427	0.482495534	0.279638672	0.097265625
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	0.233333333	0.5	0	0.5	0.333333333	0.1
Number of people e, f	3	16	0	16	3	1
Overall average h_1	0.457894737		0.5		0.275	
Total number of people	19		16		4	
Correction factor $\frac{g}{h_1}$	0.992685475		0.960795455		0.943181818	
Correction w_1, w_2	0.231626611	0.496342738	0	0.480397727	0.314393939	0.094318182
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	0.4	0.5	0	0.5	0.25	0
Number of people e, f	2	16	0	16	2	0
Overall average h_2	0.488888889		0.5		0.25	
Total number of people	18		16		2	
Correction factor $\frac{g}{h_2}$	0.929752066		0.960795455		1.0375	
Correction w_1, w_2	0.371900826	0.464876033	0	0.480397727	0.259375	0
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	0.225	0.5	0.4	0.5	0.266666667	0
Number of people e, f	4	13	1	13	3	0
Overall average h_3	0.435294118		0.492857143		0.266666667	
Total number of people	17		14		3	
Correction factor $\frac{g}{h_3}$	1.044226044		0.974720026		0.97265625	
Correction w_1, w_2	0.23495086	0.522113022	0.389888011	0.487360013	0.259375	0

Figure 11: Examination results

4.2 Analysis of students: learning stage (1)

4.2.1 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 12.

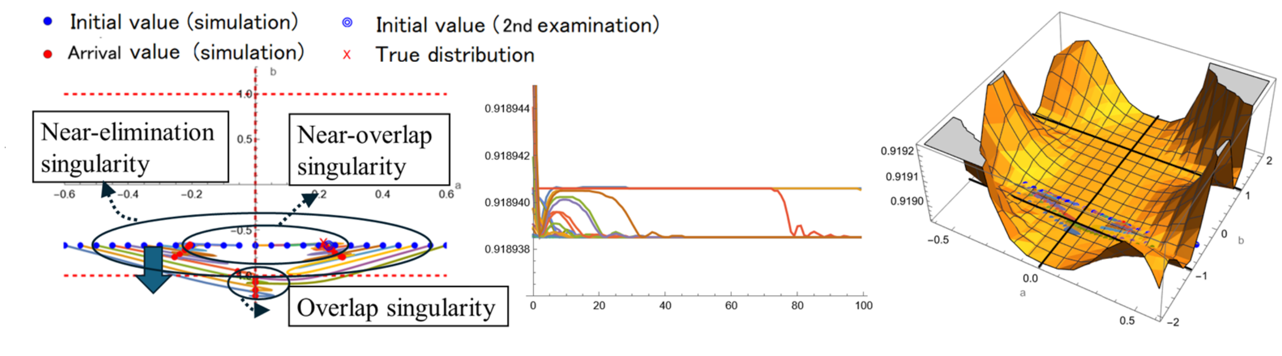

Figure 12: a varied: correct answer

The technical college overall has a smaller initial value of b than the overall high school, with learning beginning from the state $b = -0.66$, where the proportion of student group (iii) is greater than the proportion of student group (i). As the overgeneralization of combinations as permutations increases (a decreases), a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.21$. If a is sufficiently large, then weights w_1 and w_2 both equal 0.45, since $v = 0.45$, and the overgeneralization of combinations as permutations and the overgeneralization of permutations as combinations are equal (overlap singularity). Also, as the overgeneralization of permutations as combinations increases (a increases), a near-overlap singularity again occurs, due to the influence of the critical line $a = 0$, this time stagnating at $a = 0.21$. As the overgeneralization further increases (a increases further), a near-elimination singularity occurs with a large proportion of student group (iii) due to the influence of the critical line $b = -1$. Finally, weights w_1 and w_2 both equal 0.45, since $v = 0.45$, and the overgeneralization of combinations as permutations and the overgeneralization of permutations as combinations are equal (overlap singularity). Since the proportion of student group (iii) has decreased from $b = -0.65$ to $b = -0.79$, the learning of combinations progresses and a transformation of the permutational schema takes place. This can be considered a natural way of understanding in the developmental process.

4.2.2 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 13, comparing the three classes A (left), B (center), and C (right).

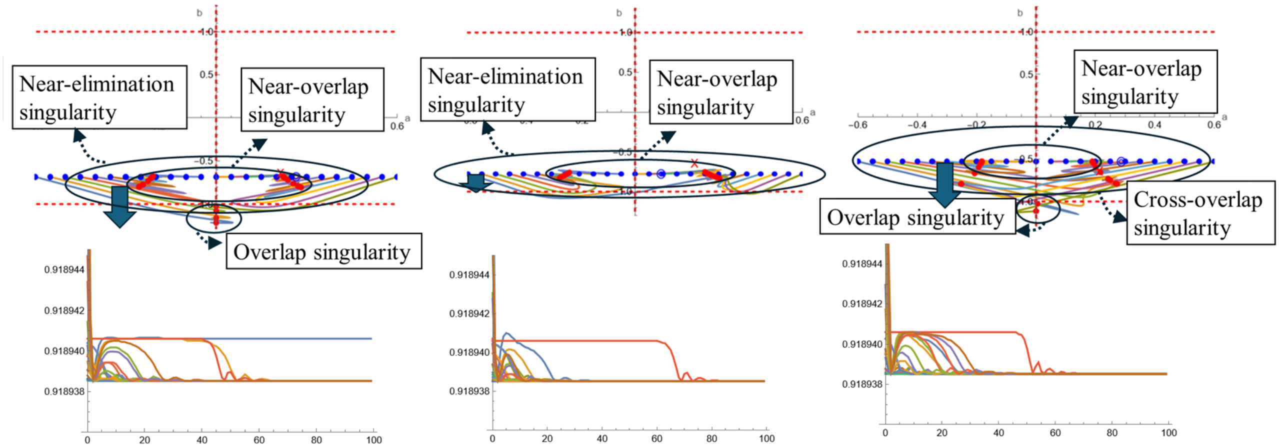

Figure 13: a varied: correct answer

(1) Class A
Learning begins from the state $b = -0.68$, where the proportion of student group (iii) is greater than the proportion of student group (i). As the overgeneralization of combinations as permutations increases (a decreases), a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.21$. If a is sufficiently large, then weights w_1 and w_2 both equal 0.45, since $v = 0.45$, and the overgeneralization of combinations as permutations and the overgeneralization of permutations as combinations are equal (overlap singularity). Also, as the overgeneralization of permutations as combinations increases (a increases), a near-overlap singularity occurs due to the influence of the critical line $a = 0$, in this case stagnating at

$a = 0.21$. As the overgeneralization further increases (a increases further), a near-elimination singularity occurs with a large proportion of student group (iii) due to the influence of the critical line $b = -1$. Since the proportion of student group (iii) has decreased from $b = -0.68$ to $b = -0.79$, the learning of combinations progresses and a transformation of the permutational schema takes place.

(2) Class B
Learning begins from the state $b = -0.77$, where the proportion of student group (iii) is greater than the proportion of student group (i). As the overgeneralization of combinations as permutations increases (a decreases), a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.24$. As the overgeneralization further progresses (a increases), a near-elimination singularity occurs with a large proportion of student group (iii), due to the influence of the critical line $b = -1$. Even if a is sufficiently large, the overgeneralization is not equal (overlap singularity). Also, as the overgeneralization of permutations as combinations increases (a increases), there is stagnating at $a = 0.28$. Since the proportion of student group (iii) does not change, no transformation of the permutation schema occurs.

(3) Class C
Learning begins from the state $b = -0.52$, where the proportion of student group (iii) is greater than the proportion of student group (i). As the overgeneralization of combinations as permutations increases (a decreases), a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.18$. Further, a cross-overlap singularity phenomenon occurs, stagnating at $a = 0.25$. Finally, weights w_1 and w_2 both equal 0.45, since $v = 0.45$, and the overgeneralization of combinations as permutations and the overgeneralization of permutations as combinations are equal (overlap singularity). Also, as the overgeneralization of permutations as combinations increases (a increases), there is stagnating at $a = -0.25, 0.18$. Since the proportion of student group (iii) has decreased from $b = -0.52$ to $b = -0.78$, the learning of combinations progresses and a transformation of the permutational schema occurs.

4.2.3 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 14.

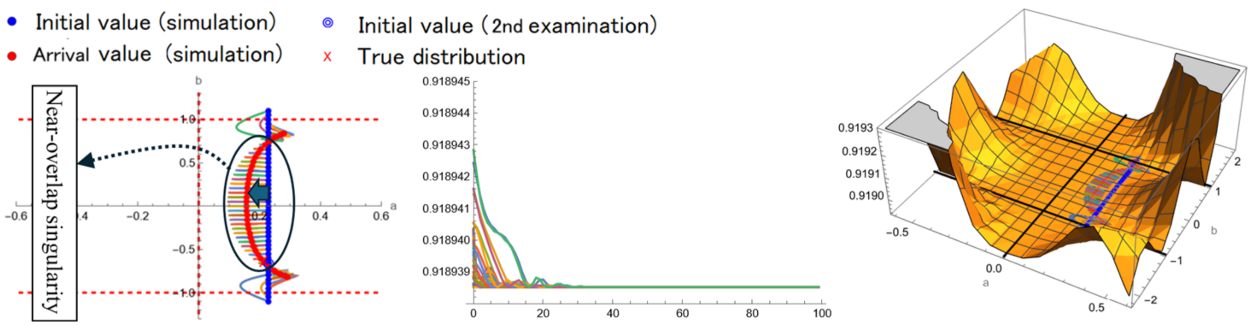

Figure 14: b varied: correct answer

The overall technical college has a bigger initial value of a than the overall high school, with learning beginning from the state $a = 0.230$, a state in which the overgeneralization of permutations as combinations is large. As b changes, a becomes small within $-0.76 \leq b \leq 0.76$ and a near-overlap singularity occurs. A positive transition occurs as the difference from the overgeneralization that occurs becomes smaller. As b continues to change, a stagnates, this time when $b < -0.76$, $0.76 < b$.

4.2.4 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 15, comparing the three classes A (left), B (center), and C (right).

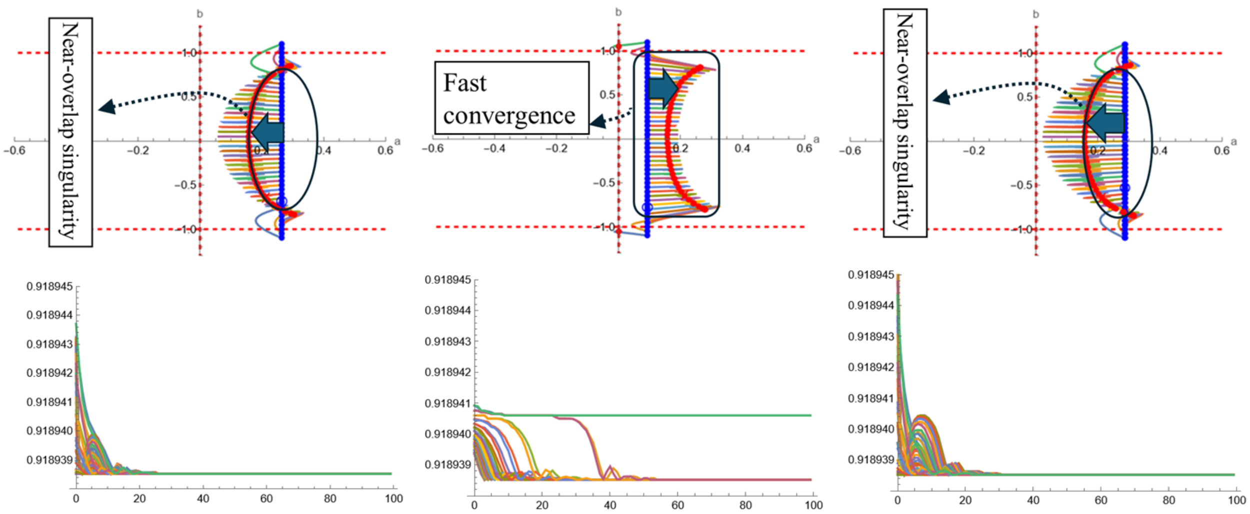

Figure 15: b varied: correct answer

(1) Class A
Learning begins from the state $a = 0.26$, a state in which the overgeneralization of permutations as combinations is large. As b changes, a becomes small within $-0.80 \leq b \leq 0.80$ and a near-overlap singularity occurs. A positive transition occurs as the difference from the overgeneralization that occurs becomes smaller. As b continues to change, this time a stagnates when $b < -0.80$, $0.80 < b$.

(2) Class B
Learning begins from the state $a = 0.09$, a state in which the overgeneralization of permutations as combinations is somewhat large. As b changes, a becomes large and converges without the influence of the critical line $a = 0$ (fast convergence). Overgeneralization of permutations as combinations is even bigger, so that a negative transition occurs.

(3) Class C
Learning begins from the state $a = 0.28$, a state in which the overgeneralization of permutations as combinations is large. As b changes, a becomes small within $-0.82 \leq b \leq 0.82$ and a near-overlap singularity occurs. A positive transition occurs as the difference from the overgeneralization that occurs becomes smaller. As b continues to change, this time a stagnates when $b < -0.82$, $0.82 < b$.

4.3 Analysis of students: learning stage (2)

4.3.1 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 16.

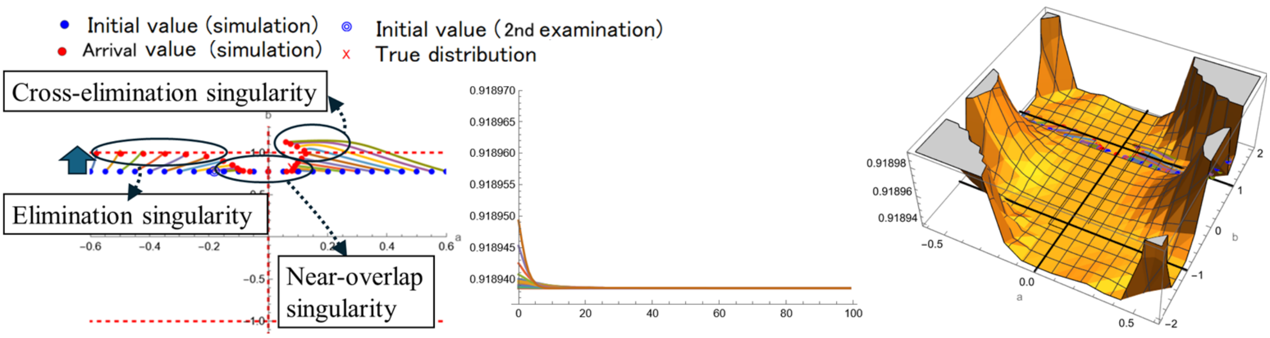

Figure 16: a varied: semi-correct answer

The overall technical college has a bigger initial value of b than the overall high school, with learning beginning from the state $b = 0.77$, where the proportion of student group (i) is greater than the proportion of student group (ii). When the overgeneralization changes, a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = \pm 0.09$. Furthermore, an elimination singularity occurs due to the influence of the critical line $b = -1$ when $a < -0.4, 0.35 < a$, resulting in only student group (i). In particular, as the overgeneralization of combinations as permutations increases (a decreases), the overgeneralization is not suppressed, and also as the overgeneralization of permutations as combinations increases (a increases), and then a cross-elimination singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = 0.07$. Since the proportion of student group (ii) has decreased from $b = 1$ to $b = 0.77$, the learning of combinations does not progress and a transformation of the permutational schema takes place.

4.3.2 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 17, comparing the three classes A (left), and B and C (right).

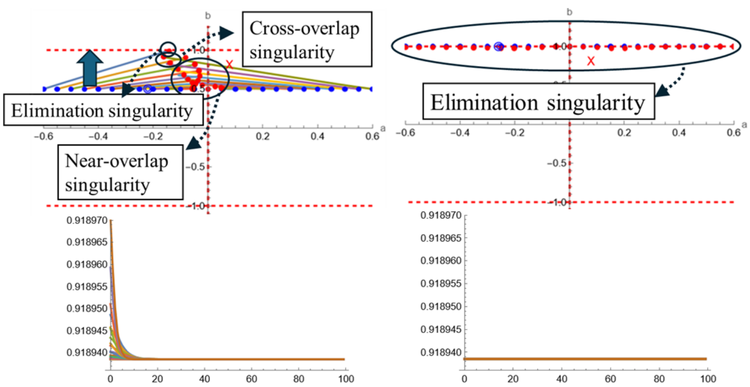

Figure 17: a varied: semi-correct answer

(1) Class A
Learning begins from the state $b = 0.5$, where the proportion of student group (i) is greater than the proportion of student group (ii). When the overgeneralization of combinations as permutations changes, a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.03$ and finally an elimination singularity occurs due to the influence of the critical line $b = 1$. When the overgeneralization of permutations as combinations changes, a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = 0.03$ and finally a cross-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.08$. Since the proportion of student group (ii) has decreased from $b = 1.0$ to $b = 0.5$, the learning of combinations does not progress and a transformation of the permutational schema takes place.

(2) Classes B and C
Learning begins from the state $b = 1.0$, which means which means learning is only in student group (i). As b changes, an elimination singularity occurs, which is only in student group (i), due to the critical line $b = 1$. The learning of combinations does not progress and a transformation of the permutational schema takes place.

4.3.3 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 18.

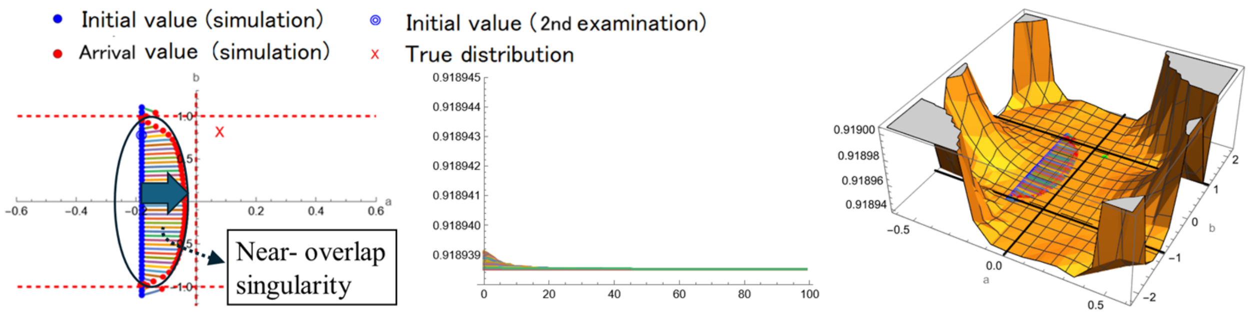

Figure 18: b varied: semi-correct answer

The overall technical college has a bigger initial value of b than the overall high school, with learning beginning from the state $a = -0.18$, where the overgeneralization of combinations as permutations is large. As b changes, a near-overlap singularity occurs due to the influence of the critical line $a = 0$, stagnating at $a = -0.03$. The overgeneralization of combinations as permutations becomes smaller, and therefore a positive transition occurs.

4.3.4 Comparing three classes

The dynamics of the simulation and the changes in learning loss are shown in Figure 19, comparing the three classes A (left), and B and C (right).

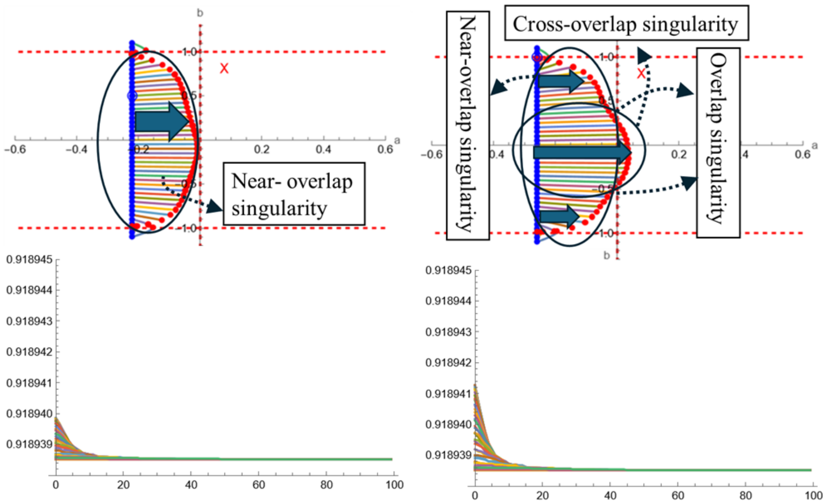

Figure 19: b varied: semi-correct answer

(1) Class A
Learning begins from the state $a = -0.22$, 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 $a = -0.01$. The overgeneralization of permutations as combinations becomes smaller, so a positive transition occurs during the learning process.

(2) Classes B and C
Learning begins from the state $a = -0.25$, where the overgeneralization of combinations as permutations is large. As b changes, a near-overlap singularity occurs due to the influence of the critical line $a = 0$ when $b < -0.48, 0.28 < b$. When $b = -0.48, 0.28$, weights w_1 and w_2 both equal 0.26, since $v = 0.26$, and the overgeneralization of combinations as permutations and the overgeneralization of permutations as combinations are equal (overlap singularity). The overgeneralization of combinations as permutations becomes smaller, so positive transitions occur. Furthermore, a cross-overlap singularity occurs, stagnating at $a = 0.035$ when $-0.48 < b < 0.28$. In addition, the overgeneralization of permutations as combinations becomes large, which leads to a negative transition.

4.4 Overall analysis of technical college and comparison of three classes

4.4.1 Correct answer

For the overall technical college, by changing the overgeneralization, a transformation of the schema of permutations occurs, and the learning of both permutations and combinations progresses. By changing the proportions of student groups, the overgeneralization of permutations as combinations becomes small, and positive transitions occur. If the proportion of student group (iii) increases, the overgeneralization of combinations as permutations becomes small, and therefore apparently correct answers are unlikely to occur.

For class A, if the overgeneralization of combinations as permutations is sufficiently large, all students are in student group (iii). By changing the proportions of any student group, the overgeneralization of permutations as combinations becomes small and positive transitions occur. For class B, even if the overgeneralization is sufficiently large, not everyone will be in student group (iii). By changing the proportions of any student group, the overgeneralization of permutations as combinations becomes big and negative transitions occur. For class C, if the overgeneralization is sufficiently large, all are in student group (iii). By changing the proportions of any student group, the overgeneralization of permutations as combinations becomes small and positive transitions occur.

In classes A, learning should be promoted so that symbolic comprehension of combinations increases by the influence of the overgeneralization of combinations as permutations. In class B, there is little change in the proportion of any student group during the learning process, but the learning proceeds with attention to the overgeneralization of permutations as combinations, and with consideration of the interrelationships. In class C, learning should be promoted so that understanding of permutations does not deteriorate with attention to the overgeneralization of permutations as combinations.

4.4.2 Semi-correct answer

For the overall technical college, learning begins where students with semi-correct comprehension exist. If the change in overgeneralization is sufficiently large, all students are in student group (i). Therefore, a transformation of the schema of permutations occurs. Even if the overgeneralization of permutations as combinations increases, the proportion of student group (ii) is not increased. Also, it is difficult to consider this as a possible error in the developmental process, since the understanding of combinations progresses little. If the proportion of student group (ii) increases, the overgeneralization of combinations as permutations is suppressed and positive transitions occur as the understanding of combinations progresses.

For class A, learning begins where the overgeneralization of combinations as permutations is large and class A includes students who are semi-correct. If the proportion of student group (ii) increases, the overgeneralization of combinations as permutations becomes small. Thus, positive transitions occur as the understanding of combinations progresses. If the overgeneralization of permutations as combinations increases, the opposite overgeneralization becomes big and the proportion of student group (ii) decreases. For classes B and C, learning begins where the overgeneralization of combinations as permutations is large. If the proportion of student group (ii) increases, the overgeneralization of combinations as permutations becomes small. Thus, positive transitions occur as the understanding of combinations progresses. If the proportion of student group (i) is close to the proportion of student group (ii), then the opposite overgeneralization becomes big, which may be affected by the overgeneralization of permutations as combinations. Students for whom the semi-correct factor is symbolic comprehension may be included.

In class A, learning can be promoted to reduce the number of students for whom the semi-correct factor is symbolic comprehension by increasing the overgeneralization of permutations as combinations. Since the overgeneralization of combinations as permutations has little effect, the learning of interrelationships should be promoted. In classes B and C, although no students have a semi-correct factor in symbolic comprehension, the permutation schema needs to be transformed. In this process, it is necessary to promote the learning of interrelationships with an emphasis on combinations and with attention to the overgeneralization of permutations as combinations.

5 Conclusion

In this paper, a learning system was constructed to simulate a loss surface defined by a neural network. We analyzed results regarding the dynamics on the loss surface and proposed learning guidance in singular regions (overlap singularity phenomenon and elimination singularity phenomenon).

First, we analyzed the understanding of symbols in the two concepts of permutations and combinations in a high school. Next, as an application of the learning system, we analyzed the understanding in a technical college. More students who answered correctly in the technical college understood both permutations and combinations than in the high school. Even with a change in the proportion of those who answered correctly, the change in overgeneralization at the technical college was less than the change in overgeneralization at the high school. Students who answered semi-correctly in the technical college made up a higher proportion of student group (i) than those in the high school. Even with a change in overgeneralization, the change in the proportion of student group (i) at the technical college was less than the change of the proportion at the high school.

Moreover, we compared three classes at the technical college and proposed learning guidance. Further data will be collected to improve the accuracy of the true distribution to ensure the visualization of comprehension.

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)