Academic Statement

Sang-Kyun Ko

Contents

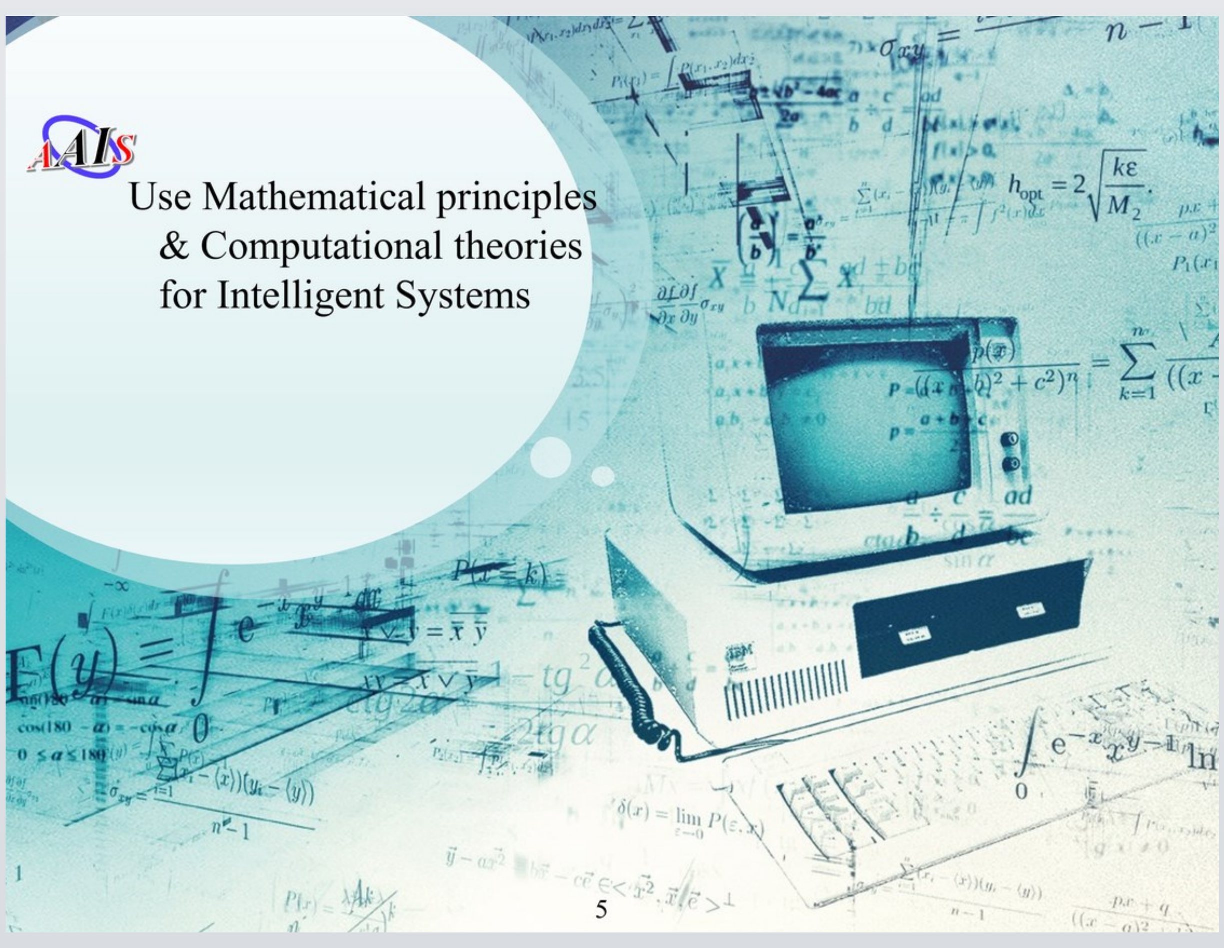
1. Overview of My Research Lab.

Lab. of Advanced Application for Intelligence Systems (AAIS Lab)

Lab. of

Advanced Application for Intelligence Systems

http://aais.hanyang.ac.kr　　　담당교수 : 노영균



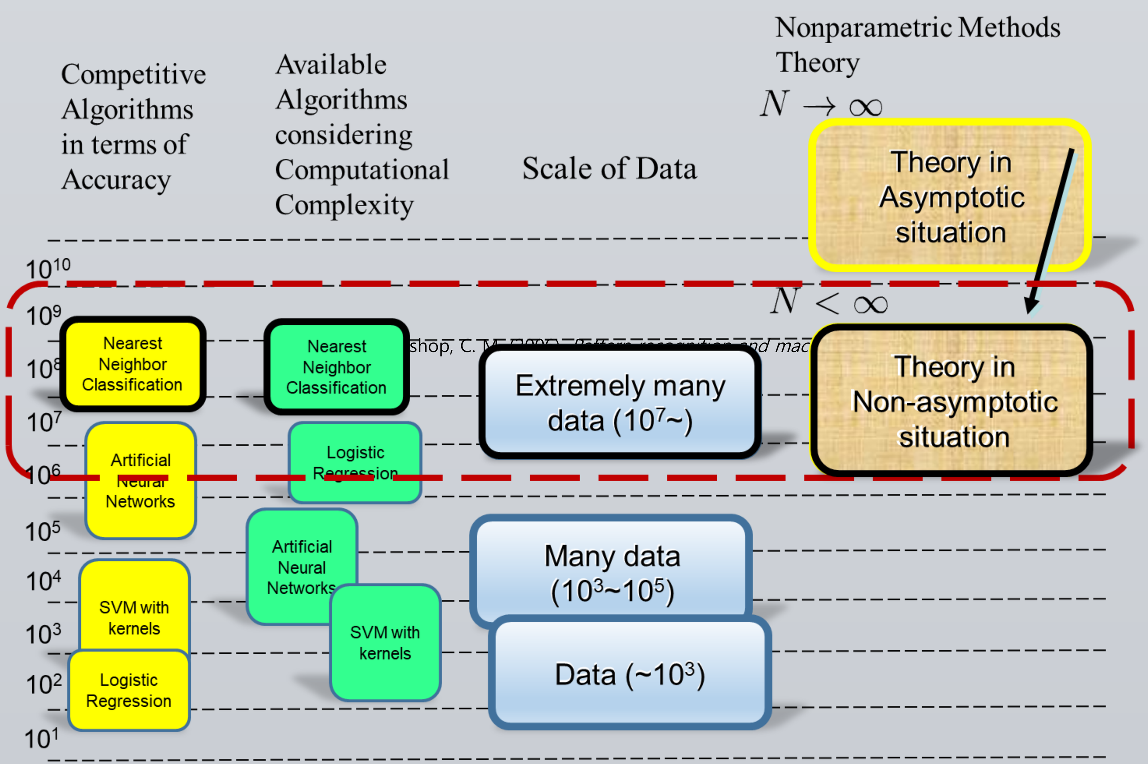
We perform principled research for learning and making intelligence systems. Our interest includes machine learning, artificial intelligence, data science, and mathematical psychology. We consider these disparate fields, individually and in combination, as advanced applications of mathematical principles and computational theory.

The Lab. of Advanced Application for Intelligence Systems started in 2019, and we are recruiting graduate students who are motivated to understand the state-of-the-art deep learning and its applications in diverse real world problems.

Students are welcome who like to solve challenging problems and enjoy to use mathematics in explaining real world problems. Please visit http://aais.hanyang.ac.kr and check out the Zoom meeting schedule.

Exploiting Nonparametric methods

	Neural Networks	Nearest Neighbor (NN) Methods
Consistency (Asymptotic)	Farago & Lugosi, 1993 (Consistency), Hornik et al., 1981 (Universal approximator)	Fix & Hodges, 1951, Cover & Hart, 1967
Non-asymptotic behavior	*Difficult to analyze*	Noh et al., 2018, Chaudhuri & Dasgupta, 2014
State-of-the-art results	Hinton & Salakhutdinov, 2006, Krizhevsky et al., 2012	*Technologies are unmatured yet*
Biological connections	Hebb, 1949, Löwel & Singer, 1992	Shepard, 1967, Nosofsky & Palmeri 1997, Noh et al., 2012



Nonparametric methods consist of basic algorithms such as nearest neighbor methods that are generally known to achieve theoretical minimum errors in asymptotic situation with infinite data. Instead of increasing the number of data to infinite, we try to modify the algorithm to exploit the known information with finite data to achieve the asymptotic results as close as possible. The research is supported by the Samsung Future Technology Foundation (삼성미래육성재단) since 2019.

Uncertainty estimation

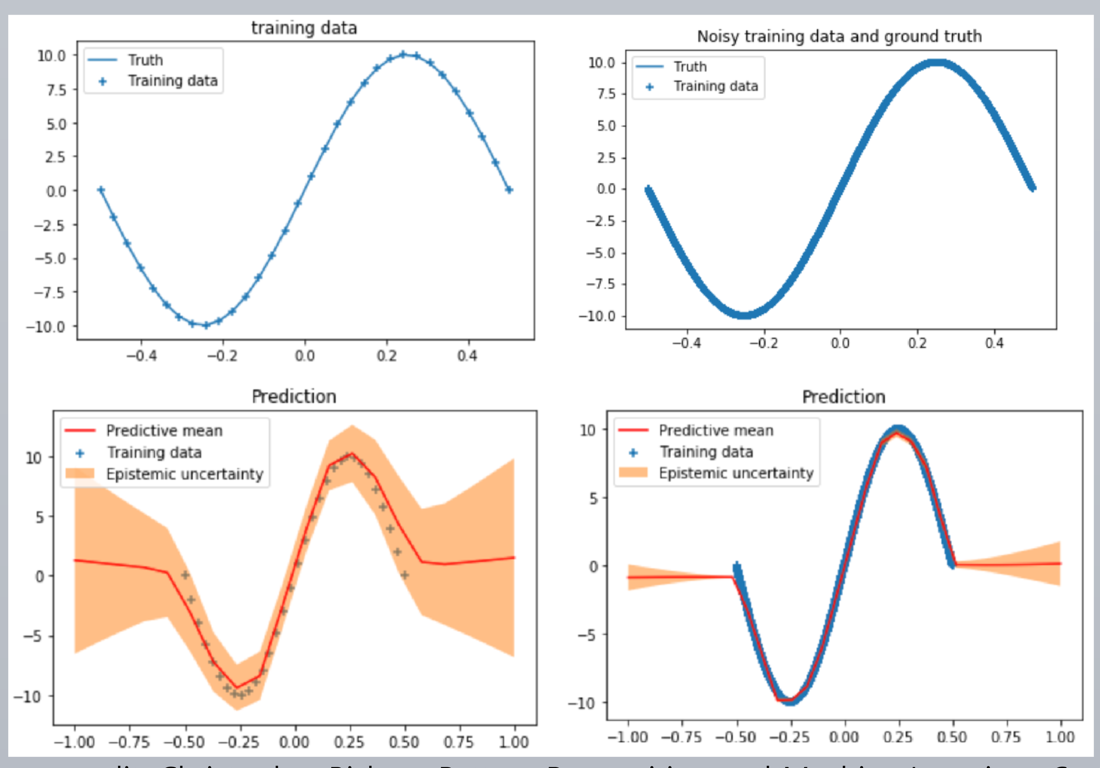


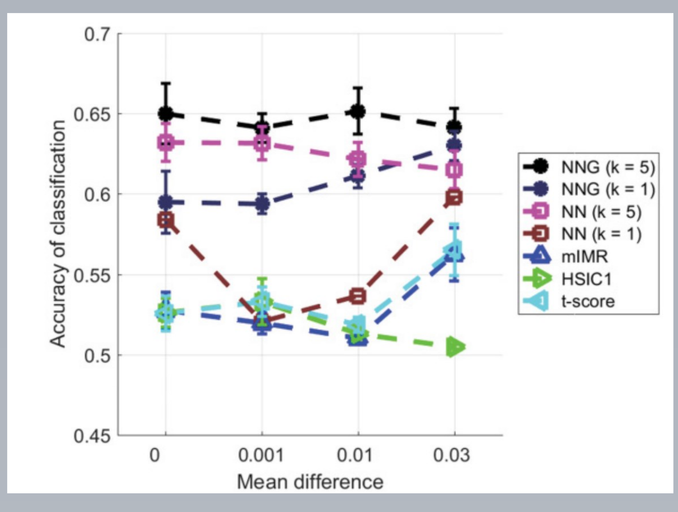
Figure credit: Christopher Bishop, Pattern Recognition and Machine Learning, *Springer*

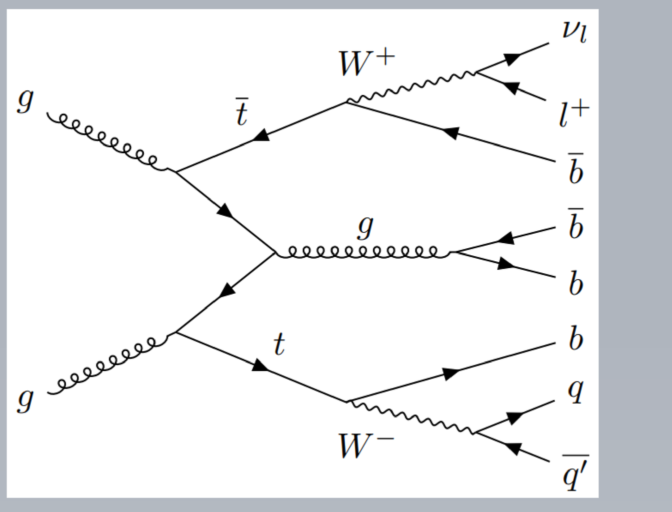
Neural Networks are weak in representing the uncertainty. Bayesian derivation deals with the belief update by data, and the application of Bayesian update naturally provide the uncertainty measure as well as the predictions.

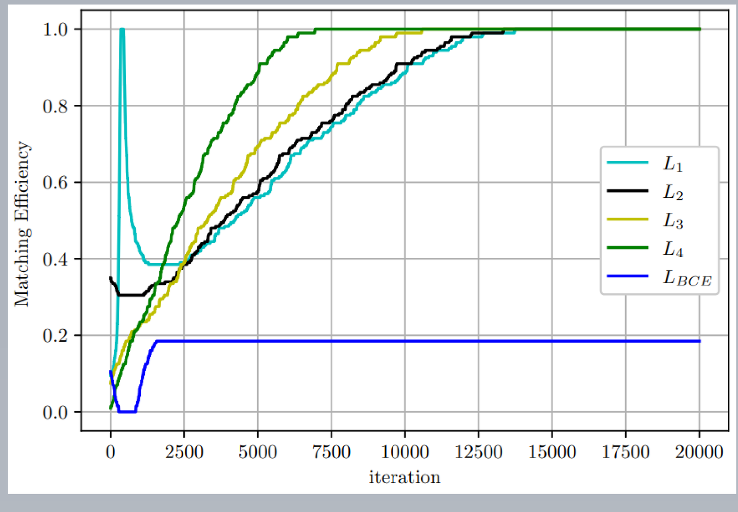
We make derivations for providing uncertainties for medical and manufacturing applications.

AI+X : Medicine, Physics, Drug discovery, and others

We have collaborations with many international groups and application domains. The collaborators include hospitals (Mayo Clinic in Rochester, USA, Korea University Guro Hospital, Kyung Hee University Dental Hospital, and Gyeongsang National University Hospital in Korea), manufacturing companies (Samsung Electronics), basic science research groups (KIAS in Korea, RIKEN-AIP in Japan, Physics department in Hanyang University), and Universities (UCSD, OSU).







Professor, Yung-Kyun Noh

Yung-Kyun Noh is an Assosiate Professor in the Department of Computer Science at Hanyang University. His research interests are learning representations and bias reduction for large-scale nonparametric methods in machine learning. He received his B.S. in Physics from POSTECH and his Ph.D. in Computer Science from Seoul National University. He worked in the GRASP Robotics Laboratory at the University of Pennsylvania as a visiting researcher. He is currently the visiting scientist at the RIKEN Center for Advanced Intelligence Project (RIKEN-AIP) in Japan since 2018, the visiting scholar at Mayo Clinic in USA since 2020, and the associate member of Korea Institute for Advanced Study (KIAS) since 2019.

면담 정보
http://aais.hanyang.ac.kr

Contact : nohyung@hanyang.ac.kr

2. Teaching Assistant Experience

Teaching Assistant Experience

2019 **Discrete Mathematics (MAT2020)**	**2020** **Automata Theory (ITE3061)**	**2020** **Artificial Intelligence (CSE4007)**	**2021** **Creative Software Programming** **(ITE1015)**

•Contributed to exam question(with solution) design and grading for **MAT2020, ITE3061 and CSE4007** in which I served as a teaching assistant.

•In **CSE4007, ITE1015**, led weekly “Lecture review”sessions following lectures, guiding students through theoretical machine learning problems.
__The lecture review involved directly teaching students.__

Teaching Assistant Experience

2019

Discrete Mathematics (MAT2020)

- A freshman-level course
- Theory-focused course
- Contributed to midterm, final exam and Quiz

MAT2020-12243 Discrete Mathematics
Mid Term
2:00pm - 3:20pm, October 29 (Tue.), 2019

Instructions

- Make sure your argument is clear to get full credit.
- Write down your student id and your name at every page of your solution.
- After finishing the quiz, submit your solution. You do not have to submit the problem set.)

Problem 1 [10pt]
(a) You are given two algorithms (Algorithm 1, Algorithm 2) for searching x in a sorted sequence of distinguished numbers $a_1, a_2, \ldots, a_n$ $(a_1 < a_2 < \ldots < a_n)$. Determine which algorithm is faster than the other considering the worst case when the size of sequence n is big. Explain why. (Note that [... er than or equal

MAT2020-12243. Discrete Mathematics
Fall 2019
Syllabus

Introduction: This course provides a variety of fundamental concepts of discrete mathematics and basic techniques used in computation. Techniques for making formal proofs will be the main topic covered in this lecture. Logics, functions, basic set theory, countability and counting arguments, mathematical induction, combinatorics, discrete probability, recursion, recurrence relation, and number theory will be taught.

Lectures: Lectures will be held on Tuesday from 2:00pm-3:30pm and on Wednesday from 10:30am-12:00pm, in lecture hall 305-207 (ITBT-207). Lectures will be given in Korean.

Teaching Assistant Experience

2020

Automata Theory (ITE3061)

- A Sophomore-level course
- Theory-focused course
- Contributed to midterm, final exam and Quiz

ITE3061-11823. Automata and Theory of Computation
Mid-term
9:00am - 10:15am, May 7 (Thu.), 2020

Instructions

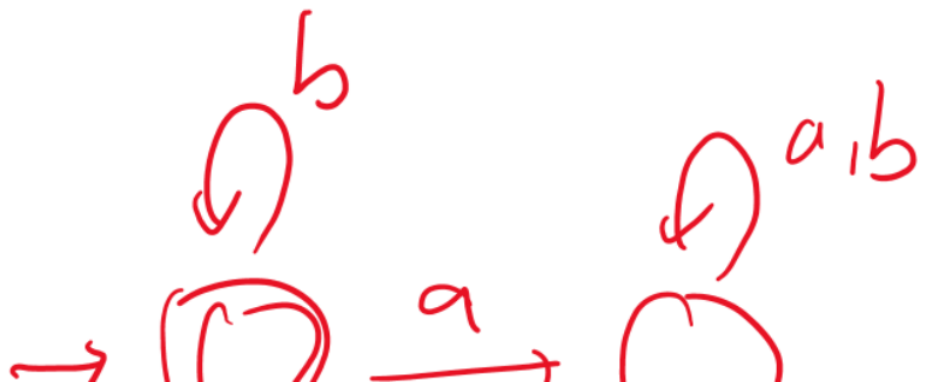
- Make sure your argument is clear to get full credit.
- Write down your student id and your name at every page of your solution.
- After finishing the quiz, please submit the solution. (When you submit the solution, you do not have to hand in the pages with problems.)
- The p

Problem 1 [15pt]

For each of the following languages on $\Sigma = \{a, b\}$, draw a minimal dfa that accepts it.

(a) All strings that have no a's.

Solution: Fig. 1.



Teaching Assistant Experience

2020

Artificial Intelligence (CSE4007)

- A Junior-level course
- Focus on both theoretical and practical aspects
- Contributed to midterm, final exam and“lecture review”
(Led theory problem-solving and practical sessions in lecture reviews)

3 2020 AI lecture Mid-term exam, class 1

CSE4007-11778. Artificial Intelligence
Mid-term solution
Instructor: Yung-Kyun Noh
10:30am - 12:30pm, October 23 (Fri.), 2020

Problem 1 [10pt]

Find the derivative of the following scalar function $f(X) = tr[ABX]$ with respect to the matrix $X \in \mathbb{R}^{p\times m}$. Here, $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ are two different matrices, and $tr[C] = \sum_{i=1}^{k} C_{ii}$ for $C \in \mathbb{R}^{k\times k}$. Use the definition of the matrix derivative, $\left[\frac{df}{dX}\right]_{ij} = \frac{df}{dx_{ij}}$ with x_{ij}, which is the (i, j)-th element of X.

Solution:

The function $f(X)$ can be rewritten as $f(X) = tr[ABX] = \sum_a \sum_b \sum_c A_{ab}B_{bc}X_{ca}$, and the (p, q)-th element of $\frac{df}{dx_{pq}}$ can be written as

(43)

CSE 4007-11779 Artificial Intelligence
Fall 2020
Syllabus

Introduction: This course provides a variety of fundamental mathematical concepts and basic techniques used in the applications of artificial intelligence. It is aimed at giving an understanding of how to build a mathematical formulation for the concept of learning-from-data. Students who are interested in application of mathematics to the modern artificial intelligence problems will need a good understanding of the basic ideas introduced here. Major topics to be covered in the class include optimization for learning, linear models, and probabilistic inference.

Teaching Assistant Experience

2021

Creative Software Programming (ITE1015)

•A Sophomore-level course

•Studied C++ theory, followed by hands-on practice in Ubuntu

•Contributed to“lecture review”
(Led practical sessions in lecture reviews)

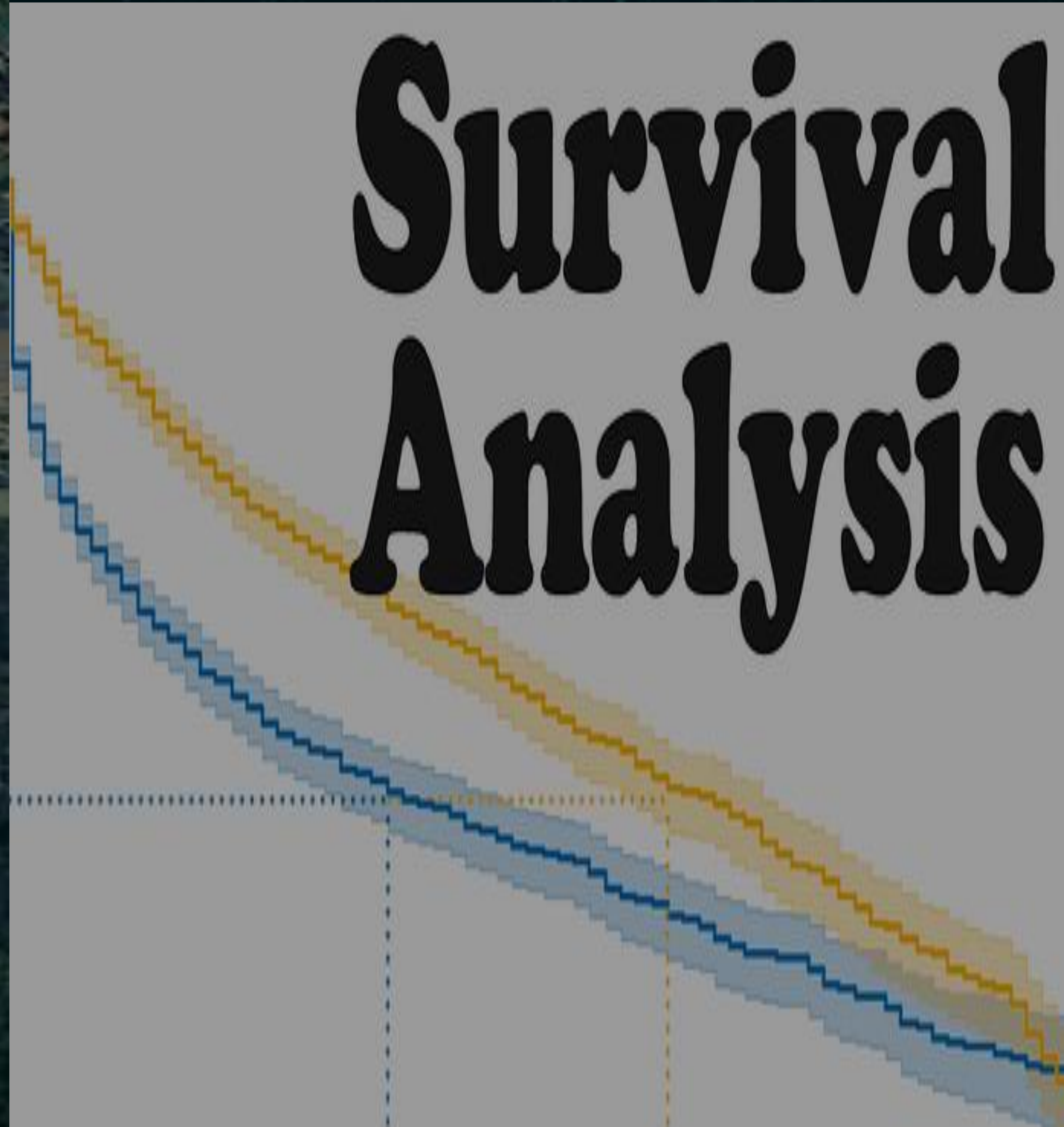
Creative Software Programming, Lab 9-1

1. Write a program that works as follows:
   A. Class C inherits from class B, class B inherits from class A.
   B. Each class has a public member function test().
      i. A::test() returns a string "A::test()".
      ii. B::test() returns a string "B::test()".
      iii. C::test() returns a string "C::test()".
   C. Create objects of class A, B, and C by new operator and put them into std::vector<A*> arr.
   D. Call the test() function of each element of arr to show the execution result as shown below. Each element of arr must be deallocated after use.
   E. Do nc
   F. Input:
   G. Outpu
   H. Files
      i. A
      ii. A

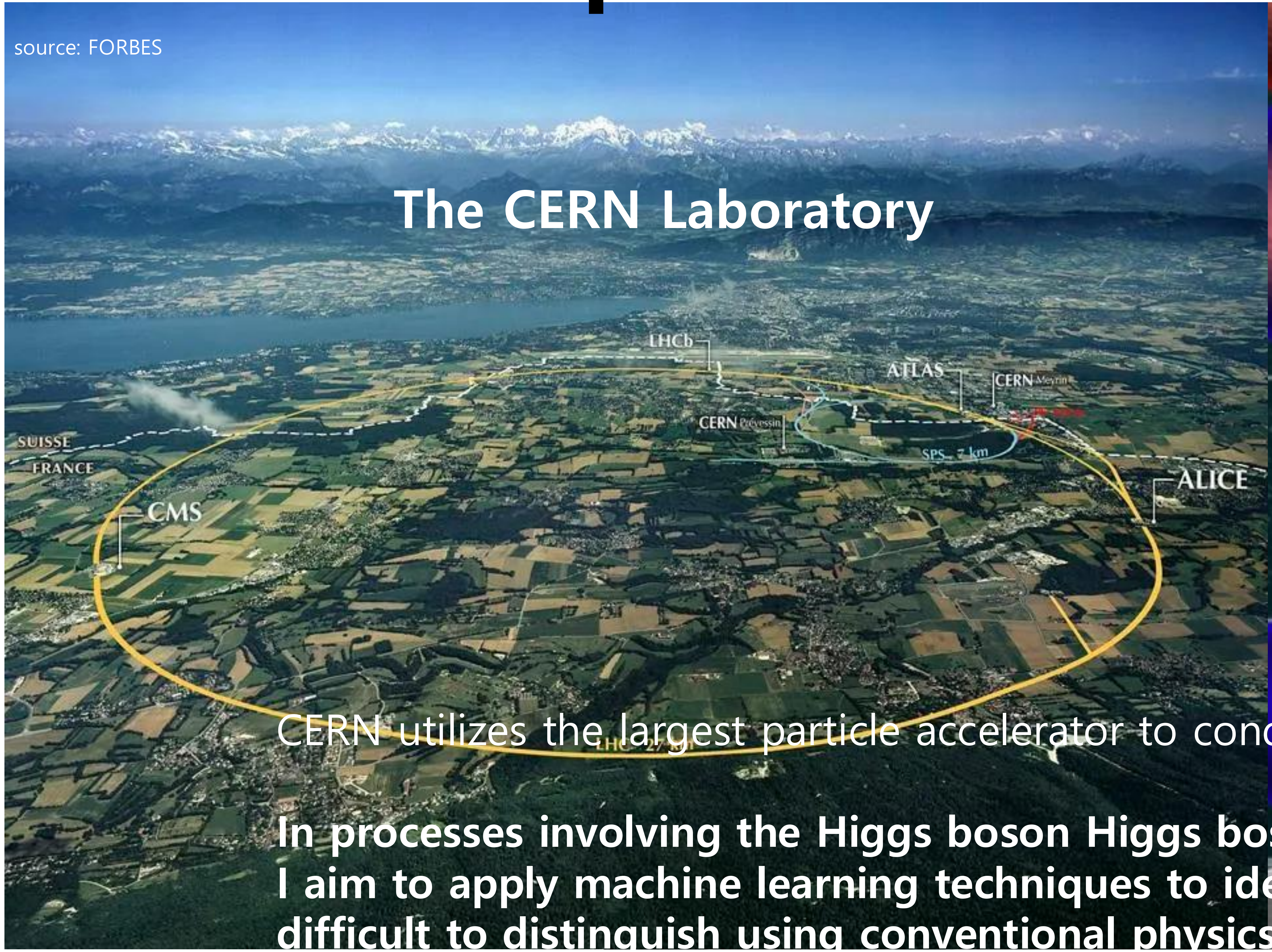
Course Overview

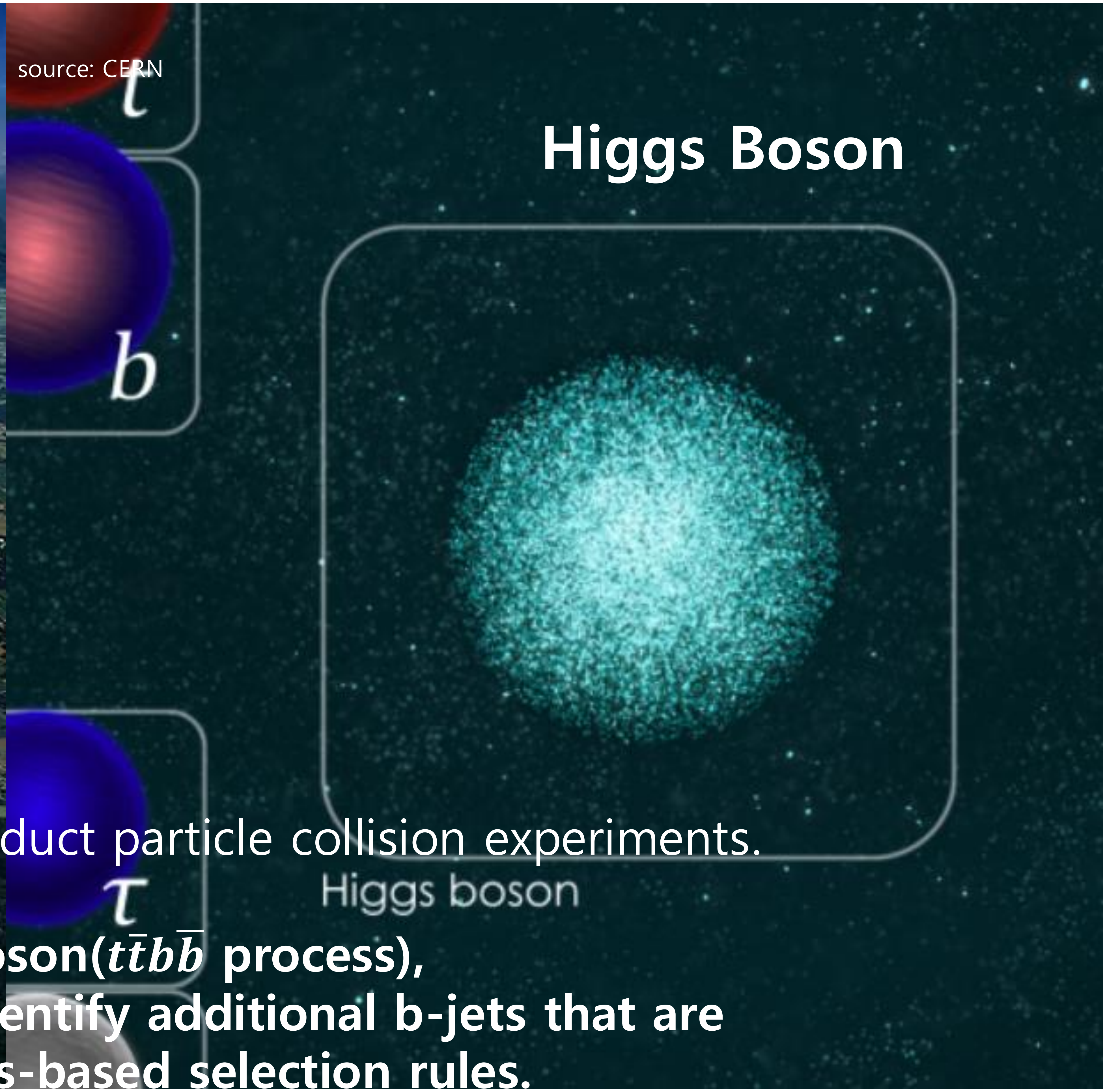
- In this course, you will
  - Learn the fundamentals of C++ language
    - key concepts of object-oriented programming such as classes, inheritance, and polymorphism
    - references, pointers, dynamic allocation
  - Practice programming skills by writing many exercise programs
  - Practice using development tools and editors in Unix/Linux environment.

3. Research Experience



Research Experience



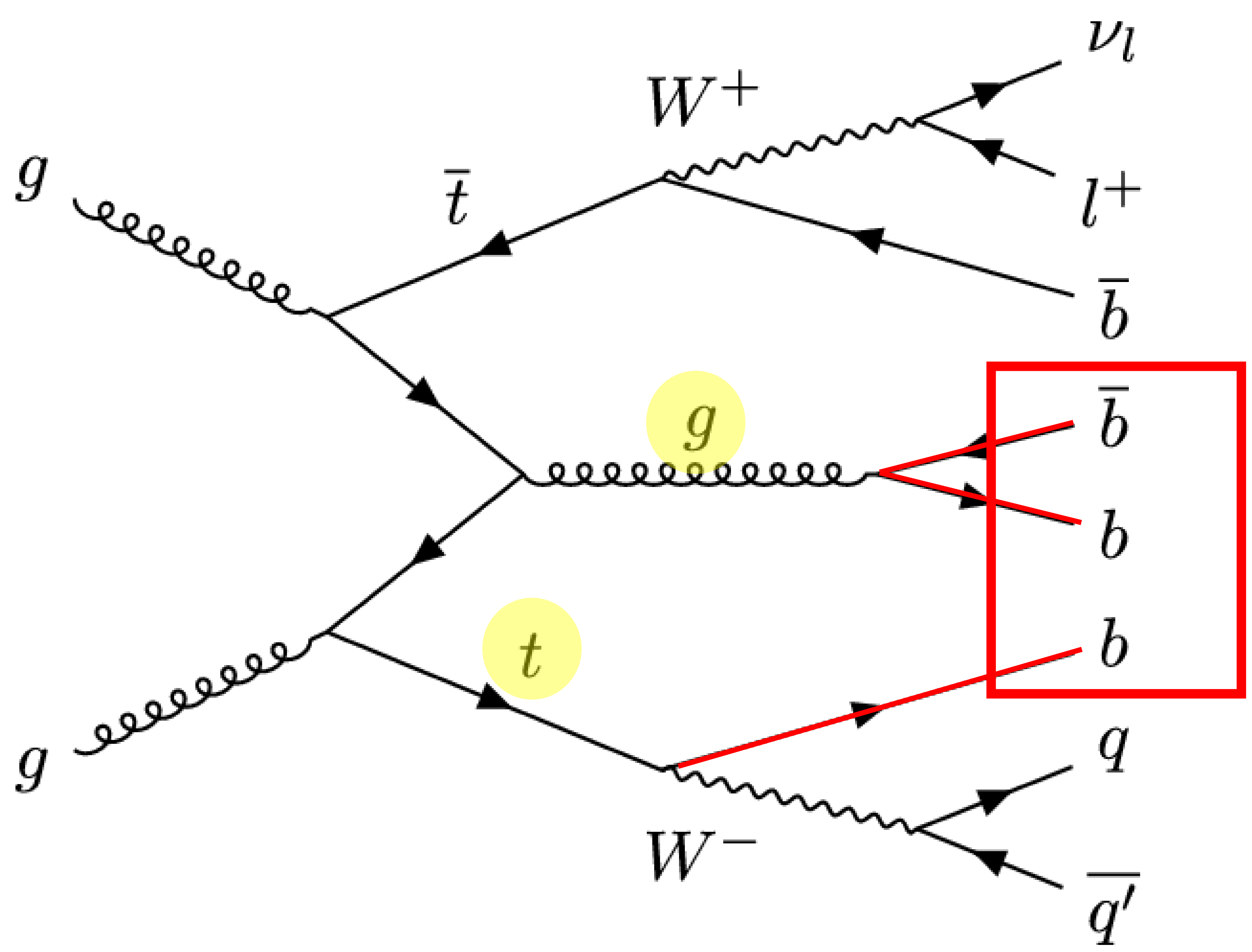


CERN utilizes the largest particle accelerator to conduct particle collision experiments.

In processes involving the Higgs boson Higgs boson($t\bar{t}b\bar{b}$ process),
I aim to apply machine learning techniques to identify additional b-jets that are difficult to distinguish using conventional physics-based selection rules.

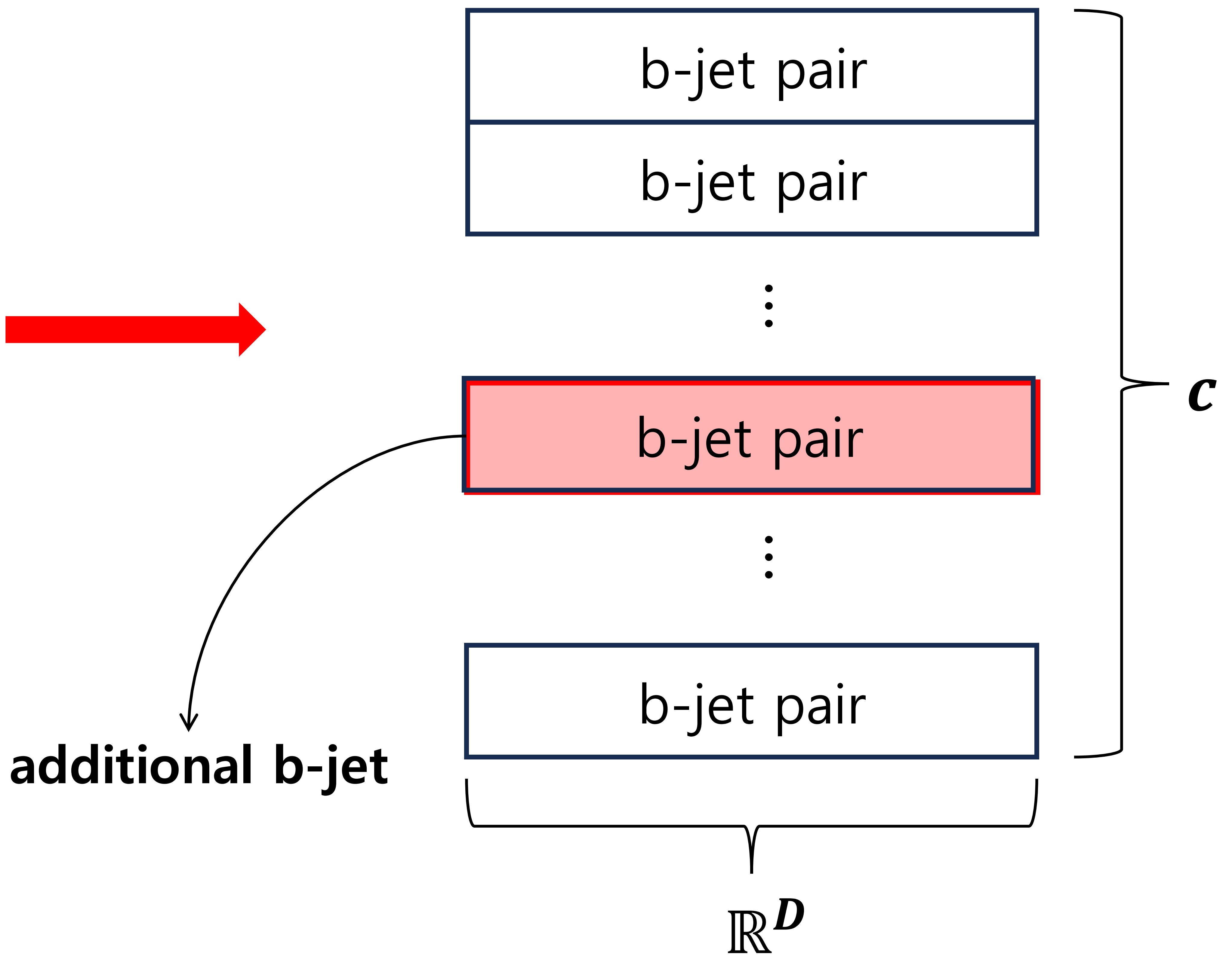
Research Experience

1. Identifying additional b-jets in $t\bar{t}b\bar{b}$ process



Feynman diagram for $t\bar{t}b\bar{b}$ process

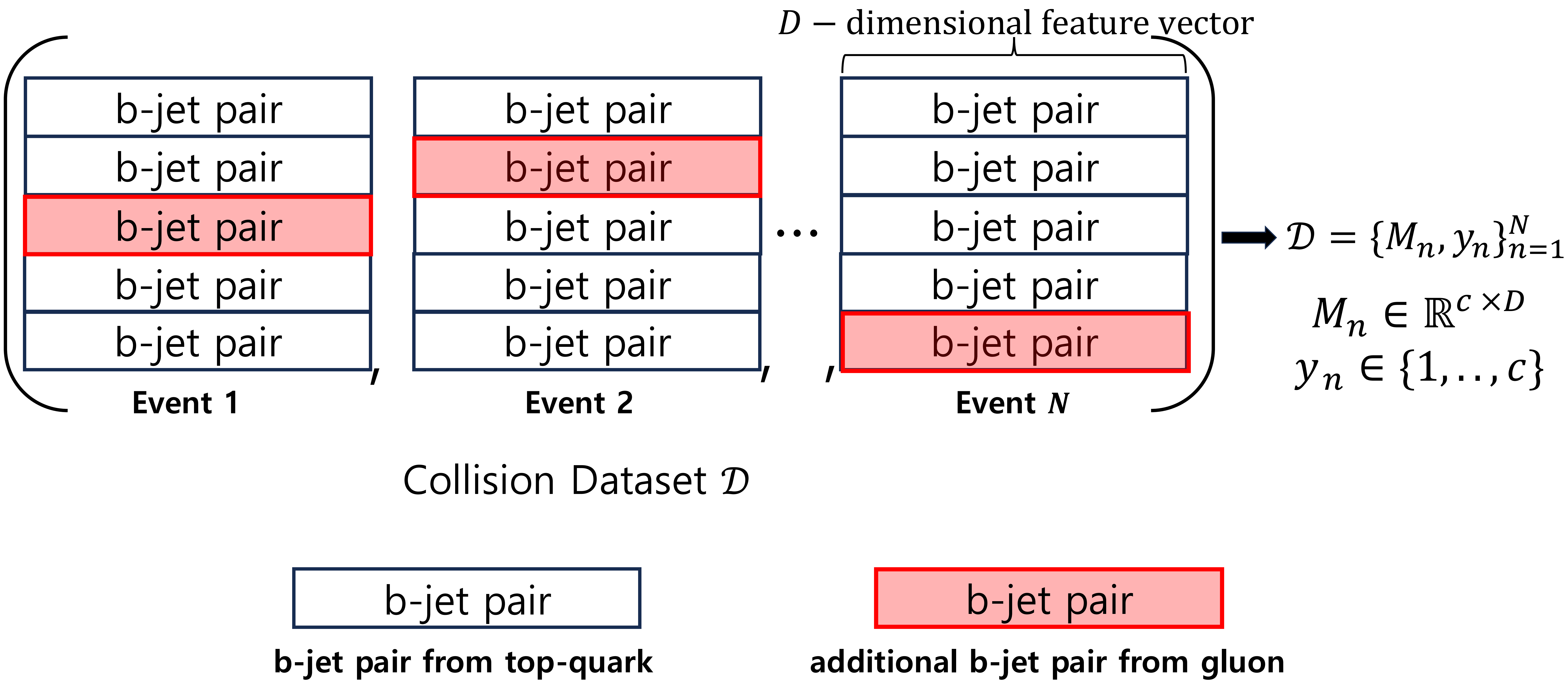
Single $t\bar{t}b\bar{b}$ event datum



A key challenge in particle physics: Identifying **additional b-jet** in single $t\bar{t}b\bar{b}$ **event**

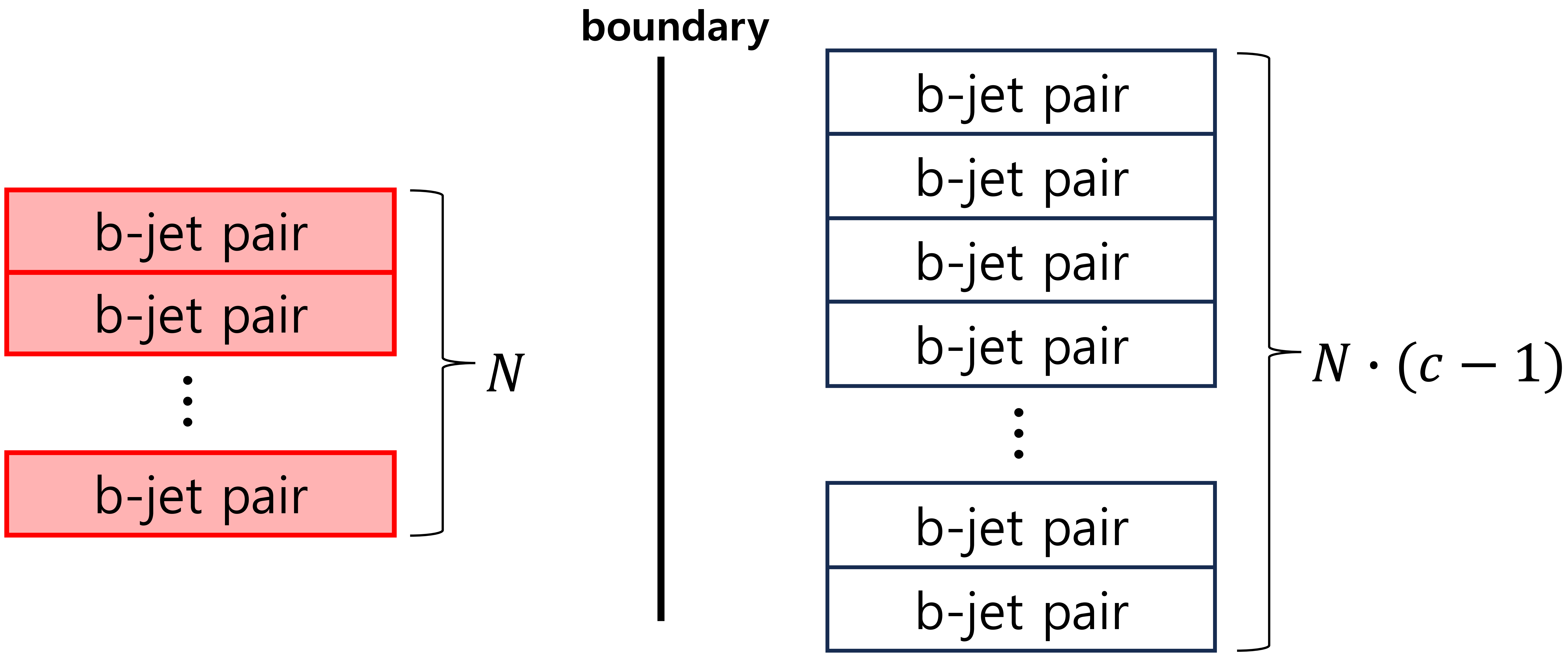
Research Experience

1. Identifying additional b-jets in $t\bar{t}b\bar{b}$ process



Research Experience

1. Identifying additional b-jets in $t\bar{t}b\bar{b}$ process

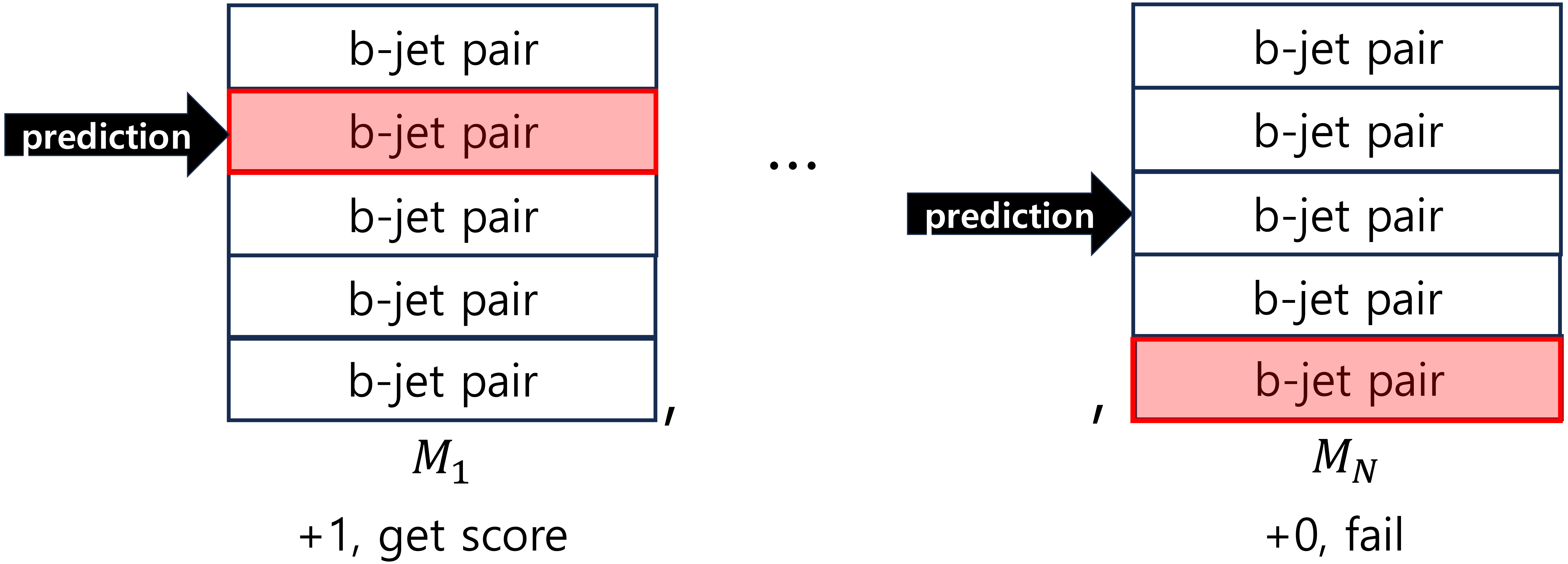


Previous Approach: Binary Classification

Minimize Binary Cross Entropy Loss

Research Experience

1. Identifying additional b-jets in $t\bar{t}b\bar{b}$ process



However, scoring rule is…

Research Experience

1. Identifying additional b-jets in $t\bar{t}b\bar{b}$ process

$$\textbf{Matching Efficiency (Identification scoring rule)} = \frac{1}{N}\sum_{n=1}^{N} \delta_{y_n \hat{y}_n}$$

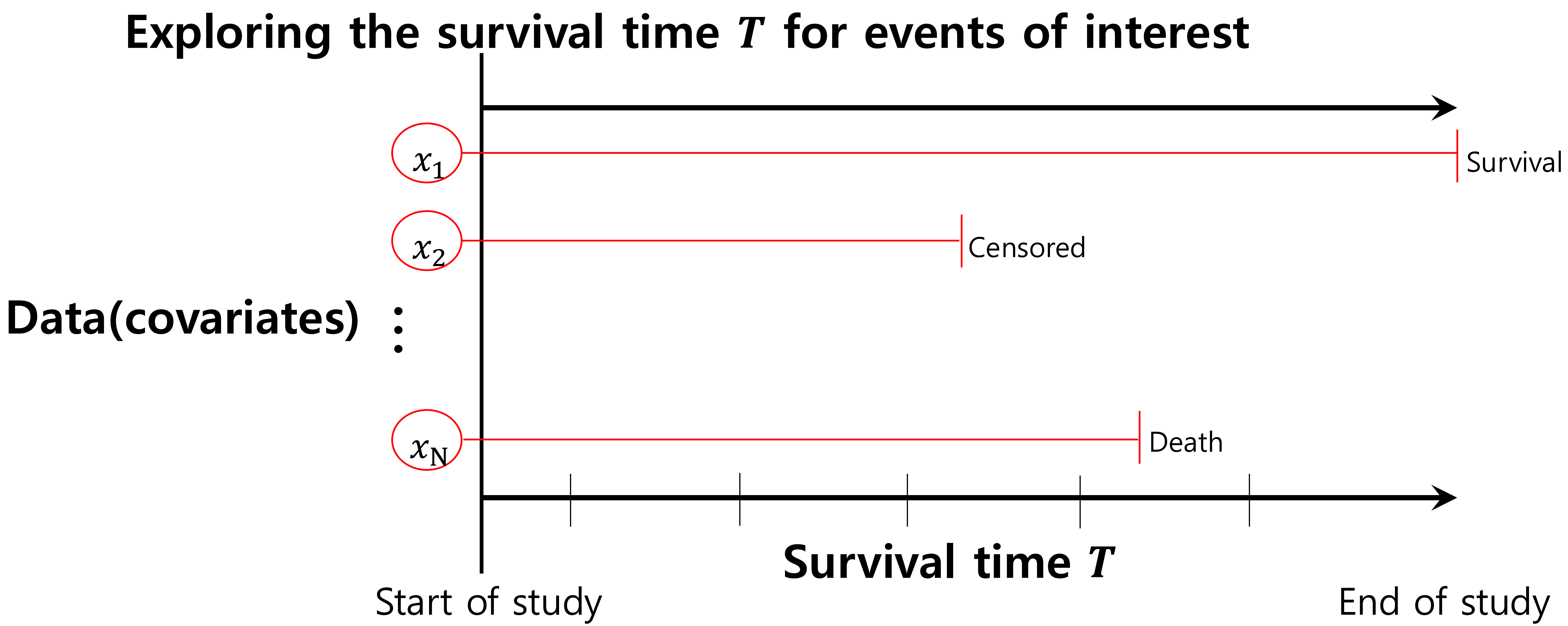
$\delta_{y_n \hat{y}_n} = 1$ if $y_n = \hat{y}_n$,
$\delta_{y_n \hat{y}_n} = 0$ otherwise

y_n: row index of additional b-jet pair (target value)
$\hat{y}_n$: index predicted on M_n

Objective: <u>Maximize</u> The Matching Efficiency

Research Experience

2. Survival Analysis



Research Experience

2. Survival Analysis

Exploring the survival time T for events of interest

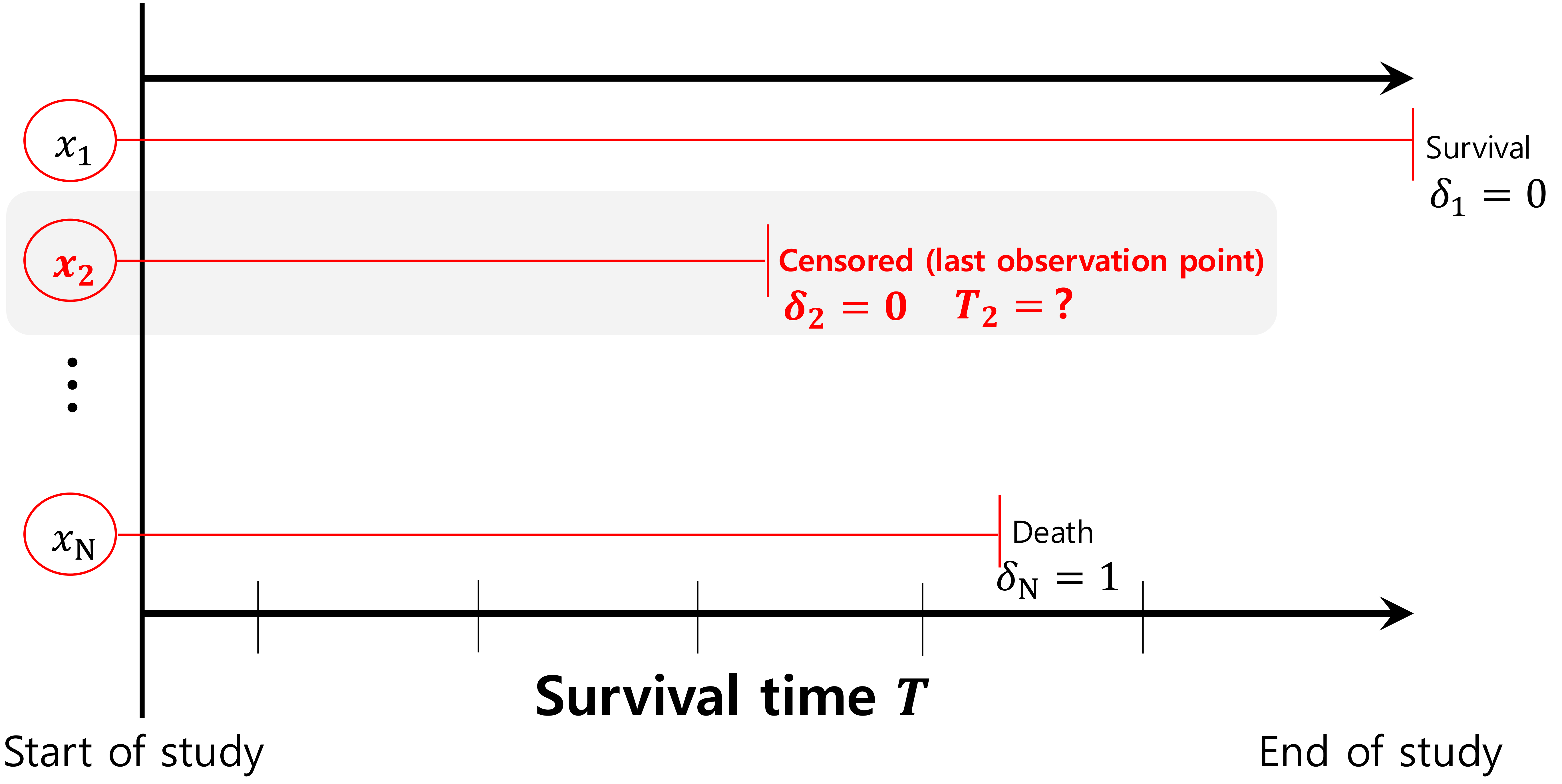
Survival Dataset

$\mathcal{D}^S = \{x_i, T_i, \delta_i\}_{i=1}^{\mathrm{N}}$

$x_i \in \mathbb{R}^D$

$T_i \in \mathbb{R} > 0$:
survival time of i-th covariate

$\delta_i \in \{0,1\}$:
$\delta_i = 0$ **if no event occurred**
$\delta_i = 1$ **otherwise**



Research Experience

2. Survival Analysis

$$\textbf{Concordance Index} = \frac{1}{|\mathcal{A}|} \sum_{i,j \in \mathcal{A}} \mathbb{I}_{f(x_i),f(x_j)}$$

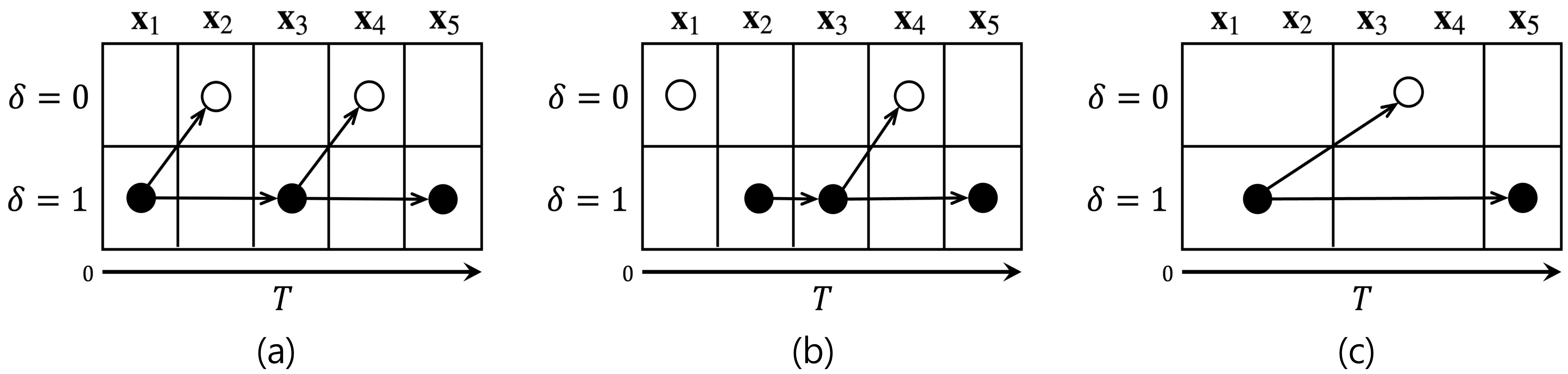
$$\mathbb{I}_{f(x_i),f(x_j)} = \begin{cases} 1, \textbf{ if } T_i < T_j \textbf{ and } f(x_i) < f(x_j) \\ 0.5, \textbf{ if } T_i = T_j \textbf{ and } f(x_i) = f(x_j) \\ \textbf{0, otherwise} \end{cases}$$

f: **survival model (for prediction of survival time T)**

$\mathcal{A}$: **set of acceptable pair (??)**

Research Experience

2. Survival Analysis



acceptable pair in (a):
(x_1, x_2), (x_1, x_3), (x_1, x_4),
(x_1, x_5), (x_3, x_4), (x_3, x_5)

acceptable pair in (b):
(x_2, x_3), (x_2, x_4), (x_2, x_5),
(x_3, x_4), (x_3, x_5),

acceptable pair in (c):
(x_1, x_2), (x_1, x_3), (x_1, x_4),
(x_1, x_5), (x_2, x_3), (x_2, x_4),
(x_2, x_5)

$\mathcal{A}$: set of acceptable pair

Research Experience

2. Survival Analysis

Problem Redefine: Identifying data has largest T in pair
(similar to additional b-jet problem)

$$\boldsymbol{\mathcal{A}}$$

$$\begin{bmatrix} (x_1, x_2) \\ (x_1, x_2) \\ \vdots \\ (x_i, x_j) \\ \vdots \end{bmatrix}$$

$$M_{i,j} = \begin{bmatrix} x_i \\ x_j \end{bmatrix}, y_{i,j} = 2$$

$$M_n = \begin{bmatrix} x_k \\ \frac{1}{2}\left(x_i + x_j\right) \\ x_k \end{bmatrix}_{k \in \{i,j\}} \quad , y_{i,j} \in \{1,2,3\}$$

Research Experience

2. Survival Analysis

Survival Data Reconstruction to Matrix:

$$\mathcal{D} = \{M_n, y_n\}_{n=1}^{N}$$

Propose Matrix c-index:

$$\textbf{Matrix Concordance Index} = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_{y_n,\hat{y}_n}$$

$$\mathbb{I}_{f(x_i),f(x_j)} = \begin{cases} 1, \textbf{ if } T_i \neq T_j \textbf{ and } y_n = \hat{y}_n \\ 0.5, \textbf{ if } T_i = T_j \textbf{ and } y_n = \hat{y}_n \\ \textbf{0, otherwise} \end{cases}$$

Research Experience

3. Integrated Problem

Identifying additional b-jets

$$\mathcal{D} = \{M_n, y_n\}_{n=1}^{N}$$

$$M_n \in \mathbb{R}^{c \times D}, y \in \{1, \dots, c\}$$

Matching Efficiency $= \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \delta_{y_n \hat{y}_n}$

$$\delta_{y_n \hat{y}_n} = \begin{cases} 1, \textbf{ if } y_n = \hat{y}_n \\ \textbf{0, otherwise} \end{cases}$$

Survival Analysis

$$\mathcal{D} = \{M_n, y_n\}_{n=1}^{N}$$

$$M_n \in \mathbb{R}^{3 \times D}, y \in \{1,2,3\}$$

Matrix Concordance Index $= \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \mathbb{I}_{y_n, \hat{y}_n}$

$$\mathbb{I}_{f(x_i), f(x_j)} = \begin{cases} 1, \textbf{ if } T_i \neq T_j \textbf{ and } y_n = \hat{y}_n \\ 0.5, \textbf{ if } T_i = T_j \textbf{ and } y_n = \hat{y}_n \\ \textbf{0, otherwise} \end{cases}$$

Research Experience

4. Methods

Maximize

- **Matching Efficiency** $= \frac{1}{N}\sum_{n=1}^{N} \delta_{y_n \hat{y}_n}$
- **Matrix Concordance Index** $= \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_{y_n, \hat{y}_n}$

However, intractable for training due to non differentiable

$\delta_{y_n \hat{y}_n}$ **and** $\mathbb{I}_{y_n, \hat{y}_n}$.

$\therefore$ **Surrogate Loss functions required!!**

Research Experience

4. Methods

Prediction Model:

$$f\colon \mathbb{R}^{c\times D} \to [0,1]_1, \dots, [0,1]_c$$

$$\hat{y}_n = \underset{k\in\{1,\dots,c\}}{arg\max} f_k(M_n), \qquad n = 1, \dots, N$$

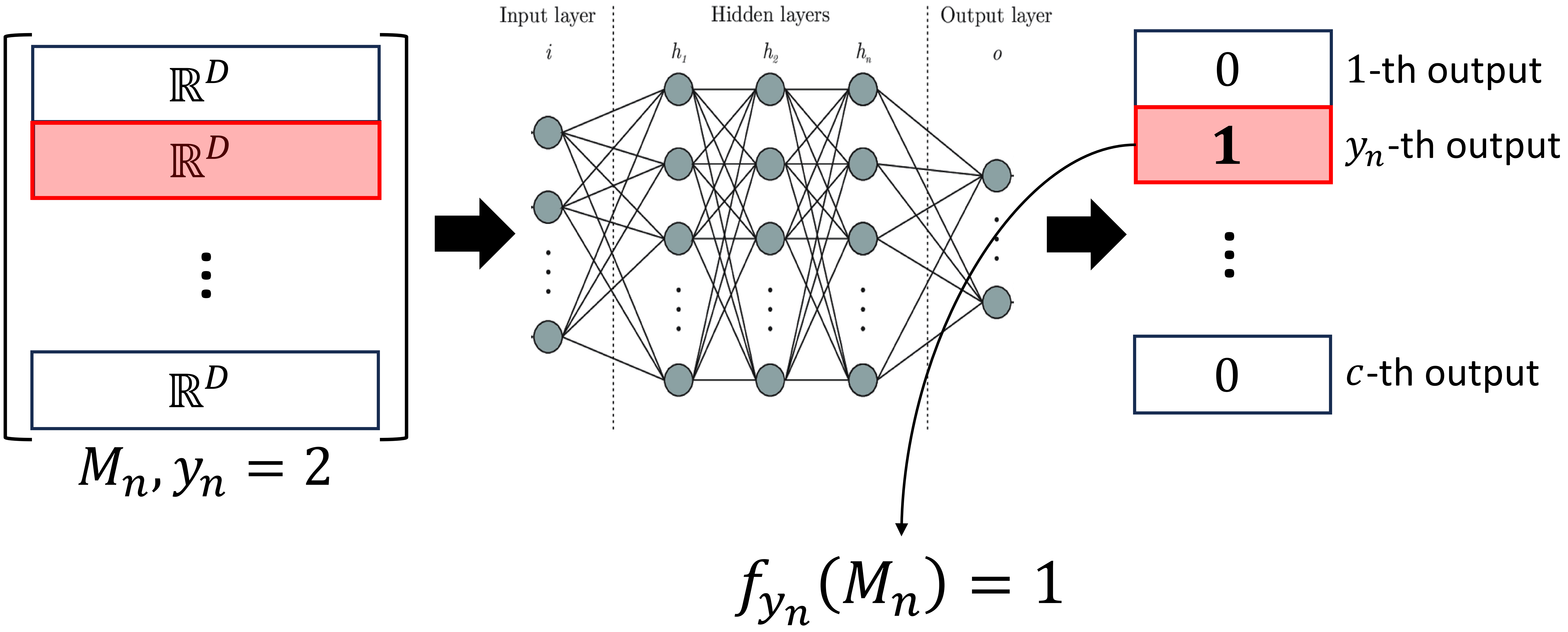
$$f_1 + f_2 +, \dots, + f_c = 1$$

Ideal Prediction Model:

$$f_{y_n}(M_n) = 1, \qquad n = 1, \dots, N$$

Research Experience

4. Methods



Research Experience

4. Methods

Ideal Prediction Model

$$f_{y_n}(M_n) = 1, \qquad n = 1, \dots, N$$

Maximize y_n-th prediction output

$$\mathcal{L}_{sur} = \frac{1}{N}\sum_{n=1}^{N} f_{y_n}(M_n)$$

Minimize Surrogate Loss

$$\mathcal{L}_{sur} = \frac{1}{N}\sum_{n=1}^{N} (1 - f_{y_n}(M_n))$$

Research Experience

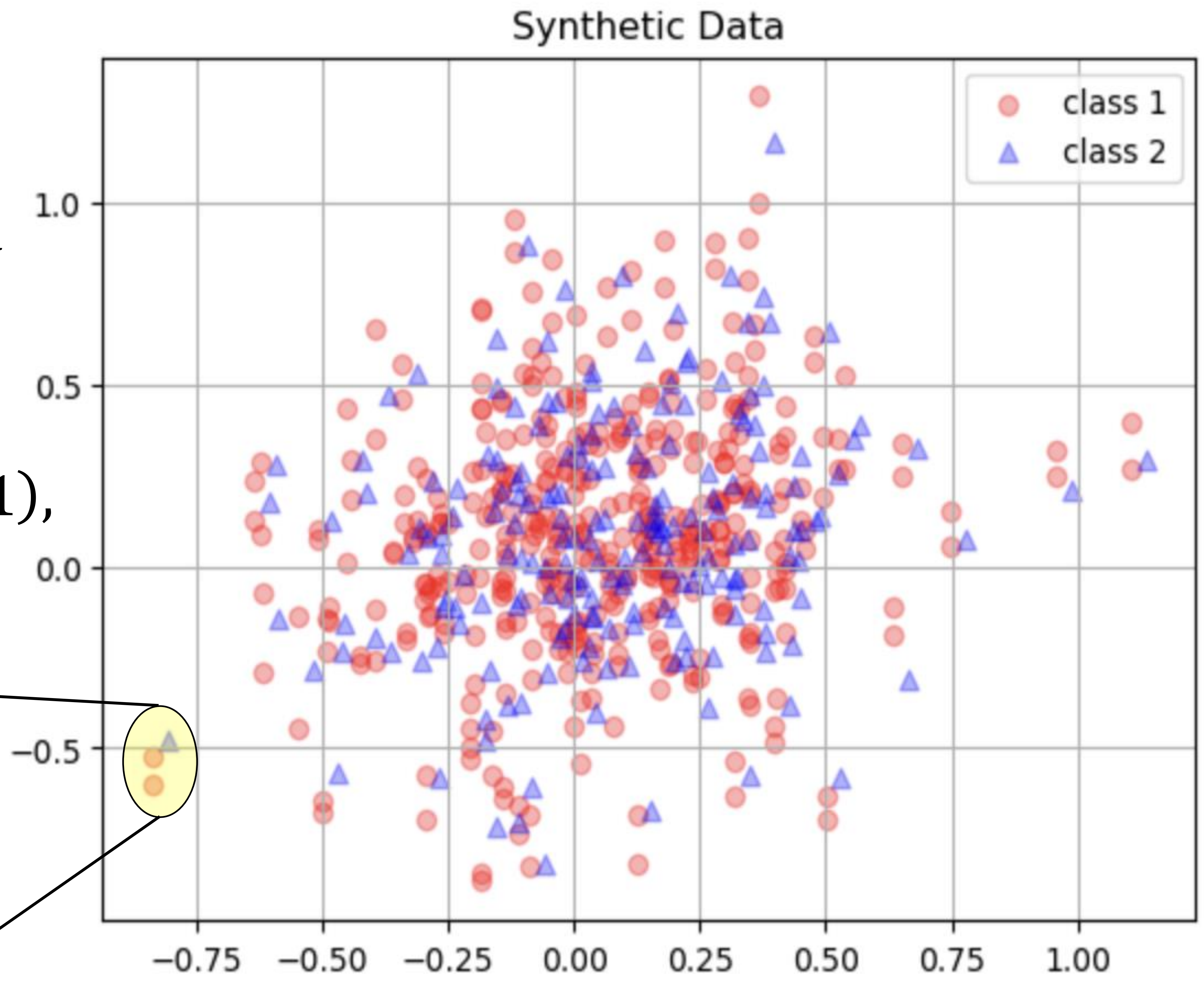
5. Experiments: Synthetic Data

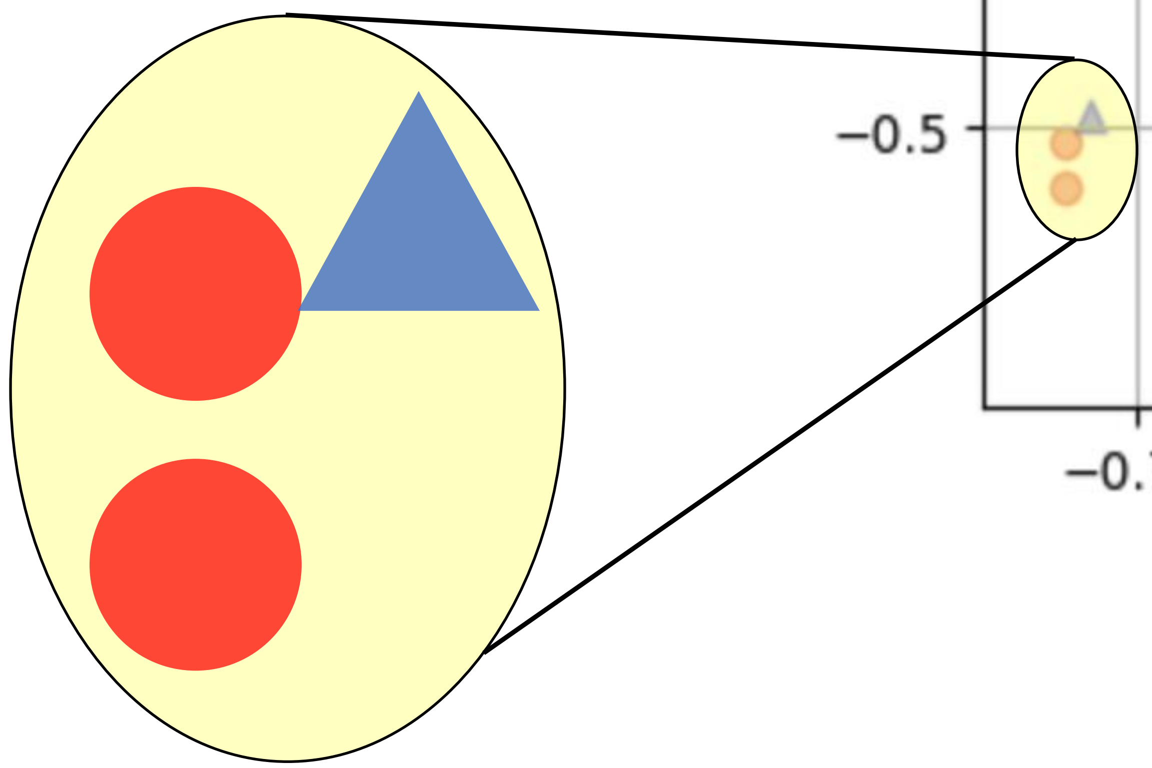
Class 2 (Signal Data)

$$\hat{x}_n \sim \mathcal{N}(0, I), n = 1, \dots, N$$

Class 1 (Background Data)

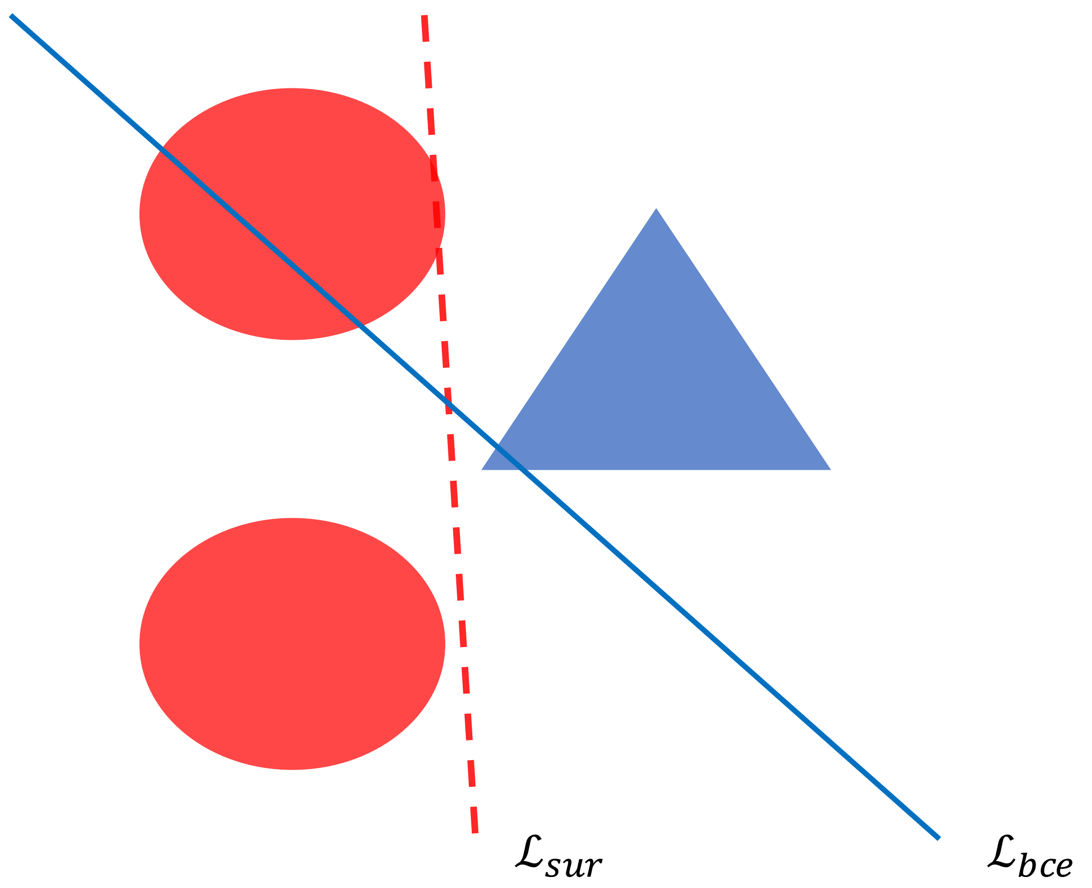
$$\hat{x}_n + (-0.03, \epsilon_k), k = 1, \dots, (c-1),$$
$$\epsilon_k \sim (0, 0.01)$$

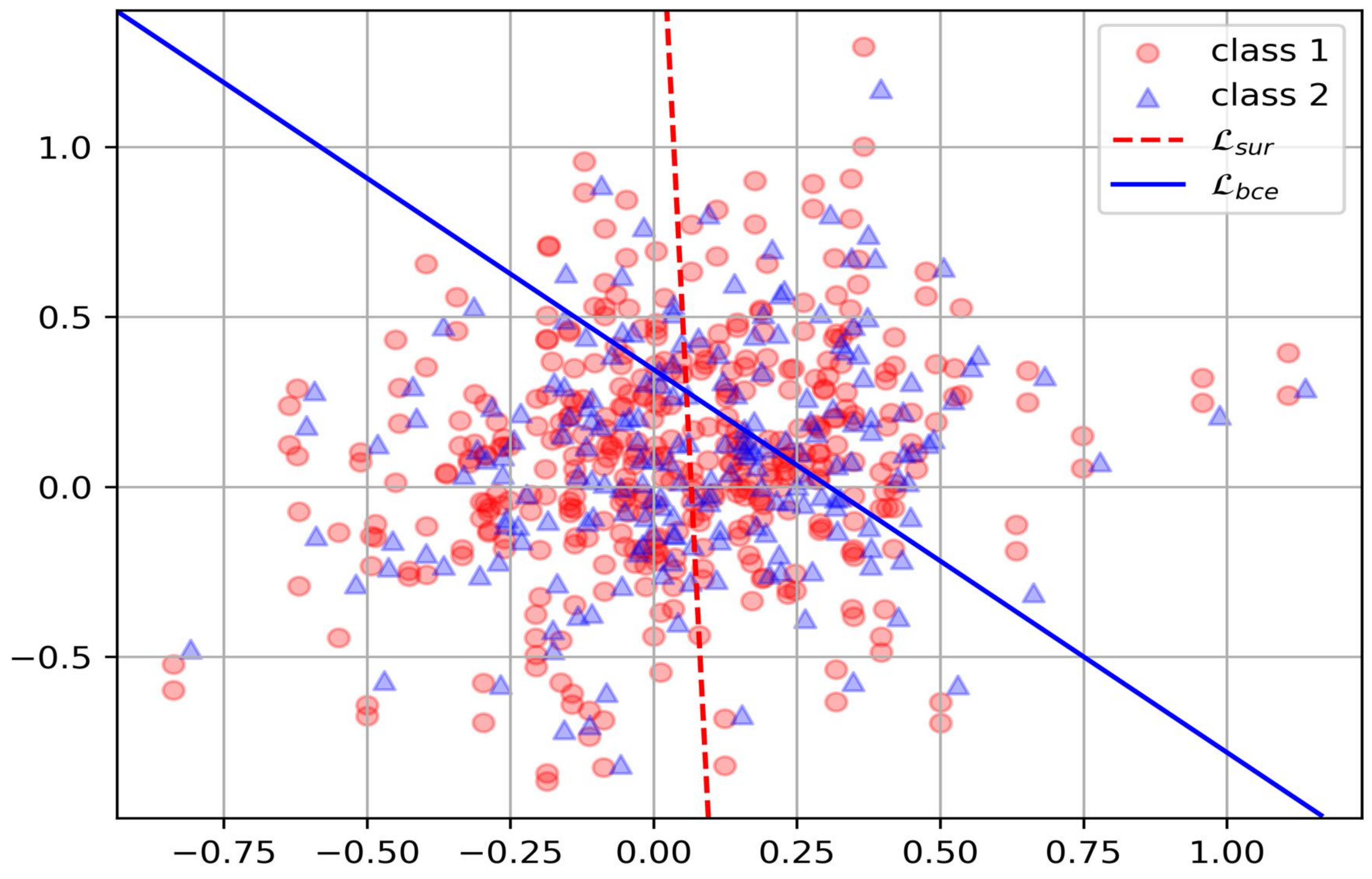


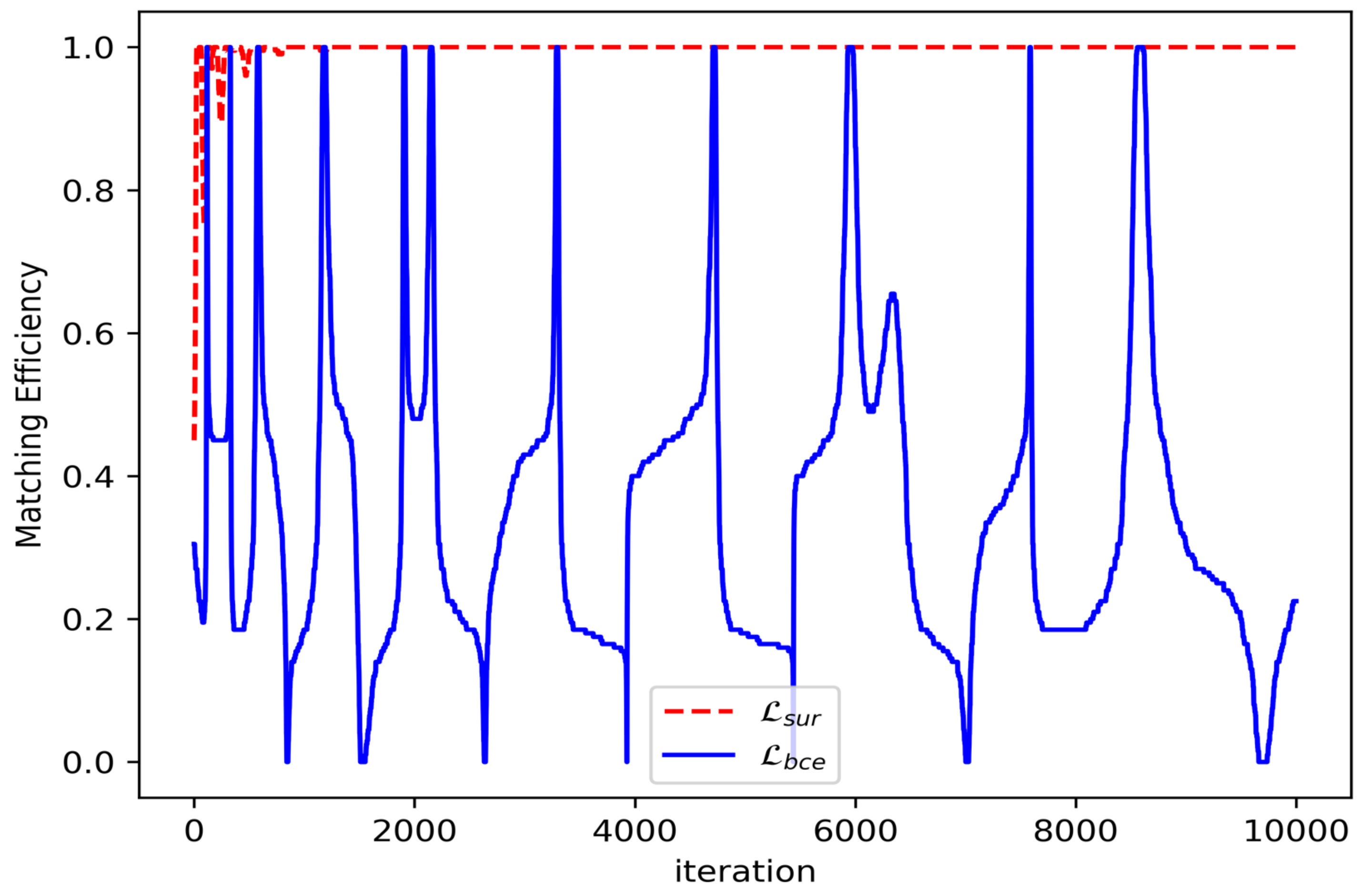


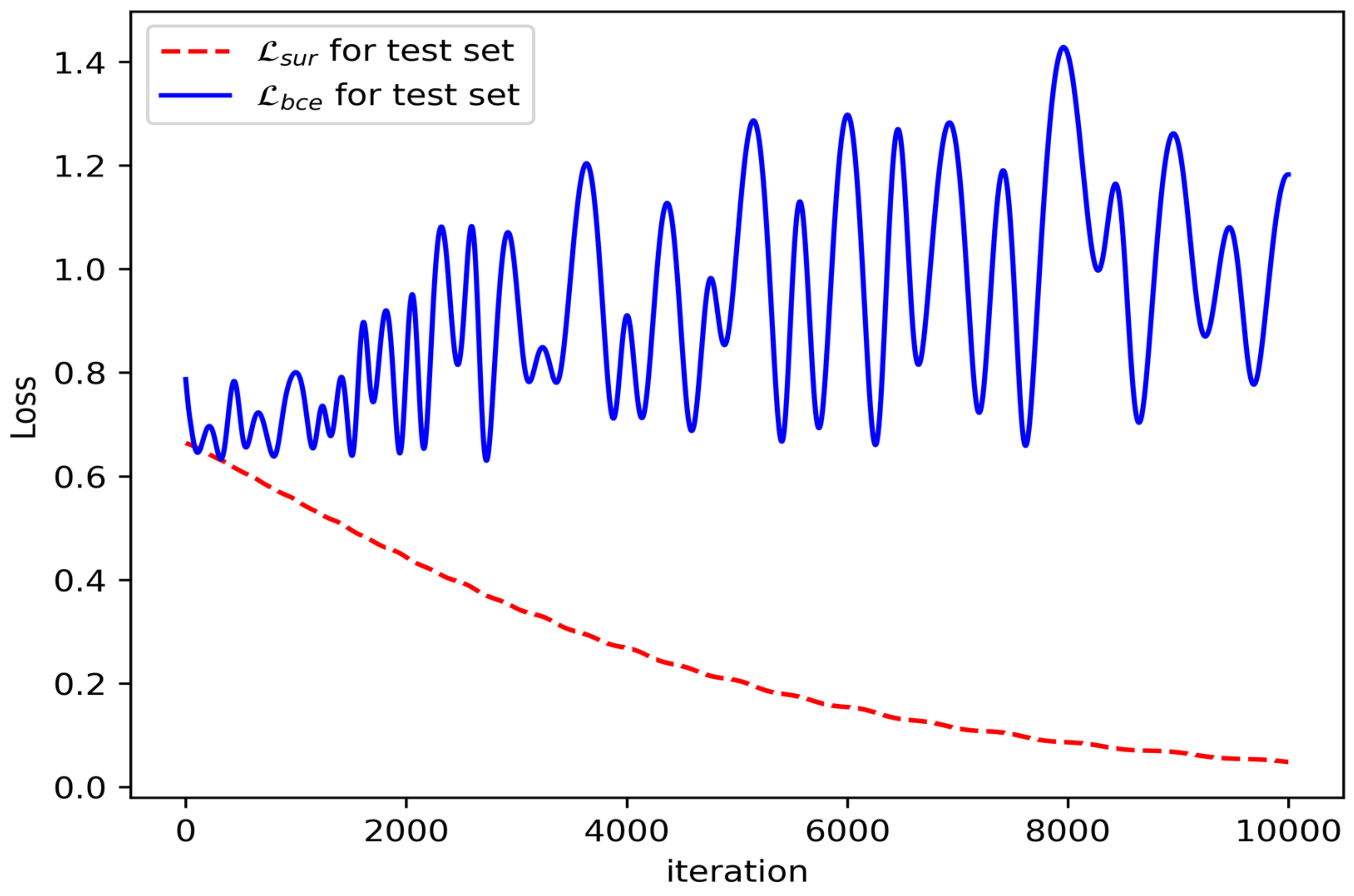
Research Experience

5. Experiments: Synthetic Data









Research Experience

5. Experiments: Simulated $t\bar{t}b\bar{b}$ Event Data

- 78 random variables
- Data Configuration 1: 161,676 training data, 40,419 validation data, 10,000 test data
- Data Configuration 2: 40,000 training data, 10,000 validation data, 10,000 test data
- Data Configuration 3: 10,000 training data, 2,500 validation data, 10,000 test data
- fully connected neural network
- hyperparameter tune to layer and node

Loss	**ME (std)**	**ME (std)**	**ME (std)**
Training data size	202,095	50,000	12,500
$\mathcal{L}_{bce}$	0.61422 (0.0034)	0.61571 (0.0046)	0.60851 (0.0068)
$\mathcal{L}_{sur}$	**0.63511** (0.0036)	**0.62949** (0.0031)	**0.61479 (0.004)**

Research Experience

5. Experiments: Survival Data

Survival Data	N of data	N(%) of censored	N of features
SUPPORT2	9105	2904 (32.2)	98
AIDS3	3985	2223 (55.8)	19
COLON DEATH	929	477 (51.3)	48

Loss	SUPPORT2	AIDS3	COLON DEATH
Partial likelihood (Cox)	0.8487(0.0146)	0.5641(0.0233)	0.6428(0.0232)
Partial likelihood (Cox Efron's)	0.8495(0.0135)	0.5638(0.0269)	0.6471(0.0324)
Ranking (Sigmoid)	0.8545(0.0137)	**0.5722**(0.0222)	**0.6536**(0.0296)
Ranking (Log Sigmoid)	0.8538(0.0147)	0.5703(0.0251)	0.6496(0.0279)
Ranking (SVM)	0.8517(0.0135)	0.5604(0.027)	0.6453(0.0232)
Ranking (Boost)	0.8537(0.0146)	0.5714(0.0195)	0.6391(0.038)
Classification (WM)	0.8542(0.014)	0.5596(0.0336)	0.6475(0.0345)
Classification (Ours)	**0.8548**(0.0135)	0.5555(0.0300)	0.6524(0.0247)
Mean of All Performance	0.8526(0.0022)	0.5647(0.0270)	0.647(0.004)

Thank You!