

A Study on Learning Mathematics for Machine Learning

Sang Pyo Jun*

Abstract

This paper is a study on mathematical aspects that can be basic for understanding and applying the contents of machine learning. If you are familiar with mathematics in the field of computer science, you can create algorithms that can diversify researches and implement them faster, so you can implement many real-life ideas. There is no curriculum standard for mathematics in the field of machine learning, and there are many absolutely lacking mathematical contents that are taught in the curriculum presented at existing universities. Machine learning now includes speech recognition systems, search engines, automatic driving systems, process automation, object recognition, and more. Many applications that you want to implement combine a large amount of data with many variables into the components that the programmer generates. In this course, the mathematical areas required for computer engineer (CS) practitioners and computer engineering educators have become diverse and complex. It is important to analyze the mathematical content required by engineers and educators and the mathematics required in the field. This paper attempts to present an effective range design for the essential processes from the basic education content to the deepening education content for the development of many researches.

▶ Keyword: Mathematics. Machine learning. Mathematics Education method. Speech recognition systems, Automatic driving. Process automation

I. Introduction

현재 컴퓨터 공학 분야에서 어느 것 보다 급격하게 다양한 영역으로 전개되고, 여러 사업에 융합 사용 되고 있는 분야는 머신러닝 분야이다. 이 머신러닝 이론은 반복적으로 데이터를 학습하고, 지능형 애플리케이션을 구축하는데 사용 될 수 있는 숨겨진 패턴을 발견하여 여러 수학 관련 분야 지식과 컴퓨터 과학 및 알고리즘이 융합된 분야이다. 이 머신러닝 이론은 엄청난 발전 가능성을 가지고 있다. 주어진 자료 속에 원리, 즉 패턴을 찾아 적절한 모델을 구축하고, 모형에 의해 다양한 미래에 대한 추천 또는 예측을 하기 위해서는 각 단계 내부에 수학적 이론이 잠재해 있는데, 수학적 분야의 이해가 완전한 이해가 될 수 있다. 알고리즘으로 구현되는 내부 동작을 잘 파악하여야 좋은 결과들을 얻을 수 있는데, 이런 기술 많은 부분들을 수학적 이론에 바탕을 두고 있다.[1,2,3] 수학적 지식을 바탕으로 지금 까지 논의되는 원리들을 이해하고, 이해된 원리를

바탕으로 새로운 분야에 융합 적용하여, 머신러닝 관련 분야의 연구에 보다 많은 것을 다양하게 표현 할 수 있다. 머신러닝의 응용분야가 많아지면서 수학의 중요성은 점점 더 강조 되고 있다. 이런 머신러닝 분야 연구에 중요한 수학 분야가 중, 고등학교에서의 기초 과정에서 부터 전문 수학분야의 석사, 박사 과정에 이르기 까지 기초 수학에서부터 세부 특화된 수학의 전문 다양한 분야의 이론까지의 수학기론의 사용이 늘어나고 있다. 이런 수학 관련 지식은 컴퓨터관련 개발자와 수학기론 함수 제공자 들 간의 별개의 학문이 아닌 융합적인 공유가 필요해 지고 있다. 지금 우리나라 컴퓨터공학에 있어 수학 교과 과정은 중, 고등학교에서는 상위학교 진학을 위한 입시용 교육으로, 대학에서는 알고리즘의 이해를 위한 수학이 아닌 응용에 필요한 라이브러리 전개의 과정을 이해하는 방법론에 의해 필요한 것을 쓰는 과정의 일부로 교육하는 정도이다.

• First Author: Sang Pyo Jun, Corresponding Author: Sang Pyo Jun
*Sang Pyo Jun (spjun7129@naver.com), Dept. of General Education, Namseoul University
• Received: 2018. 12. 12, Revised: 2019. 01. 12, Accepted: 2019. 01. 15.

일부 특성화 학생을 위한 수업으로 일반적 학생에 대한 교육은 이루어지지 않고 있다[4]. 교육 내용도 지나치게 이론적이고 실생활과는 유리된 내용의 수학 교과 내용이 부진한 학생이 입시를 위해 공부하는 과목으로 정해져 있는 실정이다. [5] 다른 나라의 경우 컴퓨터 공학의 연구에 있어 수학의 중요성을 인식하여 미국에서는 1989년 NCTM의 '학교 수학 교육과정과 평가의 기준'에서 추천하여 1991년에 나온 'Discrete Mathematics across the Curriculum, K-12'의 일련의 연구가 수행됐고, Rutgers 대학에서는 수학에 재능이 있거나 관심이 많은 학생들을 대상으로 한 과외 프로그램 'Young Scholar Program in Discrete Mathematics'를 실시하여 타 전공자에게 컴퓨터 기초수학에 대한 교육을 강화하고 있다. 네덜란드의 10, 11학년의 교수요목에는 수학의 내용이 상당한 수준까지 제시되어 있으며, 덴마크의 Gymnasium A수준에서는 프랙탈, 카오스, 선형계획법 등의 주제를 다루고 있다.[6,7,8]

우리나라에서는 일선 교육 현장에서는 컴퓨터 공학 관련 내용 공부에 필수적인 컴퓨터 관련 수학에서부터 전문적인 수학 교육을 받지 않은 수학 교사가 강의를 하거나, 컴퓨터 관련 전공자가 자기의 경험에 의해 이해한 수리적인 부분을 교육하거나, 단지 컴퓨터 관련 수학에 관심을 가지고 공부를 한 교사라서 수학의 여러 분야 전체에 익숙한 사람은 거의 없다고 할 수 있다. 머신러닝을 이해하기 위해서는 박사수준의 특성화된 수학도 필요로 하는 현재에는 다양한 이론에 대한 이해와 구현에 어려움 증가되고 있다. 현재 컴퓨터 관련 산업의 양성 기관이나 연구 기관에서 컴퓨터 공학 관련 수학 강의에 대해 통일된 교육과정이 없고 구체적으로 어떤 내용을 다루어야 하며, 그 수준을 어디에 맞추어야 하는지 또한 정하지 못하고 있다. 컴퓨터 공학 관련 수학이라는 분야가 아직 완전히 정립되지 않았고, 다루어야 하는 영역도 계속 발전 확장하고 있어 교육과 연구에 어려움이 있다.

컴퓨터 공학의 머신러닝 분야를 이해하는데 필요한 수리적인 내용을 이해하기 부족한 학생들에게 앞으로 컴퓨터 관련 분야를 전공 하는데 필요한 수학분야의 표준이 될 수 있는 내용과 수준을 제시 하여, 다양한 컴퓨터 공학 연구에 초석이 되도록 만들어져야 할 것이다.

컴퓨터 공학의 많은 발전을 위해서는 수학의 이해를 요구하는 경우가 많다. 필수적인 간결한 분야별 수학 교육 과정을 정하는 것은 시간과 비용의 낭비를 줄일 수 있을 것이다. 또 수학의 어려움에 거부감이 많은 수학에 어려움을 느끼는 사람들에게 연구개발에 최소한 알아야 하는 분야별 수학 내용을 제시하여 보다 좋은 결과를 구현하는데 도움이 되고자 한다. 본 연구에서는 컴퓨터 공학의 머신러닝의 여러 분야를 세분 공부하는데 필수적인 선행과목과 부수과목 심화 분야에 대하여 그 성격과 목표를 설명하여 머신러닝분야의 효율적인 연구가 되게 관련 내용 정리하여, 연구에 도움을 주려고 한다. 단순히 여러 가지 내용을 모아 두지 않고, 일관성 있게, 중요한 수학을 강조하며, 종합적인 표현되도록 하여 지속적인 컴퓨터 관련분야 연구에 대한 연관성을 제시 한다.

II. Contents of mathematics related to computer engineering

컴퓨터 공학 관련 수학은 다른 분야와 비교하여 상대적으로 새로운 분야이며, 최신의 수학기론을 사용하고 있으며, 고차원으로 차원도 점차 확장되고 있어, 그 내용이 타 영역과 중복되는 분야도 적고, 많은 사람이 인정하는 표준화된 교육과정도 결정하기가 어렵다. 각각 연구하고자 하는 분야에 따라 필요로 하는 수학분야도 난이도 어려워지고 있다. 컴퓨터 공학연구와 관련된 수학의 내용은 기초적인 이산수학의 분야에서 머신러닝을 연구 이해 하기 위해서는 수학의 전문화된 분야를 세분되고 다양한 분야를 알아야 한다. 기초적인 컴퓨터 공학에 공통적으로 사용하는 이산수학 책에서 다루어지고 있고, 국내외에서 공통적으로 인정하는 수학의 목표나 주제를 살펴보면 다음과 같다.

2.1. Mathematics curriculum goals in the "Mathematics Curriculum and Evaluation Criteria" common to basic computer science research

- 1) 응용 문제에 대한 상황 표현과 해석
- 2) 알고리즘의 이해, 설계 및 분석.
- 3) 셈 하는 방법을 이해하기 위한 순열과 조합.
- 4) 불확실성을 해결하기 위한 확률 문제의 해결.
- 5) 확률 변수와 확률분포함수의 이해와 적용
- 6) 선형계획법의 설계와 벡터 해석을 통한 해결.
- 7) 미분방정식을 이용한 문제의 표현과 해결 및 해석.
- 8) 제시된 문제 상황의 컴퓨터 프로그래밍화
- 9) 탐구 분석 실생활 적용[9].

2.2 Contents of mathematics proposed for computer engineering.

- 1) 그래프 이론을 이용한 알고리즘의 이해와 설계. - 그래프 구조의 기본 개념, 다이어그램화, 다양한 시스템의 행렬 표시 및 분석 및 해 찾기. 회로와 경로의 오일러 회로, 해밀턴 회로 등의 이해를 통한 최단경로문제의 알고리즘화, Network의 색칠 문제 이해와 응용, 수형도의 구조 이해
- 2) 논리, 집합의 포함배제의 원리, 셈의 합법칙, 곱의 법칙, 순열과 조합, 이항계수에 대한 이해
- 3) 확률 및 통계학에서 확률의 계산, 확률계산(조건부 확률, 베이저안 확률) 확률변수와 확률함수의 이해와 응용
- 4) 수학적 귀납법에 의한 증명과 이해, 일계 반복 관계, 수열에서 등차수열, 등비수열, 피보나치 수열등의 이해를 통한 표현, 일계 선형 점화식에서 일반항 찾기[9]
- 5) 행렬의 모델 과 확률 과정 론의 이해- Markov 연쇄, Leslie 모델, Leontidf input-output 모델[10]

2.3 Mathematics content common to domestic computer engineering

- 1) 집합의 정의와 집합의 연산. 관계의 정의와 동치관계와 반 순서 관계등의 이해, 부울 대수 함수의 이해와 응용.
- 2) 함수의 정의와 다양한 함수의 이해- 지수함수, 역삼각함수 쌍곡 함수등의 이해.
- 3) 알고리즘의 이해 와 분석, 알고리즘의 복잡성 분석
- 4) 수학적 명제에 대한 해결- 수학적 귀납법, 수학적 연역법
- 5) 확률의 경우의 수, 비둘기집이론, 순열과 조합관계, [11,12,13,14]

III. Development direction of mathematics curriculum related to machine learning

컴퓨터 공학의 머신러닝을 연구하는데 수학이 중요한 이유는 데이터의 정확도, 모델의 복잡도에 대한 해석, 변수의 선택, 검증과 표현에 대한 전략선택, 올바른 신뢰구간과 불확실성의 추정등을 복잡하고 고도화된 다양한 지식을 필요로 한다. 이들을 위해서는 현재의 기초, 응용 수학의 영역에서 더 확장하고 보다 전문화된 트랙에 적합한 내용으로 교육과정을 구성하여야 할 것이다. 공통된 영역의 수학의 기초적인 본질을 이해하고 수학 내용을 바탕으로, 기준에 존재하고 확정된 이론을 이해하고, 즉 다양한 언어들 속에 소스 파일로 구성된 라이브러리를 이해하고, 라이브러리의 이해를 통해 보다 많은 다양한 분야를 응용하여 본인이 생각하는 다양한 분야를 구현하게 해야 한다. 실생활에서의 수학의 활용성과 유용성을 느끼게 하는 공부가 되어야 하고, 컴퓨터 공학 연구의 든든한 밑거름이 되어야 한다. 컴퓨터 공학 인공지능의 여러 트랙에 필요로 하는 수학 분야를 다음과 같은 몇 개의 사항을 고려하여 설정하였다.

- 1) 수학기론의 정확한 이해를 통한 실제 현상을 문제화하는 수학적 사고 능력을 함양
- 2) 수학적 사고력에 의한 합리적 의사 결정 능력 배양
- 3) 이론에 근거한 창의적 해결 능력 증대.
- 4) 다른 수학분야와의 연계된 교육으로 앞으로 다양한 분야와의 공동연구에 필요한 사고력 증대.
- 5) 지속적 연구를 위한 폭넓은 주제에 대한 깊이가 형성.

IV. The Proposed Scheme

4.1 Mathematical fields of machine learning that are necessary in advance

머신러닝을 위해서는 현재 응용 되는 것들은 인간이 프로그래밍하고 디자인 한 것과 데이터로부터 학습된 행동의 결합된

컴퓨터공학의 지식과 기술이 필요로 하는 교육. 그리고 논리적인 참이라 명확한 증거 뿐 아니라 효과에 대한 통계학적 구성을 가지고 있어야 하고, 전통적인 소프트웨어는 성취하려는 정확한 목표에 맞추어서 단계적으로 묘사하는 프로그램을 만들었는데 머신러닝학습의 시스템에서는 최적으로 맞는 것을 묘사하도록 설계하고, 모형을 주어진 데이터를 가지고 하려고 하는 목표에 가장 최적의 것을 찾으려한다. 머신러닝 학습의 성공을 위해서는 예전과 같은 알고리즘 발전에 집중해야 하는 것이 아니라 데이터의 선택 데이터의 정제 모형 선택 통계학적 테스트 이에 필요한 부분을 세부적으로 요약해보자.

1) 선형대수학 분야.

지금은 데이터의 수집이 과거에 비해 용이하다. 수집기술도 발전하고, 저장 및 조작이 간편해 졌다. 그러나 데이터의 잘못 사용으로 구현하고자 하는 현상에 큰 왜곡을 발생하게도 한다. 그래서 구현하는 현상에 적절한 데이터 사용을 위한 사전작업이 이전 필수처럼 되어 있다. 불필요한 데이터의 정제, 비슷한 데이터의 통합, 계산 등 문제 해결을 위한 데이터의 축소 등 작업이 필수가 되어 있다. 이런 문제해결을 위해 선형대수학의 필요성은 점차 증대하고 있다.

- (1) 구현하고자 하는 현상을 설명하는데 필수적인 상관관계가 큰 변수들의
- (2) 변수들간의 독립성 확인
- (3) 독립변수들 간의 중요도 판별
- (4) 독립 변수들의 선형계획
- (5) 연립방정식의 행렬화
- (6) 연립방정식의 해를 찾는 문제
 - 기초과정에서의 벡터의 연산과 표현. 행렬의 이해와 행렬 연산, 행렬식의 계산은 필수이고, 위의 문제들을 해결 하기위해 독립 변수들 간의 핵심이 되는 변수를 찾는 주성분분석문제, 연립방정식의 행렬화. 계수행렬에 대한 분해방법론인 LU분해, QR 분해, SVD(singular value decomposition). 유일해존재확인 계산을 위한 고유값, 고유벡터 문제, Vector 공간과 표준화, 직교화 방법론등

2) 확률 및 통계 분야

주어진 자료들속의 패턴분석, 분석된 패턴의 모형 화, 모형의 모수들에 대한 가정, 모형에 대한 적합도 개선 그리고 다음 현상에 대한 추측의 정확도 증가. 등을 위한 확률과 통계부분의 중요도와 필요성은 점점더 추가되고 있다.

- (1) 주어진 Test data를 이용한 모형 설계-Modeling
- (2) 설계된 모형의 parameter 들에 대한 Estimation.
- (3) 모형 parameter 들에 대한 가설검정
- (4) Training data을 이용한 모형에 대한 개선-정확도의 증가.
- (5) 확률공리와 확률의 계산
 - 모수들의 특성을 파악 하기 위한 기댓값과 표준편차의 계산. Maximum likelihood estimation 문제, Backpropagation 시

Bayesian 해석. 사전확률과 사후확률의 계산. Markov Chain의 전이확률등의 계산. 정규분포의 해석. 정규분포의 응용.

3) 알고리즘 및 복잡도의 최적화 분야

알고리즘의 해석을 위해서는 이산수학의 기본이 되는 데이터 구조에 대한 사전지식이 필요하다. -이진트리,해싱,힙,스택 구조, Dynamic Programming, Randomized Sublinear Algorithm graphs. Primal -Dual Methods

4) 기타 수학 분야

(1) 다 변량 미분 적분학

일변수함수의 이해 -지수함수, 역삼각함수, 쌍곡함수의 이해 - Sigmoid, Tanh, ReLU, Leaky ReLU, 일변수함수의 미분과 적분 -Gradiend descent method 다변수함수의 미분과적분- 편미분의 이해 Vectoe 값 함수의 이해.

(2) 음성과 영상 해석을 위한 복소해석학의 이해

(3) 위상수학의 이해.

(4) 주파수와 주기함수의 해석을 위한 Laplace 해석과 Fourior 급수와 변환 및 해석

(5) 다양체이론과 Function spaces, Entropy 등의 최근의 수학 이론 등이 있다.

머신러닝의 공통적으로 필요한 수학 관련 분야에 대하여 기술하였다. 각 분야를 기계적인 방법으로 세분 할 수는 없지만 좀 더 세분하여 최단의 방법으로 알고리즘의 이해를 하는데 도움을 주려고 수학 관련 분야를 정리 했다.

4.1.1 Mathematics in the field of Convolution

Neural Network

여기서는 머신러닝 분야에 대한 이해를 위해서는 기초적인 미분적분학의 개념인 극한개념과 함수에 대한 이해가 필요하다. 일 변수 미적분에서 다변량에 대한 미분적분학까지 필요하다. 선형대수학의 변수들의 독립성 규명, 주성분분석, 행렬에 대한 공부는 많은 분야가 필요한데, 행렬의 곱, 행렬의 차원을 축소를 위한 해석이 필요하다. 행렬의 고유 분해 등 분해방법론과 고윳값문제에 이해가 필요하다. 독립변수간의 관계 규명의 지식이 필요하고, 통계학 분야에서는 확률의 연산, 확률변수, 정규분포의 이해, 조건부확률밀도함수, 결합 확률 밀도함수, 가 정 모형 함수의 모수 추정을 위한 회귀분석에 대한 많은 이해가 필요하다. 상관분석과 선형회귀에 모수추정과 적합 모형을 위한 테스트 등 가설검정의 이해에 대한 해석이 필요하다. 변수 상호간의 상관분석, 잡음에 대한 해석과 제거, 랜덤과정 등을 공부가 관련 있고, 알고리즘에 대한 복잡도에 대한 해석이 필요하다.

Table 1. CNN Track

Prerequisite	Major Required	Related field
Calculus	Regression analysis	Normal distribution
Linear algebra	Dervitive	Multivariable differential
Statistics	Multivariable mathematics	Dynamic programming
Discrete mathematics	Matrix operation	Complex analysis
	Vector operation	Convolution

4.1.2 Mathematics in the field of Neccurent

Neural Network

여러 가지 신기술을 다양하게 습득하여 전문화된 소프트웨어 개발 능력을 갖춘 인력을 배양하는데 목적이 있는데 데이터베이스에 대한 깊은 있어야 하고, 데이터간의 관계 흐름도 파악해야 한다. 관련성 흐름이 순차적으로 앞의 과정과 관련성을 포함하면서 만들어지는 순차적 정보처리의 개념에서는 신경망의 모든 과정에 입력과 출력이 시간대별로 상관성 가지고 있고, 이 분야에 대한 수학의 깊이가 필요하다. 집합과 함수를 기본으로 비선형함수들에 대한 해석과 신경망의 동작기법을 이해하기위해서는 행렬과 행렬식에 대한 이해가 필요하다. 행렬에서는 합성곱의 구조, 신경망과 합성곱의 연관관계, 오차가 전파되는 과정에 대한 이해가 필요하다. 다변량 벡터연산과 Kernel 함수. 차원줄이기, Support vectors machines Support vectors regression. Chain rule 에대한 정확한 연산과 이해가 필요하다.

Table 2. RNN Track

Prerequisite	Major Required	Related field
	Regression analysis	Regression analysis
Calculus		Normal distribution
Linear algebra	Neural network theory	Multivariable differential
Statistics	Matrix operation	Support vectors machines
Discrete mathematics	Multivariable Vector	Support vectors regression.
	Multivariable Statistics	Makov Chain

4.1.3 Mathematics in the field of Generative adversarial networks

이 트랙은 다양한 생명공학에 관련된 기술 그리고 그와 관련된 지식도 필요하고, 수학의 관련도 기초분야에서 상당히 높은 차원의 전문적 지식도 필요로 한다. 실생활에 가장 밀접한 것들의 표현이니까 많은 분야의 응용을 할 수 있기도 하다. 시각적인 분야인 보이는 것들 그림 사진 영상등과 청각적인 분야인 음성인식 소리 인식 등 가장 실생활과 관계가 있다. 소리나 그림등을 주기적인 진동의 표현이고 이를 컴퓨터 해석을 위해서는 변환이 필요하고, 잡음을 제거하고, 정밀성을 높여 왜곡된편

이를 줄여야 한다. 이에 불확실성이 항상 존재하므로 다변량의 통계학적 지식을 요구한다. 또 표현을 위해서는 그래픽 모형에 대한 이해 방향성그래픽과 비방향성그래픽등 지식도 필요하다. 함수해석을 위한 다양한 급수표현과 멱급수 해석도 필요 하다.

Table 3. GAN Track

Prerequisite	Major Required	Related field
		Multivariate function
Calculus	Regression analysis	Space vector
Discrete mathematics	Neural network theory	Graph theory
	Matrix operation	Linear interpolation
Statistics	Multivariable Vector	Biotechnology
Linear algebra	Multivariable Statistics	Power Serise Analysis
		Fourior Analysis

4.2. Mathematical Objectives

4.2.1 Functionalization of various patterns in materials

주어진 자료들 속에 존재하는 패턴을 분석하여, 다양한 함수로 추정한다. 추정된 함수가 적절한 가장 좋은 함수인지를 분석하고, 오차를 줄여 좋은 함수로 다양한 의사추천시스템에 적용하게 한다.

4.2.2 Motelling for solving mathematical relations

다양한 자료들 속에 있는 관계식을 찾아 특성 다항식이나 생성함수를 이용하여 일반항을 구 할 수 있게 한다.

4.2.3 Algorithm and Complexity Optimization

알고리즘의 뜻을 이해하고, 간단한 문제 해결을 위한 알고리즘을 작성 하게 한다. 알고리즘의 분석을 위해 알고리즘의 복잡도를 이해하게 한다. 최적의 알고리즘을 찾는다.

4.2.4 Regression Analysis and Model Fit Growth

회귀분석의 뜻과 용어를 이해 하게 한다. 패턴을 모델링하여 불확실한 미래 자료현상에 발생하는 문제 상황에 적용 할 수 있게 한다. 함수들의 연립방정식을 동시에 만족하는 패턴을 찾기 위해 행렬을 이용하여 적용하게 한다. 여러 함수의 행렬 전이의 최단경로 문제와 스케줄링의 문제를 행렬을 이용하여 해결할 수 있게 한다. 그래프의 평면성 판단, 평면그래프의 쌍대그래프를 이해하게 한다. 색칠 문제를 이해하고 이를 이용하여 여러 가지 문제를 해결할 수 있게 한다. 다양한 분야의 의사추천이나 ,무인주행들의 불확실성에 대한 오류를 줄이는 함수를 찾는다

4.2.5 Modeling

여러 가지 현상을 종합 하여 얻어진 자료 사이의 패턴을 분석, 함수로의 모형 화 시키는 능력이 생기게 한다.

여러 방정식의 이해, 미분방정식의 이해와 해석, 마코프 체인(Markov Chain) 모델을 이해하게 한다.

4.2.6 Bayesian probability calculation

불확실한 현상에 대한 현명한 의사결정 방법을 위해 확률의 연산, 베이지안 확률의 이해와 응용, 확률변수와 확률함수의 이해. 모수에 대한 가설 검정 기법 습득. 전이 과정에서 생기는 불확실성의 계산방법.

4.2.7 Statistics

확률변수와 정규분포, 다변량 정규분포의 이해와 응용. 표본 자료들이 주어지는 정규성의 이해. Test 자료의 정규화를 이해하게 한다.

4.2.8 Understanding the data

자료들의 독립적인 자료들의 이해. Rank등 차원과 벡터를 이용한 독립확인. 행렬의 고유분해. LU 분해법등 연립방정식의 이해와 해법. 다양한 연립방정식의 해법을 이해하게 한다.

4.2.9 Interpretation of Complex Analysis

Topology공간에 대한 해석, 3차원이상의 공간의 표현 Metric Space을 통한 변화분석.[15,16,17]

V. Simulation

컴퓨터 공학을 전공하는 학생으로 기초 미분적분학은 수강한 2학년 학생들을, 머신러닝에 선행으로 필요한 데이터 과학이라는 과목을 선택 수강한 학생과 수강하지 않은 학생들의 음성인식에 대한 텐서프로 프로그램에 소스 파일을 이해시키고, 응용 프로그램 프로젝트의 이해도를 측정하는 시험 후 성적비교에 대한 통계자료분석을 시행하였다.

대상: I 대학교에 재학 중인 공과대학 2학년 학생 100명을 단순 확률 추출법에 의하여 표본을 50명씩 구성 후 다음과 같이 교육 하였다.

집단a: 50명의 학생을 기존의 머신러닝에 선행 학습으로 데이터 과학기초과목 수강

집단b: 50명의 학생을 기존의 머신러닝에 선행 학습으로 데이터 과학기초과목 수강하지 않고 일반적 이산수학 강의 수강

모의 실험을 시행 하기 전 학생들의 사전성적을 알기위해 2학년 1학기 교과목인 통계학 과목의 성적 자료를 통계 소프트웨어인 SPSS를 사용하여 분석 하면 표 4와 같다.

Table 4. Statistics Data of Statistics

	N	Min	Max	Avg	Std. Dev.
Groupa	50	21.00	57.00	38.06	10.448
Groupb	50	22.00	59.00	41.02	9.652

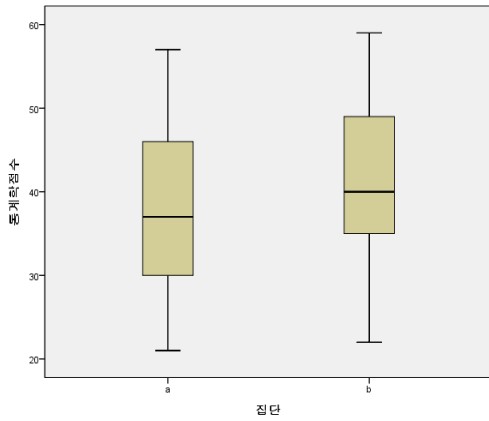


Fig. 1. Box-plot of Table 4

표 4와 그림1 에 의하면 집단b가 집단a보다 평균값이 높고, 성적의 표준편차도 작아 평균에 밀집 되어 있고, 상위 점수에 밀집 되어 있으며 전체 성적의 퍼진 정도도 집단a 보다 작다. 집단a 학생들의 언어 이해도는 집단b의 언어 이해도 보다 좋게 나타났다.

이 표본 학생들을 대상으로 머신러닝의 선행적인 교유내용인 데이터과학 기초를 수강한 학생과 수강하지 않은 학생간의 2학년 2학기 교과목인 소프트웨어 아키텍처에 대한 성적을 통계 소프트웨어인 SPSS를 사용하여 분석하면 표 5와 같다.

Table 5. Statistics Data of Software Architecture

	N	Min	Max	Avg	Std. Dev.
Group a	50	6.00	97.00	59.24	20.21
Group b	50	26.00	97.00	66.64	19.615

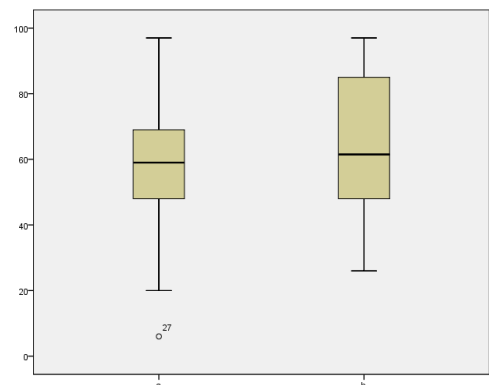


Fig. 2. Box-plot of Table 5

통계 분석한 결과 표5와 그림2에 의하면 데이터과학 기초 수강 학생의 성적이 높으나 집단a의한 평균성적과 큰 차이는

없다. 데이터 과학 기초 수강 상위의 학생의 경우 성적의 변화가 적으나, 이해력이 떨어진 학생의 성적 하락폭이 수강하지 않은 학생들의 성적 보다 상대적으로 큰 편차가 존재 한다. 또 특이하게 수강하지 않은 학생들의 성적변화에서 낮은 점수의 이상 값도 존재 한다. 데이터 과학 기초 수강 학생 성적은 중앙값을 중심으로 높은 점수의 학생과 낮은 점수의 학생의 분포가 대칭을 이루지 않고 상위 쪽으로 퍼진 정도가 길어서, 상하 대칭이 이루어지지 않고 있고, 미 수강 학생들의 성적은 대칭을 이루고 있는 것을 보아서는 성적 향상이 안된 것 같다. 데이터 과학 기초 수강생들 중 상위권 학생들은 높은 성적의 향상이 있다고 해석 할 수 있다.

각 집단 간 두 교과목의 성적 변화를 관찰하기 위해 “각 학생들간의 통계학 성적과 소프트웨어 아키텍처 성적 차이가 없다” 라는 귀무 가설을 Pairwise t-test 시행한 결과 표 6, 표 7과 같다.

Table 6. Statistical Table of $\mu_{a1} - \mu_{a2}$

	Statistical of the difference in score of group 1					
	M	S.D	Confidence interval 95%	t value	p value	
	-.580	7.502	2.322	6.7	4.15	.00

Table 7. Statistical Table of $\mu_{b1} - \mu_{b2}$

	Statistical of the difference in score of group 2					
	M	S.D	Confidence interval 95%	t value	p value	
	4.54.	7.801	-2.322	1.52	.547	.58

표6 과 표7에 의하여 보면 데이터 과학 기초 수업을 받지 않은 학생의 두 교과목의 차이가 있는지에 대한 가설검정의 결과 유의수준 0.05에서 유의확률이 0.00로 가설을 기각 할 수 없다는 결과가 나왔다. 즉 데이터 과학 기초 수강을 안 한 기본 이 산수학과 일반적인 미분적분학을 수업한 학생 후 두 교과목에 대한 학생 성적의 평균이 변화가 없다. 그러나 데이터 과학 기초에서 머신러닝에 필요한 세분화된 행렬 벡터의 연산과 개념을 선 수강한 학생들의 결과는 유의수준 0.05에서 기각 되어 학생 성적의 변화가 있었다. 즉 통계적으로 데이터 과학 기초 수업을 수강한 학생들의 두 교과목의 학생 평균 성적에 차이에 유의적인 변화가 있다고 할 수 있다.

위의 모의 실험 결과가 데이터 과학 기초 수업에 효과가 있다고 단정 할 수 는 없다. 성적의 차이가 단지 수학 교과목의 교육 효과라도 할 수 없고, 또 두 교과목의 성적을 비교 하는데 논리적 타당성도 확인 할 수 없고, 연관성도 없다 라고 할 수도 있다. 그러나 금번 연구의 실험을 통해 일부분이라도 머신러닝에 필요한 세부 분야를 선 수강하게 하면 어떤 방향에서라도 효과가 있었다는 것을 보여 주고 있으며, 또 하나 중요한 사항은 학생들이 필요성을 인식한 적극적이고 흥미로운 수업이 이루어 졌다는 점도 볼 수 있었다.

VI. Conclusions

오늘날 컴퓨터 공학의 머신러닝 분야는 급속하게 변화하고, 수많은 변수들과 자료를 이용 한다. 이러한 변화에 신속하게 대응할 수 있도록 전공 교양과 전공 이론을 바탕으로 실무적이고 전문화된 인력들에 대한 교육 및 수준이 절실히 요구되고 있다.

수학 교육은 컴퓨터 공학의 다양한 분야에서 이론 교육의 지식과 기술을 응용할 수 있는 능력을 무한대로 확장 할 수 있게 제공한다. 본 연구에서는 컴퓨터공학의 머신러닝의 세부전공트랙 별로 구분하여 기초 수학과 응용된 수학 분야를 제시하여 연구하고자 하는 부분의 연관성을 미리 파악하여 필요한 분야를 중점 교육 할 수 있는 교육 방향을 제시해 보았다. 머신러닝의 많은 분야는 점점 수학 전공지식과 깊이가 넓고 어려워지고 있다. 이 많은 분야를 다 이해 할 수는 없다. 필요한 수학분야를 집중 공부하여 전공간의 융합연구에 시너지 효과를 발휘 할 수 도 있다. 수학의 이해를 통해 기초학문과 응용학문 간의 조화를 통해 현장적응력이 우수한 전문 인력을 양성하고 기술력과 경쟁력 있는 인재 양성에 기여할 것으로 기대한다.

REFERENCES

- [1] http://www.acm.org/binaries/content/assets/education/cs2013_web_final.pdf
- [2] <http://dl.acm.org/citation.cfm?id=3289258.3277567&coll=pOrtal&dl=ACM>
- [3] <http://ko.coursera.org/specializations/mathematics-machine-learning>
- [4] Ministry of Education and Human Resources Development(1997). Mathematics Curriculum, Ministry of Education Notice No. 1997-15, 7th Mathematics Curriculum, Korean Textbook.
- [5] Ministry of Education and Human Resources Development(2001). High school curriculum commentary 5, Mathematics, Korean school curriculum.
- [6] Ministry of Education and Human Resources Development (2003). High School Discrete Mathematics, Kangwon University 1st Book Compilation Committee, Genius Education.
- [7] NCTM (1989). Curriculum and Evaluation Standards for School Mathematics.
- [8] NCTM (2000). Principles and Stangards for School Mathematics.
- [8] Mathematics Classroom, A Contemporary Approach to Teaching Grades 7-12, Brooks/Cole.
- [9] Lee.Jae-Hag (2003).However, Disciplinary Mathematics Curriculum at Teacher Training College, The Mathematical Education Society of Korea, pp.43-52,
- [10] Ministry of Education and Human Resources Development (2003). Probability and statistics of high school, Probability and Statistics of Korea 1 teacher 's book, Compilation committee, Genius education.
- [11] Lee, Jun-Yeol (2002). A Study on the Implementation Plan of the 7th Curriculum in Discrete Mathematics Education, 41 (1), pp.127-137,
- [12] Bag, Jin-Heung (1996). New Discrete Mathematics. Gyousa.
- [13] Yu Won-sik (1992). Discrete Mathematics, Seoul. Gyeongmunsa.
- [14] Lee, Seung-Woo (2008). Analysis and Suggestion of Mathematics and Statistics Related to the Development of Computer Software Course, A Series of Mathematics Education in Korea A (Mathematics Education) 47 (2) pp.225-232,
- [15] M.T.Quazi, S.C,Mukho-padhyay, N..K. Suryuyadevara, and Y.M.Huang(2012). "Towards the smart sensors based human emotion recognition" Proc of IEEE Int, Instrumentation and Measurement Technonology Conf, p2365 -2370
- [16] <https://mingrmmer.com/translation-the-mathematics-of-machine-learning>
- [17] Field Cady (2018). The Data Science Handbook, Seoul. Hanley Media

Author



Sang Pyo Jun received a Ph.D degree from Statistics of Inha University, in 2000. Now he has been working for Namseoul University, Department of General Education. His research interests include data mining, linear modeling and deep

learningg.