# **1학년 수강 과목을 통한 학부 예측 프로그램 프로젝트**[**¶**](https://render.githubusercontent.com/view/ipynb?commit=382e530d77b46a0d55fcf5331ff7b88575b99cf8&enc_url=&nwo=idebtor%2FJoyAI&path=2019-1ProjectSamples%2F8%EC%A1%B0-%EC%8B%A4%EC%8B%9C%EA%B0%84%EA%B0%90%EC%A0%95%EA%B3%BC%EC%84%B1%EB%B3%84%EC%9D%B8%EC%8B%9D%2F8%E1%84%8C%E1%85%A9_Emotion_Gender_Classification__%E1%84%91%E1%85%B3%E1%84%85%E1%85%A9%E1%84%8C%E1%85%A6%E1%86%A8%E1%84%90%E1%85%B3.ipynb&repository_id=171858338&repository_type=Repository#Emotion/Gender-Classification-%ED%94%84%EB%A1%9C%EC%A0%9D%ED%8A%B8)

### **한동대학교 18학번들의 수강과목을 통한 학부를 예측하는 AI 프로젝트**

### **모두를 위한 인공지능 Final project - 11조 윤희원(21600481)/김유빈(21900167)**

**프로젝트 소개**

MNIST BGD 모델과 한동대학교 18학번 1학년 수강과목 자료를 통해, 2학년 때 선택할 학부를 예측할 수 있다.

#### **프로젝트 목표**

1. 18학번들의 수강과목 데이터를 활용하여 2학년이 되었을때 어떤 학부를 선택할지 예측할 수 있다.
2. MNIST 모델을 사용한 학습 모델의 구조를 이해할 수 있다.
3. 18학번이 수강한 과목 데이터를 통해 어떤 학부 전공을 선택하게 될지 직접 인공지능 모델을 학습시킬 수 있다.

#### **프로젝트 개요**

1. '모두를 위한 인공지능의 활용’ 수업 github에서 JoyML11-3BatchGD.ipynb 파일을 다운로드 한다
2. MNIST BGD 모델 코드를 활용하여 우리가 실제 학습하고자하는 신경망의 형태로 변환한다.
3. 18학번들의 수강과목 데이터를 다운받아서 필요한 정보들만 추출하여 CSV 형태로 가공한다.
4. 가공된 데이터를 Jupyter Notebook으로 읽어들인 후 학습시킨다.
5. 테스트용 데이터를 활용하여 예측 정확도를 확인한다.

#### **Script**

안녕하세요 저희는 모두를 위한 인공지능의 활용 수업 중 11조 윤희원, 김유빈입니다. 저희는 1학년들의 수강과목을 통해 학부를 예측하는 프로그램을 만들고자 하는 프로젝트를 진행했습니다. 이는 한동대학교 18학번들의 수강과목을 통해 학부를 예측하는 AI 프로그램을 만드는 것인데요, 저희 프로젝트는 기본적으로 수업시간에 배운 MNIST모델을 바탕으로 만들어졌는데요 이를 데이터에 맞게 변형하여 주어진 데이터를 바탕으로 인공지능을 학습시키고 학습된 데이터를 바탕으로 추가적인 데이터를 입력하여 가공하고 주어진 데이터에 맞게 학습시키고 신경망을 통해 학습된 데이터를 통해 추가적인 학부를 얼마나 잘 예측하는지를 알아보는 것입니다.

저희의 프로젝트는 다음과 같습니다. 1. 18학번들의 수강과목을 통해 2학년이 되었을때 어떤 학부를 선택하게 될지 예측할 수 있다. 2.MNIST모델을 이용한 학습모델의 구조를 이해할 수 있다. 3. 18학번들이 수강한 과목을 통해 어떤 학부를 선택하게 될지 직접 인공지능 모델을 통해 학습을 시킬 수있다.

그래서 저희 코드를 어떻게 만들었는지 그 과정을 여러분들께 설명해드리고자 합니다. 먼저 저희 수업시간때 활용한 github를 사용하여 여러분들이 확인할 수 있듯이 활용할 듯이 deep class 속에 있는 batch gredient 를 활용하고자 합니다. 제 노트북에 이미 설치되어 있는 이 코드를 열어볼 텐데요. 해당 코드에서 저희는 MNIST에서 활용되는 경사하강법 신경망에 대해 알 수 있습니다. import 되는 joy python 에서 여러분들이 다운받으실 수있습니다. 미리 파이썬 문서를 다운로드 받아 주시고 저 같은 경우는 제가 사용할 jupiter 노트북에 담겨있는 파일 안에 미리 다운받아 놓았습니다. 여기 보이시는 mnist를 활용한 코드를 활용하여 학부 예측 프로그램을 만들어보았습니다. 우선 이 코드에서 활용되는 코드를 보면 먼저 initialize MNIST 에 사용되는 코드를 활용하여 이를 학부 예측 프로그램으로 만들었습니다. 초기에 입력층과 은닉층 사이의 가중치들로부터 오류를 수정하고 출력층과 은닉층의 사이에 초기화 하는 함수로 1~-1사이의 값이 주어지고요, 다음으로 순전파의 feedforward 함수 를 내장해두었습니다. 입력층과 은닉층간의 w 행렬곱셈을 예측값과 feed 함수는 실제값과 예측값을 y과 y^ array의 shape을 결정하는 함수입니다. y 값이 주어졌을때 역전파 과정에서 행렬의 크기가 맞지 않으면 feed 함수를 통해 그 문제를 해결시킬 수있습니다. 자세한 내용은 joypy에 들어가보면 대략 880번대에 있는데 one\_hot\_incoding 함수가 어떻게 내장되어있는지 살펴보시면 좋을 것 같습니다. 그리고 이를 통해 feed 함수 이후에 실제적으로 epoch 즉 반복되는 횟수를 바탕으로 역전파를 통해 데이터의 error들을 실제값과 예측값의 차이를 바탕으로 어떻게 가중치를 조절할지 역전파 과정이 소개되어있습니다.

이 것에 대한 자세한 내용은 교수님께서 올려주신 pdf함수를 사용하면 좋을 것 같습니다. 그 다음은 predict함수는 z1, z2, z3 순전파 함수를 epoch로 예측값을 g,g’로 시그모이드 함수와 시그모이드의 미분을 의미하며 최종적으로 evaulate 학습과정에서 학습을 최종적으로 마쳤을때 인공신경망의 정확도를 확인할 수 있습니다. 해당하는 내용을 지정해준 이후에 밑으로 내려와 실제 학습시킬 코딩을 확인하겠습니다. 앞서 말씀드린 joy, csv,numpy를 불러들이기 위해 array 수학적 연산을 쉽게 하기위해서 불러들입니다. data1는 제 노트북 기준으로 파이썬의 class1으로 가공된 데이터인데 18학번들의 수강과목들을 얻기위하여 저희들은 교무팀에 연락하여 전산팀에게 제공받은 과목에 대한 내용을 필요한 부분만 추출해 가공하여 0-200은 각각의 과목을 나타내고 있습니다. 각 숫자별로 대응되는 이에 대한 정보를 나열해놓은 것입니다. 0같은 경우는 학부선택에 있어서 거의 영향을 미치지 않는 것(창조와 진화, 공동체리더십훈련, 채플)등을 나타냈습니다. major 파일에는 각 학생들이 어떤 학부를 선택했는지 나타내고 있습니다. 이런 가공된 데이터들이 python 폴더에 있고 데이터를 바탕으로 진행할 텐데 의 과목, 학부를 넣습니다. 저희의 입력층을 (600,200) 학생의 수만큼을 0으로 초기화되도록 하고 for문으로 과목을 수강했을 경우 해당 과목을 수강했을때 index값을 1로 표현했습니다. MNIST의 경우, 다음으로 y출력층은 600으로 입력층과 같이 for문을 통하여 0으로 초기화되는 함수를 나타내고 있습니다. 미리 설정해두었던 batch gredient를 각각 입력,은닉,출력층으로 실행시키면 보시는바와 같이 1000번 트레이닝할때마다 보여주고 최종적으로 6000번의 트레이닝을 했을 때 저희가 만든 6000번 학습했을때 해당 의미있는 노드는 1로, 무의미한 노드는 0으로 표현했습니다. 현재 저희가 만든 6000번을 했을 때 약 97퍼센트의 정확성을 확인할 수 있다. data3,4로 남겨두었던 34명의 학생들을 입력하고 이를 테스트 한 결과 약 20 퍼센트의 확률로 저희는 기존의 600명의 데이터를 통한 새로운 데이터가 왔을때 어려움을 겪고있지만 self accuracy는 문제 없는것으로 보입니다. 저희는 데이터가 부족해 12개의 과목중 하나를 선택할 확률로 약 9퍼센트에 불과할텐데 20퍼센트가 보이는 것으로 보아 충분히 예측을 성공적으로 한 것으로 보입니다.