

2025년 1학기 머신러닝 중간고사 대비문제

문제 2. 아래 표와 같이 고객이 action장르와 romance장르에 각각 점수를 매겼을 때 3개의 영화 중 1개를 추천하는 시스템을 설계하고자 한다. 이때 다음을 수행 하시오. (70점)

- 문제 2-1. 정규화 된 학습용 데이터셋을 이용해 3개 SVM모델을 학습하고 각 모델 파라미터 (W, b)를 출력하시오. (25점)
 - ❖ Gradient descent를 이용해 optimal hyperplane 도출
 - ❖ Learning rate: 0.001 / Iteration: 1000 / Loss function: Hinge loss
- 문제 2-2. 정규화 된 학습용 데이터셋을 이용해 학습된 KNN 모델에서 1에서 30 사이의 홀수 값 중 최적의 K값을 선정하시오. (25점)
 - ❖ 테스트용 데이터셋을 이용해 모든 후보 K에 대해 정확도 측정 및 가장 높은 정확도를 가지는 K값 선정
 - ❖ L2 distance를 사용하여 sample 사이의 거리 측정
- 문제 2-3. 신규 데이터 (Action 26.5점, Romance 66.5점)를 SVM, KNN 모델을 이용하여 각각 추천 영화를 예측하시오. (20점)
 - ❖ SVM: 문제 2-1에서 학습한 3개 SVM 모델을 이용해 신규 데이터 입력 시 출력 값이 +1 값을 가지는 추천 영화 선정
 - ❖ KNN: 문제 2-2에서 학습한 KNN 모델과 최적의 K값을 이용해 신규 데이터에 대한 추천 영화 선정

Index	X0 (Action)	X1 (Romance)	Y (Movie)
0	49.14	79.97	노트북
1	85.28	53.01	존 워
2	33.51	23.33	레옹
...
1999	25.88	40.63	레옹

- [Label]
- ❖ 존 워 (액션 장르): 0
 - ❖ 노트북 (로맨스 장르): 1
 - ❖ 레옹 (혼합 장르): 2

참고사항

data를 A행렬로 변환

- `A = np.array(data)`

A 행렬의 형상 확인

- `size = A.shape`

A 행렬과 똑같은 크기의 1로 채워진 행렬 생성

- `arr = np.ones_like(A)`

A 행렬과 똑같은 크기의 0으로 채워진 행렬 생성

- `arr = np.zeros_like(A)`

조건에 따라 값 선택

- `A = np.where(조건, 참, 거짓)`

Gradient descent

$$\theta_{t+1} = \theta_t - \alpha \frac{\partial L}{\partial \theta_t}$$

❖ θ : *Weight / bias*

❖ α : *Learning rate*

❖ L : *Loss*

Hinge loss

$$Loss = \max(0, 1 - y_i(W^T x_i + b))$$

❖ y_i : *Label data*

❖ x_i : *Input data*

❖ W^T : *Model weight*

❖ b : *Model bias*