

1.

2

▶(3, 1) / (3,) 의 차이

[[1][2][3]]

[1 2 3]

▶행렬의 곱

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} = \begin{bmatrix} 58 \\ \end{bmatrix}$$

3.

▶파이토치에서 사용되는 자료 구조

용어	차원	Shape	사용하는 곳	예시
스칼라	0차원	()	학습률, loss	[1]
벡터	1차원	(3,)		[1, 2, 3]
행렬	2차원	(2, 3)	이미지	[[1, 2, 3], [1, 2, 3]]
텐서	3차원 이상	(2, 2, 3)	다차원 배열	[[[1, 2, 3] [1, 2, 3]],

4.

▶머신러닝 종류

지도학습 : 피쳐와 라벨이 주어진 학습

비지도학습 : 정답없이 패턴을 학습(그룹화, 이상탐지에 사용)

강화학습 : 알파고 같은 ai (혼자 실행하고 학습)

▶인공지능, 머신러닝, 딥러닝

인공지능 (AI)

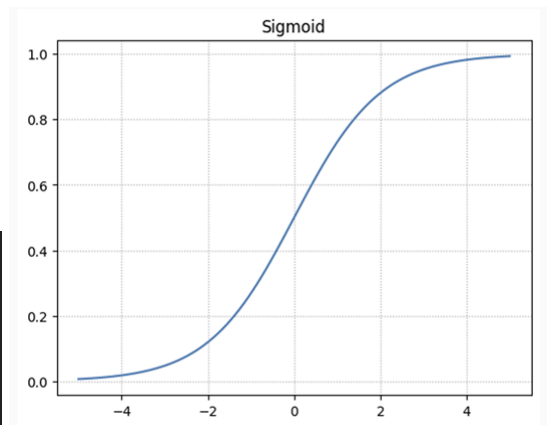
└─ 머신러닝 (ML)

└─ 딥러닝 (DL)

항목	인공지능 (AI)	머신러닝 (ML)	딥러닝 (DL)
정의	기계가 인간처럼 사고·판단·행동	데이터로부터 스스로 학습	인공신경망으로 학습을 수행하는 ML의 한 분야
방식	규칙 기반 or 학습 기반	알고리즘(결정트리, SVM, 선형회귀 등)	심층신경망(CNN, RNN, Transformer 등)
예시 기술	체스 AI, 자율주행, 로봇, 챗봇	스팸 필터, 추천 시스템, 가격 예측	음성 인식, 이미지 분류, GPT 같은 언어 모델
특징		비교적 얇은 모델	대규모 데이터 필요, 고성능 GPU 필요

►sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



0에서 1

Saturation

Not zero-centered

►Tanh

$$\delta(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

-1에서 1

Saturation

zero-centered

►ReLU

$$\delta(x) = \max(0, x)$$

0에서 ∞

Not Saturated

계산효율이뛰어남

5.

▶경사하강법

함수의최솟값또는최댓값을찾는방법



이 지점으로 수렴하도록 **learning rate**조절

종류 :

경사하강법 – 정확함(안정적임) **but** 한번에 모든 데이터를 사용해서 시간과 **ram**문제

확률적경사하강법 – 빠름 **but** 모든 데이터 사용이 아니기 때문에 정확하기 않음

미니배치경사하강법 – 빠르고 안정적임 **but** 배치에 크기 따라 모델 성능이 다름

6.

▶비용함수

손실함수를 시그마 한 것 = 비용함수의 시그마 내부는 손실함수임

회귀 예측

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$

$$J(\theta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

이진 분류

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y_i \log(h_{\theta}(x_i)) + (1 - y_i) \log(1 - h_{\theta}(x_i))]$$

$$\mathcal{L}(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

다중 분류

$$J(\theta) = -\sum_{i=1}^n y_i \log(h_{\theta}(x_i))$$

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{j=1}^C y_j \log(\hat{y}_j)$$

$$J(W, b) = \frac{1}{N} \sum_{i=1}^N \left[-\sum_{j=1}^C y_j^{(i)} \log(\hat{y}_j^{(i)}) \right]$$

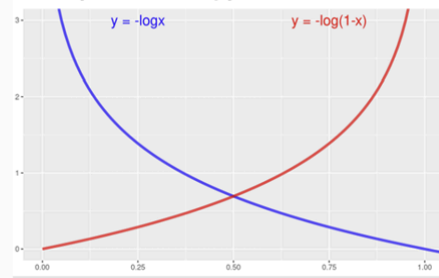
►Softmax

$$h_{\theta}(x_i) = \frac{e^{z_i}}{\sum_{j=1}^M e^{z_j}}, k = 1, 2, \dots, M$$

▶ 소프트맥스 (Softmax) 함수와 Crossentropy

$$J(\theta) = - \sum_{i=1}^n y_i \log(h_{\theta}(x_i))$$
$$\text{where, } h_{\theta}(x_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, k = 1, 2, \dots, M$$

Binary crossentropy



7.

▶선형 회귀

주로 MSE를 손실함수로 사용

$$y = wx + b$$

w : 가중치(기울기)

b : 편향

8.

▶데이터 구성

학습 데이터 : 모델을 학습

검증 데이터 : 하이퍼 파라미터 결정

테스트 데이터 : 모델을 평가

▶교차 검증

데이터를 여러 개의 부분(fold)으로 나눠서 모델을 평가하는 방법

모델이 데이터에 과적합 되는 것을 방지, 일반화된 성능 평가 가능

Fold 중 1개를 테스트 데이터 나머지를 학습데이터로 구성을 k개 fold 각각 수행

▶평가 지표

정확도 (Accuracy)

전체 중 라벨 맞춘 비율

$$\frac{TP + TN}{TP + TN + FP + FN}$$

정밀도(Precision)

예측이 양성 중 정답 비율

$$\text{Precision} = \frac{TP}{TP + FP}$$

재현율(Recall, Sensitivity)

정답이 양성 중 정답 비율

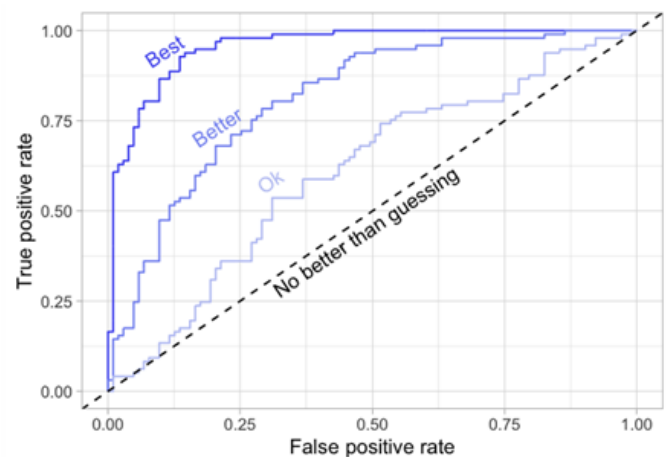
$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-점수(F1-Score)

정밀도와 재현율의 조화평균

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

AUC



9.

▶다중 분류 loss

$$L(p) = -\log(p)$$