

No. date / /

title

영어 노트

[illegible]

혼자 공부하는 머신러닝+딥러닝

혼자 공부하며 함께 만드는

혼공 용어 노트

목차

가나다 순

가중치(또는 계수) weight(또는 coefficient)	11	로지스틱 회귀 logistic regression	13
강인공지능 vs 약인공지능	06	릿지 회귀 ridge regression	12
검증 세트 validation set	16	말뭉치 corpus	26
결정 트리 Decision Tree	16	맷플롯립 matplotlib	07
결정계수 coefficient of determination	10	머신러닝과 딥러닝 machine learning과 deep learning	06
과대적합 vs 과소적합 overfitting vs underfitting	10	미니배치 경사 하강법 minibatch gradient descent	14
교차 검증 cross validation	16	배열 인덱싱 array indexing	09
군집 clustering	19	배치 경사 하강법 batch gradient descent	15
그레이디언트부스팅 gradient boosting	18	밸리드 패딩 valid padding	23
그리드 서치 Grid Search	17	변환기 transformer	12
넘파이 numpy	09	부트스트랩 샘플 bootstrap sample	17
다중 분류 multi-class classification	13	불리언 인덱싱 boolean indexing	13
다중 회귀 multiple regression	11	브로드캐스팅 broadcasting	09
다항 회귀 polynomial regression	11	비지도 학습 unsupervised learning	08
단어 임베딩 word embedding	27	샘플링 편향 sampling bias	09
데이터 전처리 data preprocessing	09	선형 회귀 linear regression	11
드롭아웃 dropout	22	셀 cell	26
딥러닝 deep learning	20	셀 상태 cell state	27
라쏘 회귀 lasso regression	12	소프트맥스 함수 softmax function	14
랜덤 서치 Random Search	17	손실 함수 loss function	15
랜덤 포레스트 Random Forest	17	순차 데이터 sequential data	25
렐루 함수 ReLU Function	21	순환 신경망 recurrent neural network, RNN	26
로지스틱 손실 함수 logistic loss function	15	스트라이드 stride	24

시계열 데이터 time series data	25	크로스엔트로피 손실 함수 cross-entropy loss function	15
시그모이드 함수 sigmoid function	13	텐서플로 TensorFlow	20
심층 신경망 deep neural network, DNN	21	토큰 token	27
앙상블 학습 ensemble learning	17	특성 feature	07
에포크 epoch	14	특성 맵 feature map	23
엑스트라 트리 extra trees	18	패딩과 세임 패딩 padding과 same padding	23
옵티마이저 optimizer	22	풀링 pooling	24
원-핫 인코딩 one-hot encoding	20	피드포워드 신경망 feedforward neural network, FFNN	25
은닉 상태 hidden state	26	필터 filter	23
은닉층 hidden layer	21	하이퍼파라미터 hyperparameter	12
이너서 inertia	19	합성곱 convolution	23
이진 분류 binary classification	06	확률적 경사 하강법 Stochastic Gradient Descent	14
인공신경망 artificial neural network, ANN	20	활성화 함수 activation function	20
인공지능 artificial intelligence	06	회귀 regression	10
적응적 학습률 adaptive learning	22	훈련 training	07
정형 데이터 vs 비정형 데이터		훈련 데이터 training data	08
structured data vs unstructured data	17	훈련 세트와 테스트 세트 train set와 test set	09
주성분 분석 principal component analysis, PCA	19	히스토그램 histogram	19
지도 학습 supervised learning	08	히스토그램 기반 그레이디언트 부스팅	
차원 축소 dimensionality reduction	19	Histogram-based Gradient Boosting	18
최대 풀링과 평균 풀링 max pooling과 average pooling	24	힌지 손실 hinge loss	15
코랩과 노트북 Colab과 Notebook	06		
콜백 callback	22		

ABC 순

activation function 활성화 함수	20	dimensionality reduction 차원 축소	19
adaptive learning 적응적 학습률	22	DNN, deep neural network 심층 신경망	21
ANN artificial neural network 인공신경망	20	dropout 드롭아웃	22
array indexing 배열 인덱싱	09	ensemble learning 앙상블 학습	17
artificial intelligence 인공지능	06	epoch 에포크	14
batch gradient descent 배치 경사 하강법	15	extra trees 엑스트라 트리	18
binary classification 이진 분류	06	feature 특성	07
boolean indexing 불리언 인덱싱	13	feature map 특성 맵	23
bootstrap sample 부트스트랩 샘플	17	FFNN feedforward neural network 피드포워드 신경망	25
broadcasting 브로드캐스팅	09	filter 필터	23
callback 콜백	22	gradient boosting 그레이디언트부스팅	18
cell 셀	26	Grid Search 그리드 서치	17
cell state 셀 상태	27	GRU Gated Recurrent Unit	28
clustering 군집	19	hidden layer 은닉층	21
coefficient of determination 결정계수	10	hidden state 은닉 상태	26
Colab과 Notebook 코랩과 노트북	06	hinge loss 힙지 손실	15
convolution 합성곱	23	histogram 히스토그램	19
corpus 말뭉치	26	Histogram-based Gradient Boosting	18
cross validation 교차 검증	16	hyperparameter 하이퍼파라미터	12
cross-entropy loss function 크로스엔트로피 손실 함수	15	inertia 이너서	19
data preprocessing 데이터 전처리	09	k-최근접 이웃 분류 vs k-최근접 이웃 회귀	10
Decision Tree 결정 트리	16	k-최근접 이웃 알고리즘	07
deep learning 딥러닝	20	k-means algorithm k-평균 알고리즘	19

lasso regression 라쏘 회귀	12	ReLU Function 렐루 함수	21
linear regression 선형 회귀	11	ridge regression 릿지 회귀	12
logistic loss function 로지스틱 손실 함수	15	RNN recurrent neural network 순환 신경망	26
logistic regression 로지스틱 회귀	13	sampling bias 샘플링 편향	09
loss function 손실 함수	15	sequential data 순차 데이터	25
LSTM Long Short-Term Memory	27	sigmoid function 시그모이드 함수	13
machine learning과 deep learning 머신러닝과 딥러닝	06	softmax function 소프트맥스 함수	14
matplotlib 맷플롯립	07	Stochastic Gradient Descent 확률적 경사 하강법	14
max pooling과 average pooling 최대 풀링과 평균 풀링	24	stride 스트라이드	24
minibatch gradient descent 미니배치 경사 하강법	14	structured data vs unstructured data	17
multi-class classification 다중 분류	13	supervised learning 지도 학습	08
multiple regression 다중 회귀	11	TensorFlow 텐서플로	20
numpy 넘파이	09	time series data 시계열 데이터	25
one-hot encoding 원-핫 인코딩	20	token 토큰	27
optimizer 옵티마이저	22	train set와 test set 훈련 세트와 테스트 세트	09
overfitting vs underfitting 과대적합 vs 과소적합	10	training 훈련	07
padding과 same padding 패딩과 세임 패딩	23	training data 훈련 데이터	08
PCA. principal component analysis 주성분 분석	19	transformer 변환기	12
polynomial regression 다항 회귀	11	unsupervised learning 비지도 학습	08
pooling 풀링	24	valid padding 밸리드 패딩	23
Random Forest 랜덤 포레스트	17	validation set 검증 세트	16
Random Search 랜덤 서치	17	weight 가중치 (또는 계수)	11
regression 회귀	10	word embedding 단어 임베딩	27

01장

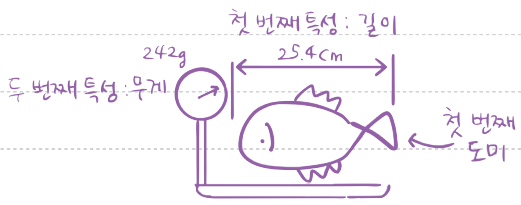
□ 인공지능	artificial intelligence	[01장 027쪽]
	학습하고 추론할 수 있는 지능을 가진 컴퓨터 시스템을 만드는 기술	
□ 강인공지능 vs		[01장 027쪽, 028쪽]
약인공지능	강인공지능은 인공일반지능이라고도 하고 사람의 지능과 유사(영화 속 전지전능한 AI)함. 약인공지능은 특정 분야에서 사람을 돕는 보조 AI(음성 비서나 자율 주행도 여기 포함)	
□ 머신러닝과	machine learning과 deep learning	[01장 028쪽, 029쪽]
딥러닝	머신러닝은 데이터에서 규칙을 학습하는 알고리즘을 연구하는 분야(대표 라이브러리는 사이킷런). 딥러닝은 인공신경망을 기반으로 한 머신러닝 분야를 일컫음(대표 라이브러리는 텐서플로)	
□ 코랩과 노트북	Colab과 Notebook	[01장 033쪽, 037쪽]
	코랩은 웹 브라우저에서 텍스트와 프로그램 코드를 자유롭게 작성 할 수 있는 온라인 에디터로 이를 코랩 노트북 또는 노트북이라 부름. 최소 단위는 셀이며 코드 셀과 텍스트 셀이 있음	
□ 이진 분류	binary classification	[01장 046쪽]
	머신러닝에서 여러 개의 종류(혹은 클래스) 중 하나를 구별해 내는 문제를 분류(classification)라고 부르며 2개의 종류(클래스) 중 하나를 고르는 문제를 이진 분류라 함	

□특성

feature

[01장 047쪽]

데이터를 표현하는 특징으로 여기서는 생선의 특징인 길이와 무게를 특성이라 함



□맷플롯립

matplotlib

[01장 048쪽]

파이썬에서 과학계산용 그래프를 그리는 대표 패키지

□k-최근접 이웃

k-Nearest Neighbors Algorithm, KNN

[01장 050쪽]

알고리즘
가장 간단한 머신러닝 알고리즘 중 하나로 어떤 규칙을 찾기보다는 인접한 샘플을 기반으로 예측을 수행함

□훈련

training

[01장 053쪽]

머신러닝 알고리즘이 데이터에서 규칙을 찾는 과정 또는 모델에 데이터를 전달하여 규칙을 학습하는 과정



02장

□ 지도 학습

supervised learning

[02장 067쪽]

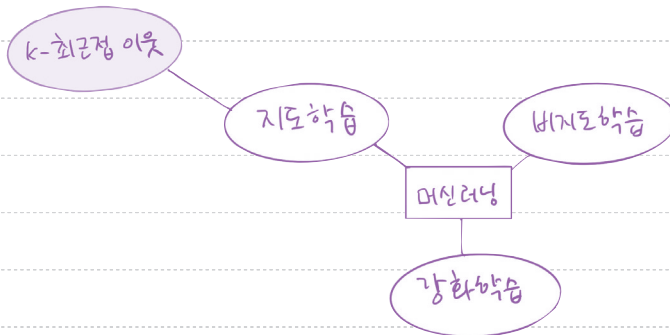
지도 학습은 입력(데이터)과 타겟(정답)으로 이뤄진 훈련 데이터가 필요하며 새로운 데이터를 예측하는 데 활용함. 1장에서 사용한 k-최근접 이웃이 지도 학습 알고리즘임

□ 비지도 학습

unsupervised learning

[02장 067쪽]

타겟 데이터 없이 입력 데이터만 있을 때 사용. 이런 종류의 알고리즘은 정답을 사용하지 않으므로 무언가를 맞힐 수가 없는 대신 데이터를 잘 파악하거나 변형하는데 도움을 줌

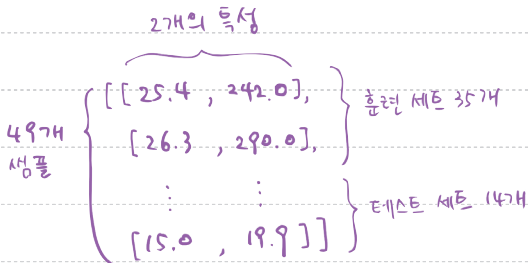


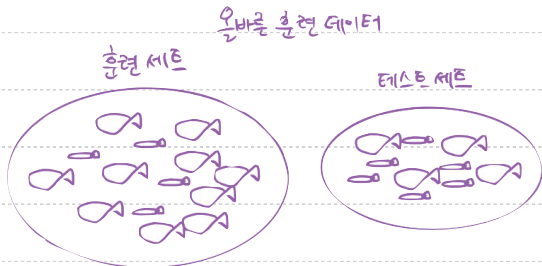
□ 훈련 데이터

training data

[02장 067쪽]

지도 학습의 경우 필요한 입력(데이터)과 타겟(정답)을 합쳐 놓은 것



□ 훈련 세트와	train set와 test set	[02장 068쪽]
테스트 세트	<p>모델을 훈련할 때는 훈련 세트를 사용하고 평가는 테스트 세트로 함. 테스트 세트는 전체 데이터에서 20~30%</p>	
□ 샘플링 편향	sampling bias	[02장 073쪽]
	<p>훈련 세트와 테스트 세트에 샘플이 고르게 섞여 있지 않을 때 나타나며 샘플링 편향이 있음. 제대로 된 지도 학습 모델을 만들 수 없음</p>	
		
□ 넘파이	numpy	[02장 073쪽]
	<p>파이썬의 대표적인 배열array 라이브러리로 고차원의 배열을 손쉽게 만들고 조작할 수 있는 간편한 도구를 많이 제공함. 공식 명칭은 NumPy</p>	
□ 배열 인덱싱	array indexing	[02장 077쪽]
	<p>넘파이 기능으로 여러 개의 인덱스로 한 번에 여러 개의 원소를 선택할 수 있음</p>	
□ 데이터 전처리	data preprocessing	[02장 099쪽]
	<p>머신러닝 모델에 훈련 데이터를 주입하기 전 가공하는 단계로 특성값을 일정한 기준으로 맞추어 주는 작업. 데이터를 표현하는 기준이 다르면 알고리즘을 올바르게 예측할 수 없음</p>	
□ 브로드캐스팅	broadcasting	[02장 100쪽]
	<p>조건을 만족하면 모양이 다른 배열 간의 연산을 가능하게 해 주는 기능</p>	

03장 ✓

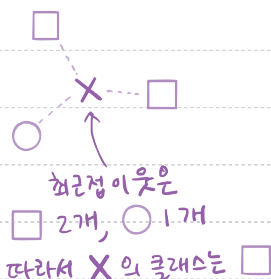
□ 회귀 regression [03장 115쪽]

클래스 중 하나로 분류하는 것이 아니라 임의의 어떤 숫자를 예측하는 문제

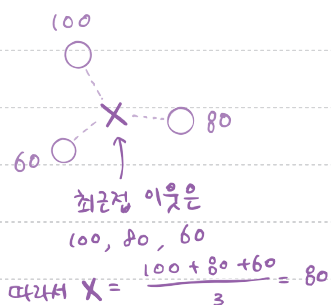
□ k-최근접 이웃 참고 용어 k-최근접 이웃 알고리즘 [03장 115쪽, 116쪽]

분류 vs k-최근접 이웃 분류
회귀 vs k-최근접 이웃 회귀

k-최근접 이웃 분류



k-최근접 이웃 회귀



□ 결정계수 (R^2) coefficient of determination [03장 120쪽]

회귀 모델에서 예측의 적합도를 0과 1 사이의 값으로 계산한 것으로 1에 가까울수록 완벽함

$$R^2 = 1 - \frac{(\text{타겟} - \text{예측})^2}{(\text{타겟} - \text{평균})^2}$$

□ 과대적합 vs overfitting vs underfitting [03장 122쪽]

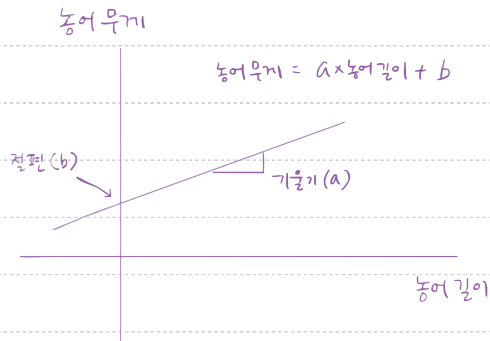
과소적합 과대적합은 모델의 훈련 세트 점수가 테스트 세트 점수보다 훨씬 높을 경우를 의미함. 과소적합은 이와 반대로 모델의 훈련 세트와 테스트 세트 점수가 모두 동일하게 낮거나 테스트 세트 성능이 오히려 더 높을 경우를 의미함

□ 선형 회귀

linear regression

[03장 135쪽]

널리 사용되는 대표적인 회귀 알고리즘으로 특성이 하나인 경우 어떤 직선을 학습하는 알고리즘(농어 무게 학습 그래프)



□ 가중치

weight (또는 coefficient)

[03장 137쪽]

(또는 계수)

선형 회귀가 학습한 직선의 기울기를 종종 가중치 또는 계수라 함

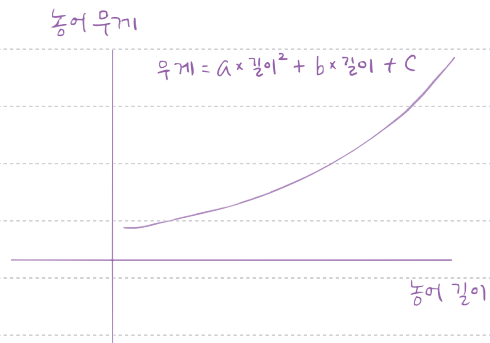
↑ 위 그림에서 기울기(a)

□ 다항 회귀

polynomial regression

[03장 139쪽]

다항식을 사용하여 특성과 타겟 사이의 관계를 나타낸 선형 회귀

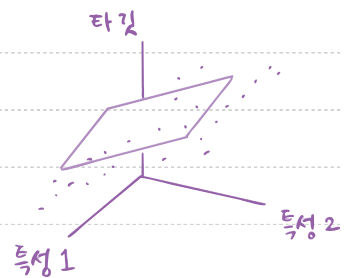


□ 다중 회귀

multiple regression

[03장 151쪽]

여러 개의 특성을 사용한 선형 회귀



□ 변환기

transformer

[03장 154쪽]

특성을 만들거나 전처리하는 사이킷런의 클래스로 타깃 데이터 없이 입력 데이터를 변환함

□ 릿지 회귀

ridge regression

[03장 160쪽]

규제가 있는 선형 회귀 모델 중 하나로 모델 객체를 만들 때 alpha 매개변수로 규제의 강도를 조절함. alpha 값이 크면 규제 강도가 세지므로 계수 값을 더 줄이고 조금 더 과소적합되도록 유도하여 과대적합을 완화시킴

□ 하이퍼파라미터

hyperparameter

[03장 161쪽]

머신러닝 모델이 학습할 수 없고 사람이 지정하는 파라미터

□ 라쏘 회귀

lasso regression

[03장 163쪽]

또 다른 규제가 있는 선형 회귀 모델로 alpha 매개변수로 규제의 강도를 조절함. 릿지와 달리 계수 값을 아예 0으로 만들 수도 있음



04장

□ 다중 분류 multi-class classification [04장 181쪽]

타깃 데이터에 2개 이상의 클래스가 포함된 문제

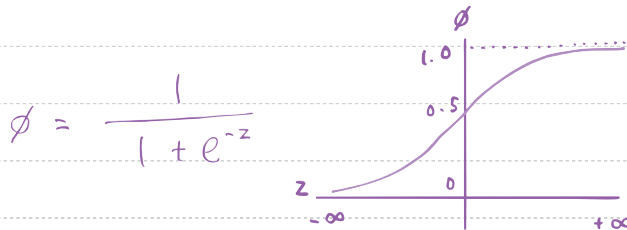
□ 로지스틱 회귀 logistic regression [04장 183쪽]

선형 방정식을 사용한 분류 알고리즘으로 선형 회귀와 달리 시그모이드 함수나 소프트맥스 함수를 사용하여 클래스 확률을 출력

□ 시그모이드 함수 sigmoid function [04장 183쪽]

 *logistic regression*

시그모이드 함수 또는 로지스틱 함수라고 부르며 선형 방정식의 출력을 0과 1 사이의 값으로 압축하며 이진 분류를 위해 사용. 이진 분류일 경우 시그모이드 함수의 출력이 0.5보다 크면 양성 클래스, 0.5보다 작으면 음성 클래스로 판단



□ 불리언 인덱싱 boolean indexing [04장 185쪽]

넘파이 배열은 True, False 값을 전달하여 행을 선택할 수 있으며 이를 불리언 인덱싱이라고 함

□ 소프트맥스 함수

softmax function

[04장 190쪽]

여러 개의 선형 방정식의 출력값을 0~1 사이로 압축하고 전체 합이 1이 되도록 만들며 이를 위해 지수 함수를 사용하기 때문에 정규화된 지수 함수라고도 함

$$s_1 = \frac{e^{z_1}}{e_sum}, s_2 = \frac{e^{z_2}}{e_sum}, \dots, s_7 = \frac{e^{z_7}}{e_sum}$$

□ 확률적 경사

Stochastic Gradient Descent

[04장 200쪽]

하강법

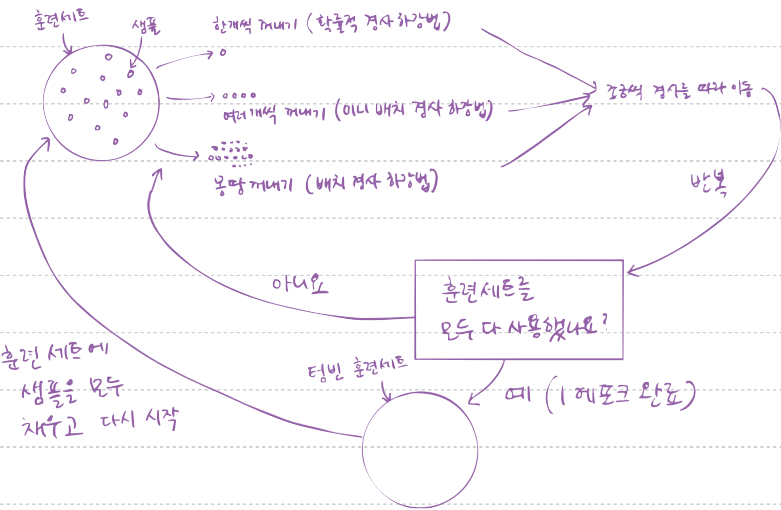
훈련 세트에서 랜덤하게 하나의 샘플을 선택하여 손실 함수의 경사를 따라 최적의 모델을 찾는 알고리즘

□ 에포크

epoch

[04장 202쪽]

확률적 경사 하강법에서 훈련 세트를 한 번 모두 사용하는 과정



□ 미니배치 경사

minibatch gradient descent

[04장 202쪽]

하강법

1개가 아닌 여러 개의 샘플을 사용해 경사 하강법을 수행하는 방법으로 실전에서 많이 사용

□ 배치 경사 하강법

batch gradient descent

[04장 202쪽]

한 번에 전체 샘플을 사용하는 방법으로 전체 데이터를 사용하므로 가장 안정적인 방법이지만 그만큼 컴퓨터 자원을 많이 사용함. 또한 어떤 경우는 데이터가 너무 많아 한 번에 전체 데이터를 모두 처리할 수 없을지도 모름

□ 손실 함수

loss function

[04장 203쪽]

어떤 문제에서 머신러닝 알고리즘이 얼마나 엉터리인지를 측정하는 기준.

□ 로지스틱 손실

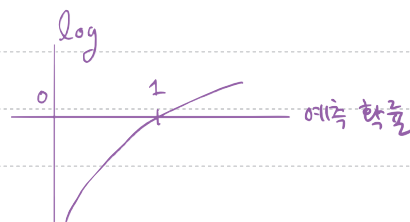
logistic loss function

이진 크로스엔트로피 손실 함수라고도 함

[04장 206쪽]

함수

양성 클래스(타겟 = 1)일 때 손실은 $-\log(\text{예측 확률})$ 로 계산하며, 1 확률이 1에서 멀어질수록 손실은 아주 큰 양수가 됨. 음성 클래스(타겟 = 0)일 때 손실은 $-\log(1 - \text{예측 확률})$ 로 계산함. 이 예측 확률이 0에서 멀어질수록 손실은 아주 큰 양수가 됨



타겟 = 1 일때

→ $-\log(\text{예측 확률})$

타겟 = 0 일때

→ $-\log(1 - \text{예측 확률})$

□ 크로스엔트로피

cross-entropy loss function

[04장 206쪽]

손실 함수

다중 분류에서 사용하는 손실 함수

□ 힙지 손실

hinge loss

[04장 212쪽]

서포트 벡터 머신(support vector machine)이라 불리는 또 다른 머신러닝 알고리즘을 위한 손실 함수로 널리 사용하는 머신러닝 알고리즘 중 하나. SGDClassifier가 여러 종류의 손실 함수를 loss 매개변수에 지정하여 다양한 머신러닝 알고리즘을 지원함



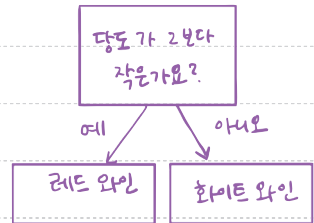
05장

□ 결정 트리

Decision Tree

[05장 226쪽]

스무고개와 같이 질문을 하나씩 던져 정답을 맞춰가며 학습하는 알고리즘으로 비교적 예측 과정을 이해하기 쉬움

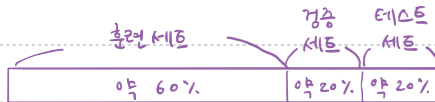


□ 검증 세트

validation set 하이퍼파라미터

[05장 243쪽]

하이퍼파라미터 튜닝을 위해 모델을 평가할 때, 테스트 세트를 사용하지 않기 위해 훈련 세트에서 다시 떼어 낸 데이터 세트



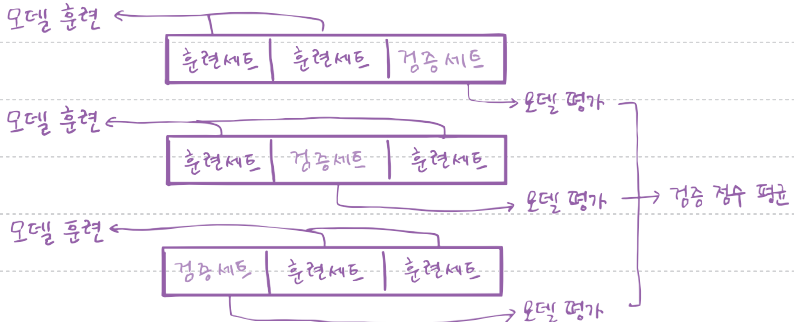
□ 교차 검증

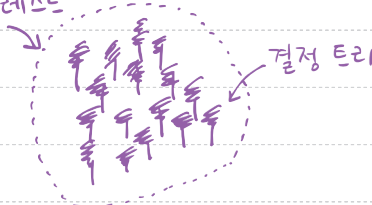
cross validation

[05장 245쪽]

훈련 세트를 여러 폴드로 나눈 다음 한 폴드가 검증 세트의 역할을 하고 나머지 폴드에서는 모델을 훈련함.

이렇게 모든 폴드에 대해 검증 점수를 얻어 평균하는 방법으로 교차 검증을 이용하면 검증 점수가 안정적이며, 훈련에 더 많은 데이터를 사용할 수 있음



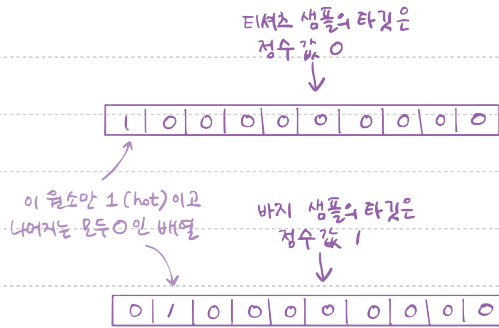
□ 그리드 서치	Grid Search	[05장 248쪽]
	하이퍼파라미터 탐색을 자동화해 주는 도구	
□ 랜덤 서치	Random Search	[05장 252쪽]
	랜덤 서치는 연속적인 매개변수 값을 탐색할 때 유용	
□ 정형 데이터 vs 비정형 데이터	structured data vs unstructured data	[05장 264쪽]
	특정 구조로 이루어진 데이터를 정형 데이터라 하고, 반면 정형화되기 어려운 사진이나 음악 등을 비정형 데이터라 함 CSV나 데이터베이스 등	
□ 앙상블 학습	ensemble learning	[05장 264쪽]
	여러 알고리즘(예, 결정 트리)을 합쳐서 성능을 높이는 머신러닝 기법	
□ 랜덤 포레스트	Random Forest	[05장 265쪽]
	대표적인 결정 트리 기반의 앙상블 학습 방법. 안정적인 성능 덕분에 널리 사용됨. 부트스트랩 샘플을 사용하고 랜덤하게 일부 특성을 선택하 여 트리를 만드는 것이 특징	
□ 부트스트랩 샘플	bootstrap sample	[05장 265쪽]
	데이터 세트에서 중복을 허용하여 데이터를 샘플링하는 방식	

06장

□ 히스토그램	histogram	[06장 294쪽]
	값이 발생한 빈도를 그래프로 표시한 것으로 보통 x축이 값의 구간(계급)이고, y축은 발생 빈도(도수)임	
		
□ 군집	clustering	[06장 298쪽]
	비슷한 샘플끼리 그룹으로 모으는 작업으로 대표적인 비지도 학습 작업 중 하나	
□ k-평균 알고리즘	k-means algorithm	[06장 304쪽]
	처음에 랜덤하게 클러스터 중심을 정하여 클러스터를 만들고 그다음 클러스터의 중심을 이동하여 다시 클러스터를 결정하는 식으로 반복해서 최적의 클러스터를 구성하는 알고리즘	
□ 이너셔	inertia	[06장 311쪽]
	k-평균 알고리즘은 클러스터 중심과 클러스터에 속한 샘플 사이의 거리를 쥌 수 있는데 이 거리의 제곱 합을 이너셔라고 함. 즉 클러스터의 샘플이 얼마나 가깝게 있는지를 나타내는 값임	
□ 차원 축소	dimensionality reduction	[06장 319쪽]
	데이터를 가장 잘 나타내는 일부 특성을 선택하여 데이터 크기를 줄이고 지도 학습 모델의 성능을 향상시킬 수 있는 방법	
□ 주성분 분석	principal component analysis, PCA	[06장 319쪽]
	차원 축소 알고리즘의 하나로 데이터에서 가장 분산이 큰 방향을 찾는 방법이며 이런 방향을 주성분이라 함. 원본 데이터를 주성분에 투영하여 새로운 특성을 만들 수 있음	

07장

□ 인공신경망	artificial neural network, ANN [07장 347쪽]
	생물학적 뉴런에서 영감을 받아 만든 머신러닝 알고리즘. 신경망은 기존의 머신러닝 알고리즘으로 다루기 어려웠던 이미지, 음성, 텍스트 분야에서 뛰어난 성능을 발휘하면서 크게 주목을 받고 있으며 종종 딥러닝이라고도 부름
□ 딥러닝	deep learning [07장 350쪽]
	딥러닝은 인공신경망과 거의 동의어로 사용되는 경우가 많으며 혹은 심층 신경망-deep neural network, DNN을 딥러닝이라고 부름. 심층 신경망은 여러 개의 층을 가진 인공신경망임
□ 텐서플로	TensorFlow [07장 350쪽]
	구글이 만든 딥러닝 라이브러리로 CPU와 GPU를 사용해 인공신경망 모델을 효율적으로 훈련하며 모델 구축과 서비스에 필요한 다양한 도구를 제공함. 텐서플로 2.0부터는 신경망 모델을 빠르게 구성할 수 있는 케라스를 핵심 API로 채택. 케라스를 사용하면 간단한 모델에서 아주 복잡한 모델까지 손쉽게 만들 수 있음
□ 활성화 함수	activation function 참고 용어 소프트맥스 함수 [07장 355쪽]
	소프트맥스와 같이 뉴런의 선형 방정식 계산 결과에 적용되는 함수
□ 원-핫 인코딩	one-hot encoding [07장 357쪽]
	타깃값을 해당 클래스만 1이고 나머지는 모두 0인 배열로 만드는 것. 다중 분류에서 크로스 엔트로피 손실 함수를 사용하려면 0, 1, 2와 같이 정수로 된 타깃값을 원-핫 인코딩으로 변환해야 함

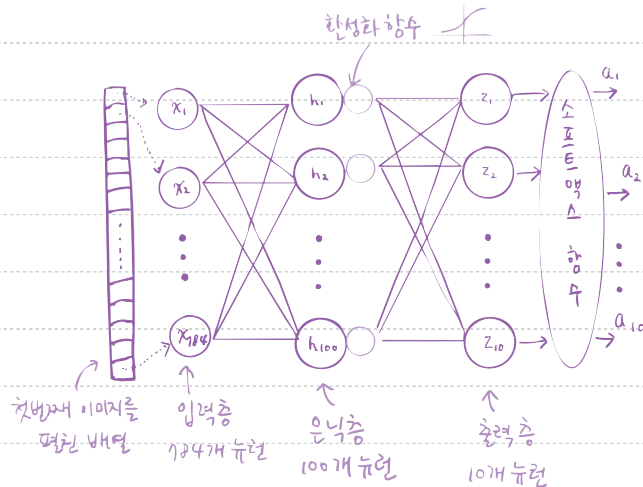


□ 은닉층

hidden layer

[07장 369쪽]

입력층과 출력층 사이에 있는 모든 층을 은닉층이라고 부름



□ 심층 신경망

deep neural network, DNN

[07장 371쪽]

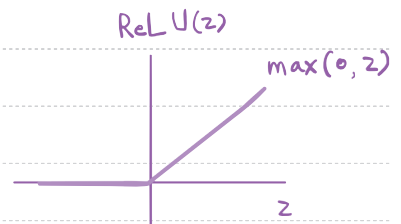
2개 이상의 층을 포함한 신경망으로 종종 다층 인공신경망, 심층 신경망, 딥러닝을 같은 의미로 사용함

□ 렐루 함수

ReLU Function

[07장 377쪽]

입력이 양수일 경우 마치 활성화 함수
가 없는 것처럼 그냥 입력을 통과시키
고 음수일 경우에는 0으로 만드는 함수



□ 옵티마이저

optimizer

[07장 381쪽]

신경망의 가중치와 절편을 학습하기 위한 알고리즘 또는 방법. 케라스에는 다양한 경사 하강법 알고리즘이 구현되어 있으며 대표적으로 SGD, 네스테로프 모멘텀, RMSprop, Adam 등이 있음

□ 적응적 학습률

adaptive learning rate

[07장 383쪽]

모델이 최적점에 가까이 갈수록 안정적으로 수렴하도록 학습률을 낮추도록 조정하는 방법. 이런 방식들은 학습률 매개변수를 튜닝하는 수고를 덜 수 있는 것이 장점

□ 드롭아웃

dropout

[07장 401쪽]

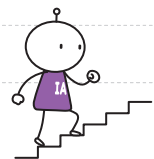
훈련 과정에서 층에 있는 일부 뉴런을 랜덤하게 꺼서(즉 뉴런의 출력을 0으로 만들어) 과대적합을 막음

□ 콜백

callback

[07장 408쪽]

케라스에서 훈련 과정 중간에 어떤 작업을 수행할 수 있게 하는 객체로 keras.callbacks 패키지 아래에 있는 클래스로 fit() 메서드의 callbacks 매개변수에 리스트로 전달하여 사용



08장

□ 합성곱

convolution

[08장 423쪽]

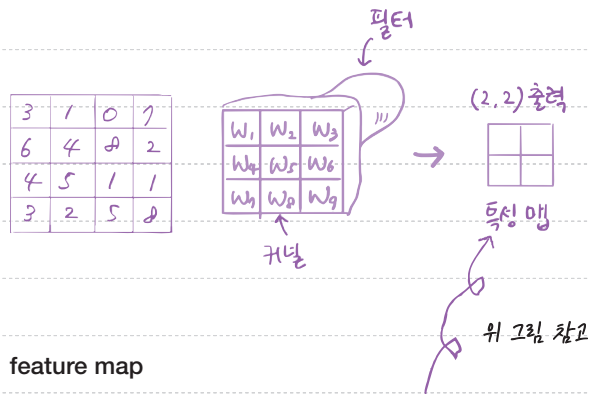
합성곱은 밀집층과 비슷하게 입력과 가중치를 곱하고 절편을 더하는 선형 계산이지만 밀집층과 달리 합성곱은 입력 데이터 전체에 가중치를 적용하는 것이 아니라 일부에 가중치를 곱함

□ 필터

filter

[08장 425쪽]

밀집층의 뉴런에 해당. 뉴런 = 필터 = 커널 모두 같은 말이라 생각해도 좋음



□ 특성 맵

feature map

[08장 427쪽]

합성곱 계산을 통해 얻은 출력을 특별히 특성 맵이라 부름

□ 패딩과

padding과 same padding

[08장 431쪽]

세임 패딩

입력 배열의 주위를 가상의 원소(보통 0)로 채우는 것을 패딩이라고 하고 합성곱 신경망에서는 세임 패딩을 많이 사용함

□ 밸리드 패딩

valid padding

[08장 431쪽]

패딩 없이 순수한 입력 배열에서만 합성곱을 하여 특성 맵을 만드는 경우이며 특성 맵의 크기가 줄어들 수밖에 없음

□ 스트라이드

stride

[08장 434쪽]

합성곱 층에서 필터가 입력 위를 이동하는 크기로 기본으로 스트라이드는 1픽셀, 즉 한 칸씩 이동함.

□ 풀링

pooling

[08장 434쪽]

합성곱 층에서 만든 특성 맵의 가로세로 크기를 줄이는 역할을 수행하지만 특성 맵의 개수는 줄이지 않음. 또한 가중치가 없는 대신 특성 맵에서 최댓값이나 평균값을 선택함

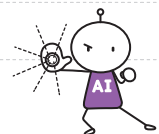
□ 최대 풀링과

max pooling과 average pooling

[08장 435쪽]

평균 풀링

풀링을 수행할 때 가장 큰 값을 고르거나 평균값을 계산하는데 이를 각각 최대 풀링과 평균 풀링이라고 부름



09장

□ 순차 데이터

sequential data

[09장 487쪽]

텍스트나 시계열 데이터와 같이 순서에 의미가 있는 데이터를 말함. 예를 들어 “I am a boy”는 쉽게 이해할 수 있지만 “boy am a I”는 말이 되지 않음

텍스트. 예) “I am a boy”
 순차 데이터 ↗
 시계열. 예) 1월 15°C, 2월 17°C, 3월 16°C, ...

□ 시계열 데이터

time series data

[09장 487쪽]

일정한 시간 간격으로 기록된 데이터

→ 주식, 일자별 날씨 등등

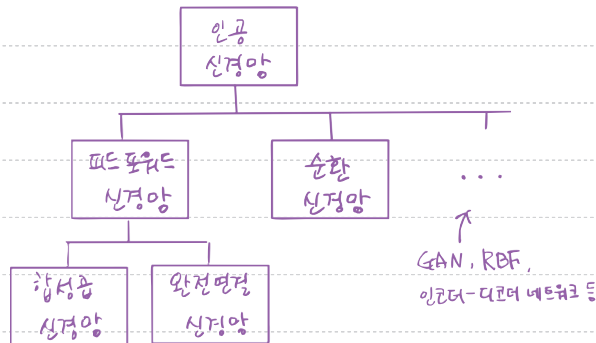
□ 피드포워드

feedforward neural network, FFNN

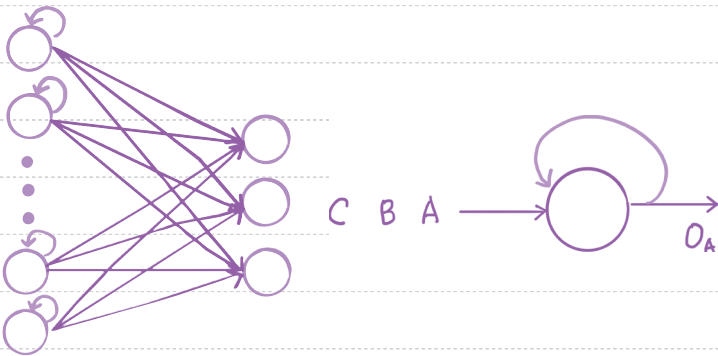
[09장 488쪽]

신경망

입력 데이터의 흐름이 앞으로만 전달되는 신경망. 완전 연결 신경망과 합성곱 신경망이 모두 피드포워드 신경망에 속함

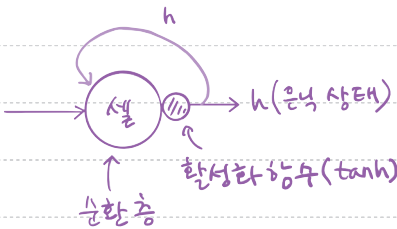


완전 연결 신경망과 거의 비슷함. 순차 데이터에 잘 맞는 인공신경망의 한 종류로
순차 데이터를 처리하기 위해 고안된 순환 층을 1개 이상 사용한 신경망



순환 신경망에서는 특별히 층을 셀이라 부르며 한 셀에는 여러 개의 뉴런이 있지만
완전 연결 신경망과 달리 뉴런을 모두 표시하지 않고 하나의 셀로 층을 표현함

순환 신경망에서는 셀의 출력을 은닉 상태라 부름. 은닉 상태는 다음 층으로 전달
될 뿐만 아니라 셀이 다음 타임스텝의 데이터를 처리할 때 재사용됨



자연어 처리 분야에서는 훈련 데이터를 종종 말뭉치라고 부름. 예를 들어 IMDB
리뷰 데이터셋이 하나의 말뭉치임

□ 토큰

token

[09장 501쪽]

일반적으로 영어 문장은 모두 소문자로 바꾸고 구두점을 삭제한 다음 공백을 기준으로 분리하는데 이렇게 텍스트에서 공백으로 구분되는 문자열 또는 단어를 토큰이라고 부름

I am a boy는 4개의 토큰. 한글은 조사가 발달해 형태소 분석을 해야 함

□ 단어 임베딩

word embedding

[09장 514쪽]

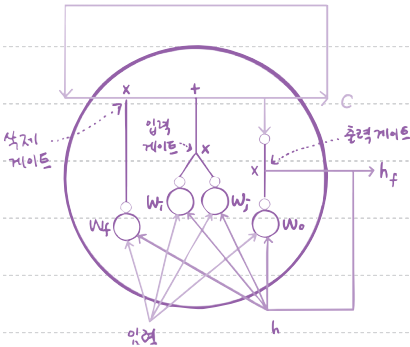
순환 신경망에서 텍스트를 처리할 때 즐겨 사용하는 방법으로 입력으로 정수 데이터를 받아 메모리를 훨씬 효율적으로 사용할 수 있음

□ LSTM

Long Short-Term Memory

[09장 527쪽]

단기 기억을 오래 기억하기 위해 고안된 순환층. 입력 게이트, 삭제 게이트, 출력 게이트 역할을 하는 작은 셀이 포함



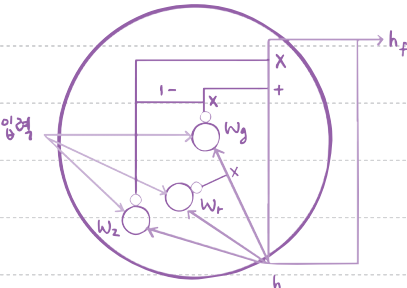
□ 셀 상태

cell state

[09장 527쪽]

LSTM 셀은 은닉 상태 외에 셀 상태를 출력. 셀 상태는 다음 층으로 전달되지 않으며 현재 셀에만 순환됨

LSTM 셀의 간소화 버전으로 생각할 수 있지만 LSTM처럼 셀 상태를 계산하지 않고 은닉 상태 하나만 포함. LSTM보다 가중치가 적기 때문에 계산량이 적지만 LSTM 못지않은 좋은 성능을 내는 것으로 알려져 있음



MEMO

Handwriting practice lines consisting of 20 sets of horizontal dashed lines.

MEMO

Handwriting practice area with horizontal dashed lines.

MEMO

Handwriting practice lines consisting of 20 sets of horizontal dashed lines.

MEMO

Handwriting practice area with horizontal dashed lines.

혼자 공부하는 사람들을 위한 용어 노트

[illegible]