

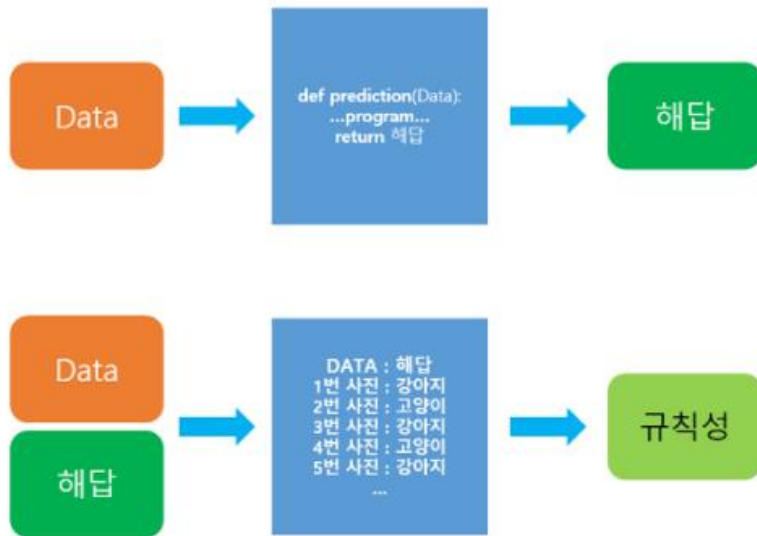


Machine Learning & Deep Learning



CUAI NLP Study
기계공학부 서혜련

Machine Learning Overview



결과를 찾는 것에 초점이 X

Data로부터 규칙성을 찾는 것에 초점!--→ 학습(Training)

Machine Learning Overview

분류(classification) or 회귀(regression)
Logistic Regression or Linear Regression

Logistic Regression

시그모이드 함수(Sigmoid function)

$$H(X) = \frac{1}{1 + e^{-(Wx+b)}} = \text{sigmoid}(Wx + b) = \sigma(Wx + b)$$

0.5 UP? 1(True) // DOWN? 0(False)로
→이진 분류 문제 사용

Cost Function

크로스 엔트로피(Cross Entropy)

$$\text{cost}(W) = -(y \log H(X) + (1 - y) \log(1 - H(X)))$$

Linear Regression

$$y = Wx + b$$

x와 y의 관계를 W와 b를 이용하여 식을 세우자! (가설)
 $H(x) = Wx + b$

Cost Function

평균 제곱 오차(Mean Squared Error, MSE)

$W, b \rightarrow \text{minimize cost}(W, b)$

→Optimizer(Gradient Descent)



Machine Learning Overview

Tensor

ndim= 차원수

shape=텐서의 크기

0차원 하나의 실수값

1차원 벡터

2차원 행렬

3차원 행렬단위로 한 번 더 배열

Machine Learning Overview

Tensor

```
d=np.array([ [1, 2, 3, 4, 5], [6, 7, 8, 9, 10], [10, 11, 12, 13, 14]],  
            [[15, 16, 17, 18, 19], [19, 20, 21, 22, 23], [23, 24, 25, 26, 27]] )  
print(d.ndim)  
print(d.shape)
```

Machine Learning Overview

Tensor

```
d=np.array([ [1, 2, 3, 4, 5], [6, 7, 8, 9, 10], [10, 11, 12, 13, 14]],  
           [[15, 16, 17, 18, 19], [19, 20, 21, 22, 23], [23, 24, 25, 26, 27]] )  
print(d.ndim)  
print(d.shape)
```

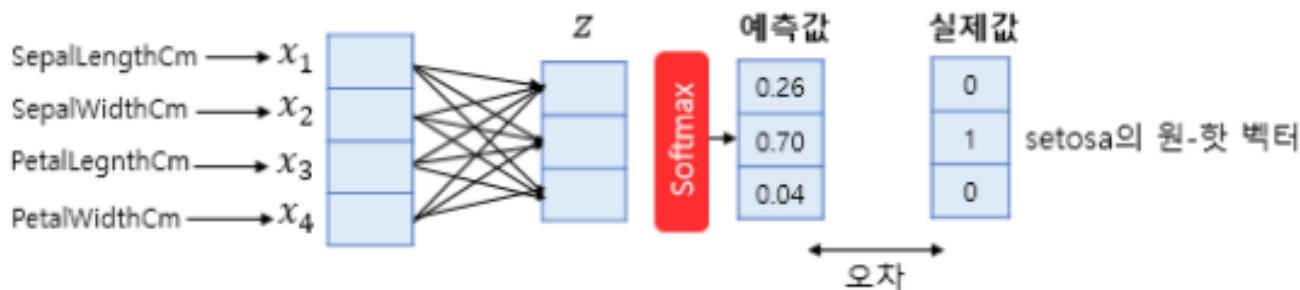
```
>>3
```

```
>>(2, 3, 5)
```

Machine Learning Overview q

Softmax Regression - 다중 클래스 분류

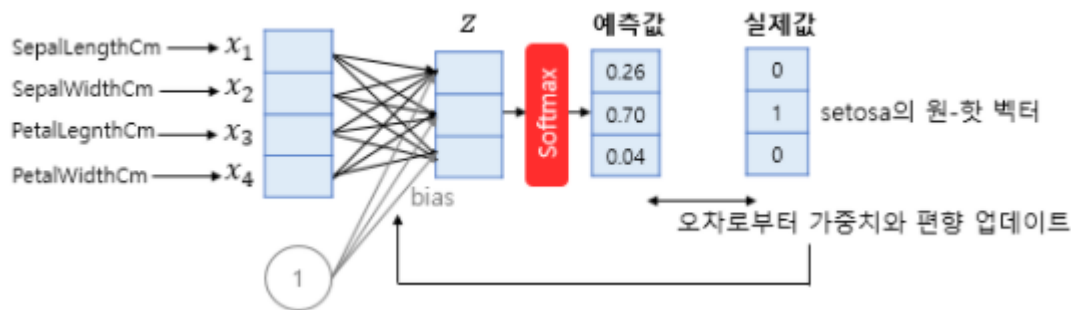
하나의 샘플 데이터에 대한 예측값으로
모든 가능한 정답지에 대해서 정답일 확률의 합이 1이 되도록 한다.



1. Vector \rightarrow Softmax input
2. Error?

Machine Learning Overview

Softmax Regression - 다중 클래스 분류



1. Vector \rightarrow Softmax input?

소프트맥스 함수의 입력 벡터 z 의 차원수만큼 결과값이 나오도록 가중치 곱을 진행
위의 그림에서 화살표는 총 ($4 \times 3 = 12$) 12

2. Error?

실제값을 원-핫 벡터로 표현하고 예측값과의 오차를 통해 W, b 를 update

Machine Learning Overview

Why One-hot encoding?

→분류 문제 모든 클래스 간의 관계를 균등하게 분배
(무작위성은 때로는 단어의 유사성을 구할 수 없다는 단점으로 언급되기도,,)

EX>

정수 인코딩을 할 때에는 class간의 연관성이 생겨날 수 있음.(실제X)

Banana :1, Tomato :2, Apple :3, ... Watermelon :10

Real=Tomato, Pre=Banana

$$SE=(2-1)^2=1$$

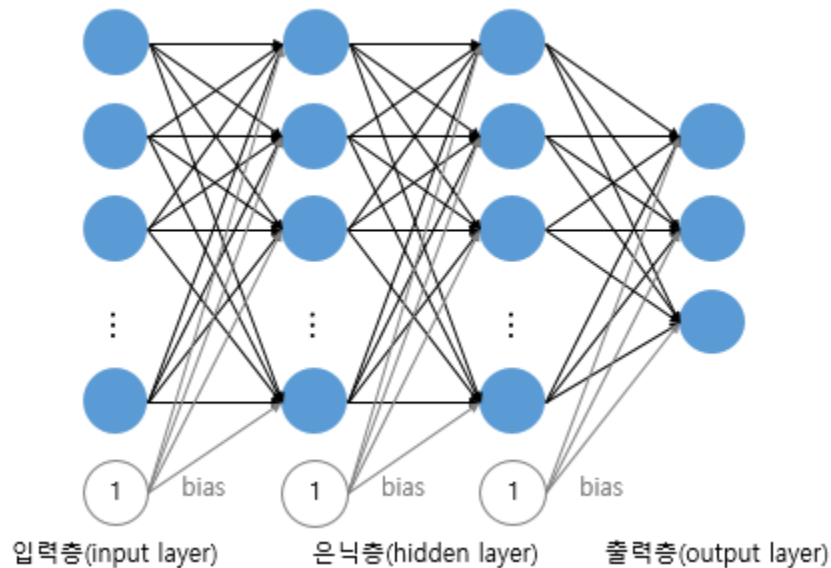
Real=Apple, Pre=Banana

$$SE=(3-1)^2=4$$

BANANA가 APPLE 보다 TOMATO에 가깝다..?

<상.식.파.괴>

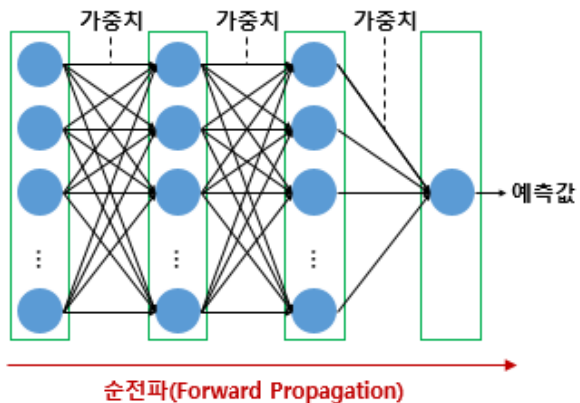
Deep Learning Overview



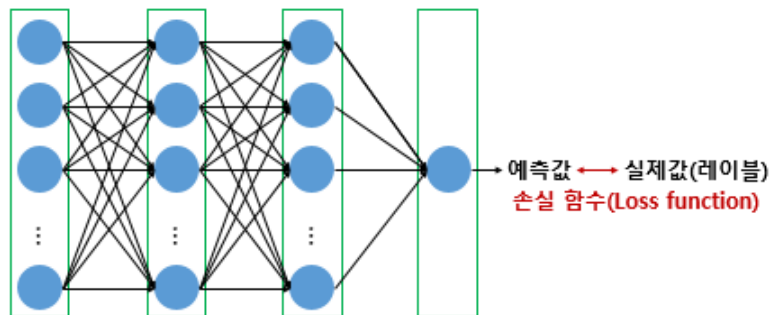
은닉층이 2개 이상인 신경망을 심층 신경망(Deep Neural Network, DNN)

Deep Learning Overview

1. 순전파(Forward Propagation)



2. 손실 함수(Loss function)



1) MSE(Mean Squared Error, MSE)

오차 제곱 평균을 의미합니다. 연속형 변수를 예측할 때 사용됩니다.

$$-\frac{1}{N} \sum (y - \hat{y})^2$$

2) 크로스 엔트로피(Cross-Entropy)

y : 실제값 (0 or 1) / \hat{y} : 예측값 (확률)

$$-\sum y \log \hat{y}$$

Deep Learning Overview _optimizer

배치 경사 하강법(Batch Gradient Descent)

한 번의 에포크에 모든 매개변수 업데이트를 단 한 번 수행

확률적 경사 하강법(Stochastic Gradient Descent, SGD)

매개변수 값을 조정 시 전체 데이터가 아니라 랜덤으로 선택한 하나의 데이터에 대해서만 계산

미니 배치 경사 하강법(Mini-Batch Gradient Descent)

정해진 양에 대해서만 계산하여 매개 변수의 값을 조정

모멘텀(Momentum)

경사 하강법에서 계산된 접선의 기울기에 한 시점(step) 전의 접선의 기울기값을 일정한 비율만큼 반영
->관성!

아다그라드(Adagrad)

각 매개변수에 서로 다른 학습률을 적용(변화 크면 학습률을 낮춤)

알엠에스프롭(RMSprop)

나중에 가서는 학습률이 지나치게 떨어진다는 단점이 있는데 이를 다른 수식으로 대체하여 이러한 단점을 개선

아담(Adam)

알엠에스프롭과 모멘텀 두 가지를 합침. 방향, 학습률 두가지 고려