@ 학습 목표
1. 컴퓨터와 뇌의 동작 방식을 비교할 수 있다.
1. 퍼셉트론의 구성요소를 이해할 수 있다.
1. 다층 신경망의 분류 원리를 이해할 수 있다.

@ 뇌와 컴퓨터의 비교
100억개의 뉴런(신경세포)
60조개의 시냅스(신경세포들 사이의 연결)
입력이 들어오면 이해할 때 까지 분산처리 방식으로 이해한다
비선형연산
분산처리를 하기 때문에 병렬처리를 한다고 볼 수 있다.

뉴런보다 빠른 연산(GHz기준으로 한 연산을 하는데 걸리는 시간은 10^{-9}초), 뉴런은 KHz단위로서 10^{-3}초
ALU를 사용하는 중앙처리방식
주로 +로 된 산술 연산(선형)를 많이 사용함
폰노이만 구조를 가지고 있기 때문에 순차 처리 방식을 사용한다고 볼 수 있다.

@ 신경회로망에 대한 연구

1940년대 - 60년대
뇌의 구조에 대한 모방이 활발했던 시기이다.
신경세포에서 공식 모형은 McCulloch and Pitts, 뉴런모델화(43년)에 발표됨. 컴퓨터를 만드는 논리 게이트 회로와 많이 닮았다. 논리합회로는 입력중에 하나라도 켜지면 출력이 켜지고, 논리곱회로는 모든 입력이 켜져야 출력이 켜짐. 맥쿨로치 신경세포는 활성화된 입력의 수가 일정 한계를 넘어서면 켜진다. 한계점에 이르면 신경세포는 논리회로처럼 작동한다.
Weinner, 사이버네틱스(48년)
Rosenblatt, 퍼셉트론(57년)

@ 생물학적 뉴런의 구조 vs 인공 뉴런
생물학적인 뉴런과의 유사성
병렬 계산(parallel computing)
분산 표현(distributed representation)

@ 다른뉴런들로부터 나오는 출력이 하나의 뉴런에 대한 입력으로 들어온다. X_1, X_2, X_3.
입력들은 가중치를 통해서 세포체(sum 함수)에 전달이 된다. 그러면 입력들이 가중치와 곱해지고 다 더해진다. 그래서 이 곱하고 더한값을 활성함수에 넣었을때 결과값이 threshold 값 이상이면 출력을 하고 이하이면 출력하지 않는다.

@ 신경망(인공뉴런) 학습에 적합한 문제
학습해야 하는 현상이 여러가지 속성($x_1, x_2, ..., x_n$)에 의해 표현 되는 경우.

$$\begin{bmatrix} feature_1 \\ feature_2 \\ ... \\ feature_n \end{bmatrix}$$

1. 학습예제에서 에러(noise)가 존재할 가능성이 경우도 신경망을 적용할 수 있다.
1. 긴 학습시간이 필요한 문제
1. 학습된 결과를 사람이 이해하는 것이 필요없는 경우

@ 퍼셉트론('57)
X입력이다. $X_1, X_2, ..., X_n$
입력은 실수 값을 갖는 벡터이다. 각각의 demension이 하나의 feature이다.
각각의 입력은 가중치 $W_1, W_1, ..., W_1$ 를 가지고 있다.
각각의 입력과 가중치가 곱해지고 다 더해진다.
이 값을 활성함수(퍼셉트론의 경우 계단함수)에 넣는다.
$o(x_1, x_2, ..., x_n) =$
1 을 출력함. if 가중치를 적용한 입력의 총합 $w_0 + w_1x_1 + ... + w_1x_1 > 0$ 일때.
-1 을 출력함. 위의 경우가 아닌 경우.

학습 : 적절한 출력이 나오도록 가중치를 변경하는 것이다.

퍼셉트론 수식으로 나태내기
$$y = f(\sum_{i=1}^{n} w_i x_i)$$
f : 활성함수.

활성함수로서 비선형함수의 종류
1. 계단함수
1. 임계논리 함수
1. 시그모이드 함수

@ 1. How do we represent the meaning of a word?
Definition of m̈eaning(from Webster dictionary)
1. the idea that is represented by a word, phrase, etc.
1. the idea that a person wants to express by using words, signs, etc.
1. the idea that is expressed in a work of writing, art, etc.

Commonest linguistic way of thinking of meaning :
1. signfier $\Leftrightarrow$ signified(idea or thing) = denotation

@ To train the model : You compute all vector gradients.
1. We often define the set of all parameters in a model in terms of one vertically long vector θ
1. In out case with d-dimensinal vector and V many words :

$$\theta = \begin{bmatrix} v_{aardvark} \\ v_a \\ ... \\ v_{zebra} \\ u_{aardvark} \\ u_a \\ ... \\ u_{zebra} \end{bmatrix} \in R^{2dV}$$

1. We then optimize these parameters.

Note : Every word has two vectors. Makes it simpler.

@Chapter 10 비모수 밀도 추정4

@ Parzen window 의 단점 :
불연속적인 모양의 밀도추정을 낳는다. 앞에서 샘플 4개를 가지고 작업을 했다. 주어진 x 값에 따라서 확률밀도가 이렇게 연속적이지 않고 히스토그램처럼 딱딱 끊어지는 확률밀도를 만들었다. 그런데 확률밀도함수는 스무드한 함수이지 딱딱 끊어지는 불연속적인 확률밀도라는건 없다. 따라서, 현실성이 떨어지는 확률밀도 함수를 만들어낸다는 것이다.

불연속성이 나타나는 이유는 Parzen window 의 한계 때문이다. 그 한계란 추정점으로부터의 거리에 관계없이 모든 점에 가중치를 같게 주는 특성을 말한다.
모든 샘플점마다 이것의 주변에 가로긴직사각형모양의 똑같은 크기의 함수들을 만들어서 더해줬었다. 샘플이 영향을 주는 영향권안에 모든 영역에다가 동일한 값을 더해준다는 의미이다.

이러한 특성으로 블록 커널 밀도 추정법 이라고 불리기도 한다.

@ 그러나 커널 밀도 추정법은 히스토그램과는 다르다.

@ 위의 불연속적인 블록 모양을 만들어내므로 커널을 바꿔보자라고 생각을 했다.
그렇게해서 나온게 스무스 커널을 이용한 커널 밀도 추정이다.
다음과 같은 조건을 만족하는 부드러운 커널로 Parzen window 가 만드는 불연속적인 모양을 극복한다.
조건: 커널함수 K(x) 를 주어진 영역 R^0 전체에 대해서 적분했을때 1이 나와야한다.
$\int_{R^0} K(x)dx = 1$

위의 조건을 만족하는 함수가 가우시안 함수이다. 부드러운 형태도 갖고있다.
부드러운 커널함수 K(x) 는 다음과 같이 단순화된 형태의 가우시안 밀도 함수로 사용된다.
단순화된 형태란 일반적인 가우시안 함수에서 분산, 공분산 행렬의 값을 단위값을 갖는 분산과 단위행렬로 가정했을때의 형태이다.
일반 가우시안 함수 : $\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}[(X-\mu)^T\Sigma^{-1}(X-\mu)]}$

$\Sigma = I$, 평균값 $\mu = 0$ 인 경우, $|\Sigma|^{\frac{1}{2}} = 1, \mu = 0, \Sigma^{-1} = 1$ 이므로 아래와 같이 된다.
$K(x) = \frac{1}{(2\pi)^{\frac{D}{2}}} e^{-\frac{1}{2}x^T x}$

@ 8개의 샘플로 확률밀도함수를 추정한게 Density estimation 이다.
그 과정은 각각의 샘플을 중심값으로 해서 가우시안 함수를 다 그려본다. 그리고 그 함수들을 다 더한다. 겹치는 곳은 피크가 된다. 그 주변은 스무드하다.

@ 밀도추정식을 수식으로 표현해보자.
커널밀도추정방법(KDE)는 스무드 커널을 사용하더라도 똑같이 $\frac{k}{NV}$ 를 사용한다.
볼륨 V는 h 의 D 제곱이다. $V = h^D$
들어오는 샘플의 갯수를 카운트 하는 k는 Parzen window 를 썼을 경우에 커널 함수가 $\sum_{i=1}^{N} K(\frac{x-x^{(i)}}{h})$ 와 같이 주어진다. 따라서, 이전에 사용했던 불연속적인 확률밀도함수를 만들어내는 커널밀도함수추정방법의 함수 $p_{KDE}(x)$ 는
$p_{KDE}(x) = \frac{1}{Nh^D} \sum_{i=1}^{N} K(\frac{x-x^{(i)}}{h})$ 이다.

간단한 형태의 가우시안 확률밀도 함수로 k 를 대체해본다.
$p_{KDEsoft} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(h\sqrt{2\pi})^D} e^{[-\frac{1}{2}(\frac{(x-x^i)}{h})^2]}$
모양을 보면 평균 μ 가 i 번째에 해당하는 샘플 $x^{(i)}$, 표준편차 σ 가 h 인 가우시안 함수로 볼 수 있다.

이 말은, 샘플하나가 있으면 샘플을 중심값, 즉, 평균으로 하고 폭이 h 인 가우시안 함수를 도입하는거다. 이렇게 샘플마다 만들어서 다 더한다. 그리고 $\frac{1}{N}$ 을 곱해서 스케일해준다. 이 결과가 커널 밀도 추정법을 사용하여 구한 확률 밀도 추정값이다.

13