



ADsP 교재 정오표



1-04. 데이터와 정보의 관계 - 문제 3



3. 아래 내용은 데이터, 정보, 지식의 차이점을 예시로 설명한 것이다.
각각의 용어와 예시가 적절하게 연결된 것은?

가. A 마트는 500원, B마트는 400원에 볼펜을 판매한다.
나. B 마트는 볼펜 가격이 싸다.
다. 상대적으로 저렴한 B 마트에서 볼펜을 사야 겠다.
라. B 마트의 다른 상품도 A 마트보다 저렴할 것이다.

- 1 데이터-(가), 정보-(라), 지식-(다)
- 2 데이터-(가), 정보-(다), 지식-(라)
- 3 데이터-(가), 정보-(나), 지식-(라)
- 4 데이터-(가), 정보-(나), 지식-(다)

B 마트의 다른 상품도 A 마트보다 저렴할 것이다 - 지혜(Wisdom)

다, 라 보기 수정
해설 수정

1-26. 의사 결정 오류 - 문제 1



1. 분석의 전형적인 의사결정 오류를 로직(논리) 오류와 프로세스 오류로 나눠 보았을 때 각 오류에 대한 설명으로 가장 부적절한 것은?

- ① 부정확한 가정을 하고 테스트를 하지 않는 것은 로직 오류이다.
- ② 결정에서 분석과 통찰력을 고려하지 않은 것은 프로세스 오류이다.
- ③ 데이터 수집이나 분석이 너무 늦어 사용할 수 없게 되는 것은 로직 오류이다.
- ④ 대안을 진지하게 고려하지 않은 것은 프로세스 오류이다.

데이터 수집이나 분석이 너무 늦어 사용할 수 없게 되는 것은 프로세스 오류이다.

4번 보기 옆 글씨 제거



File : test_011.R

```
18 r1 <- cast(a, day~month~variable)
19 r2 <- cast(a, month~variable, mean)
20 r3 <- cast(a, month~variable, mean, margins=c('grand_row', 'grand_col'))
21 r4 <- cast(a, day~month, mean, subset=variable=='ozone')
22 r5 <- cast(a, month~variable, range)
```

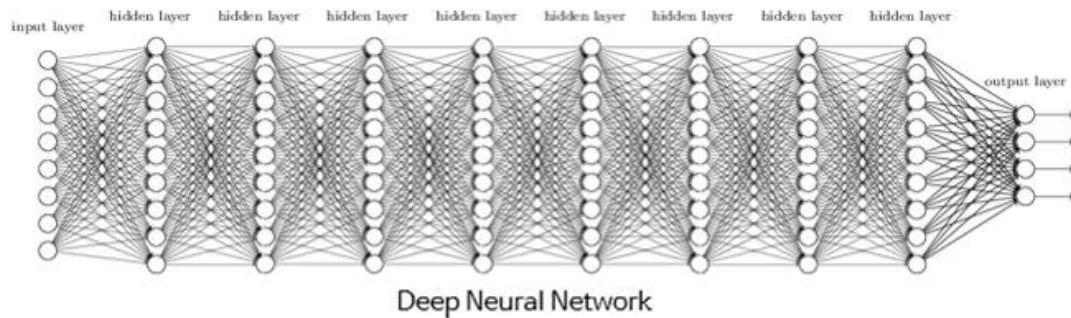
■ 결과를 출력해 구조를 확인해 볼 것!

- r1: 행을 day, 열을 month로 각 변수들을 새롭게 배치 (3차원 구조, variable은 면의 역할)
- r2: (행-month, 열-variable) 각 변수들의 month 평균
- r3: (행-month, 열-variable) 각 변수들의 month 평균을 구하고, 행과 열에 대한 소계 산출
- r4: 행을 day, 열을 month로 평균을 구함, subset기능을 사용해 ozone 변수만 처리하도록 함
- r5: (행-month, 열-variable) 각 변수들의 month range
range는 min은 '_X1', max는 '_X2' 라는 변수명을 끝에 붙여 줌

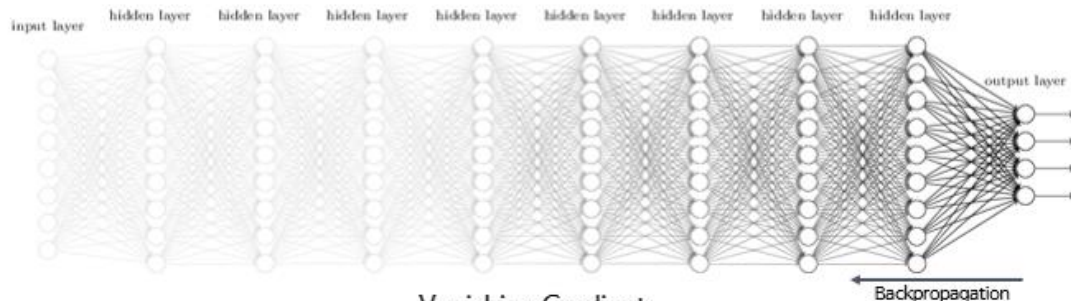
r2 <- acast 를 r2 <- cast 로 수정

3-89. 기울기 소실

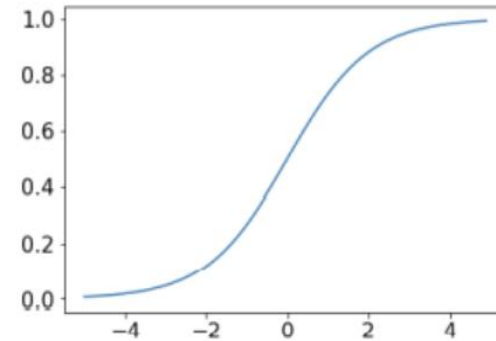
- 역전파(backpropagation) 알고리즘은 출력층(output layer)에서 입력 층(input layer)로 오차 Gradient 를 흘러 보내면서, 각 뉴런의 입력 값에 대한 손실함수의 Gradient를 계산함
- 이렇게 계산된 Gradient를 사용하여 각 가중치 매개변수를 업데이트 해 줌
- 다층신경망에서는 역전파 알고리즘이 입력층으로 갈 수록 Gradient가 점차적으로 작아져 0에 수렴하여, weight가 업데이트 되지 않는 현상
- activation function으로 sigmoid 함수를 사용할 때 발생 → 해결을 위해 ReLU 등 다른 함수 사용



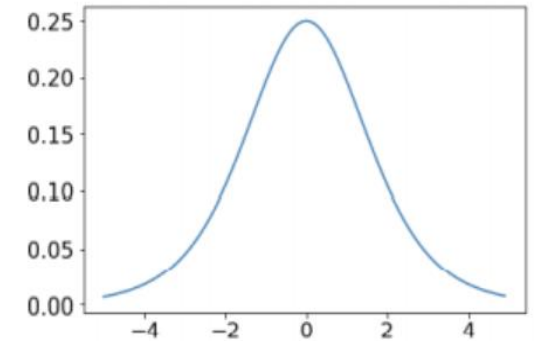
Deep Neural Network



Vanishing Gradient



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



$$S'(x) = \frac{1}{1+e^{-x}} \left(1 - \frac{1}{1+e^{-x}} \right)$$

x=0에서 기울기 최대
x가 크거나 작을 때 기울기가 0에 가까워 짐

$$W = W - \alpha \frac{d}{dw} \text{cost}(W, b)$$

3-95. 비계층적 군집(Non-Hierarchical Clustering)

제목 수정
3-90.

비계층적 군집 - 분할적 군집 방법

k-중심 군집

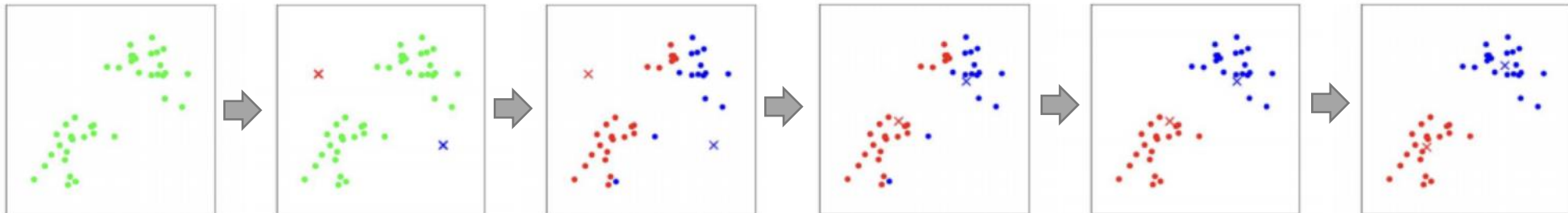
k-means

Nbclust 패키지를
통해 군집 수에 대
한 정보 참고

- **k-mean 방법은 사전에 군집의 수 k를 정해 주어야 함 (k : hyperparameter)**
- 군집수 k가 원데이터 구조에 적합하지 않으면 좋은 결과를 얻을 수 없음
- 알고리즘이 단순하며 빠르게 수행되어 계층적 군집보다 많은 양의 자료를 처리
- k-means 군집은 잡음이나 이상값에 영향을 받기 쉬움
- k-means 분석 전에 이상값을 제거하는 것도 좋은 방법
- 평균 대신 중앙값을 사용하는 k-medoids 군집을 사용할 수 있음

k-means 절차

1. 초기 군집의 중심으로 k개의 객체를 임의로 선택한다
2. 각 자료를 가장 가까운 군집의 중심에 할당한다
3. 각 군집 내의 자료들의 평균을 계산하여 군집의 중심을 갱신한다
4. 군집 중심의 변화가 거의 없을 때까지 2, 3을 반복한다



출처 : <http://stanford.edu/~cpiech/cs221/img/kmeansViz.png>