

- 주성분 분석 (Principal Component Analysis)은 다차원의 데이터가 주어졌을 때, 원래의 데이터가 가장 이상적인 원래의 차원보다 낮은 차원의 데이터를 찾아내는 방법이다. 차원 축소라고도 한다.
- 원리 : N개의 M차원 데이터가 있을 때 각각의 데이터는 규칙을 따르고 있다. 이런 규칙은 "물리적 / 사회적" 원리에 의해 형성될 수 있다. 이런 원인과 규칙을 찾아내는 것이 주성분 분석이다.
- 바로 축소되는 양치만 추출된 데이터의 차이에 따라 데이터 분포는 데이터는
- "잠재 변수 (Latent factor)"라 한다.

- 잠재변수와 추출데이터가 연속관계로 되어있다고 가정한다.
- 이렇게 표현된 추출데이터 x_i 의 각 원소를 선형조합 하면 그 뒤에 숨은 잠재변수 u_i 의 값을 계산할 수 있다고 가정할 수 있다.

$$u_i = W^T x_i$$

↪ 가중치 벡터

- 예를 들어 꽃잎 길이와 꽃잎 폭을 선형 조합 하여 꽃의 크기를 나타내는 어떤 잠재변수는 찾아내는 것을 할 수 있다.

$$u_i = W_1 x_{i,1} + W_2 x_{i,2}$$

↪ 2차원은 1차원으로 축소

• 차원 축소 방법

- 차원 축소 방법은 다차원 벡터는 G 낮은 차원의 벡터공간에 투영하는 방법으로 생각할 수 있다. 이는 "low-rank 근사"라고도 할 수 있다.