[SNU DS^2] Data Mining - Linear Algebra

March 22, 2018

1. 선형 변환과 행렬

   - 함수 $f : \mathbb{R}^m \to \mathbb{R}^n$ 이 모든 $a, b \in \mathbb{R}, x, y \in \mathbb{R}^m$ 에 대해 다음 성질을 만족하면 f를 선형 변환이라고 한다.

   $$f(a \cdot x + b \cdot y) = a \cdot f(x) + b \cdot f(y)$$

   - 임의의 $n \times m$ 행렬 A에 대해, 함수 $f_A : \mathbb{R}^m \to \mathbb{R}^n$ 를 $f_A(x) = A \cdot x$ 로 정의 하면 f_A 는 선형 변환이 된다.
   - **임의의 선형 변환 $f : \mathbb{R}^m \to \mathbb{R}^n$ 는 $f = f_A$ 를 만족하는 $n \times m$ 행렬 A가 존재한다.**
   - 따라서, 행렬은 선형 변환 함수와 동일하게 생각할 수 있다.
   - ex) 행렬의 곱셈은 교환이 안된다 = 함수의 합성은 교환이 안된다.

2. 행렬식 (determinant) 숫자의 절대값과 비슷하다. 그리고 절대값은 '크기'를 의미한다.

   - 정사각 행렬 : 행의 개수와 열의 개수가 일치하는 행렬.
   - 정사각 행렬 A의 행렬식은 $det(A)$ 혹은 $|A|$ 로 표기한다.
   - 2×2 행렬 A $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 의 행렬식

   $$det(A) = ad - bc$$

   - $m \times m$ 행렬 A의 행렬식 $det(A)$의 대의적 정의 :

     $\mathbb{R}^m$ 상의 집합 V에 대해

   $$\frac{\{y : y = Ax \text{ for } x \in V\}\text{의 부피}}{V\text{의 부피}}$$

3. 선형 독립, 선형 종속, 기저

- 벡터 $v_1, \cdots, v_m \in \mathbb{R}^n$ 이 다음 식을 만족하면 선형 종속 (linearly dependent) 이라 한다. 임의의 실수 $a_1, \cdots, a_m \in \mathbb{R}$ 이 존재해서 어떤 $1 \le j \le m$ 에 대해 $a_j \neq 0$ 이 성립하며

$$\sum_{j=1}^{m} a_j \cdot v_j = 0$$

  이 성립한다.

- 다시 말해, 어떤 벡터 $v_j, 1 \le j \le m$ 이 다른 벡터의 선형 결합으로 표현되면 선형 종속이라고 한다. 선형종속인 경우 하나의 피쳐를 다른 피쳐의 선형결합으로 만들 수 있다는 뜻이며 데이터가 겹친다는 의미다. 즉, 어떤 데이터는 필요 없다.

- 벡터 $v_1, \cdots, v_m \in \mathbb{R}^n$ 이 선형 종속이 아닌 경우 선형 독립 (linearly independent) 이라고 한다.

- $m \times m$ 정사각 행렬 A 에 대해 다음이 성립한다.
  - 모든 $x \in \mathbb{R}^m \neq 0$ 에 대해 $A \cdot x \neq 0$ 이 성립한다
    $\leftrightarrow$ A 의 모든 열벡터는 선형 독립이다 $\leftrightarrow$ $det(A) \neq 0$ 이 경우 원을 선으로 표현하는 것은 불가능하다?
  - 어떤 $x \in \mathbb{R}^m \neq 0$ 가 존재해서 $A \cdot x = 0$ 이 성립한
    $\leftrightarrow$ A 의 모든 열벡터는 선형 종속이다 $\leftrightarrow$ $det(A) = 0$ 이 경우, 원을 점으로도 표현할 수 있다?

- m차원 벡터 $x_1, \cdots, x_n$ 이 모든 벡터 $y \in \mathbb{R}^m$ 을 선형으로 표현 가능한 경우, 즉 어떤 상수 $a_1, \cdots, a_m$ 이 존재해서 $y = \sum_{j=1}^{m} a_j x_j$ 가 성립하는 경우 m차원 벡터 $x_1, \cdots, x_n$ 이 $\mathbb{R}^m$ 을 span 한다고 한다.

- $\mathbb{R}^m$ 을 span하는 m차원 벡터 $x_1, \cdots, x_n$ 이 선형 독립인 경우, $x_1, \cdots, x_n$ 을 기저 (basis) 라고 하며 $m = n$ 이 성립한다.

4. 벡터의 내적 (dot product) 및 수직관계

- $x = (x_1, \cdots, x_m)'$, $y = (y_1, \cdots, y_m)'$ 를 m 차원 벡터라 하자.

- 차원이 같은 두 벡터 간의 내적은 다음과 같이 정의된다.

$$x \cdot y = x'y = \sum_{j=1}^{m} x_j \cdot y_j$$

- 벡터의 크기 (norm) 은 다음과 같이 정의한다.

$$|x| = \sqrt{\sum_{j=1}^{m} x_j^2} = \sqrt{x \cdot x}$$

- 두 벡터 x, y간의 사잇각을 θ라 하면 다음 식이 성립한다.

$$x \cdot y = |x| \cdot |y| \cdot cos(\theta)$$

- 두 벡터 x, y가 수직관계에 있는 경우, 즉 사잇각이 90도인 경우, $\cos 90 = 0$이므로 $x \cdot y = 0$이 성립한다.
- 벡터 $x_1, \cdots, x_n$이 직교한다 (orthogonal)는 것은 이들 중 아무렇게나 두 순서쌍을 뽑아도 두 벡터가 수직관계에 있다는 것을 의미한다.
- 직교하는 벡터 $x_1, \cdots, x_n$은 선형 독립이다.
- m차원 벡터 $x_1, \cdots, x_m$이 직교하는 경우 이들은 m차원 기저 (basis)를 이룬다.
- 벡터 $x_1, \cdots, x_n$이 orthonormal 한다는 것은 이들이 직교함 (orthogonal)과 동시에 이들의 크기가 모두 1임을 의미한다. 수직이면서도 거리가 1인 것: orthonormal

5. 고유값 (eigenvalue)와 고유벡터 (eigenvector)

- 임의의 $m \times m$ 정사각 행렬 A에 대해 벡터 $x \in \mathbb{R}^m$과 실수 λ가 존재해서 $Ax = \lambda x$를 만족하면 λ를 (x에 대응되는) A의 고유값, x를 (λ에 대응되는) A의 고유벡터라고 한다.
- A가 $m \times m$ 대칭행렬인 경우, m개의 고유값과 이에 대응되는 고유벡터쌍 $(\lambda_1, e_1), \cdots, (\lambda_m, e_m)$이 존재하며, 이 때 고유벡터들은 orthnormal하게 선택할 수 있다.
- $m \times m$ 대칭행렬 A의 m개의 고유값과 이에 대응되는 고유벡터쌍이 $(\lambda_1, e_1), \cdots, (\lambda_m, e_m)$인 경우, 다음 식이 성립한다 (스펙트럴 분해)

$$A = P \Lambda P' = \lambda_1 e_1 e_1' + \cdots \lambda_m e_m e_m'$$

이 때 $\Lambda = diag(\lambda_1, \cdots, \lambda_m), P = [e_1, \cdots, e_m]$이며, 각 쌍의 순서가 바뀌어도 위 식은 성립한다.

6. 다변량 변수의 분산

- 확률벡터 $X = [X_1, \cdots, X_m]$의 공분산 행렬은 다음과 같이 정의된다.

$$Cov(X) = \Sigma = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_m) \\ Cov(X_1, X_2) & Var(X_2) & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ Cov(X_m, X_1) & \cdots & \ddots & Var(X_m) \end{bmatrix}$$

- 임의의 벡터 $a = (a_1, \cdots, a_m)', b = (b_1, \cdots, b_m)'$에 대해 다음이 성립한다.

$$Var(a'X) = Var(a_1X_1 + \cdots a_mX_m) = a'\Sigma a$$
$$Cov(a'X, b'X) = a'\Sigma b$$

- 공분산 행렬 Σ는 대칭행렬이며, 따라서 스펙트럴 분해가 가능하다. Σ의 m개의 고유값과 이에 대응되는 고유벡터쌍을 고유값의 크기 순으로 $(\lambda_1, a_1), \cdots, (\lambda_m, a_m)$이라 하자. (즉, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$이 성립).
- X의 총 변동 $= Var(X_1) + Var(X_2) + \cdots Var(X_m)$이라고 정의하자. 이 때 X의 총 변동 $= \lambda_1 + \cdots \lambda_m$이 성립한다.

7. 주성분 분석

- 주성분 분석의 목적 : 가능한 적은 변수로 기존의 다변량 벡터 $X = [X_1, \cdots, X_m]$이 설명하는 정보를 최대한 설명하기.
- 주성분의 정의
  - 첫번째 주성분 : 변수들의 전체 분산 중 가장 큰 부분을 설명할 수 있는 선형 결합
  - 두번째 주성분 : 첫번째 주성분과 독립이면서 첫번째 주성분에 의해 설명되지 않은 잔여 분산을 최대한 설명할 수 있는 선형 결합
  - 세번째 주성분 : ...
- **(주성분 분석 알고리즘)** $C_1 = a_1'X = a_{11}X_1 + a_{m1}X_m, C_2 = a_2'X, \cdots, C_m = a_m'X$ 라 하면, C_i가 X의 i번째 주성분이 되며, i번째 주성분의 분산은 λ_i와 동일하다. $(\rightarrow \lambda_i \geq 0)$