[K-MOOC] 인공지능에 필요한 기초수학 입문(High school)

(Introductory Math for Artificial Intelligence)

이상구 | 성균관대학교

11주차. 기댓값, 분산, 공분산, 상관계수, 공분산 행렬

* 기댓값, 분산, 표준편차 : 확률변수의 기댓값(expectation)은 확률적 사건에 대한 평균값으로, 사건이 일어나서 얻는 값과 그 사건이 일어날 확률을 곱한 것을 모든 사건에 대해 합한 값이다. 확률변수의 분산(variance)은 그 확률변수가 기댓값으로부터 얼마나 떨어진 곳에 분포하는지를 가늠하는 수이고, 표준편차(standard deviation)는 분산의 양의 제곱근으로 정의된다.

* 결합확률분포(joint probability distribution) : 확률변수가 두 개 이상 있는 경우에는 각각의 확률변수에 대한 확률분포 이외에도 확률분포 쌍이 가지는 복합적인 확률분포이다.

* 주변확률분포란 결합 확률분포에서 하나의 확률변수만 고려한 확률분포를 뜻한다.

* 공분산(covariance) : 확률변수 간의 상관관계를 알기 위한 개념으로, 확률변수 X와 Y의 공분산은 다음과 같이 정의된다.

$$Cov(X, Y) = \sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)] = E(XY) - \mu_x\mu_y$$

* 상관계수 : 확률변수의 절대 크기에 영향을 받지 않도록 공분산을 각 확률변수의 표준편차로 나누어 표준화 시킨 것. 확률변수 X와 Y 사이의 상관계수(correlation)는 다음과 같이 정의된다.

$$Corr(X, Y) = \rho = \frac{Cov(X, Y)}{S(X)S(Y)} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{E(X-\mu_x)^2 \cdot E(Y-\mu_y)^2}}$$

* 공분산 행렬 : p개의 확률변수를 $\{X_1, \ldots, X_p\}$에 대한 공분산 행렬(covariance matrix)은 (i, j)성분이 $i \neq j$일 때는 i번째 확률변수 x_i와 j번째 확률변수 x_j 사이의 공분산 σ_{ij}으로, $i = j$일 때는 i번째 확률변수의 분산 $\sigma_{ii} = \sigma_i^2$으로 하는 $p \times p$ 행렬로 정의하고 Σ로 표기한다.

$$\Sigma = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_p) \\ Cov(X_2, X_1) & Var(X_2) & \cdots & Cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & Cov(X_p, X_2) & \cdots & Var(X_p) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}$$