

확률변수와 확률분포

real function (실험수)

■ 확률변수 X

$$X(\xi) = x$$

■ 확률분포

$$P(X = x)$$

■ CDF (Cumulative Distribution Function : 누적분포함수)

$$F_X(X) = P[X \leq x], \quad -\infty < x < \infty$$

$$0 \leq F_X(X) \leq 1$$

$$\lim_{x \rightarrow \infty} F_X(X) = 1$$

$$\lim_{x \rightarrow -\infty} F_X(X) = 0$$

$$F_X(b) = \lim_{h \rightarrow 0} F_X(b+h) = F_X(b^+)$$

$$F_X(a) \leq F_X(b), \quad \text{if } a \leq b$$

■ PDF (Probability Density Function : 확률밀도함수:연속)

$$f_X(X) = \frac{dF_X(X)}{dx}$$

$$f_X(X) > 0$$

$$P[a < X < b] = \int_a^b f_X(X) dx$$

$$F_X(X) = \int_{-\infty}^X f_X(X) dx$$

$$1 = \int_{-\infty}^{\infty} f_X(X) dx$$

$$f_X(x|A) = \frac{d}{dx} F_X(x|A)$$

$$F_X(x|A) = \frac{P[\{X < x\} \cap A]}{P[A]}, \quad \text{if } P[A] > 0$$

■ PMF (Probability Mass Function : 확률질량함수:이산)

$$f_x(x) = \frac{\Delta F_x(x)}{\Delta x}$$

■ 기대값 : 확률변수의 평균

$\bar{x}$, μ (평균), s (분산), σ (표준편차)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{x} = \frac{1}{n} \sum_{\text{all } x} n_x x = \sum_{\text{all } x} x \frac{n_x}{n}$$

$$\frac{n_x}{n} \text{ (상대빈도)}$$

$$E[X] = \mu = \sum_{\text{all } x} x p(x)$$

$$E[X] = \mu = \int_{-\infty}^{\infty} x f_x(x) dx$$

■ 확률변수의 분산과 표준편차

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s^2 = \sum_{\text{all } x} (x_i - \bar{x})^2 \frac{n_x}{n-1}$$

$$\sigma^2 = \sum_{\text{all } x} (x_i - \bar{x})^2 p(x)$$

$$\text{Var}[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f_x(x) dx$$

$$\text{STD}[X] = \text{Var}[X]^{1/2}$$

Vector Random Variables

S_j : Random Point

Joint CDF (결합 누적분포함수), Joint PDF (결합 확률밀도함수)

$$F_X(x) = P\{X \leq x\}, F_Y(y) = P\{Y \leq y\}$$

$$F_{X,Y}(x, y) = P\{X \leq x, Y \leq y\}, P\{X \leq x, Y \leq y\} = P(A \cap B)$$

$$X = [x_1 \ x_2 \ \dots \ x_N]^T$$

■ 결합 누적분포함수

$$F_X(X) = P_X[\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\} \cap \dots \cap \{X_N \leq x_N\}]$$

■ 결합 확률밀도함수

$$f_X(X) = \frac{\partial^N F_X(X)}{\partial x_1 \partial x_2 \dots \partial x_N}$$

marginal (주변) 확률밀도함수

$$(\text{Random Vector}) X = [x_1 \ x_2]^T, f_{x_1 \ x_2}(x_1 \ x_2)$$

$$f_{x_1}(x_1) = \int_{x_2=-\infty}^{x_2=+\infty} f_{x_1 \ x_2}(x_1 \ x_2) dx_2$$

Random Vector

■ 평균 벡터

$$E[X] = [E[X_1] \ E[X_2] \ \dots \ E[X_N]]^T = [\mu_1 \ \mu_2 \ \dots \ \mu_N] = \mu$$

■ 공분산 벡터

$$\begin{aligned} \text{COV}[X] &= \Sigma = E[(X - \mu)(X - \mu)^T] \\ &= E \begin{bmatrix} E[(X_1 - \mu)(X_1 - \mu)] & \dots & E[(X_1 - \mu)(X_N - \mu)] \\ \vdots & \ddots & \vdots \\ E[(X_N - \mu)(X_1 - \mu)] & \dots & E[(X_N - \mu)(X_N - \mu)] \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \dots & c_{1N} \\ \vdots & \ddots & \vdots \\ c_{1N} & \dots & \sigma_N^2 \end{bmatrix} \end{aligned}$$

공분산 행렬

만약, x_i, x_k 가 모두 증가하면, $c_{ik} > 0$

만약, x_k 가 증가할 때, x_i 감소한다면, $c_{ik} < 0$

만약, x_i, x_k 가 무상관 (uncorrelated) 이라면 $c_{ik} = 0$

| c_{ik} | < $\sigma_i \sigma_k$, 여기서 σ_i 는 x_i 의 표준편차이다.

$$c_{ii} = \sigma_i^2 = \text{VAR}(x_i)$$

$$c_{ii} = \sigma_i^2, \quad c_{ik} = \rho_{ik} \sigma_i \sigma_k$$

σ_{ik} (상관계수)

$$(\text{공분산행렬}) \Sigma = E[(X - \mu)(X - \mu)^T] = E[XX^T] - \mu\mu^T = S - \mu\mu^T$$

$$\text{(자기상관행렬)} \mathbf{S} = \mathbf{E}[\mathbf{X}\mathbf{X}^T] = \begin{bmatrix} \mathbf{E}[x_1 x_1] & \cdots & \mathbf{E}[x_1 x_N] \\ \vdots & \ddots & \vdots \\ \mathbf{E}[x_N x_1] & \cdots & \mathbf{E}[x_N x_N] \end{bmatrix}$$

$$\text{(공분산행렬)} \mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{R}\mathbf{\Gamma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{bmatrix} \cdot \begin{bmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1N} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & 0 & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{bmatrix}$$

$\mathbf{\Gamma}$ 가 특징들의 스칼라를 포함

$\mathbf{R}$ (상관행렬) 은 특징들 간의 근본적인 관계 정보를 보유

만약, $\mathbf{E}[x_i x_k] = \mathbf{E}[x_i] \mathbf{E}[x_k]$ 이면, 두개의 랜덤변수 $x_i x_k$ 는 무상관

만약, $\mathbf{P}[x_i x_k] = \mathbf{P}[x_i] \mathbf{P}[x_k]$ 이면, 두개의 랜덤변수 $x_i x_k$ 는 독립

Gaussian Distribution (가우시안 분포)

$$g_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

invertible positive definite (양의 정부호) 대칭행렬

$x^T \Sigma^{-1} x$ 의 형태

d차원 ($d \times 1$) 특징벡터 x , ($d \times 1$) μ , ($d \times d$) Σ

$$g_{(\mu, \Sigma)}(x) = \frac{1}{(\sqrt{2\pi})^d \sqrt{\det[\Sigma]}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$\mu = \mathbf{E}[x], \Sigma = \mathbf{E}[(x - \mu)(x - \mu)^T]$$

가우시안 확률밀도함수 $N(\mu, \Sigma)$

$x \sim N(\mu, \Sigma)$

~ 분포를 가진다.

■ Central Limit Theorem (중심 극한 정리)

$$\hat{\mu}_N \approx N\left(\mu, \frac{\sigma}{\sqrt{N}}\right)$$

■ 완전 공분산 Gaussian 형태

Σ (공분산 행렬) 이면 확률밀도함수

$D + \frac{D(D+1)}{2}$ 개의 파라미터 필요

학습 Data 로 부터 μ , Σ 추정

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n : \text{MATLAB mean (X)}$$

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{x}_n - \hat{\mu}) (\mathbf{x}_n - \hat{\mu})^T : \text{MATLAB cov (X)}$$

■ 대각 공분산 Gaussian 형태

Σ (공분산 행렬) 가 대각 공분산 행렬이면 확률밀도함수

2 D개의 파라미터 필요

학습 Data로 부터 μ , (D × 1) 분산 벡터 $\sigma^2 = \text{diag}(\Sigma)$

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n : \text{MATLAB mean (X)}$$

$$\hat{\sigma}_i^2 = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{x}_{in} - \hat{\mu}_i)^2 : \text{MATLAB diag (cov (X))}$$

■ 구형 공분산 Gaussian 형태

Σ (공분산 행렬) 가 $\Sigma = \sigma^2 \mathbf{I}$ 의 형태

D + 1 개의 파라미터 필요

학습 Data 로 부터 μ 와 스칼라 분산값 σ^2 추정

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n : \text{MATLAB mean (X)}$$

$$\hat{\sigma}_i^2 = \frac{1}{D(N-1)} \sum_{n=1}^N \|\mathbf{x}_{in} - \hat{\mu}_i\|^2 : \text{MATLAB mean (diag (cov (X)))}$$

$$\|\mathbf{X}\|^2 \approx \mathbf{X}^T \mathbf{X}$$