PRML Chapter.1-2

Information Theory + The Gaussian Distribution

2017010698
수학과 오서영

1-6. Information Theory

이산 랜덤 변수 : x
우리가 x 의 구체적인 값을 관찰하는 경우
얼마만큼의 information 를 얻는지 정량화
Information? : 학습에 있어 필요한 놀람의 정도 (degree of surprise)

(EX) 잘 일어날 것 같지 않은 사건을 관찰하는 경우,
빈번하게 일어나는 사건보다 더 많은 정보를 취득했다고 고려
-> 따라서 항상 발생하는 일이라면, 사건 발생 후 얻는 정보의 양은 0

정보의 양을 h(x) 라고 정의 -> 확률 함수의 조합으로 표현
정보의 값은 0 이상
-> 0~1 사이의 확률 값에 log를 붙이는 경우 앞에 음수가 붙어야 한다.
낮을 확률 값을 가지는 사건이 높은 정보량

$$h(x) = -\log_2 p(x)$$

1-6. Information Theory

b=2 -> h(x) 의 기본 단위가 비트(bit)
랜덤 변수 하나를 송신자가 수신자에게 전달한다고 가정
-> 이 때 전송되는 데이터 양의 평균을 고려 -> **entropy**
엔트로피는 평균 정보량을 의미, p(x) 인 분포에서 h(x)함수의 기대값

$$\mathrm{H}[x] = -\sum_{x} p(x) \log_2 p(x).$$

(EX) 랜덤 변수 x 가 8 개의 가능한 값을 가지는 경우
각각의 경우가 발현될 확률이 모두 동일하게 1/8 인 경우
(확률 분포가 Uniform 분포)
하나의 데이터 x 를 전송하기 위해 필요한 평균 비트 수는 3이 된다.
- Solution)

$$\mathrm{H}[x] = -8 \times \frac{1}{8} \log_2 \frac{1}{8} = 3 \text{ bits}.$$

1-6. Information Theory

$$\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right)$$

$$\mathrm{H}[x] = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{16}\log_2\frac{1}{16} - \frac{4}{64}\log_2\frac{1}{64} = 2 \text{ bits.}$$

Non-Uniform 분포의 엔트로피가
Uniform 분포의 엔트로피보다 낮다

2-3. The Gaussian Distribution

Gaussian Distribution

$$N(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

입력 변수가 D 차원인 벡터
-> 다변량 가우시안 분포

$$N(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

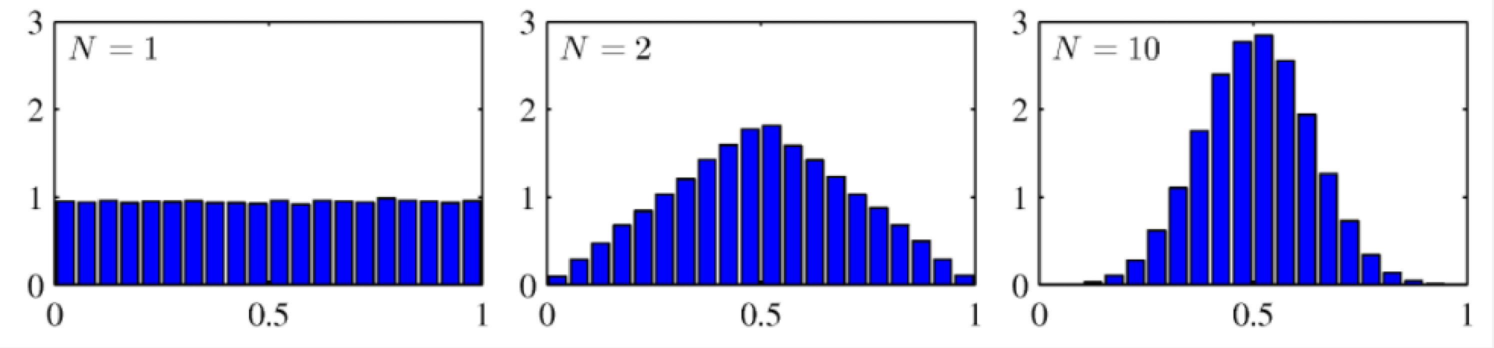
μ 는 D 차원의 평균 벡터
Σ 는 D×D 크기의 공분산 행렬

2-3. The Gaussian Distribution

중심 극한 정리

동일한 확률 분포를 따르는 N 개의 독립 확률 변수의
평균 값과 합은 N 이 충분이 크다면 가우시안 분포를 따름.
연속형 확률 변수에서 가우시안 분포는
변수의 엔트로피를 최대화하는 분포

(EX) [0,1] 범위에서 균등분포(uniform distribution)를
가지는 결과의 평균 값의 히스토그램



-> N 이 커질수록 정규 분포의 모양

2-3. The Gaussian Distribution

가우시안 분포의 기하학적인 형태

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

x 에 대한 가우시안의 함수적 종속성은
exp 지수부에 등장하는 quadratic 에 있음

$$\Delta^2 = (\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

Δ : μ 에서 x 까지의 Mahalanobis distance
만약 공분산이 identity matrix 인 경우 이 값은
Euclidean distance 와 동일해진다.
-> 평균과의 거리를 측정할 때 분산도를 고려한다는 의미

2-3. The Gaussian Distribution

공분산 행렬 Σ 가 symmetric matrix
-> 고유벡터(eigen-vector) 식을 적용할 수 있다.

$$\mathbf{\Sigma u}_i = \lambda_i \mathbf{u}_i$$

-> 공분산 행렬의 고유벡터 (orthonormal) , 고유값을 정의
즉 u : 단위 직교 벡터

$$\mathbf{u}_i^T \mathbf{u}_j = I_{ij}$$

$$I_{ij} = \begin{cases} 1 & if\ i = j \\ 0 & otherwise \end{cases}$$

공분산 행렬을 고유벡터로 표현 -> 선형결합 형태로 표현가능

$$\mathbf{\Sigma} = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

$$\mathbf{\Sigma}^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^T$$

2-3. The Gaussian Distribution

quadratic 식에 이 값을 대입

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

$$y_i = \mathbf{u}_i^T(\mathbf{x} - \boldsymbol{\mu})$$

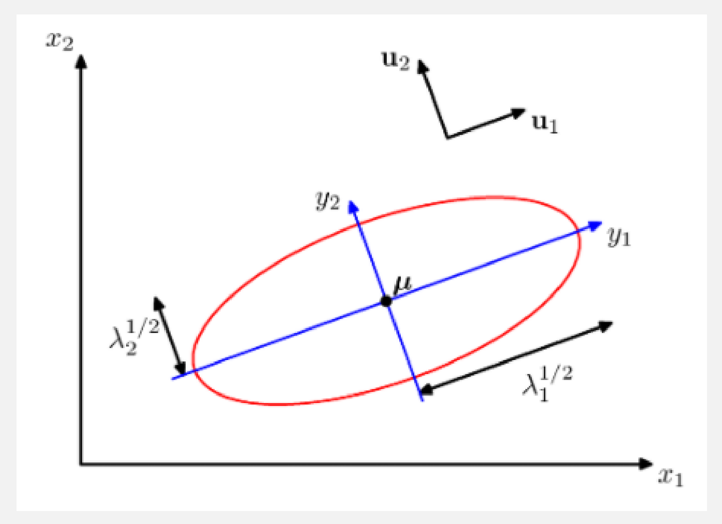
$$\mathbf{y} = \mathbf{U}(\mathbf{x} - \boldsymbol{\mu})$$

y 해석하기

y 를 x 에 대한 좌표 변환
-> 가우시안 함수의 원점을
μ 로 옮기고 고유 벡터를 축으로
회전 변환하는 식

U : 직교 행렬
UTU=I, UUT=I 를 만족
(주축 변환)

2-3. The Gaussian Distribution



Red : 타원 표면으로 동일한
확률 값을 가지는 위치가 된다.

u 는 고유 벡터로 정의되며
이 축으로 타원이 형성된다.
-> 고유벡터를 축으로 하는 타원에
대한 공분산은 고유값 λ 의 영향

그림에서 붉은 선처럼 동일한 상수 확률
값을 가지는 위치에서

$\Delta^2 = \sum_D \frac{y_i^2}{\lambda_i}$ 도 동일한 상수 값을 가지게됨

만약 고유값 λ 가 모두 양수
-> 컨투어는 타원형태가 된다.
그림 -> 가우시안에서는
고유값이 모두 양수 (positive definite)

타원의 중심이 μ 가 되고 고유값
$\lambda_i^{1/2}$ 스케일로 타원이 형성하게 된다.