LIKE LION

보충 자료

기초 선형대수 및 확률

* 전체 내용을 다루는게 아닌, 강의에 필요한 부분만 다루겠습니다.
자세히 설명하지 않은 부분은 “아 이런게 있구나~” 하고 넘어가시면 됩니다.

강사 최윤호

LIKE LION

선형대수

Scalars (스칼라)

- 스칼라는 하나의 숫자
- 정수, 실수, 유리수 등
- 이탤릭체로 표시합니다.

$$a\,,\,n\,,\,x$$

Vectors (벡터)

- 벡터는 숫자들의 1-D array (소문자, 볼드체로 표시)

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

- 실수, binary (0 또는 1), 정수 등 일 수 있음
- n차원의 실수 벡터의 집합을 다음과 같이 나타냄

$$\mathbb{R}^n$$

Matrices (행렬)

- 행렬은 숫자들의 2-D array

Row

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}.$$

Column

- A 라는 행렬 (대문자, 볼드체로 표시)이 m개의 row (행), n개의 column (열)을 가지고 있다는 것을 다음과 같이 나타냄:

$$\boldsymbol{A} \in \mathbb{R}^{m \times n}$$

LIKE LION

Tensors (텐서)

- 텐서는 숫자들의 array인데,
  - 0차원의 스칼라일 수도 있고
  - 1차원의 벡터일 수도 있고,
  - 2차원의 행렬일 수도 있고
  - n차원의 array일 수도 있음.

Matrix Transpose (전치)

- 행이 열로 가도록, 열이 행으로 가도록 함. (주대각선 (i=j)에 거울처럼 반사 시킨다고 생각해도 됨)

$$(\boldsymbol{A}^\top)_{i,j} = A_{j,i}.$$

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow \boldsymbol{A}^\top = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

- 행렬 곱의 전치

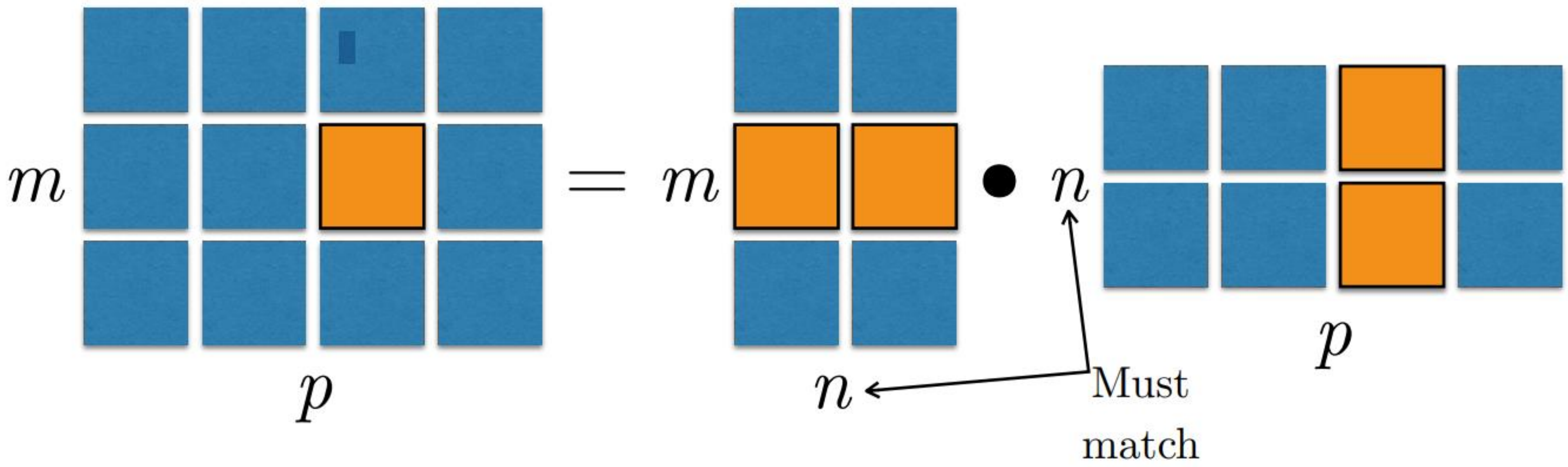
$$(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$$

LIKE LION

Matrix Product (행렬 곱)

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}.$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}.$$



Matrix Product (행렬 곱)

$$A(B + C) = AB + AC.$$

$$A(BC) = (AB)C.$$

$$AB \neq BA$$

Identity Matrix (단위 행렬)

- 주 대각선의 성분 (i=j)이 1, 나머지는 0인 정사각행렬 (m=n)

$$\boldsymbol{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- 단위행렬은 어떤 벡터를 곱해도 해당 벡터를 변형시키지 않음

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}.$$

Systems of Equations (선형 방정식 계)

- m개의 선형 방정식을 아래와 같이 행렬 곱을 이용하여 나타낼 수 있음. (세미콜론: slicing)

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

expands to

$$\boldsymbol{A}_{1,:}\boldsymbol{x} = b_1$$

$$\boldsymbol{A}_{2,:}\boldsymbol{x} = b_2$$

$$\ldots$$

$$\boldsymbol{A}_{m,:}\boldsymbol{x} = b_m$$

Systems of Equations (선형 방정식 계)

- 선형 방정식 계는
  - 해가 없을 수도 있고
  - 여러 개 일수도 있고
  - 하나의 해만 존재할 수도 있음. (행렬 A의 역행렬이 존재할 때)

Matrix Inversion (역행렬)

- 역행렬: 원행렬에 곱했을 때 단위행렬이 되는 행렬

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

- 이를 통해 선형 방정식 계를 해결

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

- 역행렬이 존재하지 않는 경우
  - 열이 행보다 많거나, 행이 열보다 많거나
  - 서로 비슷한 행/열들을 가지고 있을 때 (용어: 선형 종속) (선형 독립이어야 역행렬 존재!)

LIKE LION

Norms

- 벡터가 얼마나 큰지 나타내는 함수: zero vector와의 거리로 볼 수도 있음

  - L^p norm

$$||\boldsymbol{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

  - Most popular norm: L2 norm, p=2
  - L1 norm, p=1: $||\boldsymbol{x}||_1 = \sum_i |x_i|.$
  - Max norm, infinite p: $||\boldsymbol{x}||_\infty = \max_i |x_i|.$

특별한 성질의 행렬과 벡터

- Unit vector (단위 벡터)

$$||\boldsymbol{x}||_2 = 1.$$

- 대칭행렬

$$\boldsymbol{A} = \boldsymbol{A}^\top.$$

- 직교 행렬

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}.$$

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$$

LIKE LION

Eigendecomposition (고윳값 분해)

- n x n 행렬 $\boldsymbol{A}$의 고윳값 λ와 고유벡터 $\boldsymbol{v}$ 의 정의

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

- 각각 n개의 고윳값과 고유벡터를 얻었을 때, $Av_i = \lambda_i v_i \text{ for } i = 1, 2, \cdots, n$

$$V = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times n} \qquad AV = \begin{bmatrix} | & | & & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \\ | & | & & | \end{bmatrix}$$

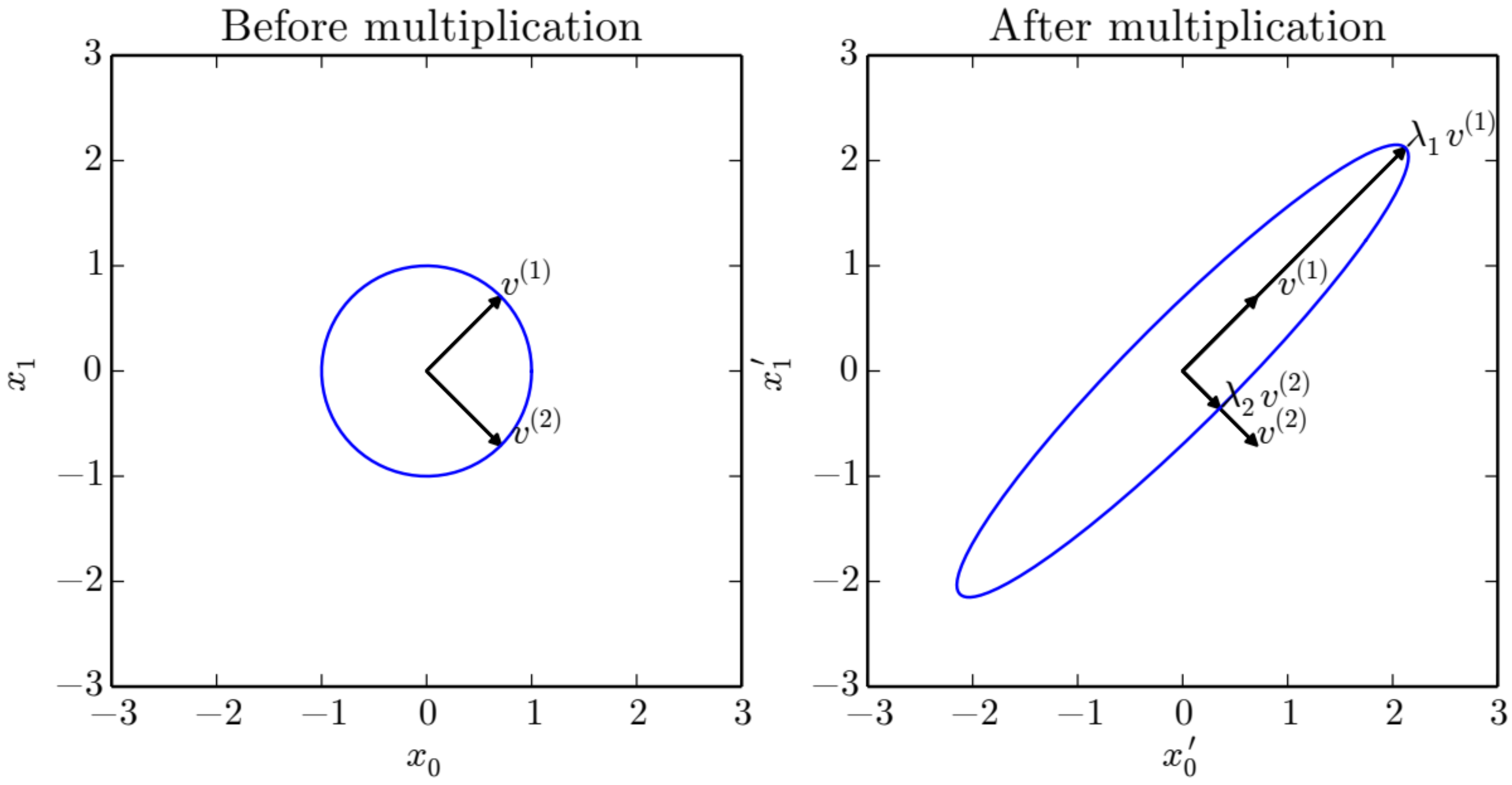
- $\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ 라고 하면 맨 위의 정의를 다음과 같이 쓸 수 있다. $AV = V\Lambda$

특히 고유벡터들이 선형 독립하다면: $A = V\Lambda V^{-1}$

LIKE LION

Eigendecomposition (고윳값 분해)

- 단위 원 위의 단위 벡터들에 행렬 A를 곱하면, A의 고유벡터 v 방향으로 각각의 고유값으로 스케일링하는 식으로 공간을 변형함



LIKE LION

SVD, Trace, Determinant (Optional)

- Singular Value Decomposition (특이값 분해): 고윳값 분해와 비슷하나 정사각행렬일 필요 없는 더 일반적인 방법.

$$A = UDV^\top.$$

  - **A**: m x n, **U**: m x m, **D**: m x n, **V**:, n x n

- Trace (대각합): 주 대각선의 성분을 모두 더한 것.

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i \boldsymbol{A}_{i,i}.$$

$$\mathrm{Tr}(\boldsymbol{AB}) = \mathrm{Tr}(\boldsymbol{BA})$$

- Determinant (행렬식): 정사각행렬 **A**의 행렬식 det(**A**)는 스칼라 값으로, **A**의 모든 고윳값들의 곱이다.

확률

LIKE LION

Random Variable (확률변수, r.v.)

- 확률적으로 다른 값들을 가질 수 있는 변수

- Discrete variable: 셀 수 있는 값들 (불연속) 중 하나를 가질 수 있는 변수. ex) 내일의 강수 여부 X
- Continuous variable: 연속적인 값을 가질 수 있는 변수 ex) 내일의 강수량 Z

LIKE LION

PMF, PDF

- Probability Mass Function (확률질량함수): discrete r.v.가 어떤 값을 가질 확률을 계산하는 함수

$$\forall x \in \mathrm{x}, 0 \leq P(x) \leq 1.$$

$$\textstyle\sum_{x \in \mathrm{x}} P(x) = 1$$

- ex) uniform distribution: $P(\mathrm{x} = x_i) = \dfrac{1}{k}$

- Probability Density Function (확률밀도함수): continuous r.v.가 어떤 값을 가질 확률을 계산하는 함수

$$\forall x \in \mathrm{x}, p(x) \geq 0$$

$$\int p(x)dx = 1.$$

- ex) 구간 [a,b] 사이의 uniform distribution: $u(x; a, b) = \frac{1}{b-a}$

Marginal Probability

- Sum rule:

  - discrete r.v. :

$$\forall x \in \mathrm{x}, P(\mathrm{x}=x)=\sum_{y} P(\mathrm{x}=x, \mathrm{y}=y).$$

  - continuous r.v. :

$$p(x)=\int p(x, y) d y.$$

Conditional Probability

- $\mathrm{x} = x$ 가 주어졌을 때, $\mathrm{y} = y$ 일 확률

$$P(\mathrm{y} = y \mid \mathrm{x} = x) = \frac{P(\mathrm{y} = y, \mathrm{x} = x)}{P(\mathrm{x} = x)}.$$

Chain Rule of Probability

- 여러 개의 r.v.에 대한 확률분포가 주어졌을 때 (joint probability), 한 r.v.에 대한 조건부확률들로 이를 분해 가능

$$P(\mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(n)}) = P(\mathrm{x}^{(1)}) \Pi_{i=2}^{n} P(\mathrm{x}^{(i)} \mid \mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(i-1)}).$$

- 예시

$$\begin{aligned} P(\mathrm{a}, \mathrm{b}, \mathrm{c}) &= P(\mathrm{a} \mid \mathrm{b}, \mathrm{c}) P(\mathrm{b}, \mathrm{c}) \\ P(\mathrm{b}, \mathrm{c}) &= P(\mathrm{b} \mid \mathrm{c}) P(\mathrm{c}) \\ P(\mathrm{a}, \mathrm{b}, \mathrm{c}) &= P(\mathrm{a} \mid \mathrm{b}, \mathrm{c}) P(\mathrm{b} \mid \mathrm{c}) P(\mathrm{c}). \end{aligned}$$

LIKE LION

Independence (독립)

- independence (독립): x와 y가 독립이라면 아래의 식 성립, 또한 아래의 식 성립하면 x와 y는 독립

$$\forall x \in \mathrm{x}, y \in \mathrm{y},\ p(\mathrm{x}=x, \mathrm{y}=y) = p(\mathrm{x}=x)p(\mathrm{y}=y).$$

- conditional independence: z가 주어졌을 때 x와 y가 조건부 독립, 또한 아래의 식 성립하면 조건부 독립

$$\forall x \in \mathrm{x}, y \in \mathrm{y}, z \in \mathrm{z},\ p(\mathrm{x}=x, \mathrm{y}=y \mid \mathrm{z}=z) = p(\mathrm{x}=x \mid \mathrm{z}=z)p(\mathrm{y}=y \mid \mathrm{z}=z).$$

LIKE LION

Expectation (기댓값)

- Discrete r.v.

$$\mathbb{E}_{\mathrm{x}\sim P}[f(x)] = \sum_{x} P(x) f(x)$$

- Conditional r.v.

$$\mathbb{E}_{\mathrm{x}\sim p}[f(x)] = \int p(x) f(x) dx$$

- 선형성

$$\mathbb{E}_{\mathrm{x}}[\alpha f(x) + \beta g(x)] = \alpha \mathbb{E}_{\mathrm{x}}[f(x)] + \beta \mathbb{E}_{\mathrm{x}}[g(x)]$$

Variance

- Variance (분산) : 평균(기댓값)에 대해 확률분포가 퍼져있는 정도

$$\mathrm{Var}(f(x)) = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right]$$

- Covariance (공분산): 두 r.v.의 상관 관계를 알 수 있음 (양수: 양의 상관관계, 음수: 음의 상관관계)

$$\mathrm{Cov}(f(x), g(y)) = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])(g(y) - \mathbb{E}[g(y)])\right].$$

- 공분산 행렬

$$\mathrm{Cov}(\mathbf{x})_{i,j} = \mathrm{Cov}(\mathrm{x}_i, \mathrm{x}_j).$$

LIKE LION

대표적인 확률분포

- 베르누이 분포: x는 binary random variable로 0 또는 1의 값을 가질 수 있음. x=1일 확률을 ϕ 라고 했을 때,

$$P(\mathrm{x}=1)=\phi$$

$$P(\mathrm{x}=0)=1-\phi$$

$$P(\mathrm{x}=x)=\phi^{x}(1-\phi)^{1-x}$$

$$\mathbb{E}_{\mathrm{x}}[\mathrm{x}]=\phi$$

$$\mathrm{Var}_{\mathrm{x}}(\mathrm{x})=\phi(1-\phi)$$

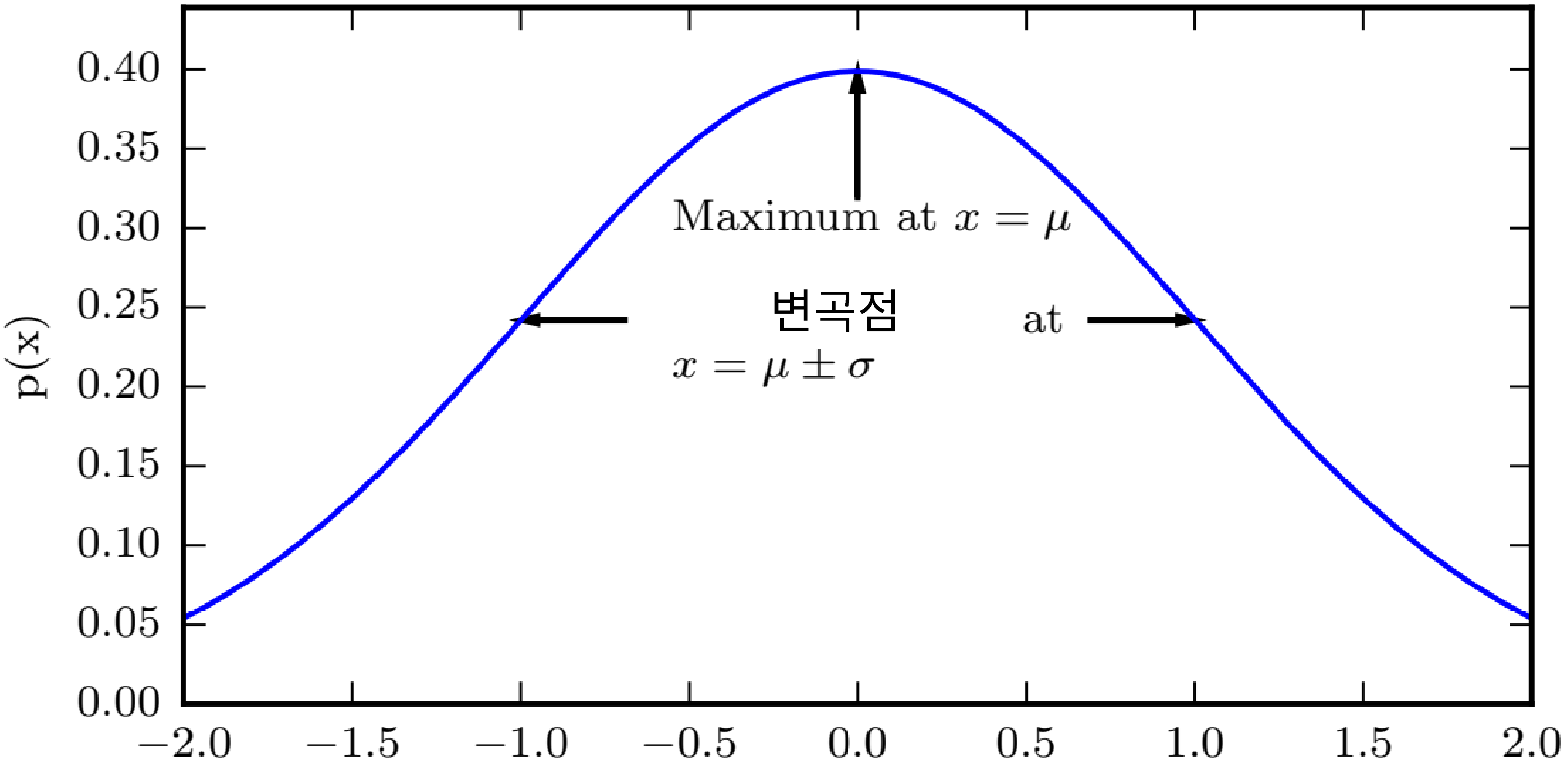
- 가우시안 분포: mean과 variance로 parameterized됨.

$$\mathcal{N}(x;\mu,\sigma^2)=\sqrt{\frac{1}{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

- 다변수 가우시안 분포: mean vector와 covariance matrix로 parameterized됨.

$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})=\sqrt{\frac{1}{(2\pi)^n\det(\boldsymbol{\Sigma})}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

가우시안 분포



0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
p(x)
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
Maximum at $x = \mu$
변곡점
at
$x = \mu \pm \sigma$

LIKE LION

Bayes' Rule

- 베이즈 룰: 조건부 확률 + 체인 룰

$$P(\mathrm{x} \mid \mathrm{y}) = \frac{P(\mathrm{x})P(\mathrm{y} \mid \mathrm{x})}{P(\mathrm{y})}.$$

정보 이론 (Optional)

- Information Theory
  - information: 확률이 낮은 사건이 일어나는 것이 더 유용한 정보임. ex) 해가 동쪽에서 뜬다 → 유용한 정보 X

$$I(x) = -\log P(x).$$

  - Entropy: 분포 P에서 얻을 수 있는 정보의 기댓값. 전체 확률 분포에 대한 uncertainty (불확실성)의 양으로 볼 수 있음.

$$H(\mathrm{x}) = \mathbb{E}_{\mathrm{x}\sim P}[I(x)] = -\mathbb{E}_{\mathrm{x}\sim P}[\log P(x)].$$

  - KL divergence: 항상 0보다 크고, 두 분포 P와 Q가 같을 때만 0 → 두 분포가 얼마나 다른지 나타낼 때 사용

$$D_{\mathrm{KL}}(P\|Q) = \mathbb{E}_{\mathrm{x}\sim P}\left[\log \frac{P(x)}{Q(x)}\right] = \mathbb{E}_{\mathrm{x}\sim P}\left[\log P(x) - \log Q(x)\right].$$