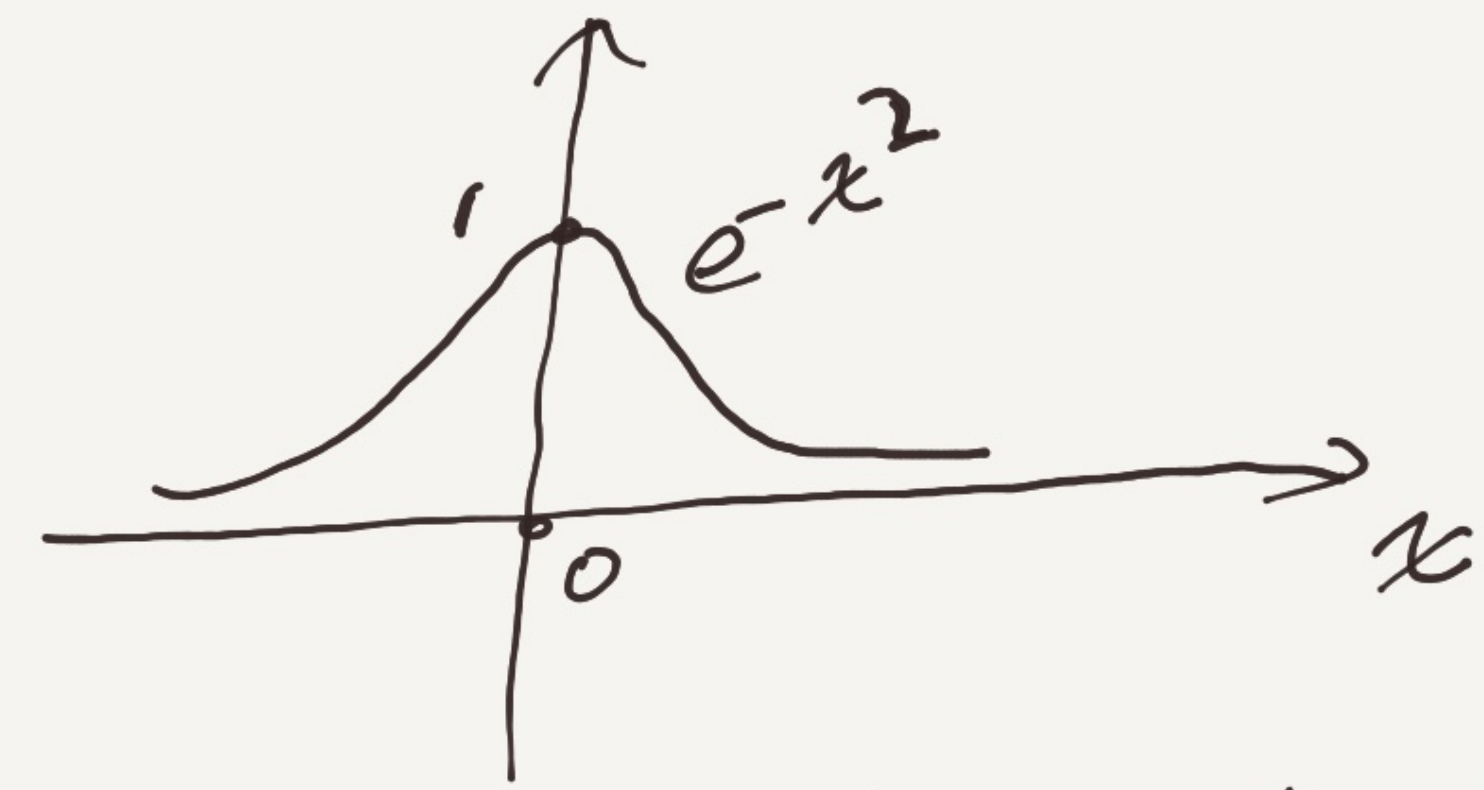
Lecture 3. Functions

8/28/23

1. Gaussian function

① Base form

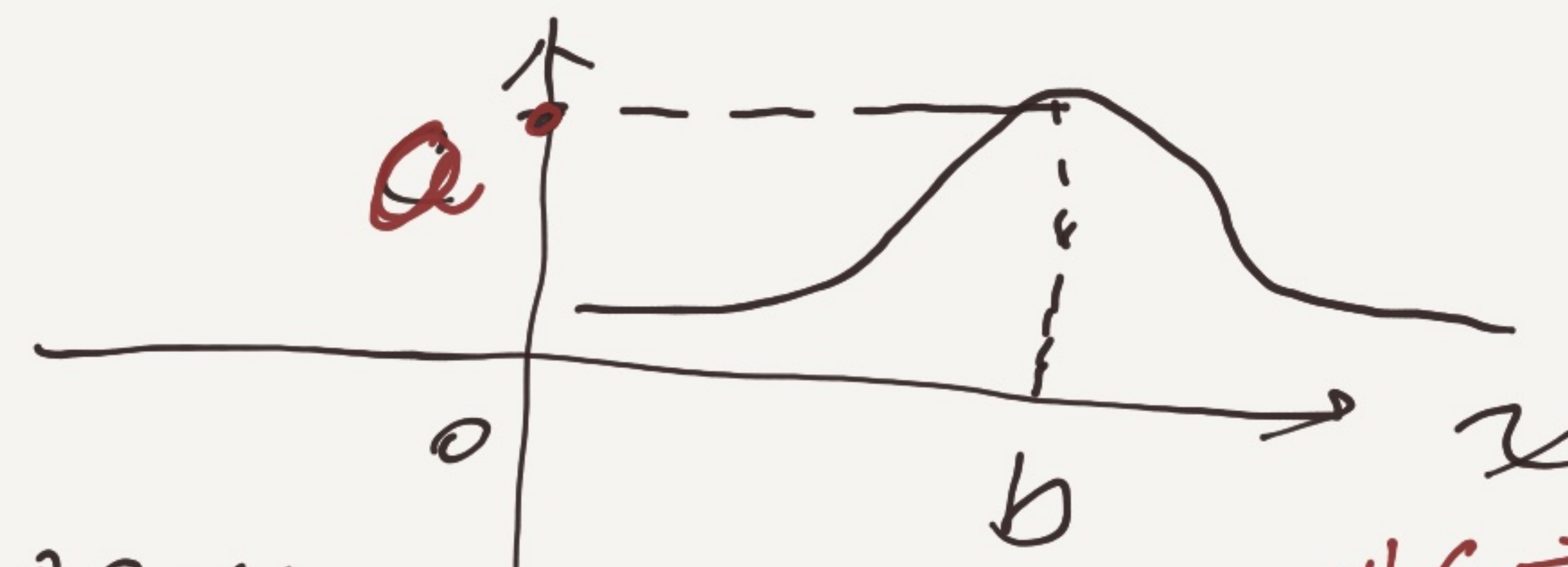
$$f(x) = e^{-x^2}$$



symetric "bell shape"

② parametric extension

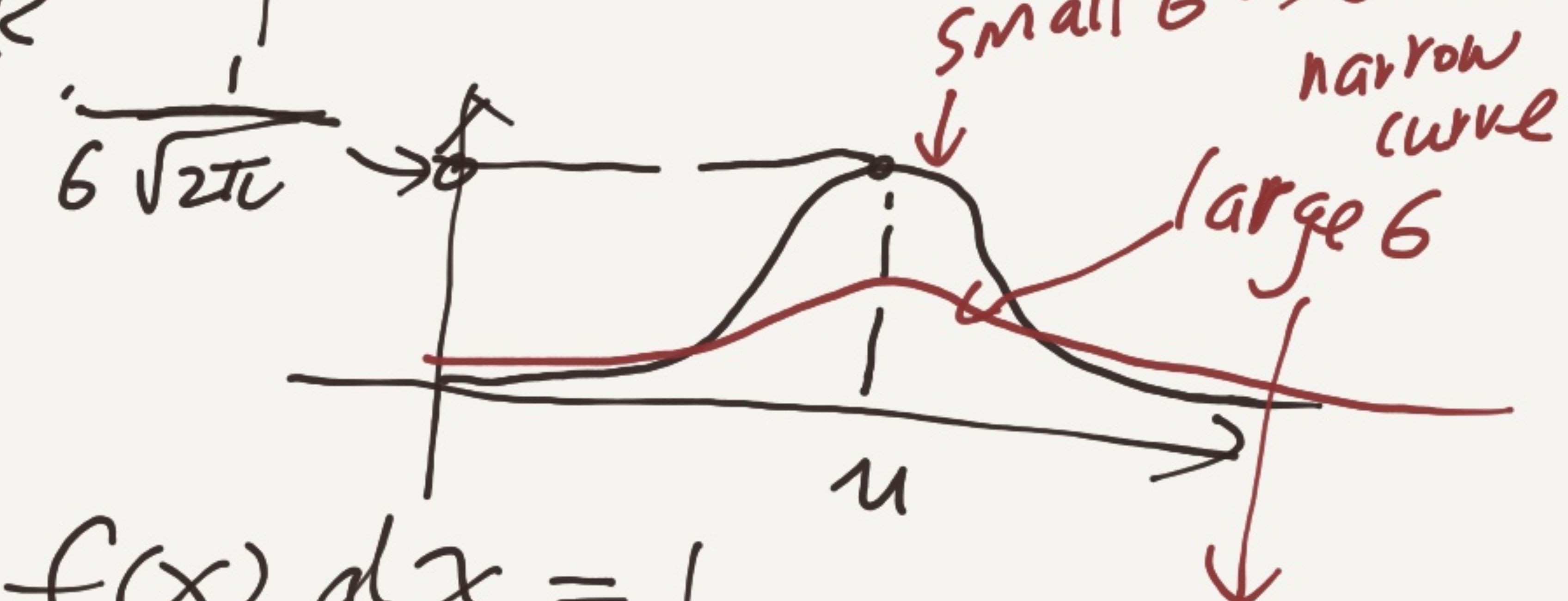
$$f(x) = a \cdot e^{-\frac{(x-b)^2}{c^2}}$$



a: height of peak.

b: position of the center of the peak

c: width of the bell



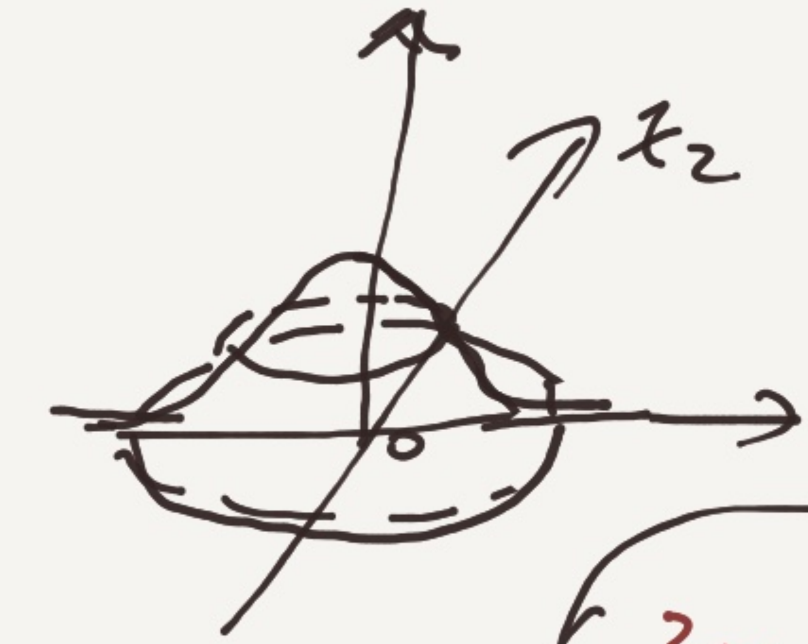
③ Gaussian distribution.

$$f(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{a} \cdot e^{-\frac{(x-\overset{b}{u})^2}{\underset{c}{2\sigma^2}}}$$

$$\int f(x)\,dx = 1$$

$$f(x) \in [0, 1]$$

⑥ 2D Gaussian functions



1) $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$: $f(x) = e^{-(x_1^2 + x_2^2)}$

$$= e^{-\underline{x^T \cdot x}}$$

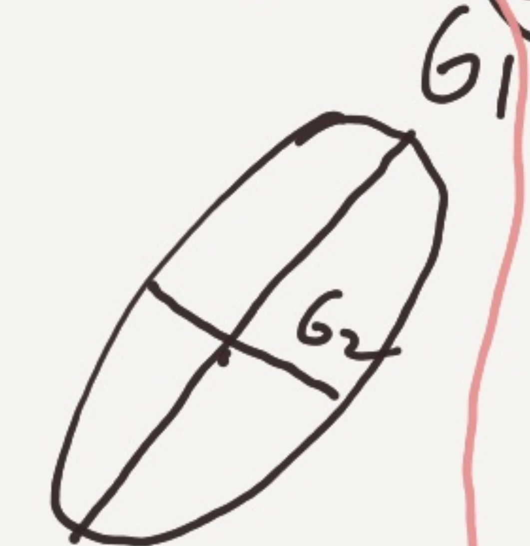
2) $f(x) = A\, e^{-\frac{1}{2}\cdot\left(\frac{(x_1-u_1)^2}{\sigma_1^2} + \frac{(x_2-u_2)^2}{\sigma_2^2}\right)} \Longrightarrow \frac{\sigma_2^2 (x_1-u_2)^2 + \sigma_1^2 (x_2-u_2)^2}{\sigma_1^2 \cdot \sigma_2^2}$

$$= \left(\frac{x_1-u_1}{\sigma_1^2}, \frac{x_2-u_2}{\sigma_2^2}\right) \cdot \overbrace{\begin{bmatrix} x_1-u_1 \\ x_2-u_2 \end{bmatrix}}^{x-u}$$

$$= (x-u)^T \cdot \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \cdot (x-u)$$

$$= (x-u)^T \Sigma^{-1} (x-u)$$

$b = u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ $\quad \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix}$



$$= A\, e^{-\frac{1}{2}(x-u)^T \Sigma^{-1} (x-u)}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}_{2\times 2} \qquad \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}$$

$$\det(\Sigma) = \sigma_1^2 \cdot \sigma_2^2 - 0 \cdot 0 = (\sigma_1 \sigma_2)^2$$

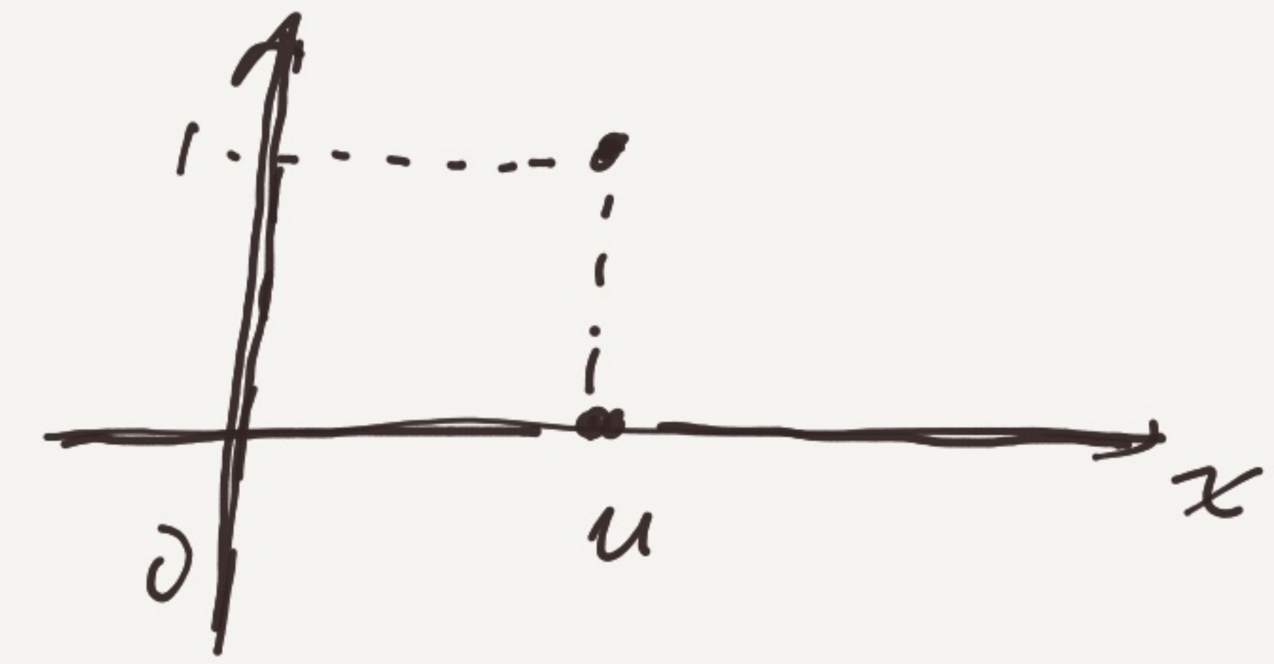
3) 2D/~~k~~D Gaussian distribution

$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$ $\quad f(x) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} e^{-\frac{1}{2}(x-u)^T \Sigma^{-1} (x-u)}$

$|\Sigma|$ or $\det(\Sigma)$: determinant of Σ.

2. Dirac delta function

$$\delta(x-u) = \begin{cases} 1 & \text{if } x=u \\ 0 & \text{otherwise} \end{cases}$$



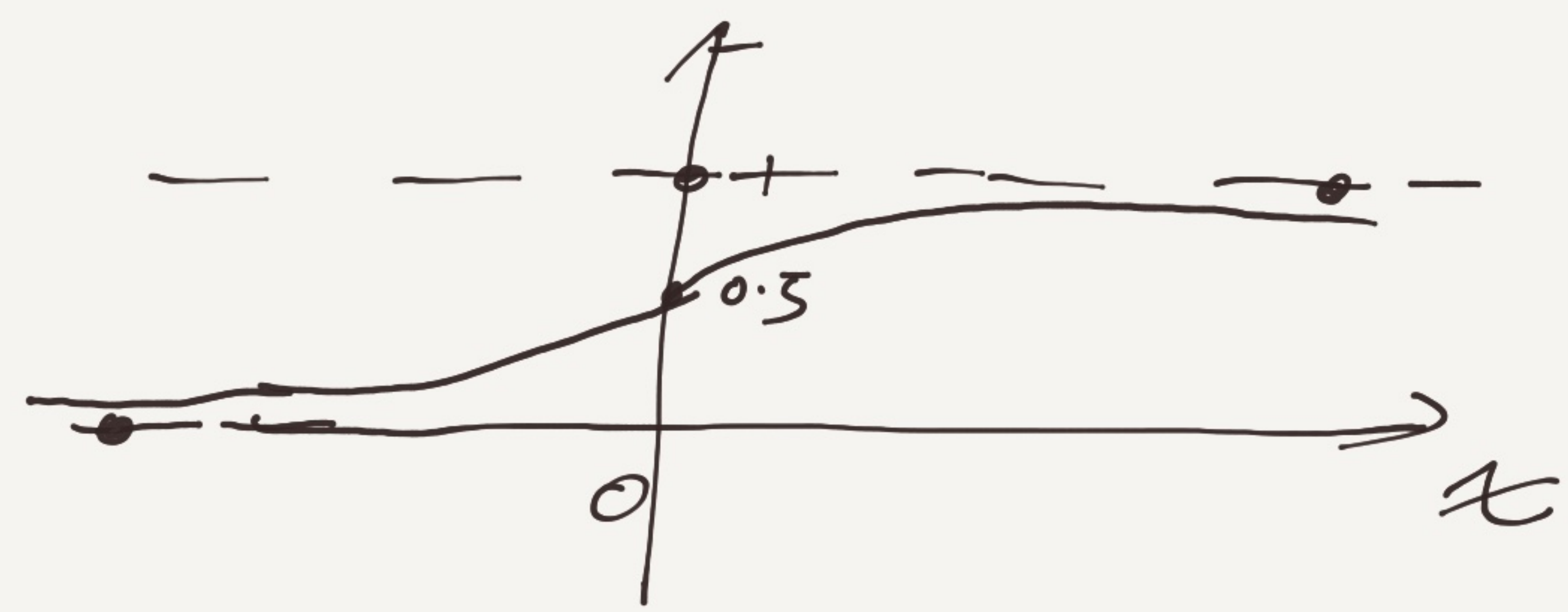
$$\delta(x-u)\cdot f(x) = \begin{cases} f(u) & \text{if } x=u. \\ 0 & \text{otherwise.} \end{cases}$$

$D(x^n) = n.x^{n-1}$

3. logistic sigmoid function

$D(\frac{1}{x}) = D(x^{-1})$
$= -x^{-2}$
$= \frac{-1}{x^2}$

$$sigm(x) = \frac{1}{1+e^{-x}}$$



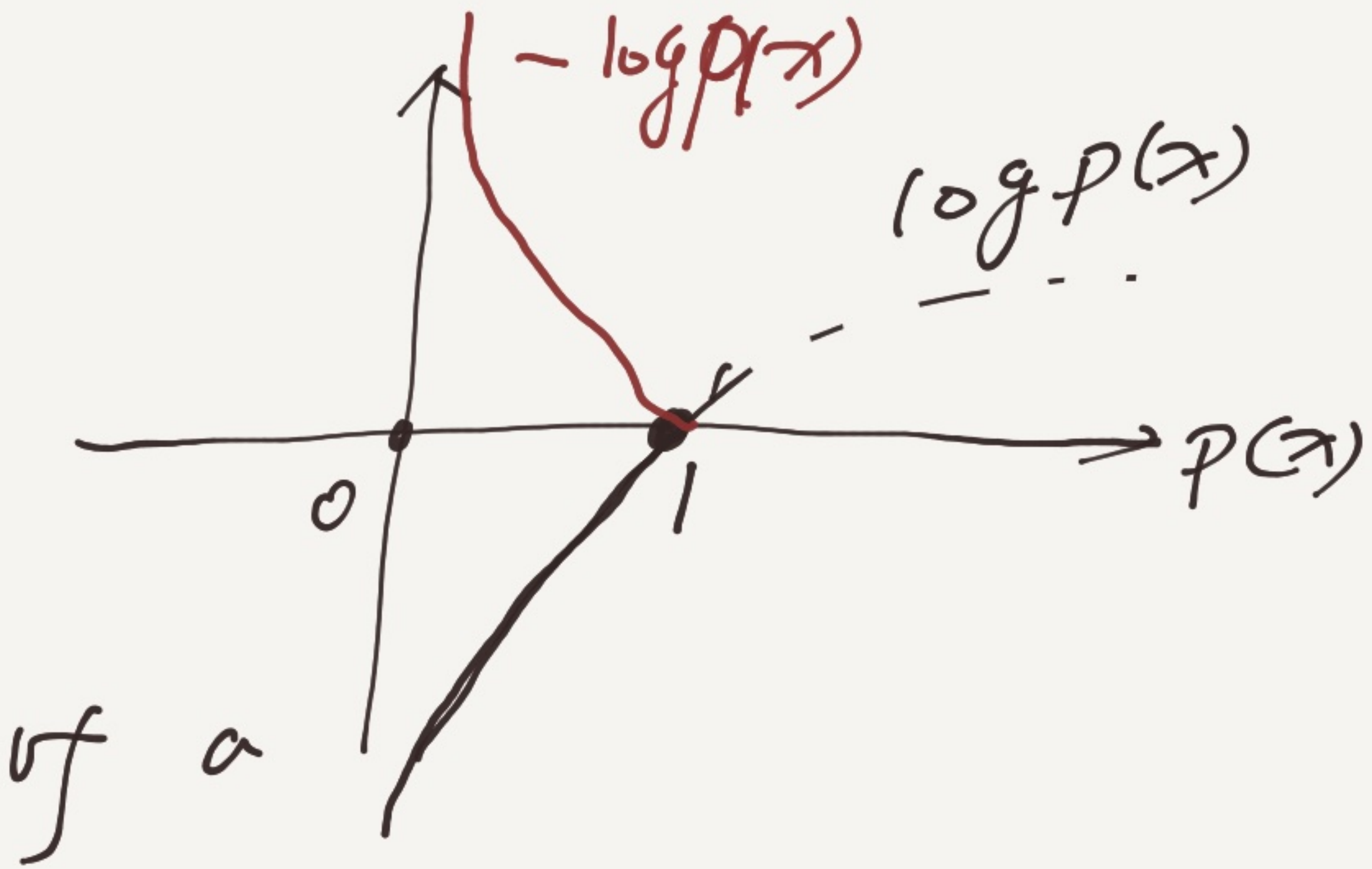
① bounded: $sigm(x) \in (0, 1)$

② smooth, continuous and have derivatives at any point.

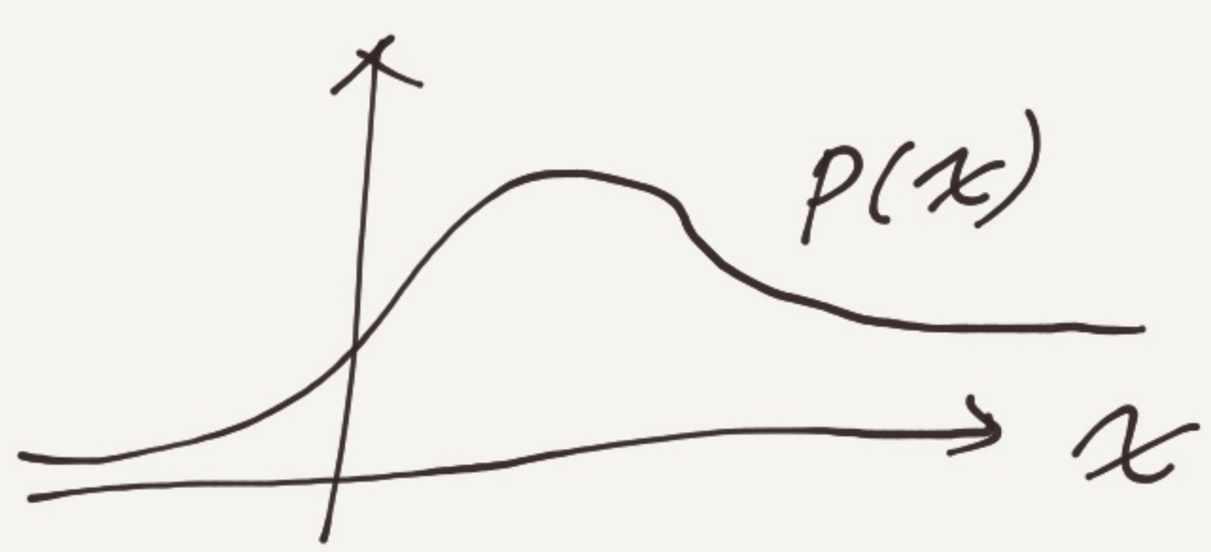
③ $D(sigm(x)) = D((1+e^{-x})^{-1}) = \frac{e^{-x}}{(1+e^{-x})^2} = sigm(x)\cdot(1-sigm(x))$

3. Self-information measures the amount of information of a prob.

$p(x) \in [0, 1]$ $\quad I(x) = -\log p(x)$



4. Shanon entropy measures the information of a distribution



$$H(x) = E_{x \sim p(x)} (I(x))$$

$$= \sum_i p(x_i) I(x_i)$$

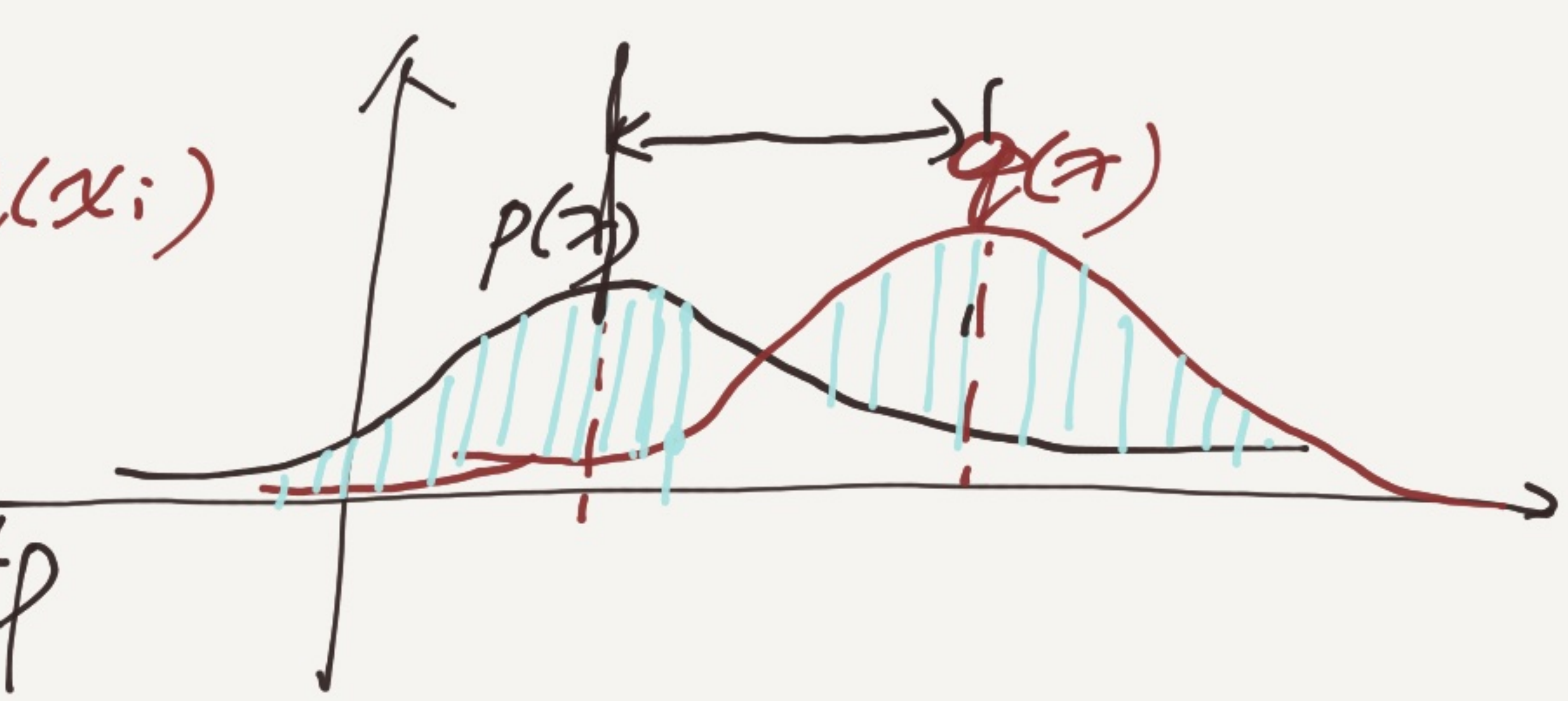
$$= -\sum_i p(x_i) \cdot \log p(x_i)$$

5. Cross-entropy: difference between two distributions

1) $\int (p(x) - q(x))^2 dx$ $\quad H(p,q) = -\sum_i p(x_i) \cdot \log q(x_i)$

2) Correlation ??

3) distance between the peaks $d = \mu_q - \mu_p$



6. KL-divergence (Kullback-Leibler) measures the divergence of two distributions

$KL(p\|q)$ divergence from p to q. (relative entropy)

$$= \sum_i p(x_i) \cdot \log \frac{p(x_i)}{q(x_i)}$$

$\log \frac{1}{x} = -\log x$

$$= -\sum_i p(x_i) \cdot \log \frac{q(x_i)}{p(x_i)}$$

$\log \frac{x}{y} = \log x - \log y$

$$= \underbrace{-\sum_i p(x_i) \cdot \log q(x_i)} - \underbrace{\left[-\sum_i p(x_i) \log p(x_i)\right]}$$

$$= H(p, q) - H(p)$$

Jensen-Shannon Divergence:

$D_{JS}(p\|q) = \frac{1}{2} KL(p\|m) + \frac{1}{2} KL(q\|m)$

$(m = \frac{1}{2}(p+q))$

symmetric version,

1) $KL(p\|q) \geq 0$ non-negative

$KL(p\|q) = 0$ if and only if $p = q$.

2) KL is asymmetric: $KL(p\|q) \neq KL(q\|p)$

↳ KL is not a distance metric.