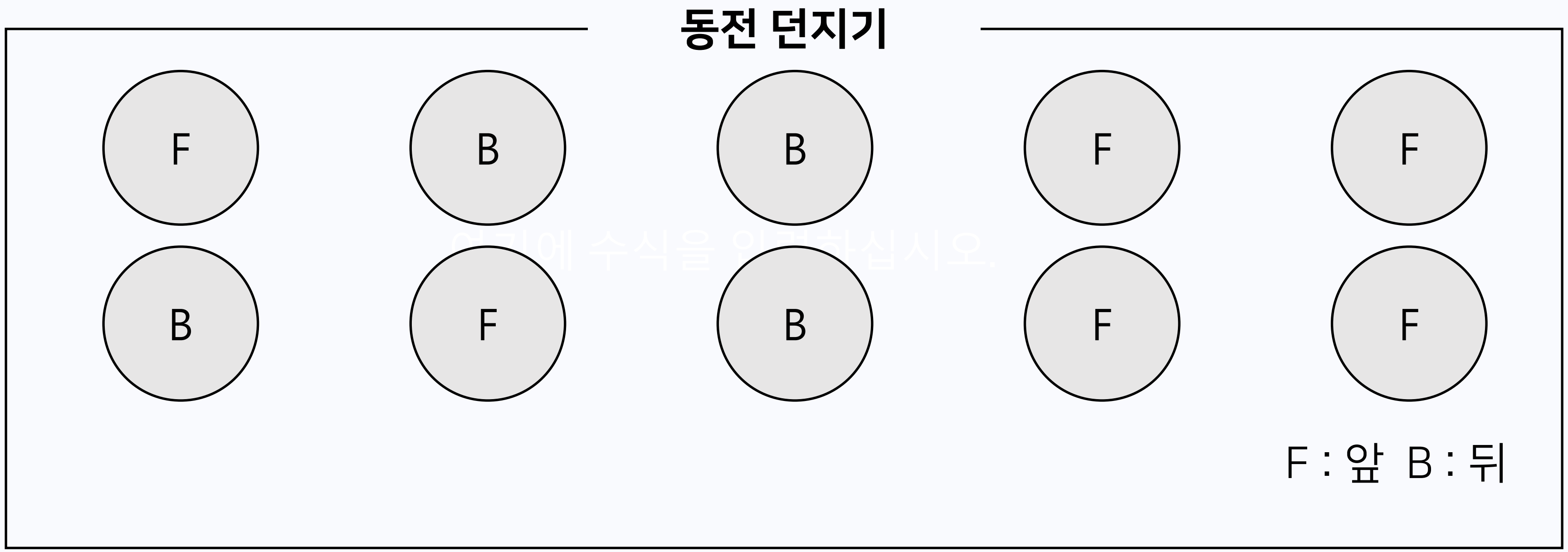
Hoeffding inequality

정영석

Contents

Frequency

- **동전 던지기!**



$P(x_f)=0.6?$

Contents

Frequency

- **P.A.C (Probably Approximate Correct)**

✓ μ : 어떤 확률변수 x의 ' 모든 ' 사건에 대한 frequency (동전 앞면 확률 : ½)

✓ $\hat{\mu}$: 어떤 확률변수 x의 '일부' 사건에 대한 frequency (동전 앞면 확률 : ¾(?))

↓

$\hat{\mu}$와 μ는 같다고 표현할 수 없지만 아마도 대략적으로 근사 한다고 할 수 있음.

↓

어떤 확률 변수 x에 대해 sample의 크기를 얼마정도로 해야 전체에 대한 확률을 근사할까?
(Hoeffding inequality)

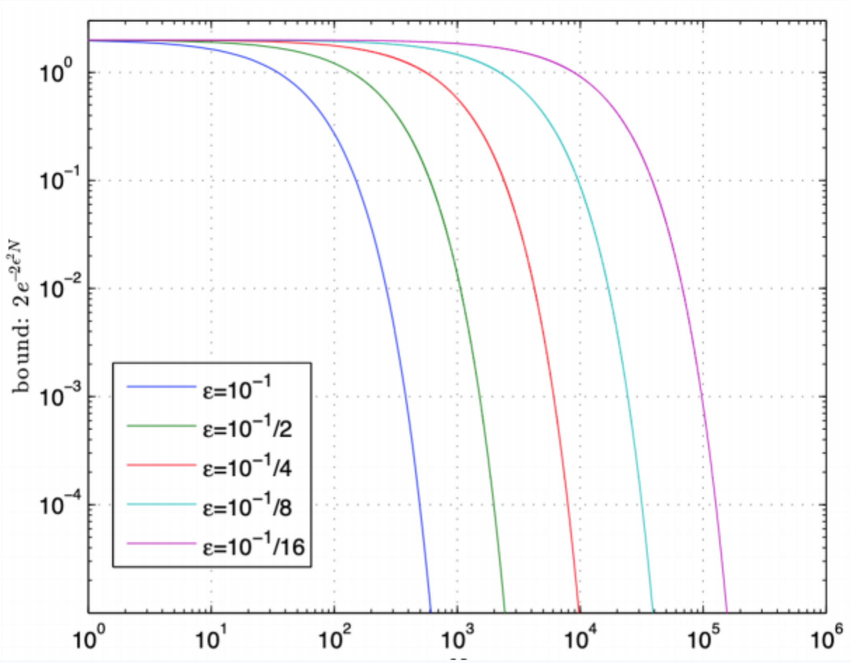
Contents

Heffding

- **정의**

$P(|\hat{\mu} - \mu| > \epsilon) \leq \ 2e^{-2\epsilon^2 N}$ (ϵ : margin, N : # of samples)

✓ P : 어떤 확률변수에 대한 확률



Contents

Hoeffding

- **동전던지기**

  ✓ 동전던지기를 몇 번 해야 1/2에 근사할까?

$$P(|\hat{\mu} - \mu| > \epsilon) \leq 2e^{-2\epsilon^2 N}$$

  - N : 1000, ϵ : 0.05 → 99%

$$\mu - 0.05 \leq \hat{\mu} \leq \mu + 0.05$$

  - N : 500, ϵ : 0.05 → 84%

$$\mu - 0.05 \leq \hat{\mu} \leq \mu + 0.05$$

Contents

Hoeffding

- **기계학습에서의 Hoeffding Inequality**

✓ 사건에 대해 μ를 알아내는 것은 어려움

✓ 사건에 대해 X→Y mapping하는 함수 f를 찾는 것 역시 어려움.

∴ 이를 보완하기 위해 확률변수를 어떤 hypothesis(h)에 대한 에러율로 치환함

- True $\text{if } f(x) \, ! = \; h(x)$
- False $\text{if } f(x) == \; h(x)$

Contents

Hoeffding

- **기계학습에서의 Hoeffding Inequality**

✓ $\hat{\mu} = \mathrm{E_{in}(h)}$

✓ $\mu = \mathrm{E_{out}(h)}$

$$\mathrm{P}(|\mathrm{E_{in}(h)} - \mathrm{E_{out}(h)}| > \epsilon) \leq \ 2\mathrm{e}^{-2\epsilon^2 \mathrm{N}}$$

❖ N이 증가하면, $\mathrm{E_{in}(h)}$이 $\mathrm{E_{out}(h)}$을 근사할 확률이 높아짐

→ N이 충분할 경우, sample에 대한 에러를 최소로 하는 h는 out of sample 에 대한 오류도 최소로 할 확률이 높아짐.

Hoeffding

- **기계학습에서의 Hoeffding Inequality**

$$
\begin{aligned}
|\mathrm{E}_{\text{in}}(g) - \mathrm{E}_{\text{out}}(g)| > \epsilon \quad \Longrightarrow \quad & |\mathrm{E}_{\text{in}}(h_1) - \mathrm{E}_{\text{out}}(h_1)| > \epsilon \\
\text{or } & |\mathrm{E}_{\text{in}}(h_2) - \mathrm{E}_{\text{out}}(h_2)| > \epsilon \\
& \cdots \\
\text{or } & |\mathrm{E}_{\text{in}}(h_M) - \mathrm{E}_{\text{out}}(h_M)| > \epsilon
\end{aligned}
$$

implication

$$
\begin{aligned}
\mathbb{P}[|\mathrm{E}_{\text{in}}(g) - \mathrm{E}_{\text{out}}(g)| > \epsilon] \quad \leq \quad \mathbb{P}[\quad & |\mathrm{E}_{\text{in}}(h_1) - \mathrm{E}_{\text{out}}(h_1)| > \epsilon \\
\text{or } & |\mathrm{E}_{\text{in}}(h_2) - \mathrm{E}_{\text{out}}(h_2)| > \epsilon \\
& \cdots \\
\text{or } & |\mathrm{E}_{\text{in}}(h_M) - \mathrm{E}_{\text{out}}(h_M)| > \epsilon]
\end{aligned}
$$

Hoeffding

- 기계학습에서의 Hoeffding Inequality

$$\begin{aligned}\mathbb{P}[|\mathrm{E}_{\mathrm{in}}(g) - \mathrm{E}_{\mathrm{out}}(g)| > \epsilon] \quad \leq \quad \mathbb{P}[\quad & |\mathrm{E}_{\mathrm{in}}(h_1) - \mathrm{E}_{\mathrm{out}}(h_1)| > \epsilon \\ \text{or } & |\mathrm{E}_{\mathrm{in}}(h_2) - \mathrm{E}_{\mathrm{out}}(h_2)| > \epsilon \\ & \cdots \\ \text{or } & |\mathrm{E}_{\mathrm{in}}(h_M) - \mathrm{E}_{\mathrm{out}}(h_M)| > \epsilon]\end{aligned}$$

Union Bound

$$\mathrm{P}(|\mathrm{E}_{\mathrm{in}}(\mathrm{g}) - \mathrm{E}_{\mathrm{out}}(\mathrm{g})| > \epsilon) \leq \sum P(|E_{\mathrm{in}}(\mathrm{h_m}) - \mathrm{E}_{\mathrm{out}}(\mathrm{h_m})| > \epsilon)$$

$$\leq 2\mathrm{M}\mathrm{e}^{-2\epsilon^2 \mathrm{N}}$$

M : # of hypothesis,
complexity of the Optimal hypothesis

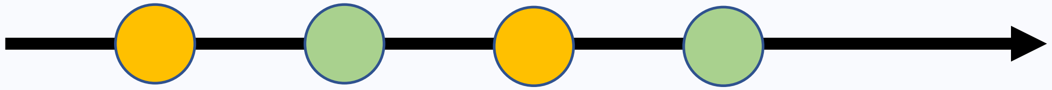
Contents

VC

- **VC(Vapnik-Chervonenkis) Dimension**

  ✓ 정의 : 모델이 분류 가능한 경우의 수(최대 데이터의 개수).

    ❖ 선형 모델의 경우 차원수 + 1개의 데이터의 모든 경우에 대해 구분이 가능함



Hoeffding

- **Complexity관점의 Hoeffding Inequality**

$$P(|E_{in}(h) - E_{out}(h)| > \epsilon) \leq 2e^{-2\epsilon^2 N}$$

✓ M ↑ : $E_{in}(h)$ ↓ , $E_{out}(h)$ ↑

✓ M ↓ : $E_{in}(h)$ ↑ , $E_{out}(h)$ ↓