Sequence to Sequence: Encoder
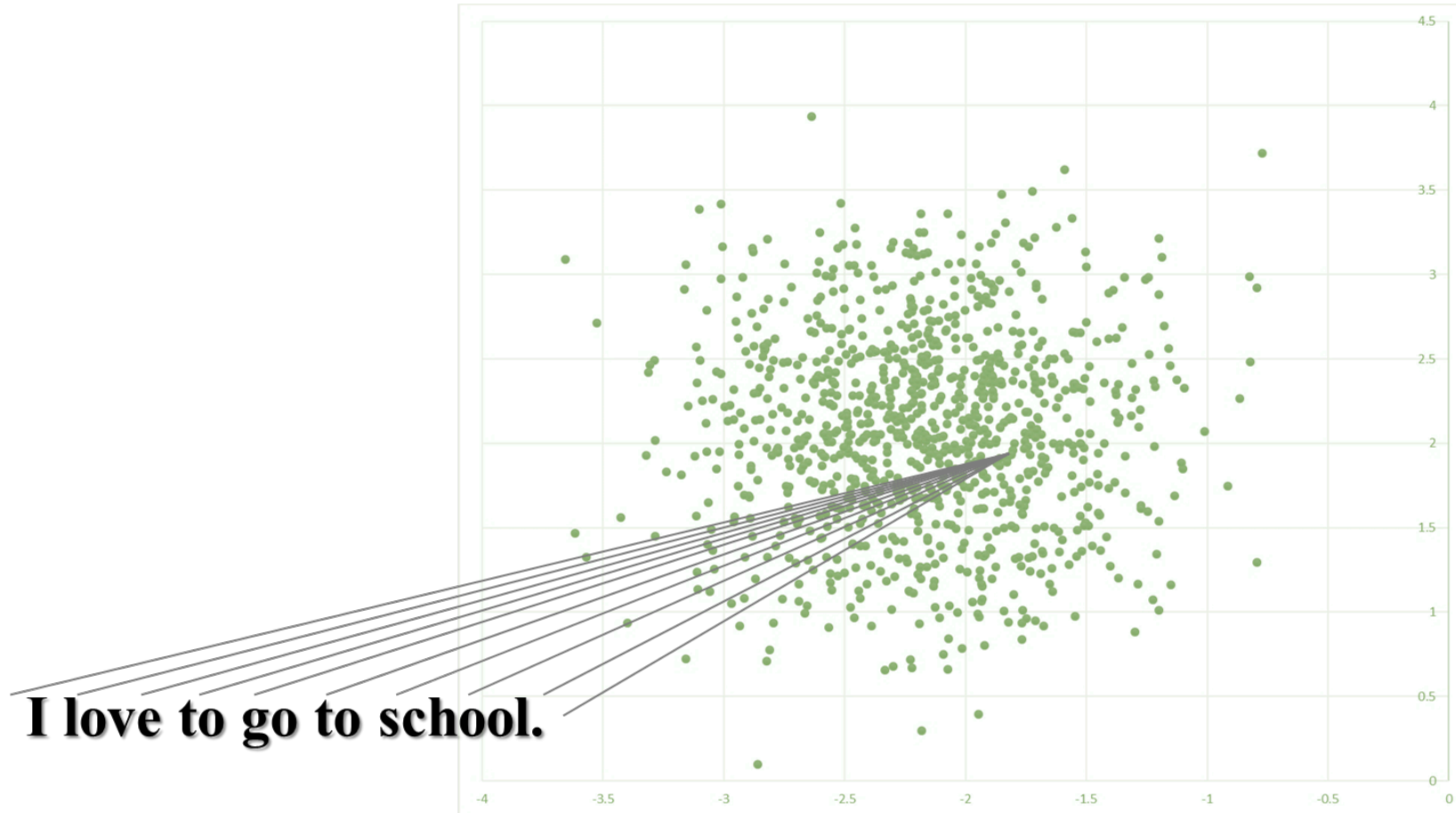
Ki Hyun Kim

nlp.with.deep.learning@gmail.com
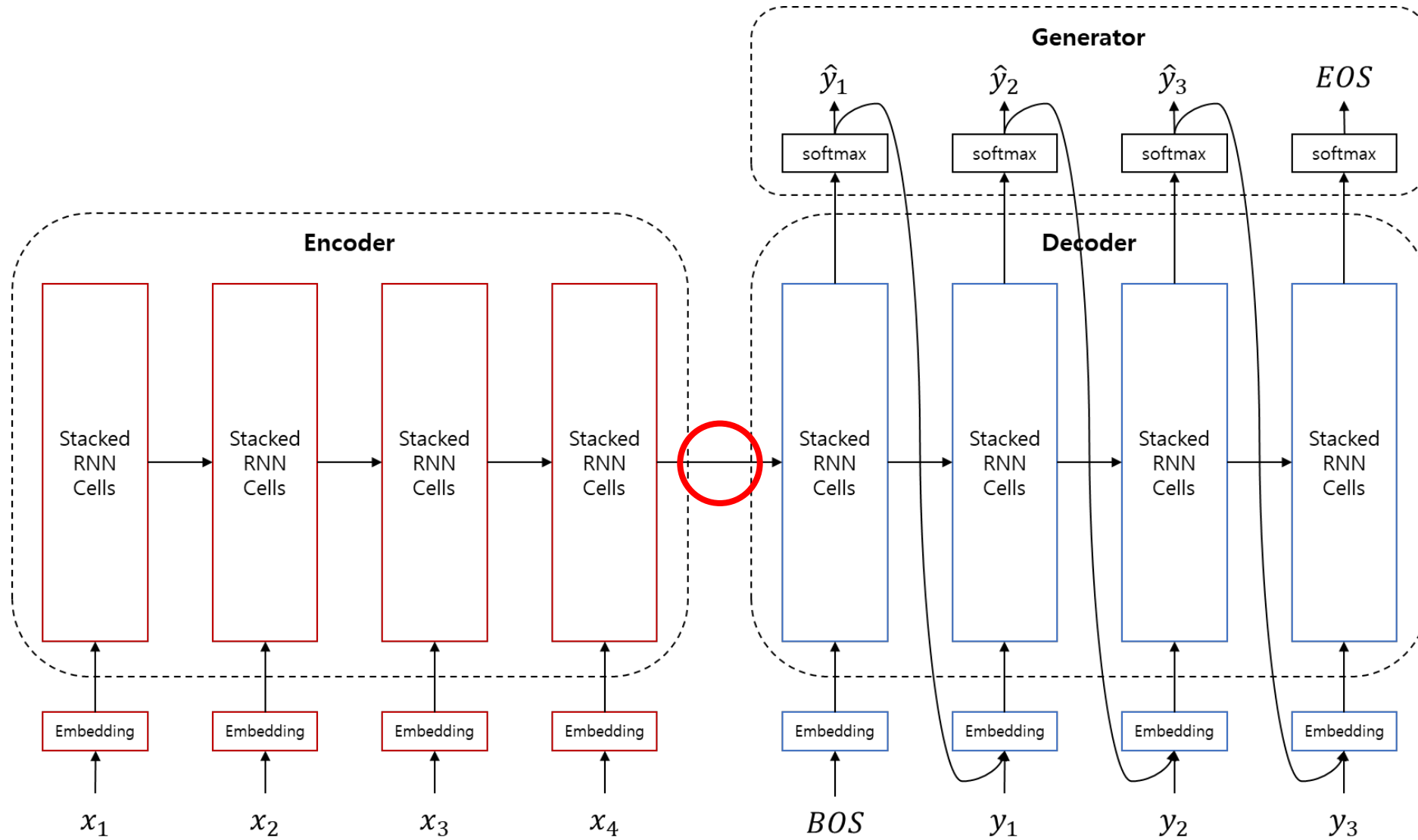
Fast campus

Sequence to Sequence

Encoder

I love to go to school.
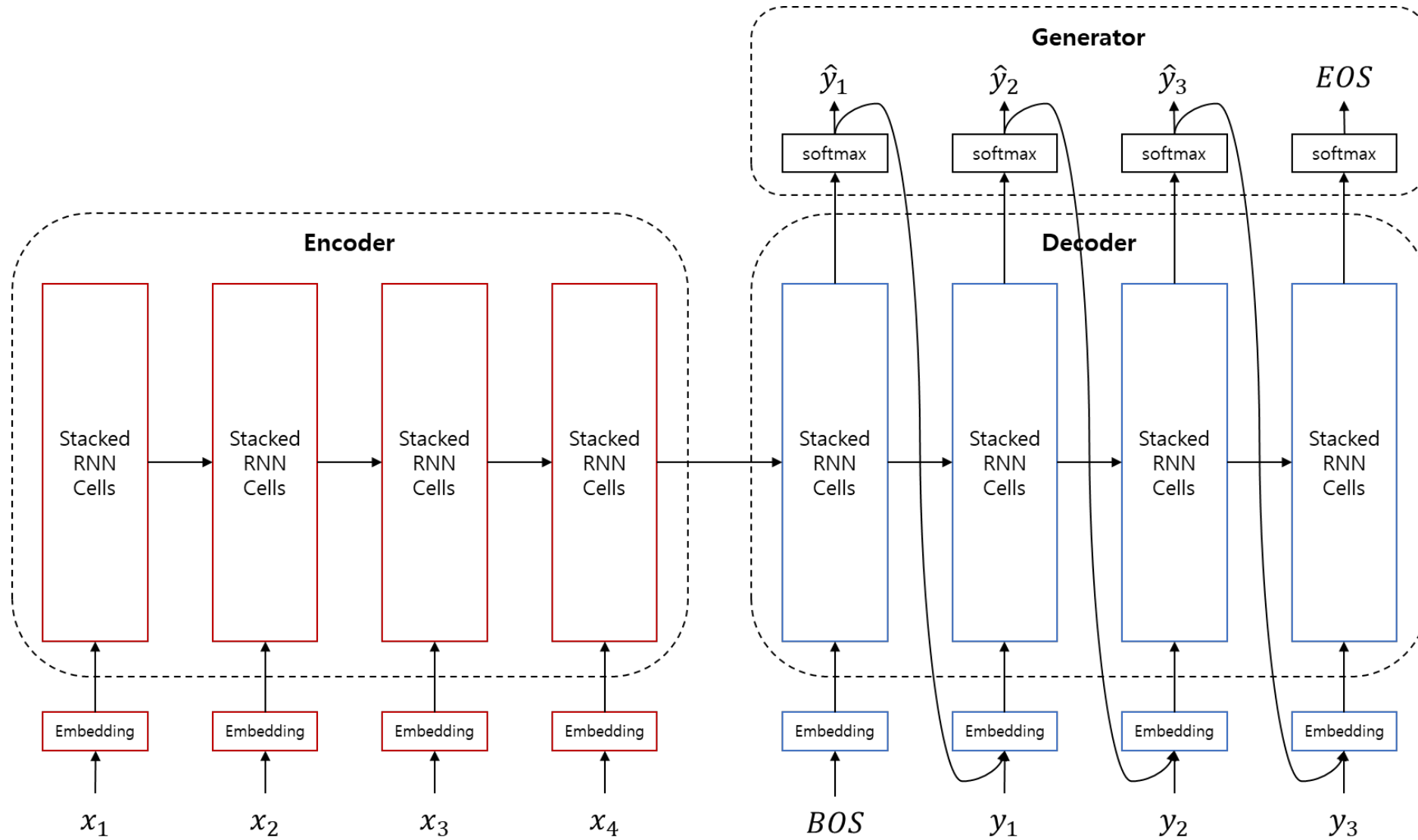
Sequence to Sequence

Equations

- Given dataset,

$$\mathcal{D} = \{x^i, y^i\}_{i=1}^N$$

$$x^i = \{x_1^i, \cdots, x_m^i\} \text{ and } y^i = \{y_0^i, y_1^i, \cdots, y_n^i\},$$

$$\text{where } y_0 = <\text{BOS}> \text{ and } y_n = <\text{EOS}>.$$

- Get hidden states of encoder

$$h_t^{\text{enc}} = \text{RNN}_{\text{enc}}(\text{emb}_{\text{enc}}(x_t), h_{t-1}^{\text{enc}}), \text{ where } h_0^{\text{enc}} = 0.$$

$$h_{1:m}^{\text{enc}} = [h_1^{\text{enc}}; \cdots; h_m^{\text{enc}}],$$

$$\text{where } h_t^{\text{enc}} \in \mathbb{R}^{\text{batch\_size} \times 1 \times \text{hidden\_size}} \text{ and } h_{1:m}^{\text{enc}} \in \mathbb{R}^{\text{batch\_size} \times m \times \text{hidden\_size}}.$$

- If we use bi-directional RNN,

$$h_t^{\text{enc}} \in \mathbb{R}^{\text{batch\_size} \times 1 \times (2 \times \text{hidden\_size})} \text{ and } h_{1:m}^{\text{enc}} \in \mathbb{R}^{\text{batch\_size} \times m \times (2 \times \text{hidden\_size})}.$$

Fast campus

Summary

- Encoder는 <u>source 문장을 압축한 context vector</u>를 decoder에게 넘겨준다.

- Encoder는 train/test 시에 항상 문장 전체를 받음
  - Encoder 자체만 놓고 보면 non-auto-regressive task.
  - 따라서 <u>bi-directional RNN 사용</u> 가능