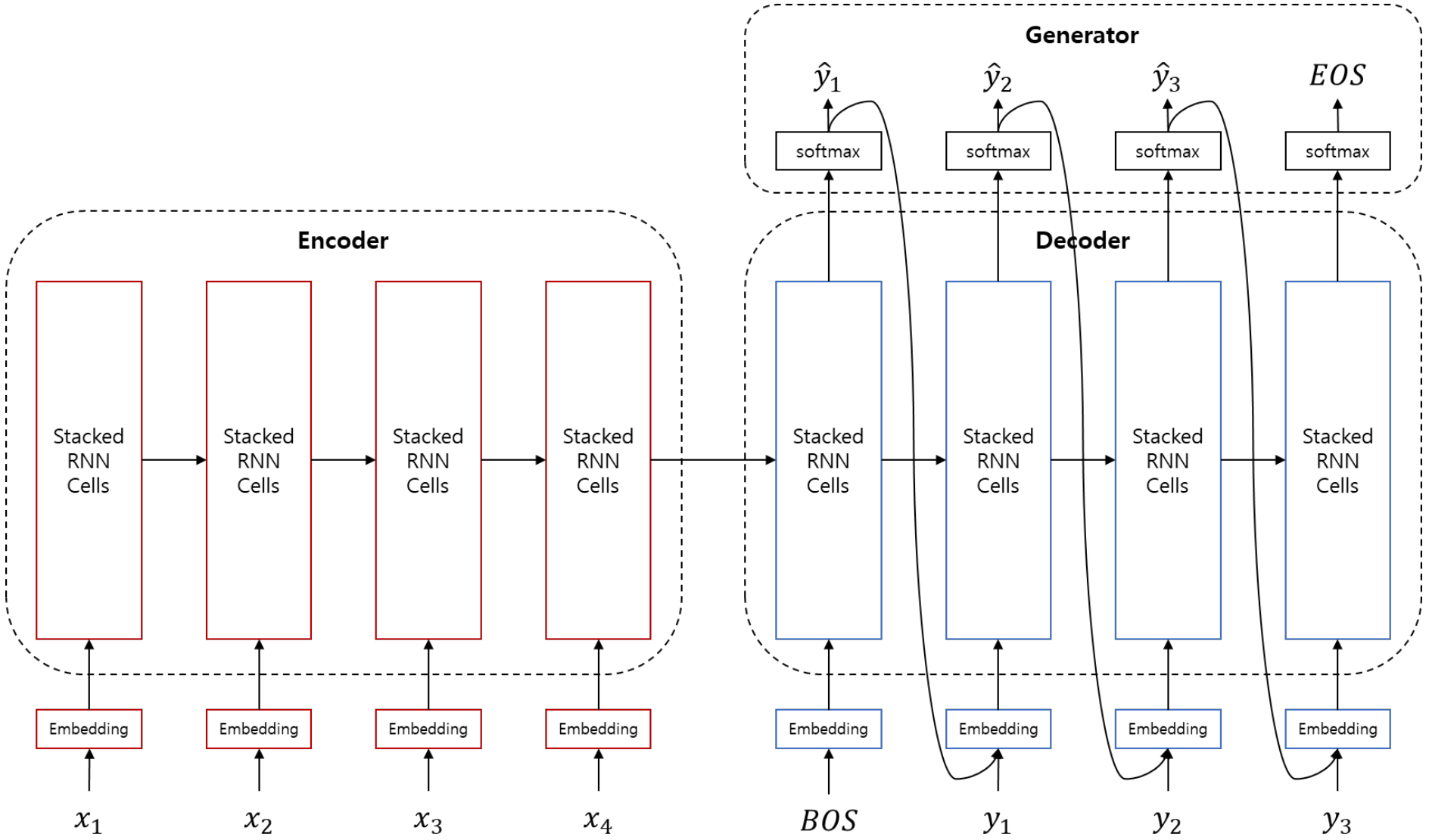
Sequence to Sequence: Decoder

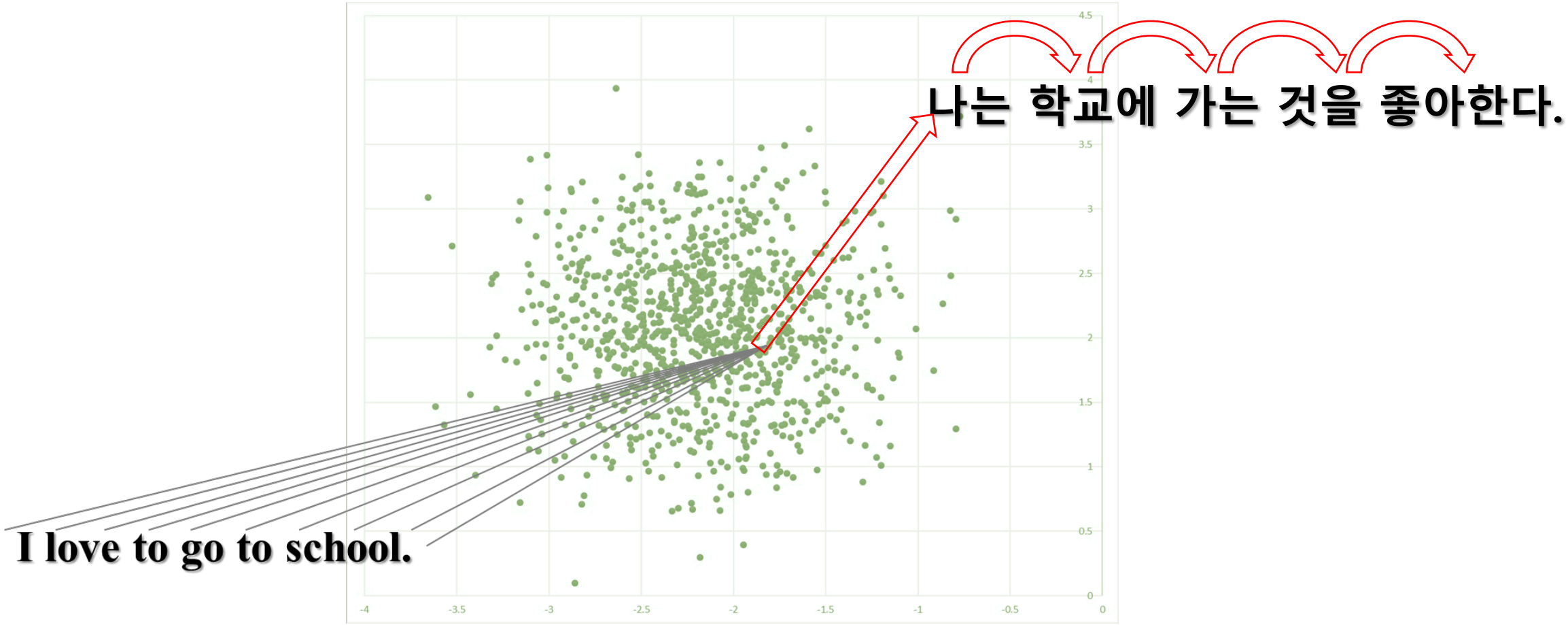
Ki Hyun Kim

nlp.with.deep.learning@gmail.com

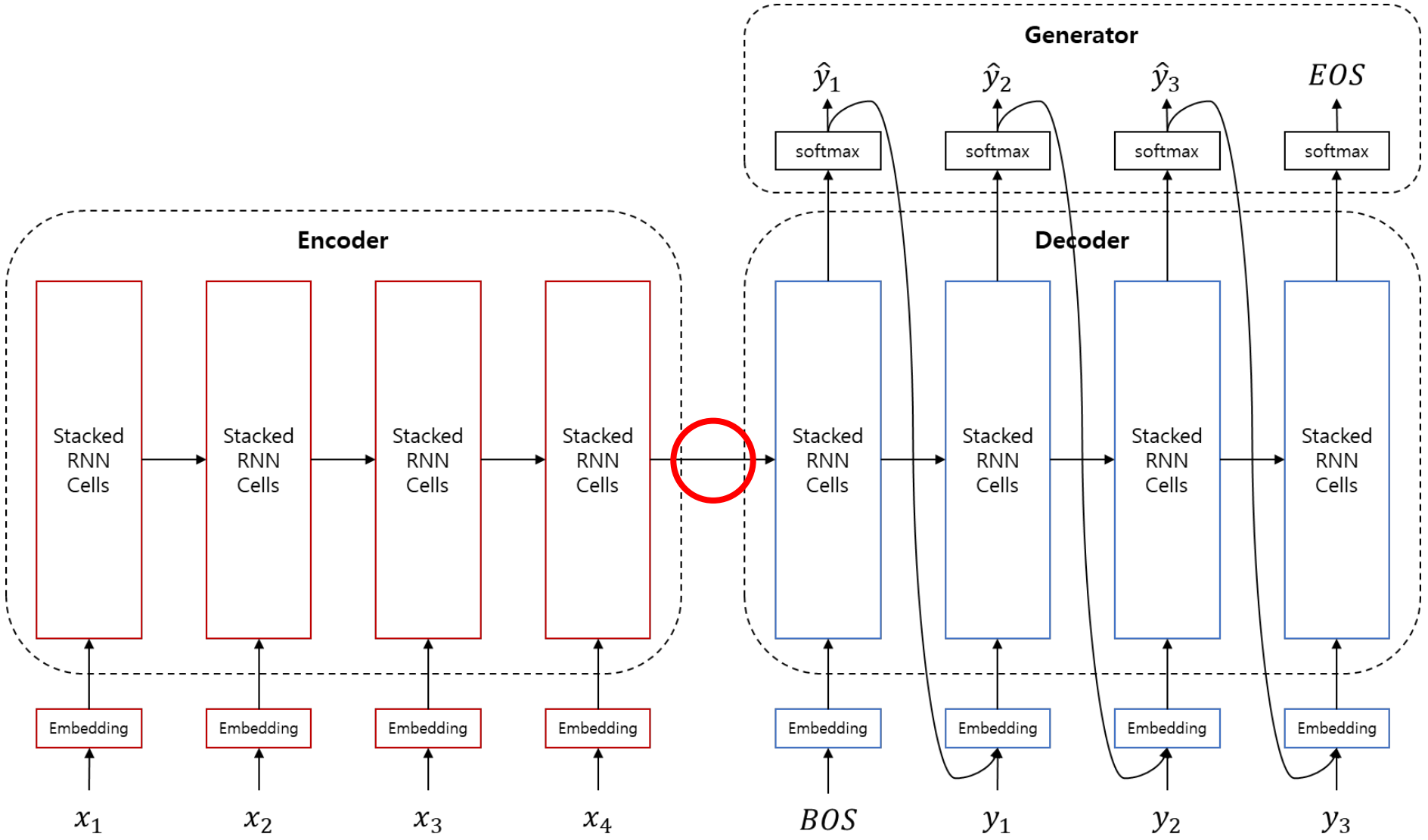
Sequence to Sequence



Decoder



Sequence to Sequence



Equations

- Given dataset,

$$\mathcal{D} = \{x^i, y^i\}_{i=1}^N$$
$$x^i = \{x_1^i, \cdots, x_m^i\} \text{ and } y^i = \{y_0^i, y_1^i, \cdots, y_n^i\},$$
$$\text{where } y_0 = <\text{BOS}> \text{ and } y_n = <\text{EOS}>.$$

- We can get hidden state of decoder,

$$h_t^{\text{dec}} = \text{RNN}_{\text{dec}}(\text{emb}_{\text{dec}}(\hat{y}_{t-1}), h_{t-1}^{\text{dec}}),$$
$$\text{where } h_0^{\text{dec}} = h_m^{\text{enc}}.$$

$$h_{1:n}^{\text{dec}} = [h_1^{\text{dec}}; \cdots; h_n^{\text{dec}}]$$

Summary

- 디코더는 conditional language model이라고 볼 수 있음
  - 인코더로부터 문장을 압축한 context vector를 바탕으로 문장을 생성

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^{N} \log P(y^i | x^i; \theta)$$

$$= \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{j=1}^{n} \log P(y_j^i | x^i, y_{<j}^i; \theta)$$

- Auto-regressive task에 속하므로, uni-directional RNN을 사용