

Wrap-up: Text Classification

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

In Probabilistic Perspective,

- 문장이 주어졌을 때, 문장이 속할 클래스의 확률 분포 함수를 approximate.

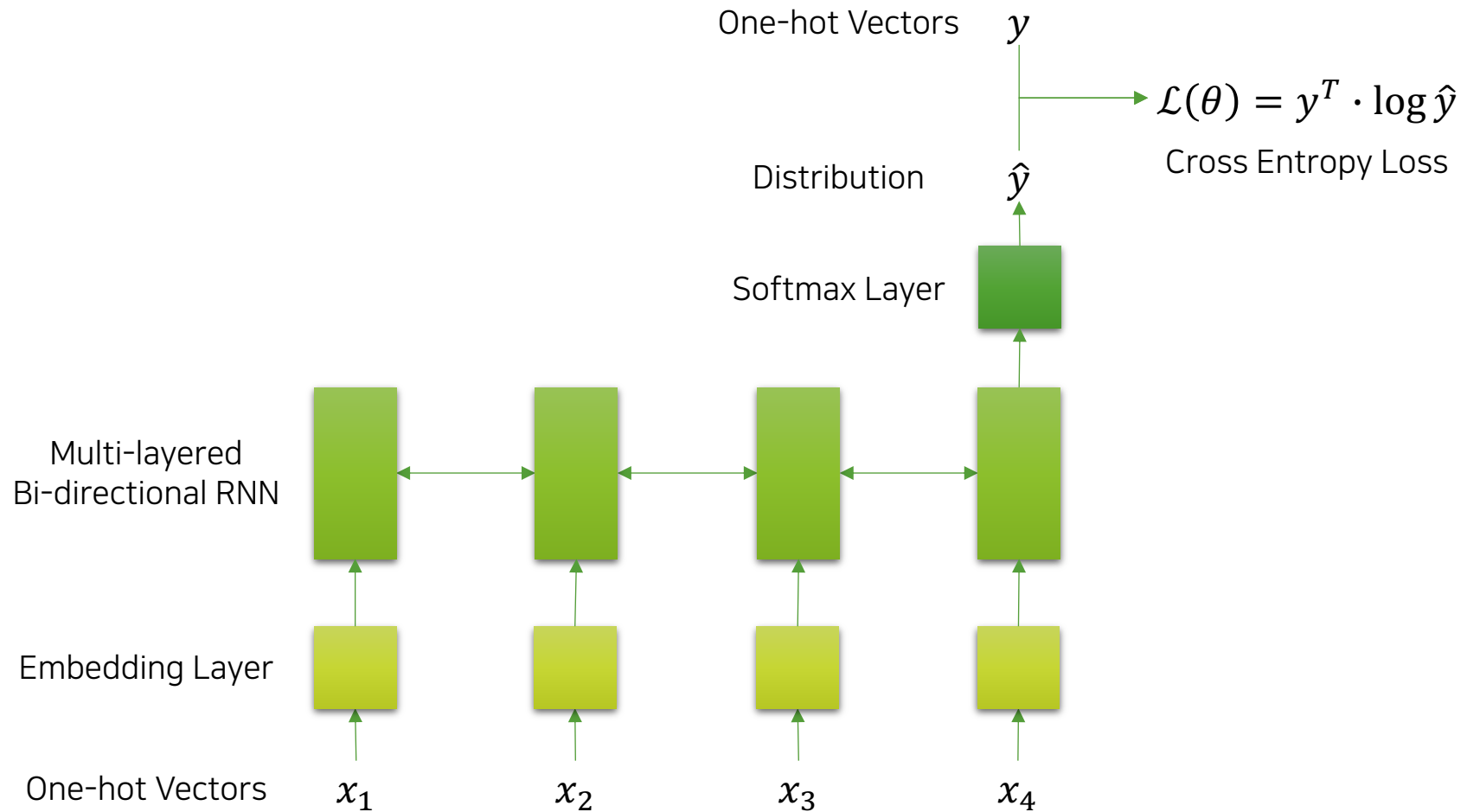
$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_{\mathbf{x} \sim P(\mathbf{x})} \left[\mathbb{E}_{y \sim P(y|\mathbf{x})} [\log P(y|\mathbf{x}; \theta)] \right]$$

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$$

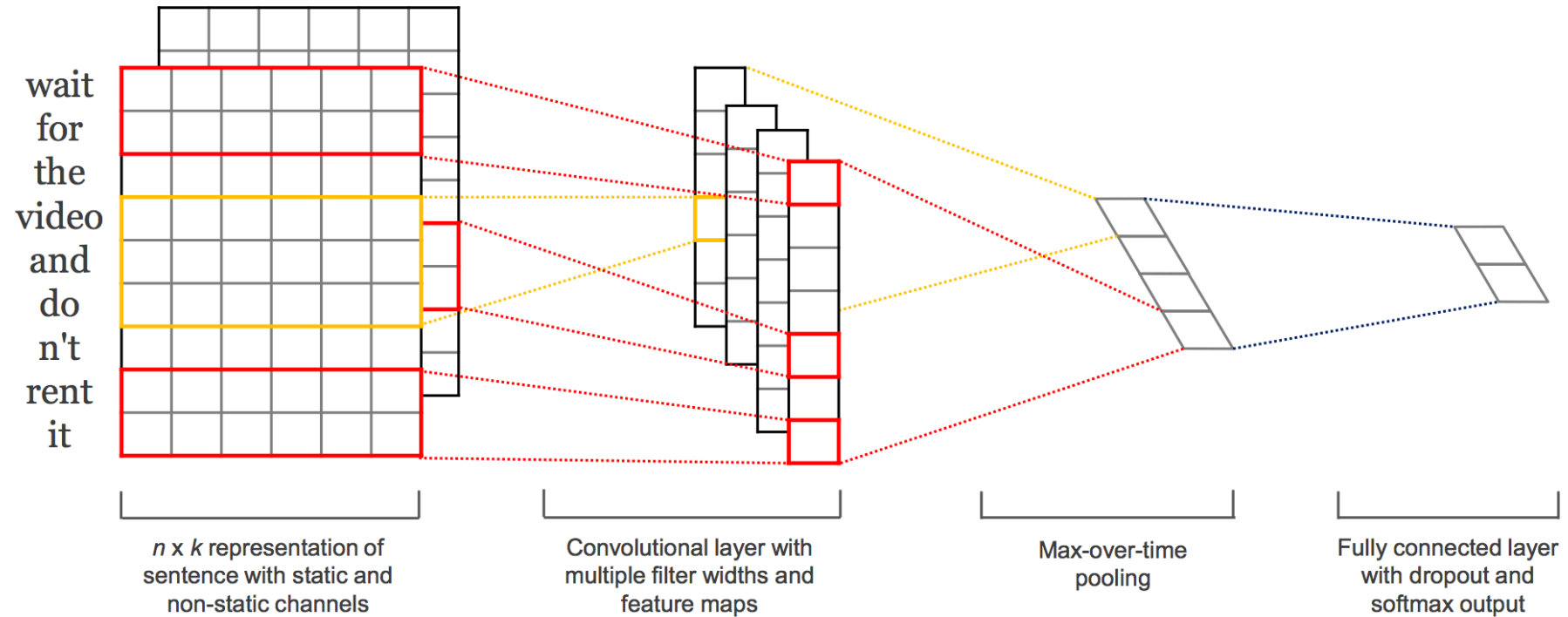
$$\begin{aligned} \hat{\theta} &= \operatorname{argmax}_{\theta \in \Theta} \sum_{i=1}^N \log P(y_i | x_i; \theta) \\ &= \operatorname{argmin}_{\theta \in \Theta} - \sum_{i=1}^N \log P(y_i | x_i; \theta) \end{aligned}$$

$$\begin{aligned} \mathcal{L}(\theta) &= - \sum_{i=1}^N \log P(y_i | x_i; \theta) \\ \theta &\leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}(\theta) \end{aligned}$$

How the Classifier Works



Text Classification with CNN [Kim, 2014]



Conclusion

- 신경망은 텍스트를 입력으로 받아 context vector로 인코딩
 - RNN의 경우, 단어의 출현 여부와 순서에 따른 정보를 종합적으로 활용
 - CNN의 경우, 문구의 출현 여부를 종합적으로 활용
- Context vector는 y 를 예측하기 위한 정보를 담고 있을 것
- 자연어 생성 미리보기:
 - Sequence-to-Sequence encoder의 경우, 문장을 생성하기 위한 context vector 생성