

Greedy Decoding

- In each step, the model computes a distribution over the vocabulary V (given source \mathbf{x} , the previous outputs h , and the model parameters θ).

$$p(y|h) = g(\mathbf{x}, h, \theta)$$

- In greedy decoding:

$$y^* = \operatorname{argmax}_{y \in V} p(y|h)$$