deeplearning.ai

# Sequence to sequence models

---

# Refinements to beam search

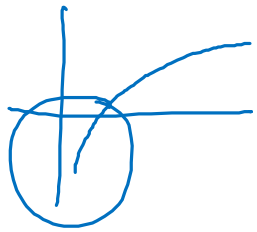
# Length normalization

$$\arg\max_y \prod_{t=1}^{T_y} P(y^{<t>} | x, y^{<1>}, \ldots, y^{<t-1>})$$

$$\arg\max_y \sum_{t=1}^{T_y} \log P(y^{<t>} | x, y^{<1>}, \ldots, y^{<t-1>})$$

$$\frac{1}{T_y^{\alpha}} \sum_{t=1}^{T_y} \log P(y^{<t>} | x, y^{<1>}, \ldots, y^{<t-1>})$$

$\alpha = 0.7$ $\alpha = 1$ $\alpha = 0$



Andrew Ng

# Beam search discussion

large B: better result, slower

small B: worse result, faster

Beam width B?

1 -> 3 -> 10, 100, 1000, -> 3000

Unlike exact search algorithms like BFS (Breadth First Search) or DFS (Depth First Search), Beam Search runs faster but is not guaranteed to find exact maximum for $\arg\max_y P(y|x)$.

Andrew Ng