LM evaluation using perplexity

17

$$PPL(\mathbf{X}^{(i)}) = \exp\left\{ -\frac{1}{T}\sum_{t=1}^{T} \log P\left(\mathbf{x}^{(t)}{}_i \,|\, \mathbf{x}^{(t-n)}{}_i \ldots \mathbf{x}^{(t-1)}{}_i ; \boldsymbol{\theta} \right) \right\}$$

$$= \exp\left\{ H(P_D, P_{\boldsymbol{\theta}}) \right\}$$

Language modeling (LM) Leaderboard

LM evaluation using perplexity

$$PPL(\mathbf{X}^{(i)}) = \exp\left\{ -\frac{1}{T}\sum_{t=1}^{T} \log P\left(\mathbf{x}^{(t)}{}_i \,|\, \mathbf{x}^{(t-n)}{}_i \ldots \mathbf{x}^{(t-1)}{}_i ; \boldsymbol{\theta} \right) \right\}$$
$$= \exp\left\{ H(P_D, P_{\boldsymbol{\theta}}) \right\}$$

Language modeling (LM) Leaderboard

Transfer learning

- Neural language modeling provides a means to extract semantically rich features from text data. How is this then applied towards downstream tasks?
- Transfer learning:
  - LMs are first trained on next-word or next-sentence prediction tasks on a very large corpus of text (e.g., wikitext 103).
  - The output layer, which is used to map onto the word prediction output space, is removed, and replaced with a randomly initialized projection matrix mapping to the output space of the particular prediction task of interest. It is trained on a much lower number of examples from that task.