

1. 損失関数は次式のとおりとする (スライド p.16)

$$E(\mathbf{W}, \mathbf{W}') = -\log L = -\sum_{j=1}^V t_j \left(u_j - \log \sum_{j'=1}^V \exp(u_{j'}) \right)$$

次式の導出の過程を示せ。

$$\frac{\partial E}{\partial u_j} = -(t_j - y_j)$$

$$\frac{\partial}{\partial u_j} \log \sum_{j'} \exp(u_{j'}) = \frac{\exp(u_j)}{\sum_{j'} \exp(u_{j'})} = y_j \text{ を使う.}$$

2. 単語の one-hot-vector の平均でコンテキストを表現する場合と、埋め込みベクトルの平均で表現する場合を次の観点を盛り込んで比較して論じなさい。

- one-hot-vector の平均は、コンテキスト中の単語の出現頻度を表す。
- \mathbf{x}_i と \mathbf{x}_j ($i \neq j$) は意味に依らず、全てが互いに直交しており、互いの距離は $\sqrt{2}$ で一定
 - 初めて見る文章のコンテキストと似たコンテキストを学習データ中から探すことができるだろうか。
- 一方の単語の埋め込みベクトル \mathbf{h} は方向が意味を反映している。
- one-hot-vector を単語表現に使うと、コンテキスト間の類似度評価が極めて困難になりそうであることを説明する。

- The boy has a pen.
- The boy has a pencil.
- The boy has a spoon.
- The boy has a bucket.
- The girl has a pen.
- The king has a pen.
- The rabbit has a pen.

いびつ

$$\|\mathbf{x}_{x1} - \mathbf{x}_{xj}\| = 1$$

持っているもの、
主語の違いを評価できない。

例えば、文脈に依る識別器を構築すると
 $P(y | \mathbf{x}_{ctx})$ を推定する必要が有る。

$\mathbf{x}_{ctx} = \sum \mathbf{x}_i$ が one-hot vector の
得られただけ。 $\|\mathbf{x}_{ctx1} - \mathbf{x}_{ctx2}\|$ が
文脈の類似度を反映しないため。

$P(y | \mathbf{x}_{ctx})$ の推定は \mathbf{x}_{ctx} が完全に一致
する学習文脈を必要としない。