

lec1

1. Word2Vec 目标函数

(1) 似然函数

$$L(\theta) = \prod_{t=1}^T \prod_{-m \leq j \leq m, j \neq 0} P(w_{t+j} | w_t; \theta)$$

(2) 负对数似然损失函数

$$J(\theta) = -\frac{1}{T} \log L(\theta) = -\frac{1}{T} \sum_{t=1}^T \sum_{-m \leq j \leq m, j \neq 0} \log P(w_{t+j} | w_t; \theta)$$

2. Softmax 条件概率

$$P(o | c) = \frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)}$$

为什么这样记？

1. $u_o^T v_c = uv = \sum u_i v_i$, T 是转置
2. \exp 取指数把里面的 uv 转化为正数
3. 概率的范围是 $[0, 1]$, 分母是所有的可能性, 归一化常数

其中:

- (c) : 中心词(变量, 对应向量 (v_c))
- (o) : 上下文词(固定观测值, 对应向量 (u_o))
- (V) : 词汇表(常数集合)

3. 记下来计算梯度(对 (v_c) 求偏导)

(1) 展开对数概率

$$\frac{\partial J}{\partial v_c} = \frac{\partial}{\partial v_c} \left[-\log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)} \right] = -\frac{\partial}{\partial v_c} \left[u_o^T v_c - \log \sum_w \exp(u_w^T v_c) \right]$$

(2) 计算第一项梯度

$$\frac{\partial}{\partial v_c} (u_o^T v_c) = u_o$$

(3) 计算第二项梯度

应用链式法则, 注意 (v_c) 是变量, 其他是常数, $\log(x)$ 的导数是 $\frac{1}{x}$

$$\begin{aligned} \frac{\partial}{\partial v_c} \left(\log \sum_w \exp(u_w^T v_c) \right) &= \frac{\sum_w \exp(u_w^T v_c) u_w}{\sum_w \exp(u_w^T v_c)} \\ &= \sum_w \left(\frac{\exp(u_w^T v_c)}{\sum_w \exp(u_w^T v_c)} \right) u_w, \text{ 注意上文的softmax函数的定义带入这里} \\ &= \sum_w P(w | c) u_w \end{aligned}$$

(4) 合并结果

$$\frac{\partial J}{\partial v_c} = -1 \times (u_o - \sum_w P(w | c) u_w)$$

4. 物理意义

- $(-u_o)$: 真实上下文词的方向(正向推动)
- $(\sum_w P(w | c) u_w)$: 模型预测的期望方向(负向修正)

5. 参数更新规则

$$v_c \leftarrow v_c - \eta \left(-u_o + \sum_w P(w | c) u_w \right)$$

其中 (η) 为学习率.