1. Word2Vec 目标函数

(1) 似然函数

$$L(heta) = \prod_{t=1}^T \prod_{-m < j < m, j
eq 0} P(w_{t+j} \mid w_t; heta)$$

(2) 负对数似然损失函数

$$J(heta) = -rac{1}{T} \log L(heta) = -rac{1}{T} \sum_{t=1}^{T} \sum_{-m \leq j \leq m, j
eq 0} \log P(w_{t+j} \mid w_t; heta)$$

2. Softmax 条件概率

$$P(o \mid c) = rac{\exp\left(u_o^T v_c
ight)}{\sum_{w \in V} \exp\left(u_w^T v_c
ight)}$$

为什么这样记?

$$1.u_o^Tv_c=uv=\sum u_iv_i,T$$
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

$$\left[rac{\partial J}{\partial v_c} = rac{\partial}{\partial v_c} \Biggl[-\lograc{\expig(u_o^T v_cig)}{\sum_w \expig(u_w^T v_cig)} \Biggr] = -rac{\partial}{\partial v_c} \Biggl[u_o^T v_c - \log\sum_w \expig(u_w^T v_cig) \Biggr]$$

(2) 计算第一项梯度

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

(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 $\frac{1}{x}$

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

(4) 合并结果

$$rac{\partial J}{\partial v_c} = -1 imes (u_o - \sum_w P(w \mid c) u_w)$$

4. 物理意义

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

5. 参数更新规则

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

其中 (η) 为学习率.