parameter and the	and summer who	and the same	Name of Street,	-
Page	NO.			
ALC: UNIVERSITY OF	and the same of the same of	the same of the same	1 16-3	Market State of the State of th
Dote				
A STATE OF THE PARTY.				

The objective described in the negative sampling part is:

10g t (Vivo Vut) + & Ewi ~ Ph(w) [log t (-Vivi Vwx)]

Here, to wo is a context word. WI is the target word and wi's are to random words chosen on the basis of the distribution Ph(w), that gives:

Ph(wi) = 1- | t f(wi) is frequof wi.

BASICALLY,

we take a real context word from the target words neighborhood, and do gradient ascent for  $log(\tau(Vw^T_v)) = y(say)$ .

dy = 1 (Vwo Vwr) (1-o (Vwo Vwr)). VwI d Vwo Vwo Vwr)

34 = (1-+ (VTWO VWI)). VWI

So, the parameter update for Vous =

\$ Vwo = Vwo + & (1-o(Vtwo Vwz)). Vwz

Same for the context-vector derivative and update:
dy = (1- \sqrt{VwoVwI}). \sqrt{wo}

d \sqrt{will}

EVWI = VWE + a (1- T (VWO VWI)) VWO

x is learning rate.

Now, for the negative samples: The gradient steps will be in the opposite direction, because the expressions differ only by av-10001 (00) 8-3-3 3 + (300/ 201/) TO 00 For each 'fake', 'false' context word:

harved and is Ed. Asyon produces to a low rest of and

MESONO DISTORDING CONTRA STORE STATE Vwi = Vwi - a (10 - o (Vwy Vwi)) Vwi

VWE = VWE - a (1- J (VW VWi)) VWi