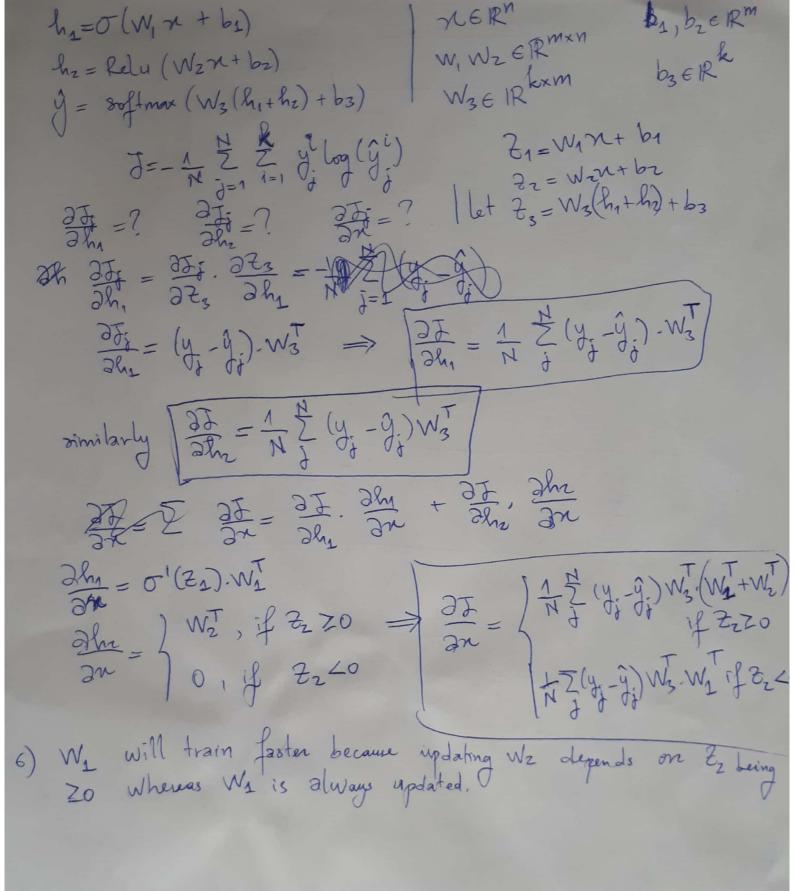
CS224n: Deep Learning for NLP Pratical Midterm#1 WERMXN (1) h1 = 0 (Wx1+6) ; h2 = 0 (Wx2+6) berm J= = 1 | h1-h2 | + = | WILF NI, NZER" JE = 5. JE = 5. J Could be rewritten as: J= 1/2 (hn-hz) + 2 Z; Wj. Wj $\Rightarrow \frac{\partial J}{\partial W} = \frac{1}{2} \frac{\partial \left[(h_1 - h_2) \cdot (h_1 + h_2) \right]}{\partial (h_1 - h_2)} \frac{\partial (h_1 - h_2)}{\partial W} + \frac{1}{2} \frac{\partial}{\partial W} \left(Z_j W_j^T W_j^T \right)$) 3 = 1 (h1-h2) [0 (Wx1+b)x[-0 (Wx2+b)x[] + 1/2 W imilarly; [] = 1 (h, -h2)[J'(WM,+b) - J'(WMz+b)] M=: M - 4 gr p= 2 p - 4 3 p 4) This model doesn't make use of the contextual information in surrounding words 3 60 parameters



b) a)
$$\frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial Z^{(t)}} \cdot \frac{\partial Z^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}} = \frac{\partial J^{(t)}}{\partial J_{t}} \cdot \frac{\partial J^{(t)}}{\partial J_{t}$$