עיבוד שפה טבעית - תרגיל בית 2

מגישים:
יותם מנה, 204717862
גיא אזוב, 312567654

שאלה 1: בדפים הסרוקים

שאלה 2:
סעיפים a,b: בדפים הסרוקים

סעיף d:

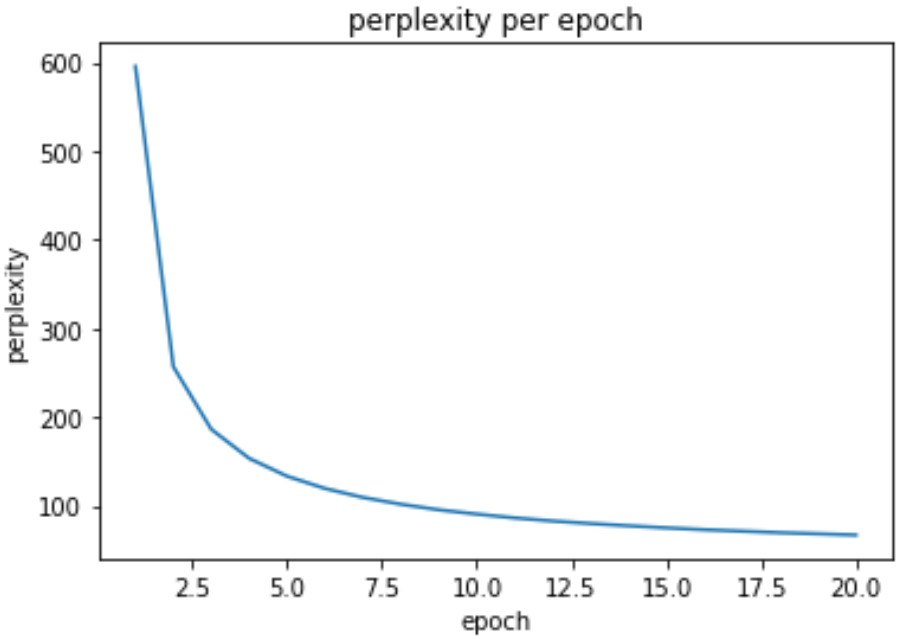
תוצאת הריצה בשלב ה eval:

Perplexity = 112.967665327491

שאלה 3:

סעיף a: בדפים הסרוקים

סעיף c:
גרף ערכי ה perplexity לאורך האימון:



ערך ה perplexity הסופי: 104.709

שאלה 1 b:

λ_1 \ λ_2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	208	124	101	87	78	72	68	64	62	61	
0.1	111	87	76	69	64	60	57	56	55		
0.2	92	75	67	62	58	55	54	53			
0.3	83	68	61	57	54	53	53				
0.4	77	64	58	55	53	52					
0.5	75	61	56	54	53						
0.6	74	60	56	54							
0.7	75	60	57								
0.8	80	63									
0.9	93										
1											

Best Parameters: $\lambda_1 = 0.4$

$\lambda_2 = 0.5$

Perplexity = 52.43

$$CE(y,\hat{y}) = -\sum_i y_i \log(\hat{y}_i) \qquad (a) \quad 2.$$

$$\text{Softmax}(\theta)_i = \frac{\exp(\theta_i)}{\sum_j \exp(\theta_j)}$$

$$\frac{\partial CE(y,\hat{y})}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\left(-\sum_k y_k \log(\hat{y}_k)\right) =$$

כלל השרשרת

$$= \frac{\partial\left(-\sum_k y_k \log(\hat{y}_k)\right)}{\partial \log(\hat{y}_k)} \cdot \frac{\partial \log(\hat{y}_k)}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial \theta_i} =$$

$$= -\sum_k \frac{\partial(y_k \log(\hat{y}_k))}{\partial \log(\hat{y}_k)} \cdot \frac{1}{\hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial \theta_i} =$$

$$= -\sum_k y_k \cdot \frac{1}{\hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial \theta_i} = \ldots$$

$$\frac{\partial \hat{y}_k}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \frac{e^{\theta_k}}{\sum_j e^{\theta_j}}$$

נפרק לשני מקרים:

$k=i$:

$$\frac{\partial}{\partial \theta_i} \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} = \frac{e^{\theta_i} \sum_j e^{\theta_j} - e^{\theta_i} e^{\theta_i}}{\left(\sum_j e^{\theta_j}\right)^2} = \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} \cdot \frac{\sum_j e^{\theta_j} - e^{\theta_i}}{\sum_j e^{\theta_j}} =$$

(במכנה ובמונה מופיע θ_i)

$$= \hat{y}_i \cdot (1 - \hat{y}_i)$$

$k \neq i$:

$$\frac{\partial}{\partial \theta_i} \frac{e^{\theta_k}}{\sum_j e^{\theta_j}} = \frac{-e^{\theta_k} \cdot e^{\theta_i}}{\left(\sum_j e^{\theta_j}\right)^2} = -\frac{e^{\theta_k}}{\sum_j e^{\theta_j}} \cdot \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} = -\hat{y}_k \hat{y}_i$$

(רק במכנה מופיע θ_i)

על הכל ביחד:

$$\frac{\partial}{\partial \theta_i} \frac{e^{\theta_k}}{\sum_j e^{\theta_j}} = \begin{cases} \hat{y}_i(1-\hat{y}_i), & k=i \\ -\hat{y}_k \hat{y}_i, & k \neq i \end{cases} = \hat{y}_k(\delta_{ki} - \hat{y}_i)$$

(δ_{ki} — פונ' הדלתא של קרונקר)

ולסיכום נקבל -

$$\ldots = -\sum_k y_k \cdot \frac{1}{\hat{y}_k} \cdot \hat{y}_k(\delta_{ki} - \hat{y}_i) =$$

$$= -y_i \cdot (1-\hat{y}_i) + \sum_{k \neq i} y_k \hat{y}_i = -y_i + y_i \hat{y}_i + \sum_{k \neq i} y_k \hat{y}_i =$$

$$= -y_i + \hat{y}_i \left(y_i + \sum_{k \neq i} y_k \right) =$$

$$= \hat{y}_i \underbrace{\left(\sum_k y_k \right)}_{\text{1 הוקטור}} - y_i = \boxed{\hat{y}_i - y_i}$$

$$\frac{\partial J}{\partial x} = \frac{\partial CE(y,\hat{y})}{\partial (hW_2+b_2)} \cdot \frac{\partial (hW_2+b_2)}{\partial (\sigma(xW_1+b_1))} \cdot \frac{\partial (\sigma(xW_1+b_1))}{\partial (xW_1+b_1)} \cdot \frac{\partial (xW_1+b_1)}{\partial x} = (\hat{y}-y) \cdot \frac{\partial [\sigma(xW_1+b_1)W_2+b_2]}{\partial (\sigma(xW_1+b_1))} \cdot \sigma'(xW_1+b_1) \cdot W_1^T$$

$$= (\hat{y}-y) \cdot W_2^T \cdot \sigma'(xW_1+b_1) \cdot W_1^T =$$

$$= \boxed{(\hat{y}-y) W_2^T \left[\sigma(xW_1+b_1) \cdot (1-\sigma(xW_1+b_1)) \right] W_1^T}$$

(3) (a)

$$h_{drop_i} = \begin{cases} \gamma \cdot h_i & \text{w.p. } 1-p_{drop} \\ 0 & \text{w.p. } p_{drop} \end{cases}$$

$$E_{p_{drop}}[h_{drop}]_i = [1-p_{drop}] \cdot \gamma h_i = h_i \Rightarrow$$

$$\Rightarrow \gamma = \frac{1}{1-p_{drop}}$$