

NLP- Exercise 2

-1-

2(a)

$$J = ce(y, \hat{y})$$

1.20

$y_k = 1$ \rightarrow \sim

$$\frac{\partial J}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \left(- \sum_j y_j \log \hat{y}_j \right) \quad i=k \quad \text{only}$$

$$= \frac{\partial}{\partial \theta_i} \left(- y_i \log \hat{y}_i \right) = \frac{\partial}{\partial \theta_i} \left(- \log \hat{y}_i \right) \rightarrow$$

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

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

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

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

-2-

$$- \left(-1 - \frac{1}{\sum_j e^{\theta_j}} \cdot e^{\theta_i} \right) =$$

$$= - \left(1 + \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} \right) = -1 + \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} =$$

$$= \text{softmax}(\theta)_i - 1$$

$i \neq k$ or

$$\frac{\partial J}{\partial \theta_i} = \frac{\partial (-\log \hat{y}_k)}{\partial \theta_i} =$$

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

-3-

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

$$= - \left(0 - \frac{1}{\sum_j e^{-\theta_j}} e^{\theta_i} \right) = \hat{y}_i$$

$i \neq k$ or $i = k$

$$\frac{\partial \tilde{J}}{\partial \theta} = \hat{y} - y$$

$(b_1) \quad b_1(x_1, x_2, x_3) \quad p_k, p_l$

2(b)

$$\frac{\partial J}{\partial x} = \frac{\partial J}{\partial \theta} \cdot \frac{\partial \theta}{\partial h} \cdot \frac{\partial h}{\partial x} \quad \text{chain rule}$$

$$\begin{cases} \theta = h \cdot w_2 + b_2 \\ J = CE(y, \hat{y}) \end{cases} \quad \text{also}$$

$$\frac{\partial J}{\partial \theta} = \hat{y} - y \quad \text{c (a) from y'}$$

$$\frac{\partial \theta}{\partial h} = \frac{\partial (h \cdot w_2 + b_2)}{\partial h} = w_2^T$$

we want to find the w_2 for each x ?
 f the model output \hat{y} and D_h
 also f_1, \dots, f_{D_y} - 1 per each

1. D_h and f_i to
 $w_2 \rightarrow i$ the output f_i to w_2 to find

$$\frac{\partial f_i}{\partial h_j} = \frac{\partial}{\partial h_j} \sum_{k=1}^n h_k \cdot W_{2,k,i} = W_{2,j,i}$$

כל הכנסות מתכנסות ב-1. $k=j$

דובלי אקסון את ההיסטוריה של $W_{2,j,i}$

$$\frac{\partial h}{\partial x}$$

$$h = \sigma(xW_1 + b_1)$$

$$z = xW_1 + b_1$$

למשל

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial z} \cdot \frac{\partial z}{\partial x}$$

$$\frac{\partial h}{\partial z} = \sigma'(z) = \sigma(z) \cdot (1 - \sigma(z))$$

פונקציית סיגמויד (הזניחה בדרך כלל)
sigmoid

$$\frac{\partial z}{\partial x} = W_1^T$$

צירוף של הקשר

בטל את ההפרה

$$\frac{\partial J}{\partial x} = (\hat{y} - y) \cdot W_2^T \cdot \sigma(xW_1 + b_1) \cdot$$

$$(1 - \sigma(xW_1 + b_1)) \cdot W_1^T$$

Ngrams model results

#trigrams: 413540

#bigrams: 122930

#unigrams: 2000

#tokens: 1118296

#perplexity: 51.05675514045975 , lambda1: 0.5 , lambda2: 0.4

Search grid: step 0.05, best perplexity is at the bottom

#perplexity: 189.62443159556764 , lambda1: 0 , lambda2: 0

#perplexity: 135.0828858966259 , lambda1: 0 , lambda2: 0.05

#perplexity: 115.66484646581385 , lambda1: 0 , lambda2: 0.1

#perplexity: 103.46797270445407 , lambda1: 0 , lambda2: 0.15

#perplexity: 94.72955732912476 , lambda1: 0 , lambda2: 0.2

#perplexity: 88.03387679231152 , lambda1: 0 , lambda2: 0.25

#perplexity: 82.68776158211935 , lambda1: 0 , lambda2: 0.3

#perplexity: 78.30017872189315 , lambda1: 0 , lambda2: 0.35

#perplexity: 74.62991657721106 , lambda1: 0 , lambda2: 0.4

#perplexity: 71.51899907118825 , lambda1: 0 , lambda2: 0.45

#perplexity: 68.85982814105435 , lambda1: 0 , lambda2: 0.5

#perplexity: 66.5776351064429 , lambda1: 0 , lambda2: 0.55

#perplexity: 64.62074925082231 , lambda1: 0 , lambda2: 0.6

#perplexity: 62.955423270958484 , lambda1: 0 , lambda2: 0.65

#perplexity: 61.56395735765189 , lambda1: 0 , lambda2: 0.7

#perplexity: 60.44630257811496 , lambda1: 0 , lambda2: 0.75

#perplexity: 59.62750429425663 , lambda1: 0 , lambda2: 0.8

#perplexity: 59.179578905105075 , lambda1: 0 , lambda2: 0.85

#perplexity: 59.29324643858427 , lambda1: 0 , lambda2: 0.9

#perplexity: 60.62724974413923 , lambda1: 0 , lambda2: 0.95

#perplexity: 121.44796881034364 , lambda1: 0.05 , lambda2: 0

#perplexity: 102.65800802771979 , lambda1: 0.05 , lambda2: 0.05

#perplexity: 92.48423296093252 , lambda1: 0.05 , lambda2: 0.1

#perplexity: 85.24454998981624 , lambda1: 0.05 ', lambda2: 0.15
#perplexity: 79.6745458198154 , lambda1: 0.05 ', lambda2: 0.2
#perplexity: 75.20075099235922 , lambda1: 0.05 ', lambda2: 0.25
#perplexity: 71.50757863430135 , lambda1: 0.05 ', lambda2: 0.3
#perplexity: 68.40204377113223 , lambda1: 0.05 ', lambda2: 0.35
#perplexity: 65.75811583974462 , lambda1: 0.05 ', lambda2: 0.4
#perplexity: 63.49007404127168 , lambda1: 0.05 ', lambda2: 0.45
#perplexity: 61.53852851587068 , lambda1: 0.05 ', lambda2: 0.5
#perplexity: 59.862795683070296 , lambda1: 0.05 ', lambda2: 0.55
#perplexity: 58.43698159103913 , lambda1: 0.05 ', lambda2: 0.6
#perplexity: 57.24882251689967 , lambda1: 0.05 ', lambda2: 0.65
#perplexity: 56.301581024136496 , lambda1: 0.05 ', lambda2: 0.7
#perplexity: 55.62128347993121 , lambda1: 0.05 ', lambda2: 0.75
#perplexity: 55.27739244675994 , lambda1: 0.05 ', lambda2: 0.8
#perplexity: 55.45012063438885 , lambda1: 0.05 ', lambda2: 0.85
#perplexity: 56.75772466865094 , lambda1: 0.05 ', lambda2: 0.9
#perplexity: 104.21014060885702 , lambda1: 0.1 ', lambda2: 0
#perplexity: 90.29523997117539 , lambda1: 0.1 ', lambda2: 0.05
#perplexity: 82.63468183834307 , lambda1: 0.1 ', lambda2: 0.1
#perplexity: 77.02742296395861 , lambda1: 0.1 ', lambda2: 0.15
#perplexity: 72.62231893934889 , lambda1: 0.1 ', lambda2: 0.2
#perplexity: 69.02960194278027 , lambda1: 0.1 ', lambda2: 0.25
#perplexity: 66.03083754196426 , lambda1: 0.1 ', lambda2: 0.3
#perplexity: 63.490308663968385 , lambda1: 0.1 ', lambda2: 0.35
#perplexity: 61.31855306729383 , lambda1: 0.1 ', lambda2: 0.4
#perplexity: 59.45486352169986 , lambda1: 0.1 ', lambda2: 0.45
#perplexity: 57.858233933043444 , lambda1: 0.1 ', lambda2: 0.5
#perplexity: 56.502797965854086 , lambda1: 0.1 ', lambda2: 0.55
#perplexity: 55.37633072569653 , lambda1: 0.1 ', lambda2: 0.6
#perplexity: 54.481896294327946 , lambda1: 0.1 ', lambda2: 0.65
#perplexity: 53.84476753605702 , lambda1: 0.1 ', lambda2: 0.7
#perplexity: 53.53242617200751 , lambda1: 0.1 ', lambda2: 0.75
#perplexity: 53.71983021681823 , lambda1: 0.1 ', lambda2: 0.8
#perplexity: 55.00683155278933 , lambda1: 0.1 ', lambda2: 0.85

#perplexity: 94.23161700053606 , lambda1: 0.15 ', lambda2: 0
#perplexity: 82.57225825956435 , lambda1: 0.15 ', lambda2: 0.05
#perplexity: 76.23535593745912 , lambda1: 0.15 ', lambda2: 0.1
#perplexity: 71.5536385385448 , lambda1: 0.15 ', lambda2: 0.15
#perplexity: 67.84378656534138 , lambda1: 0.15 ', lambda2: 0.2
#perplexity: 64.79831434528553 , lambda1: 0.15 ', lambda2: 0.25
#perplexity: 62.24570033121368 , lambda1: 0.15 ', lambda2: 0.3
#perplexity: 60.0797196427379 , lambda1: 0.15 ', lambda2: 0.35
#perplexity: 58.23100446516654 , lambda1: 0.15 ', lambda2: 0.4
#perplexity: 56.65373240130592 , lambda1: 0.15 ', lambda2: 0.45
#perplexity: 55.319169442917264 , lambda1: 0.15 ', lambda2: 0.5
#perplexity: 54.213225940112 , lambda1: 0.15 ', lambda2: 0.55
#perplexity: 53.33755840629845 , lambda1: 0.15 ', lambda2: 0.6
#perplexity: 52.71606888954287 , lambda1: 0.15 ', lambda2: 0.65
#perplexity: 52.4143485002371 , lambda1: 0.15 ', lambda2: 0.7
#perplexity: 52.6035606908306 , lambda1: 0.15 ', lambda2: 0.75
#perplexity: 53.87147632892331 , lambda1: 0.15 ', lambda2: 0.8
#perplexity: 87.50877140144976 , lambda1: 0.2 ', lambda2: 0
#perplexity: 77.13804551389606 , lambda1: 0.2 ', lambda2: 0.05
#perplexity: 71.62571328971343 , lambda1: 0.2 ', lambda2: 0.1
#perplexity: 67.54988430860072 , lambda1: 0.2 ', lambda2: 0.15
#perplexity: 64.31265483495633 , lambda1: 0.2 ', lambda2: 0.2
#perplexity: 61.651264959856995 , lambda1: 0.2 ', lambda2: 0.25
#perplexity: 59.42123401331949 , lambda1: 0.2 ', lambda2: 0.3
#perplexity: 57.53441664948557 , lambda1: 0.2 ', lambda2: 0.35
#perplexity: 55.934770623799835 , lambda1: 0.2 ', lambda2: 0.4
#perplexity: 54.58755157417025 , lambda1: 0.2 ', lambda2: 0.45
#perplexity: 53.474881235585215 , lambda1: 0.2 ', lambda2: 0.5
#perplexity: 52.59582374379511 , lambda1: 0.2 ', lambda2: 0.55
#perplexity: 51.972256788749434 , lambda1: 0.2 ', lambda2: 0.6
#perplexity: 51.66775122111883 , lambda1: 0.2 ', lambda2: 0.65
#perplexity: 51.85042087663324 , lambda1: 0.2 ', lambda2: 0.7
#perplexity: 53.09961744497666 , lambda1: 0.2 ', lambda2: 0.75
#perplexity: 82.65536812474502 , lambda1: 0.25 ', lambda2: 0

#perplexity: 73.08707059394953 , lambda1: 0.25 ', lambda2: 0.05
#perplexity: 68.1344680263809 , lambda1: 0.25 ', lambda2: 0.1
#perplexity: 64.4874160753521 , lambda1: 0.25 ', lambda2: 0.15
#perplexity: 61.5959936516965 , lambda1: 0.25 ', lambda2: 0.2
#perplexity: 59.22433642822513 , lambda1: 0.25 ', lambda2: 0.25
#perplexity: 57.24531729287139 , lambda1: 0.25 ', lambda2: 0.3
#perplexity: 55.583378390503356 , lambda1: 0.25 ', lambda2: 0.35
#perplexity: 54.192925184086654 , lambda1: 0.25 ', lambda2: 0.4
#perplexity: 53.04955647471897 , lambda1: 0.25 ', lambda2: 0.45
#perplexity: 52.148133251122125 , lambda1: 0.25 ', lambda2: 0.5
#perplexity: 51.507589042522106 , lambda1: 0.25 ', lambda2: 0.55
#perplexity: 51.189103436682984 , lambda1: 0.25 ', lambda2: 0.6
#perplexity: 51.35810034822876 , lambda1: 0.25 ', lambda2: 0.65
#perplexity: 52.58803287991109 , lambda1: 0.25 ', lambda2: 0.7
#perplexity: 79.02928641104755 , lambda1: 0.3 ', lambda2: 0
#perplexity: 69.9729070376946 , lambda1: 0.3 ', lambda2: 0.05
#perplexity: 65.41987801336134 , lambda1: 0.3 ', lambda2: 0.1
#perplexity: 62.0920362882769 , lambda1: 0.3 ', lambda2: 0.15
#perplexity: 59.46705814842095 , lambda1: 0.3 ', lambda2: 0.2
#perplexity: 57.326252808026005 , lambda1: 0.3 ', lambda2: 0.25
#perplexity: 55.554703734051635 , lambda1: 0.3 ', lambda2: 0.3
#perplexity: 54.08692770902022 , lambda1: 0.3 ', lambda2: 0.35
#perplexity: 52.88735625640967 , lambda1: 0.3 ', lambda2: 0.4
#perplexity: 51.94407423472364 , lambda1: 0.3 ', lambda2: 0.45
#perplexity: 51.271615031669754 , lambda1: 0.3 ', lambda2: 0.5
#perplexity: 50.92804567539584 , lambda1: 0.3 ', lambda2: 0.55
#perplexity: 51.07613211955486 , lambda1: 0.3 ', lambda2: 0.6
#perplexity: 52.28530317588607 , lambda1: 0.3 ', lambda2: 0.65
#perplexity: 76.2861977788883 , lambda1: 0.35 ', lambda2: 0
#perplexity: 67.54532309316419 , lambda1: 0.35 ', lambda2: 0.05
#perplexity: 63.28621138463526 , lambda1: 0.35 ', lambda2: 0.1
#perplexity: 60.20503064010318 , lambda1: 0.35 ', lambda2: 0.15
#perplexity: 57.79439011366239 , lambda1: 0.35 ', lambda2: 0.2
#perplexity: 55.84705743844988 , lambda1: 0.35 ', lambda2: 0.25

#perplexity: 54.25775301780426 , lambda1: 0.35 ', lambda2: 0.3
#perplexity: 52.97072326163178 , lambda1: 0.35 ', lambda2: 0.35
#perplexity: 51.96268680054395 , lambda1: 0.35 ', lambda2: 0.4
#perplexity: 51.241288146841626 , lambda1: 0.35 ', lambda2: 0.45
#perplexity: 50.860174154453524 , lambda1: 0.35 ', lambda2: 0.5
#perplexity: 50.97908654476532 , lambda1: 0.35 ', lambda2: 0.55
#perplexity: 52.16484068083213 , lambda1: 0.35 ', lambda2: 0.6
#perplexity: inf , lambda1: 0.35 ', lambda2: 0.65
#perplexity: 74.22636363152093 , lambda1: 0.4 ', lambda2: 0
#perplexity: 65.65364479569932 , lambda1: 0.4 ', lambda2: 0.05
#perplexity: 61.61417574457767 , lambda1: 0.4 ', lambda2: 0.1
#perplexity: 58.72969609289352 , lambda1: 0.4 ', lambda2: 0.15
#perplexity: 56.49903089743215 , lambda1: 0.4 ', lambda2: 0.2
#perplexity: 54.72308715228743 , lambda1: 0.4 ', lambda2: 0.25
#perplexity: 53.305674507272954 , lambda1: 0.4 ', lambda2: 0.3
#perplexity: 52.203013967437315 , lambda1: 0.4 ', lambda2: 0.35
#perplexity: 51.41126684144457 , lambda1: 0.4 ', lambda2: 0.4
#perplexity: 50.97726926981079 , lambda1: 0.4 ', lambda2: 0.45
#perplexity: 51.05675514045975 , lambda1: 0.4 ', lambda2: 0.5
#perplexity: 52.21489019654714 , lambda1: 0.4 ', lambda2: 0.55
#perplexity: inf , lambda1: 0.4 ', lambda2: 0.6
#perplexity: 72.73085476755635 , lambda1: 0.45 ', lambda2: 0
#perplexity: 64.20449224793792 , lambda1: 0.45 ', lambda2: 0.05
#perplexity: 60.329894488138066 , lambda1: 0.45 ', lambda2: 0.1
#perplexity: 57.607445943235994 , lambda1: 0.45 ', lambda2: 0.15
#perplexity: 55.53574568562644 , lambda1: 0.45 ', lambda2: 0.2
#perplexity: 53.922461197611526 , lambda1: 0.45 ', lambda2: 0.25
#perplexity: 52.68248870435064 , lambda1: 0.45 ', lambda2: 0.3
#perplexity: 51.79117508010556 , lambda1: 0.45 ', lambda2: 0.35
#perplexity: 51.28395908468968 , lambda1: 0.45 ', lambda2: 0.4
#perplexity: 51.310465200925755 , lambda1: 0.45 ', lambda2: 0.45
#perplexity: 52.43457210051572 , lambda1: 0.45 ', lambda2: 0.5
#perplexity: inf , lambda1: 0.45 ', lambda2: 0.55
#perplexity: 71.73216091858131 , lambda1: 0.5 ', lambda2: 0

#perplexity: 63.141481512118055 , lambda1: 0.5 ', lambda2: 0.05
#perplexity: 59.39001865493764 , lambda1: 0.5 ', lambda2: 0.1
#perplexity: 56.80675206736001 , lambda1: 0.5 ', lambda2: 0.15
#perplexity: 54.88526944518292 , lambda1: 0.5 ', lambda2: 0.2
#perplexity: 53.4410948364399 , lambda1: 0.5 ', lambda2: 0.25
#perplexity: 52.40714435589907 , lambda1: 0.5 ', lambda2: 0.3
#perplexity: 51.79790196501015 , lambda1: 0.5 ', lambda2: 0.35
#perplexity: 51.7525038119608 , lambda1: 0.5 ', lambda2: 0.4
#perplexity: 52.83284883870939 , lambda1: 0.5 ', lambda2: 0.45
#perplexity: inf , lambda1: 0.5 ', lambda2: 0.5
#perplexity: 71.2007353718514 , lambda1: 0.55 ', lambda2: 0
#perplexity: 62.43546591573319 , lambda1: 0.55 ', lambda2: 0.05
#perplexity: 58.77503283632245 , lambda1: 0.55 ', lambda2: 0.1
#perplexity: 56.319222282268925 , lambda1: 0.55 ', lambda2: 0.15
#perplexity: 54.55380386551248 , lambda1: 0.55 ', lambda2: 0.2
#perplexity: 53.30814497472289 , lambda1: 0.55 ', lambda2: 0.25
#perplexity: 52.55319194334002 , lambda1: 0.55 ', lambda2: 0.3
#perplexity: 52.40799044430007 , lambda1: 0.55 ', lambda2: 0.35
#perplexity: 53.429561802645935 , lambda1: 0.55 ', lambda2: 0.4
#perplexity: inf , lambda1: 0.55 ', lambda2: 0.45
#perplexity: 71.14080307449925 , lambda1: 0.6 ', lambda2: 0
#perplexity: 62.0810208146342 , lambda1: 0.6 ', lambda2: 0.05
#perplexity: 58.48804465973374 , lambda1: 0.6 ', lambda2: 0.1
#perplexity: 56.16159982259757 , lambda1: 0.6 ', lambda2: 0.15
#perplexity: 54.58097820941368 , lambda1: 0.6 ', lambda2: 0.2
#perplexity: 53.60858101536627 , lambda1: 0.6 ', lambda2: 0.25
#perplexity: 53.31985083275255 , lambda1: 0.6 ', lambda2: 0.3
#perplexity: 54.25876047140675 , lambda1: 0.6 ', lambda2: 0.35
#perplexity: inf , lambda1: 0.6 ', lambda2: 0.4
#perplexity: 71.59384750178913 , lambda1: 0.65 ', lambda2: 0
#perplexity: 62.098112648224905 , lambda1: 0.65 ', lambda2: 0.05
#perplexity: 58.55925708642692 , lambda1: 0.65 ', lambda2: 0.1
#perplexity: 56.386251116654215 , lambda1: 0.65 ', lambda2: 0.15
#perplexity: 55.06552643923591 , lambda1: 0.65 ', lambda2: 0.2

#perplexity: 54.559282963998015 , lambda1: 0.65 ', lambda2: 0.25
#perplexity: 55.375641377873876 , lambda1: 0.65 ', lambda2: 0.3
#perplexity: inf , lambda1: 0.65 ', lambda2: 0.35
#perplexity: 72.65202752799017 , lambda1: 0.7 ', lambda2: 0
#perplexity: 62.540497582966715 , lambda1: 0.7 ', lambda2: 0.05
#perplexity: 59.05996788309772 , lambda1: 0.7 ', lambda2: 0.1
#perplexity: 57.111031226168144 , lambda1: 0.7 ', lambda2: 0.15
#perplexity: 56.24818120277747 , lambda1: 0.7 ', lambda2: 0.2
#perplexity: 56.870533859951394 , lambda1: 0.7 ', lambda2: 0.25
#perplexity: 74.48996649238694 , lambda1: 0.75 ', lambda2: 0
#perplexity: 63.51696323967588 , lambda1: 0.75 ', lambda2: 0.05
#perplexity: 60.139531517348914 , lambda1: 0.75 ', lambda2: 0.1
#perplexity: 58.6126425031242 , lambda1: 0.75 ', lambda2: 0.15
#perplexity: 58.89894858667646 , lambda1: 0.75 ', lambda2: 0.2
#perplexity: 77.43962875715775 , lambda1: 0.8 ', lambda2: 0
#perplexity: 65.24487380188029 , lambda1: 0.8 ', lambda2: 0.05
#perplexity: 62.13783410607519 , lambda1: 0.8 ', lambda2: 0.1
#perplexity: 61.75478399589083 , lambda1: 0.8 ', lambda2: 0.15
#perplexity: 82.18946639873849 , lambda1: 0.85 ', lambda2: 0
#perplexity: 68.21025327886099 , lambda1: 0.85 ', lambda2: 0.05
#perplexity: 66.08945993911992 , lambda1: 0.85 ', lambda2: 0.1
#perplexity: 90.45303607601097 , lambda1: 0.9 ', lambda2: 0
#perplexity: 73.8689919845812 , lambda1: 0.9 ', lambda2: 0.05
#perplexity: 108.514079035969 , lambda1: 0.95 ', lambda2: 0
#best perplexity: 50.860174154453524 , lambda1: 0.35 , lambda2: 0.5

Neural_Net:

#params: 104550
#train examples: 1118296
training took 0 seconds
dev perplexity : 112.96766532749089