Home Assignment 2: Language Models

Daniel Solomon ID: 206335283
Gal Ron ID: 304856529

April 24, 2017

Problem 1.b. *ngram_model.py* output is:

```
C:\Windows\System32\cmd.exe
C:\Users\galr\Dropbox\My Documents\workspace\NLP Course\ex2\assignment2\code>python ngram_model.py
#trigrams: 413540
#bigrams: 122930
#unigrams: 2000
#tokens: 1118296
#perplexity: 51.752503812
#Performing grid search on lambda values...

             l1=0.00 l1=0.05 l1=0.10 l1=0.15 l1=0.20 l1=0.25 l1=0.30 l1=0.35 l1=0.40 l1=0.45 l1=0.50 l1=0.55 l1=0.60 l1=0.65 l1=0.70 l1=0.75 l1=0.80 l1=0.85 l1=0.90 l1=0.95
l2=0.00      189.62  121.45  104.21  94.23   87.51   82.66   79.03   76.29   74.23   72.73   71.73   71.20   71.14   71.59   72.65   74.49   77.44   82.19   90.45   108.51
l2=0.05      135.08  102.66  90.30   82.57   77.14   73.09   69.97   67.55   65.65   64.20   63.14   62.44   62.08   62.10   62.54   63.52   65.24   68.21   73.87   -
l2=0.10      115.66  92.48   82.63   76.24   71.63   68.13   65.42   63.29   61.61   60.33   59.39   58.78   58.49   58.56   59.06   60.14   62.14   66.09   -       -
l2=0.15      103.47  85.24   77.03   71.55   67.55   64.49   62.09   60.21   58.73   57.61   56.81   56.32   56.16   56.39   57.11   58.61   61.75   -       -       -
l2=0.20      94.73   79.67   72.62   67.84   64.31   61.60   59.47   57.79   56.50   55.54   54.89   54.55   54.58   55.07   56.25   58.90   -       -       -       -
l2=0.25      88.03   75.20   69.03   64.80   61.65   59.22   57.33   55.85   54.72   53.92   53.44   53.31   53.61   54.56   56.87   -       -       -       -       -
l2=0.30      82.69   71.51   66.03   62.25   59.42   57.25   55.55   54.26   53.31   52.68   52.41   52.55   53.32   55.38   -       -       -       -       -       -
l2=0.35      78.30   68.40   63.49   60.08   57.53   55.58   54.09   52.97   52.20   51.79   51.80   52.41   54.26   -       -       -       -       -       -       -
l2=0.40      74.63   65.76   61.32   58.23   55.93   54.19   52.89   51.96   51.41   51.28   51.75   53.43   -       -       -       -       -       -       -       -
l2=0.45      71.52   63.49   59.45   56.65   54.59   53.05   51.94   51.24   50.98   51.31   52.83   -       -       -       -       -       -       -       -       -
l2=0.50      68.86   61.54   57.86   55.32   53.47   52.15   51.27   50.86   51.06   52.43   -       -       -       -       -       -       -       -       -       -
l2=0.55      66.58   59.86   56.50   54.21   52.60   51.51   50.93   50.98   52.21   -       -       -       -       -       -       -       -       -       -       -
l2=0.60      64.62   58.44   55.38   53.34   51.97   51.19   51.08   52.16   -       -       -       -       -       -       -       -       -       -       -       -
l2=0.65      62.96   57.25   54.48   52.72   51.67   51.36   52.29   -       -       -       -       -       -       -       -       -       -       -       -       -
l2=0.70      61.56   56.30   53.84   52.41   51.85   52.59   -       -       -       -       -       -       -       -       -       -       -       -       -       -
l2=0.75      60.45   55.62   53.53   52.60   53.10   -       -       -       -       -       -       -       -       -       -       -       -       -       -       -
l2=0.80      59.63   55.28   53.72   53.87   -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -
l2=0.85      59.18   55.45   55.01   -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -
l2=0.90      59.29   56.76   -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -
l2=0.95      60.63   -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -

#minimal perplexity 50.86 obtained for lambda1=0.35, lambda2=0.50, lambda3=0.15
#perplexity for training only unigrams (lambda1=0, lambda2=0, lambda3=1) is: 189.62
#perplexity for training only bigrams is undefined - there are some bigrams never seen during training
```

Problem 2.a. We want to derive $CE(y, \hat{y}) = -\Sigma_i y_i log(\hat{y}_i)$ with respect to vector θ. First let's derive it for each co-ordinate:

$$\frac{\partial CE}{\partial \theta_k} = -\frac{\partial log(\hat{y}_i)}{\partial \theta_k} = -\frac{\partial}{\partial \theta_k} log(\frac{\exp \theta_i}{\Sigma_j \exp \theta_j})$$

Assume $k = i$, we get:

$$-\frac{\frac{\exp(\theta_k)\Sigma_j \exp(\theta_j) - \exp(\theta_k)^2}{(\Sigma_j \exp(\theta_j))^2}}{\frac{\exp(\theta_k)}{\Sigma_j \exp(\theta_j)}} = -\frac{\Sigma_j \exp(\theta_j) - \exp(\theta_k)}{\Sigma_j \exp(\theta_j)} = \frac{\exp(\theta_k) - \Sigma_j \exp(\theta_j)}{\Sigma_j \exp(\theta_j)} = \frac{\exp \theta_k}{\Sigma_j \exp \theta_j} - 1 = \hat{y_k} - 1.$$

Assume $k \neq i$, we get:

$$-\frac{\frac{-\exp(\theta_i)\exp(\theta_k)}{(\Sigma_j \exp(\theta_j))^2}}{\frac{\exp(\theta_k)}{\Sigma_j \exp(\theta_j)}} = \frac{\exp \theta_i}{\Sigma_j \exp \theta_j} = \hat{y}_i.$$

Bottom line we can generalize it as:

$$\frac{\partial CE}{\partial \theta} = \hat{y} - y.$$

Problem 2.b. For one's comfort let's denote $z_1 = xW_1 + b_1$ and $z_2 = h(W_2 + b_2)$. Using chain rule we get:

$$\frac{\partial CE}{\partial x} = \frac{\partial CE}{\partial z_2}\frac{\partial z_2}{\partial x} \overset{(i)}{=} (\hat{y} - y)\frac{\partial z_2}{\partial h}\frac{\partial h}{\partial x} \overset{(ii)}{=} (\hat{y} - y)W_2^T\frac{\partial h}{\partial z_1}\frac{\partial z_1}{\partial x} \overset{(iii)}{=} (\hat{y} - y)W_2^T \circ \sigma'(z_1)W_1^T.$$

Explanations:
(i) Left derivative calculated by 2.a, right derivative chain rule.
(ii) Left derivative calculated by simple calculus, right derivative chain rule.
(iii) Left derivative calculated by simple calculus so as right derivative (notice that $\circ$ sign represents multiplication member by member as we saw in class).

Problem 2.d. *neural_lm.py* output is:
params: 104550
train examples: 1118296
training took 324 seconds
dev perplexity : **112.967665327**