NLP - Ex2

Itay Sofer – 201507357

Or Perel - 200732444

<u>Q1</u>

b. #trigrams: 413540

#bigrams: 122930

#unigrams: 2000

#tokens: 1118296

#perplexity for lambda1(left) lambda2(top) grid search:

 000.000
 000.100
 000.200
 000.300
 000.400
 000.500
 000.600
 000.700
 000.800
 000.900

 000.000
 189.624
 115.665
 094.730
 082.688
 074.630
 068.860
 064.621
 061.564
 059.628
 059.293

 000.100
 104.210
 082.635
 072.622
 066.031
 061.319
 057.858
 055.376
 053.845
 053.720
 -01.000

 000.200
 087.509
 071.626
 064.313
 059.421
 055.935
 053.475
 051.972
 051.850
 -01.000
 -01.000

 000.300
 079.029
 065.420
 059.467
 055.555
 052.887
 051.272
 051.076
 -01.000
 -01.000
 -01.000

 000.400
 074.226
 061.614
 056.499
 053.306
 051.411
 051.057
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01.000
 -01

#best coefficients (lambda1, lambda2) are (0.4, 0.5) with perplexity 51.0567551405

$$CE(y,\hat{y}) = -\sum_{j} y_{j} \log(\hat{y}_{j})$$

$$= -\log(\hat{y}_{j})$$

$$= -\log(\int SOH(nax(\Theta)_{j}))$$

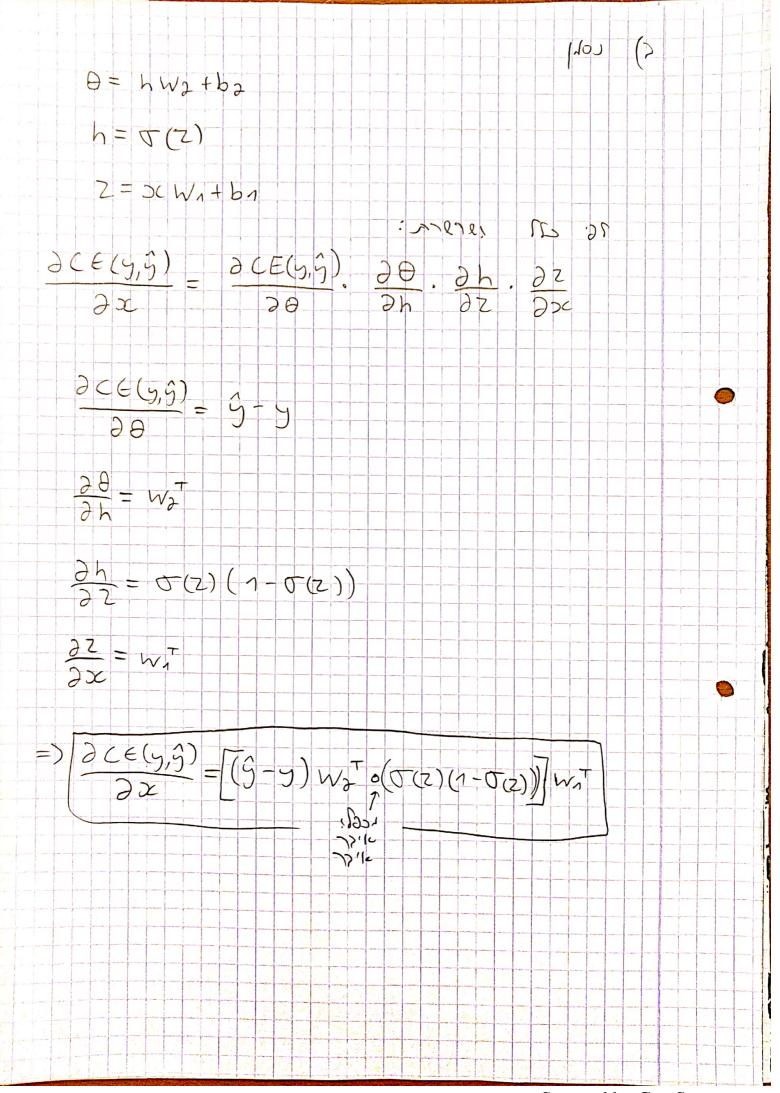
$$= -\log(\int (\exp(\Theta_{j})) + \log(\sum_{j} \exp(\Theta_{j}))$$

$$= -\log(\exp(\Theta_{j})) + \log(\sum_{j} \exp(\Theta_{j}))$$

$$= -\log(\exp(\Theta_{j})) + \log(\sum_{j} \exp(\Theta_{j}))$$

$$= -\frac{1}{2} \log(\exp(\Theta_{j})) + \log(\sum_{j} \exp(\Theta_{j}))$$

$$= -\frac{1}{2} \exp(\Theta_{j}) + \log(\sum_{j}$$



d.

#params: 104550

#train examples: 1118296

training took 0 seconds (re-run for evaluation after training)

dev perplexity: 112.967665327

$E_{Resp} [h_{resp}]_{i} = P_{stop} \cdot [rohi] + (1 - P_{stop})[rohi] $ $= (1 + P_{stop}) \cdot rh;$ $E_{trop} [h_{stop}]_{i} = h;$ $(1 - P_{stop}) \cdot rh;$ $(1 - P_{s$	$ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla A_i = A_i \\ & = (1 - \rho_{arop}) \cdot \nabla A_i = A_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla A_i = A_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla A_i = A_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla A_i = A_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla A_i = A_i \end{aligned} $	$ \begin{aligned} & = (1 - \rho_{drop}) \cdot \delta \cdot h; \\ &$
$ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \\ & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \\ & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \end{aligned} $ $ \end{aligned} $ $ \begin{aligned} & = (1 - \rho_{arop}) \cdot \nabla \cdot h_i \end{aligned} $ $ \end{aligned} $	$ \begin{aligned} & = (N - P_{drop}) \cdot \nabla \cdot \lambda_i \\ & = (N - P_{drop}) \cdot \nabla \cdot \lambda_i \end{aligned} $ $ \begin{aligned} & = (N - P_{drop}) \cdot \nabla \cdot \lambda_i \\ & = (N - P_{drop}) \cdot \nabla \cdot \lambda_i \end{aligned} $ $ \begin{aligned} & = (N - P_{drop}) \cdot \nabla \cdot \lambda_i \\ & = \lambda_i \end{aligned} $ $ \begin{aligned} & = (N - P_{drop}) \cdot \nabla \cdot \lambda_i \\ & = \lambda_i \end{aligned} $ $ \end{aligned} $	$ \begin{aligned} & = (1 + \rho_{arop}) \cdot \nabla \cdot h \cdot \\ & = (1 + \rho_{arop}) \cdot \nabla \cdot h$
$[h_{1}, h_{2}] = [h_{2}, h_{3}] = [h_{3}, h_{3}] + [h_{4}, h_{2}] = [h_{3}, h_{3}] = [h_{4}, h_{2}] = [h_{4}, h_{3}] = [h_{4}, h_{4}] = [h_{$	$[h_{1}cop] := p_{1}cop \cdot [r \circ hi] + (n - p_{1}cp)[r \circ h \cdot hi]$ $= (n - p_{2}cop) \cdot r \cdot hi$ $= (n - p_{3}cop) := hi$ $(p_{1}cop) \cdot r \cdot hi$ $= (n - p_{3}cop) := hi$ $(1 - p_{3}cop) := hi$ $(1 - p_{3}cop) := hi$	$[h_{1}cop] := p_{1}cop \cdot [cohi] + (1-p_{1}coh)[cohhi] $ $= (1-p_{2}cop) \cdot bhi$ $= hi$ $= hi$ $= hi$
$[h_{sop}]_{i} = \rho_{stop} \cdot [r \circ h_{i}] + (1 - \rho_{stop})[r \circ 1 \cdot h_{i}] $ $= (1 - \rho_{stop}) \cdot r \cdot h_{i}$ $= (1 - \rho_{stop}) \cdot r \cdot h_{i}$ $= h_{i}$ $\vdots \mid 0 \mid $	$[h_{Jrop}] = P_{Jrop} \cdot [rohi] + (1 + P_{Jrop})[rohh] $ $= (1 + P_{Jrop}) \cdot rohi$ $= h_{Jrop} = h_{Jrop} $	$[h_{Jrop}]_{i} = p_{Jrop} \cdot [r \cdot o \cdot h_{i}] + (1 - p_{Jrop})[r \cdot 1 \cdot h_{i}] $ $= (1 - p_{Jrop}) \cdot r \cdot h_{i}$ $= (1 - p_{Jrop}) \cdot r \cdot h_{i}$ $= h_{i}$ $= h_{i}$ $= h_{i}$
(x - x - y) = (x - y - y) = (x - y	$ (x - y) ^{2} = (y - y) ^{2} = (y$	$ \mathcal{L}_{\text{cop}} = P_{\text{stop}} \cdot [\mathcal{L}_{\text{coh}} - h_{i}] + (1 - P_{\text{stop}})[\mathcal{C}_{\text{ch}} - h_{i}] $ $= (1 - P_{\text{stop}}) \cdot \mathcal{L}_{\text{ch}} \cdot h_{i}$ $= \mathcal{L}_{\text{cop}} \mathcal{L}_{\text{stop}} \mathcal{L}_{\text{stop}} \mathcal{L}_{\text{ch}} \cdot h_{i} $ $= \mathcal{L}_{\text{cop}} \mathcal{L}_{\text{stop}} \mathcal{L}_{\text{stop}} \mathcal{L}_{\text{ch}} \cdot h_{i} $ $= \mathcal{L}_{\text{cop}} \mathcal{L}_{\text{stop}} \mathcal{L}_{\text{stop}} \mathcal{L}_{\text{ch}} \cdot h_{i} $ $= \mathcal{L}_{\text{cop}} \mathcal{L}_{\text{stop}} \mathcal{L}_{$
$ \begin{aligned} & = & P_{arop} \cdot \left[r \cdot o \cdot h_{i} \right] + (1 - P_{arop}) \left[r \cdot n \cdot h_{i} \right] \\ & = & (1 + P_{arop}) \cdot r \cdot h_{i} \\ & = & h_{i} \end{aligned} $ $ \begin{aligned} & = & h_{i} \\ & = & h_{i} \end{aligned} $ $ \begin{aligned} & = & h_{i} \\ & = & h_{i} \end{aligned} $ $ \begin{aligned} & = & h_{i} \\ & = & h_{i} \end{aligned} $	$P_{Arop} \cdot [r \circ hi] + (1 - P_{Arop})[r \circ h \cdot hi]$ $= (1 - P_{Arop}) \cdot r \cdot hi$ $= [r \circ p [h_{Arop}] := hi$ $r \lambda_{i} = \lambda_{i}$ $r \lambda_{i} = \lambda_{i}$	$ \begin{aligned} & \langle P \rangle_{i} = P_{arop} \cdot [r \circ h_{i}] + (1 - P_{arop})[r \circ h_{i}] \end{aligned} $ $ \begin{aligned} & = (1 - P_{arop}) \cdot [r \circ h_{i}] + (1 - P_{arop})[r \circ h_{i}] + (1 - P_{arop})[r \circ h_{i}] \end{aligned} $ $ \begin{aligned} & = (1 - P_{arop}) \cdot [r \circ h_{i}] + (1 - P_{arop})[r \circ h_{i}] + (1 - P_{arop})[r$
$ \begin{aligned} & = \rho_{arop} \cdot \left[ro \cdot h_{i} \right] + (1 - \rho_{arop}) \left[ro \cdot h_{i} \right] \\ & = (1 + \rho_{arop}) \cdot ro \cdot h_{i} \\ & = h_{i} \\ \lambda_{i} & = \lambda_{i} \end{aligned} $ $ \begin{aligned} & = \lambda_{i} \\ & = \lambda_{i} \end{aligned} $ $ \begin{aligned} & = \lambda_{i} \\ & = \lambda_{i} \end{aligned} $	$ \begin{aligned} & = P_{stop} \cdot \left[r \cdot o \cdot h_{i} \right] + (1 - P_{stop}) \left[r \cdot 1 \cdot h_{i} \right] \\ & = (1 - P_{stop}) \cdot \delta \cdot h_{i} \\ & = h_{i} \text{else} \end{aligned} $ $ \begin{aligned} & = h_{i} \\ & = h_{i} \end{aligned} $	$ \begin{aligned} & = P_{Jrop} \cdot \left[\mathcal{V} \circ - h_{i} \right] + (1 - P_{Jrop}) \left[\mathcal{V} \cdot 1 \cdot h_{i} \right] \\ & = (1 - P_{Jrop}) \cdot \mathcal{V} \cdot h_{i} \\ & = h_{i} \text{eld} \\ \lambda_{i} = \lambda_{i} \end{aligned} $
$= P_{drop} \cdot [V \circ hi] + (1 - P_{drop})[V \cdot 1 \cdot hi]$ $= (1 - P_{drop}) \cdot V \cdot hi$ $= hi$ $= hi$ $= hi$	$= P_{arop} \cdot [r \circ hi] + (1 - P_{arop})[r \cdot hi] $ $= (1 - P_{arop}) \cdot r \cdot hi$ $= hi$ $= hi$ $= hi$	$= (1 - P_{Jrop}) \cdot (1$
Parop. [8.0-hi] + (1-parop)[8.1.hi] (1-parop) · 8.hi Elsop [harop]; = hi : [38]	Parap. $[V \circ hi] + (1-parap)[V \circ 1 \cdot hi]$ $(1-parap) \cdot V \cdot hi$ $E V \circ p \cap h \cap$	Parop. [8.0-hi] + (1-parop)[8.1.h] (1-parop). 8.h; Elsop [harop]; = h; : 108
Lop [hrop]; = h; ειρρ Ενορ [hrop]; = h; ειρρ : [ος]	2 (Δορ [h 1 ο ρ]; = h; (1 - ρ 1 ο ρ); = h; (1) ρ (1)	$L_{\text{lop}} \cdot [0 \cdot h] + (1 - p_{\text{lop}})[0 \cdot h]$ $= L_{\text{lop}} \cdot [h_{\text{lop}}] = h;$ $= h;$ $= h;$ $= h$
Parop) · D · h; Elrop [hrop] := h; 2178	$P_{Jrop} = h_i$	ρ·[r·ο·hi] + (1-ρμορ)[σ·1·h] Εμορ [hμορ]; = h; εμορ [hμορ]; = h; εμορ [hμορ]; = h;
$[r \circ h:] + (1-p_{1}n_{p})[r \cdot h:] $	$[r \circ h:] + (1-p_{pop})[r \cdot h \cdot h]$ $[r \circ h:] + (1-p_{pop})[r \cdot h \cdot h]$ $[r \circ h:] + (1-p_{pop})[r \cdot h \cdot h]$ $[r \circ h:] + (1-p_{pop})[r \cdot h \cdot h]$	$[r \circ h] + (1 - \rho_{ln} \varphi) [r \cdot 1 \cdot h] $ $[r \circ h] + (1 - \rho_{ln} \varphi) [r \cdot 1 \cdot h] $ $[r \circ h] + (1 - \rho_{ln} \varphi) [r \cdot 1 \cdot h] $ $[r \circ h] + (1 - \rho_{ln} \varphi) [r \cdot 1 \cdot h] $
$[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $ $[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $ $[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $ $[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $ $[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $ $[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $ $[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $ $[r \circ h :] + (1 - \rho_{1} \circ \rho) [r \cdot h \cdot h] $	$[r \circ -h :] + (1-p_{nop})[\sigma \cdot 1 \cdot h] $ $[r \circ -h :] + (1-p_{nop})[\sigma \cdot 1 \cdot h] $ $[r \circ -h :] + (1-p_{nop})[\sigma \cdot 1 \cdot h] $ $[r \circ -h :] + (1-p_{nop})[\sigma \cdot 1 \cdot h] $	$[r \circ h:] + (1-\rho_{\mu\nu\rho})[\sigma \cdot h:] $
0-h;] + (1-pmp)[0-1.h] (12 7-h; [hmp]; = h; (17)	0-hi] + (1-p,np)[0.1.h] (12 7.h: [h,np]:=h: [217 22	$0 \cdot h : \int + (1 - \rho_{1} \cdot \rho_{2}) \left[\sigma \cdot A \cdot h \cdot \right] $ $(1 \cdot h) \cdot \left[h \cdot \rho_{2} \cdot \rho_{3} \right] = h \cdot \left[h \cdot \rho_{3} \cdot \rho_{3} \right] $
$h_{i}] + (1-p_{i},p_{i})[0.1.h_{i}] $ h_{i} h_{i} h_{i} h_{i} h_{i}	$h: J + (1 - \rho_{J \cap p}) [0 \cdot 1 \cdot h \cdot] $ $h: h: h: en p: e$	$h:] + (1-\rho_{jnop})[o.1.h] $ $h: $ $h: $ $h_{jnop}]: = h: $ lip_{jo}
$i \int + (1 - \rho_{J} \circ \rho) \left[\circ \circ \cdot 1 \cdot h \right] $ $h : \qquad $	$i \int + (1 - \rho_{1} \rho_{2}) [0.1.h] $ h_{i} $\mu_{0} = h_{i}$ $2 \ln 2$	$i] + (1 - \rho_{J \cap Q}) [$
$+(1-\rho_{jnop})[\sigma\cdot 1\cdot h]$ $[[c]$ $\rho]_{i}=h_{i}$ $(1)\rho$	$+(1-\rho_{J},\rho_{0})[\sigma,\Lambda,h]$ $[(1-\rho_{J},\rho_{0})]=h;$ $[(1-\rho_{J},\rho_{0})]=h;$	$+(1-\rho_{jnop})[\sigma\cdot 1\cdot h]$ $[c]$ $[c]$ $[c]$
$\frac{1}{1} = h;$ $\frac{1}{1} = h;$ $\frac{1}{1} = h;$	$\frac{1}{1} = h;$	$\frac{1}{1} = h;$ $\frac{1}{1} = h;$ $\frac{1}{1} = h;$
1-Purap)[0.1.h.] (1c) = h; (1r)	1-Purp)[0.1.h] =h;	1-Purp)[0.1.h.] =h; (12
Punp)[0.1.h] h;	Punp)[0.1.h] (12 h; (17 2)	Punp)[0.1.h] h; enp
(R) [0.1.h]	(1c) [0.1.h]	(R) [0.1.h]
(/c 21) = (/c 21) = 2	(K) (K) (K) (K)	(k (k (n)
217 P	[(/ (/ (/ (/ (/ (/ (/ (/ (/ (217 P
(K (L)	(n.h.) (k	(L)
1. h.] (/c	1. h.]	1. h.] (k
h.] (k	h] (k	h.] (12
J (12		17 20
12	172	12

<u>Q3</u>

c.

```
Perplexity for each epoch:
[598.2366782129152, 258.35566638496755, 186.748500751412,
153.8955940469948, 133.7895546230422, 119.46139157102863,
109.51243179110595, 101.60273449526015, 95.69344691128141,
90.6552240888448, 86.34052603632608, 82.98170482530692, 80.04136627427236,
77.28627441036822, 75.2353598328523, 73.01505584565827, 71.32777081434989,
69.88417041995659, 68.37327530689748, 67.18151719599369]
```

Finished training

Epoch: 20 Validation Perplexity: 104.119