

1) QUESTION 1:

(a)

$$\sum_{i=1}^{\infty} P(\text{infinite sentence } i) =$$

$$\Rightarrow P(\text{infinite sentence } i) = \underset{j=1}{\prod} P(w_{ij} = \text{start}) \prod_{j=1}^{\infty} P(w_{ij} | P_{w_{ij-1}}) =$$

w_{ij} : $i \in \text{Coen} \cup j \in \text{נוסף}$

$$\textcircled{1} = \prod_{j=1}^{\infty} P(w_j | w_{j-1}) < \prod_{j=1}^{\infty} P(w_s | w_k) \stackrel{\textcircled{2}}{=} \prod_{j=1}^{\infty} g \stackrel{\textcircled{3}}{=} 0$$

$$\sum_{i=1}^{\infty} P(\text{infinite sentence } i) < \sum_{i=1}^{\infty} \prod_{j=1}^{\infty} g < \sum_{n=1}^{\infty} g^n = \lim_{n \rightarrow \infty} \frac{n}{g^n} \stackrel{\text{e בערך}}{\downarrow} 0$$

הוכחה סימetric

* נציין כי נעה כוותחיתר בדרכו אחרת.

ונכון לכך, וכאן כוותחוויות יונאנס stop

$$\forall \omega_i \sum_{j=1}^{n-1} P(w_j | \omega_i) + P(\text{stop} | \omega_i) = 1 \quad \text{ולכן, } \sum_{j=1}^{n-1} P(w_j \neq \text{stop} | \omega_i) < 1$$

ולכן $0 \leq P(w_s | \omega_k) < 1$ נוצרת קיון כי (ω_k, ω_s)

כך. מכאן $0 \leq P(w_s | \omega_k) < 1$ נוצרת קיון כי (ω_k, ω_s)

כך. מכאן $0 \leq P(w_s | \omega_k) < 1$ נוצרת קיון כי (ω_k, ω_s)

(b) $W = \{\text{"START"}, \text{"STOP"}, \text{"x"}\}$

• *po Cadeu, para d'Idre*

$$P(\text{STOP} | w_1, \dots, w_n) = \frac{1}{n+1} \quad \forall n \in \mathbb{N}$$

$$P(E | \omega_1, \dots, \omega_n) = \frac{n}{n+1}$$

(כ) אם זכרת מה שכתוב בפונט נתקדם

הנוכחות שדרוגן כ'אלה' מושג ערך נקי של ריבוי וריבוי.

$$\lim_{n \rightarrow \infty} P(\text{START}, w_1 = "E", w_2 = "E", \dots, w_n = "E") = \lim_{n \rightarrow \infty} \left(\frac{n}{n+1}\right)^n =$$

$$\left(\frac{n+1-1}{n+1}\right)^n = \frac{\left(1 - \frac{1}{n+1}\right)^{n+1}}{\left(1 - \frac{1}{n+1}\right)} = \frac{\frac{1}{e}}{1} = \frac{1}{e}$$

2) QUESTION 2 -

a) Unigram language model:

Given w_1, \dots, w_n (וְהַלְאָנֶם w_1, \dots, w_n הַיְהוּ בְּפִרְשֵׁת עֲדָמָה) where w_i (וְהַלְאָנֶם w_i הַיְהוּ בְּפִרְשֵׁת עֲדָמָה) is the i^{th} word in the sequence. The probability of w_i is given by $P(w_i | w_1, \dots, w_{i-1})$. The probability of the entire sequence is given by $P(w_1, \dots, w_n)$.

$$P(w_1, \dots, w_n) = P(w_1) P(w_2 | w_1) \dots P(w_n | w_1, \dots, w_{n-1})$$

$$P(w_i) = \frac{\text{count}(w_i)}{\text{total words}} = \frac{\text{count}(w_i)}{M} \quad w_i \in \{w_1, \dots, w_n\}$$

בכדי w_i ב- w_j כה-זאת count(w_i) מוגדר כמספר הפעם בה נתקל ב- w_i . count(w_j) מוגדר כמספר הפעם בה נתקל ב- w_j . count(w_i, \dots, w_n) מוגדר כמספר הפעם בה נתקל ב- w_i, \dots, w_n .

$$P(w_1, \dots, w_n) = \prod_{j=1}^n P(w_j)$$

He went where there were more opportunities.

ג'ונליזציונ

: המבנה של פונקציית הסתברות

count(were) > count(where) \Rightarrow פונקציית הסתברות מילאה יותר ב-were מאשר(where). \therefore פונקציית הסתברות מילאה יותר ב-were מאשר(where).

count(where) > count(were) \Rightarrow פונקציית הסתברות מילאה יותר ב-where מאשר(were). \therefore פונקציית הסתברות מילאה יותר ב-where מאשר(were).

51c count (where) = count (were) פ"ק"ה א' נתקן -
! 0.5 נסמן בפונט **①** כ where מילוי פונט 0.5
were כ **②** מילוי פונט 0.5 נסמן בפונט **③** כ
ו^תז' ח' כ' $(0.5)^2 = \underline{0.25}$ כ' 0.25 נסמן בפונט **④**

Bigram language model: P (b)

רַבְנִים וְכָסְפָנִים גַּם מֵהֶם וְקַיְמָם

: $\beta_{K-1} \mid \beta_{K-1} \omega_{K-1} = \omega_1$

$$P(W_k = w_j | W_{k-1} = w_i) = \frac{\text{count}(w_j, w_i)}{\sum_{j' \in [n]} \text{count}(w_i, w_{j'})}$$

הכו w_j, w_i הינה סט נספנ' ב- NCF .
בנוסף, w_j, w_i נספנ' ב- count .

וילא. וילא אם יש לנו מילון של כל המילים בפונטיקה
 וילא. וילא אם יש לנו מילון של כל המילים בפונטיקה $p(w_j | w_i) = 0$
 • על כל w_1, \dots, w_n כמו בסיס

ולא הולא went were יכולות נסיעה על כל המילים
 ו הולא were לא יופיע בדרכו על כל המילים וילא.
 • ערך נסיעות Back off models גויס בסיס ולא -

- If the bigram (x_{m-1}, x_m) didn't appear enough times, we back-off to conditioning only on x_{m-1} :

lecture 2 : N

: מילוי

$$p(w_i | w_{i-1}) = \begin{cases} \text{Use bigram prob , if } \text{counter}(w_{i-1}, w_i) > 0 \\ p(w_i) \quad \text{else} \end{cases}$$

וילא או מילוי לא, אם סינס יש לנו מילוי כל מילא
 מילוי מילוי סינס מילא מילא מילא. מילוי מילוי סינס

3) QUESTION 3 -

a)

$$(*) N = 0 \cdot N_0 + 1 \cdot N_1 + 2 \cdot N_2 + \dots + C_{\max} \cdot N_{C_{\max}}$$

$$\sum_{c=1}^{C_{\max}} N_c \cdot P_c = \sum_{c=1}^{C_{\max}} \frac{(c+1) N_{c+1}}{N_c N} \cdot \cancel{N_c} = \sum_{c=1}^{C_{\max}} \frac{N_{c+1} (c+1)}{N} =$$

$$\frac{1}{N} \sum_{c=1}^{C_{\max}} N_{c+1} (c+1) = \frac{1}{N} \left(N_2 \cdot 2 + N_3 \cdot 3 + \dots + \underbrace{N_{C_{\max}+1} (C_{\max}+1)}_{N - N_1 \cdot 1} \right) =$$

$$\frac{1}{N} \left(N_2 \cdot 2 + N_3 \cdot 3 + \dots + N_{C_{\max}} C_{\max} \right) = \frac{\overset{\circ}{N} - N_1}{N} = 1 - \frac{N_1}{N}$$

(b) NON C(ω) ! ISN'T IT JUST A PICTURE IN THE BOOK ?
 מילוי הינה לא מושג בפונקציית נספחים, אלא בפונקציית נספחים שמשתמשה בפונקציית נספחים.

$$q_{\text{add-1}}(\omega) = \frac{c(\omega) + 1}{\sum_{\omega'} (c(\omega') + 1)} = \frac{c(\omega) + 1}{\sum_{\omega'} c(\omega') + |V|}$$

$c+1$ אנו מודדים את סכום נספחים הינה
 $\sum_{\omega'} c(\omega') + |V|$

$$\mu = \frac{c(\omega)}{\sum_{\omega'} c(\omega')} < \frac{c(\omega) + 1}{\sum_{\omega'} c(\omega') + |V|}$$

$$c(\omega) \left(\sum_{\omega'} c(\omega') + |V| \right) < (c(\omega) + 1) \sum_{\omega'} c(\omega')$$

$$C(\omega) \cdot \sum_{\omega'} C(\omega') + C(\omega) \cdot |V| < C(\omega) \sum_{\omega'} C(\omega') + \sum_{\omega'} C(\omega')$$

$$C(\omega) \cdot |V| < \sum_{\omega'} C(\omega') / |V| \leq \bar{C}(\omega)$$

$$\frac{C(\omega)}{\sum_{\omega'} C(\omega')} < \frac{1}{|V|}$$

כ' שפה. נניח מעתה ש $C(\omega)$ נקי סוד על V , אוניברסיטאי ש $\frac{1}{|V|}$ נקרא smoothed estimate, ו $\bar{C}(\omega)$ נקרא

נניח כ' ש $\bar{C}(\omega)$ מוגדר כך, $\frac{C(\omega)}{\sum_{\omega'} C(\omega')}$

היפוך ימינו ש $\bar{C}(\omega)$ מוגדר כ'

$$\bullet \bar{C}(\omega) = \frac{1}{|V|}$$

(c) $\rho \sigma \text{ נס} \text{ מפה } \rho \text{ ש } \rho'' \text{ נס} \text{ היפר}$
 $\text{היפר } \frac{C}{N} < \frac{(C+1) N_{C+1}}{N_C \cdot N}$

$$\frac{C}{N} > \frac{(C+1) N_{C+1}}{N_C \cdot N} \quad \text{הנ} \text{ היפר} \text{ מפה} \sqrt{N}$$

$$: k = \frac{C_0}{N} \sqrt{N}$$

הנ היפר מפה $\frac{C_0+1}{N}$ מפה מפה \sqrt{N} :

$$\frac{(C_0+2) N_{C_0+2}}{N_{C_0+1}} < \frac{C_0+1}{1}$$

✳ $(C_0+2) N_{C_0+2} < (N_{C_0+1})(C_0+1)$

מינימום $\frac{C_0+2}{N_{C_0+1}}$ מפה, מפה מפה מפה מפה
 $\rho'' \text{ נס} 2 \quad ! \quad \rho'' \text{ נס} 4 \quad \text{היפר} \sqrt{N} \quad 50$
 $C_0=2 : \text{היפר} \rho'' \text{ נס} 3 \quad \text{היפר} \rho'' \text{ נס} 2$

$$(2+2) \cdot N_{2+2} < N_{2+1} \cdot (2+1)$$

$$4 \cdot 50 < 2 \cdot 3$$

$$200 < 6$$

היפר מפה מפה מפה מפה מפה מפה מפה
 מפה מפה מפה מפה מפה מפה מפה מפה
 מפה מפה מפה מפה מפה מפה מפה מפה
 מפה מפה מפה מפה מפה מפה מפה מפה

4) QUESTION 4 8

(a) ההוכחה מוכיחת כי לא ניתן לחלק את המילים בcount(w_{i1}, w_{i2}, w_j)
בנוסף לכך, סימן ה=count(w_{i1}, w_{i2}, w_j)

$$P(X_m = w_j | X_{m-1} = w_{i1}, X_{m-2} = w_{i2}) = \frac{\text{count}(w_{i2}, w_{i1}, w_j)}{\sum_j \text{count}(w_{i2}, w_{i1}, w_j)}$$

: $P(w_{i1}, w_{i2}, \dots, w_{in}) = S_i$ לפיכך מוכיחים

$$P(w_{i1}, w_{i2}, \dots, w_{in}) = P(w_{i1}) \cdot P(w_{i2} | w_{i1}) \cdot \prod_{j=3}^n P(w_{ij} | w_{ij-2}, w_{ij-1})$$

(b) "How many words are there?" כ"א, גמרנו שcount(w_{i1}, w_{i2}, ..., w_{in}) מוכיח שw_{i1}, w_{i2}, ..., w_{in} מופיעות מחרוזת אחת. כ"א, מוכיח שw_{i1}, w_{i2}, ..., w_{in} מופיעות מחרוזת אחת. כ"א, מוכיח שw_{i1}, w_{i2}, ..., w_{in} מופיעות מחרוזת אחת.

"Steven speaks a lot!"

מי שSteven מזכיר נגין שSPINDE מוכיח שSPINDE מזכיר Steven בcount(w_{i1}, w_{i2}, ..., w_{in}). "Steven" מזכיר SPINDE בcount(w_{i1}, w_{i2}, ..., w_{in}). "Steven" מזכיר SPINDE בcount(w_{i1}, w_{i2}, ..., w_{in}).

(c) "The group of tourists was heading south."

מי שgroup of students מזכיר SPINDE מוכיח שgroup of students מזכיר SPINDE. "group of students" מזכיר SPINDE בcount(w_{i1}, w_{i2}, ..., w_{in}). "group of students" מזכיר SPINDE בcount(w_{i1}, w_{i2}, ..., w_{in}). "group of students" מזכיר SPINDE בcount(w_{i1}, w_{i2}, ..., w_{in}). "group of students" מזכיר SPINDE בcount(w_{i1}, w_{i2}, ..., w_{in}). "group of students" מזכיר SPINDE בcount(w_{i1}, w_{i2}, ..., w_{in}).

נורספְּרָדָה עַל "גִּירְגִּירָה" בְּמִינְגָּן, שֶׁשְׁמֵלְכָה וְעַמְּקָם
וְעַמְּקָם, מִן-טַבְּרָא 4-gram מִינְגָּן תְּמִימָה בְּמִינְגָּן 3
. "גָּרְגָּרָה" בְּ "גָּרְגָּרָה"

(5) Question 5 :

valid valid valid valid valid valid valid

- ① The cat jumped suddenly walked out the door.
 invalid

: פ' ג' נס צי ל' פ' כ' ו' נ' פ' ק' ו' פ' ג' נ'

- The cat slept jumped suddenly
 - She suddenly walked out the door.

2

The bag of clothes were, thrown from the window.

invalid

: **P'is'n j'el'e** ↳ **p'is'nac'ip p'is'nac'ip p'is'nac'ip**

The bag of clothes is next to you.

Bags of clothes were thrown from the window.

- ③ "The cat was sat on the mat is sleeping"
'piptp for corn'

: p'f'n r'c' n'c' p'f'one p"p'p'p' p'Coen

The cat was sat on the mat.

The cat on the mat is sleeping.

Practical Part :

```
Task 2:  
I have a house in the  
Task 3-a:  
Bigram probability for 'Brad Pitt was born in Oklahoma': -inf  
Bigram probability for 'The actor was born in USA': -29.712  
Task 3-b:  
The perplexity of both sentences: inf  
Task 4:  
Linear Interpolation with weights 1/3 and 2/3  
The probability for 'Brad Pitt was born in Oklahoma': -36.209  
The probability for 'The actor was born in USA': -31.027  
The interpolated perplexity of both sentences: 271.240
```

Notes on implementation details and results:

- 1) All probabilities reported above are log probabilities.
- 2) To calculate a sentence's interpolated probability, we first *applied the interpolation formula on the original probability* of the pair of tokens/ single token then we took the log of the result. Then for the whole sentence's log probability, we summed the above explained results for each pair of consecutive tokens/ single token.
- 3) To compute the perplexity for the interpolated model, we took the negative average of the log probabilities for all sentences and applied exponentiation.