

203765698 10111 - 1 01211 - 1 826 201 201

לעתה נסמן ביגראם $p_{\gamma}(w_i, w_{i+1})$ ו- $\sum_{w \in V} p_{\gamma}(w, w) = 1$ (I)

$$\forall w \in V \quad \sum_{w' \in V} p_{\gamma}(w, w') = 1 \quad (\text{II})$$

$$\forall w \in V \quad p_{\gamma}(w, \text{STOP}) > 0 \quad (\text{III})$$

$$\forall n \quad \sum_{k=1}^{\infty} \sum_{w_1, w_k \in V} p_{\gamma}(\text{START}, w_1, \dots, w_k, \text{STOP}) = 1 \quad (\text{IV})$$

לעתה נסמן ביגראם $p_{\gamma}(w_i, w_{i+1})$ ו- $\sum_{w \in V} p_{\gamma}(w, w) = 1$ (I)

$$V = \{w_1, w_2, \dots, w_n\}$$

נניח. ($w_i \neq \text{STOP}$) $\Rightarrow p_{\gamma}(w_i, w_{i+1}) > 0$

פ.מ. $p_{\gamma}(\text{START}, w_1, \dots, w_n, \text{STOP}) = p_{\gamma}(\text{START}, w_1, \dots, w_{n-1}, \text{STOP}) + p_{\gamma}(w_n, \text{STOP})$

$$0 < p < 1 \quad \text{ס.} \quad p = \min_{1 \leq i \leq n} p_{\gamma}(\text{STOP} | w_i) \quad (\text{No. 1})$$

$\text{START}, w_{i_1}, w_{i_2}, \dots, w_{i_k}$: מוגדרות $i_1 < i_2 < \dots < i_k$

$$p_{\gamma}(i_t \geq k) \leq p_{\gamma}(w_{i_1} \neq \text{STOP} \wedge w_{i_2} \neq \text{STOP} \wedge \dots \wedge w_{i_k} \neq \text{STOP}) =$$

$$= (1 - p_{\gamma}(\text{STOP} | w_{i_1})) \cdot \dots \cdot (1 - p_{\gamma}(\text{STOP} | w_{i_k})) \leq (1 - p)^{k-1}$$

$$(1 - p)^{\frac{k-1}{k-1}} \rightarrow 0 \quad \text{ס.} \quad 0 < p < 1 \quad \text{ס.} \quad \text{No. 2}$$

ו.ג. $\lim_{k \rightarrow \infty} p_{\gamma}(i_t \geq k) = 0$

ולכן $p_{\gamma}(i_t \geq k) = 0$

STOP תְּנִפְתָּחָה נִפְתָּחָה עַל

$\text{START}, w_1, \dots, w_k$ DANN SIE $k \in \mathbb{N}$ SIE $\rightarrow w_k$ SIE

$$P_{\gamma}(\text{STOP} \mid \text{START}, w_{i_1}, \dots, w_{i_k}) > 0 \quad ; \quad i_1, \dots, i_k \in N$$

START, W_1, \dots, W_k ↗ 20 מילון כ' ↘ מילון מילון ↗ 120 ↘ 10

$$P_g(\text{STOP} | \text{START}, w_{i_1}, \dots, w_{i_k}) = \frac{1}{(k+n)^2}$$

$$P_{\delta}(w_i | \text{START}, w_1, \dots, w_{i-1}) = \frac{1}{n} \cdot \left(1 - \frac{1}{(k+1)^2}\right) \quad \text{STOP} \neq w_i \in V$$

|V|=n 100%

ללאם כ' (טווין) פולימר נייר ערך קבוצתית וריאנטים של מילויים

$$\begin{aligned}
 P_{\delta}(\text{START}, w_{i_1}, \dots, w_{i_K}) &= P_{\delta}(w_{i_2} \neq \text{STOP} \wedge w_{i_3} \neq \text{STOP} \wedge \dots \wedge w_{i_K} \neq \text{STOP}) = \\
 &= \left(1 - P_{\delta}(\text{STOP} | \text{START}, w_{i_1})\right) \cdot \dots \cdot \left(1 - P_{\delta}(\text{STOP} | \text{START}, w_{i_2}, \dots, w_{i_{K-1}})\right) = \\
 &= \left(1 - \frac{1}{2^2}\right) \cdot \left(1 - \frac{1}{3^2}\right) \cdot \dots \cdot \left(1 - \frac{1}{(K+1)^2}\right) = \prod_{i=1}^K \left(1 - \frac{1}{(K+i)^2}\right)
 \end{aligned}$$

$$\lim_{k \rightarrow \infty} \left(1 - \frac{1}{(k+1)^2}\right) \xrightarrow{k \rightarrow \infty} \frac{1}{2}$$

STOP reading book stories in Hebrew with us

828 N.Y. 1000 P.M. 1971 Nov 11, 1971

D 05:17 V P'RN 31K 178 (2

For NDC Spelling correction 78% > Unigram 70% RN

P3 : W_1, W_2, \dots, W_k ED GDN $\rightarrow P$

$$P_\delta(W_1, W_2, \dots, W_{1k}) = \prod_{i=1}^k P_\delta(W_i)$$

$$P_g(W) = \frac{c(w)}{|D|} \quad w \in V \quad \text{so } c(w)$$

$c'(w) > 0$ ו- $c(w) = 0$ מינימום נסיבי.

... שאלת היברידיים נוגע בפ' נס' = 10) פ' נס' נוגע בפ' נס' = 10)

Spalling corrector $\geq w_1 w_2 \dots w_k$ (see 178, 18)

Next we want to insert in $i \in \{1, \dots, k\}$ in π , σ

· Wi; (-S where, were)

$D = \{w_1, \dots, w_k \mid 1 \leq i \leq k\}$. $w_i = \begin{cases} w_i & \text{if } i \in \{1, \dots, k\} \\ \text{"where" or "what", else} & \end{cases}$

$$|D'| = 2^{\ell} \quad \text{and} \quad D' \subseteq D \quad \text{sk}$$

W_1^*, \dots, W_n^* First term (Gen) $\approx 15\%$ Spallling corrected \rightarrow 5%

$$w_1^*, \dots, w_n^* = \arg \max_{w_1, \dots, w_k \in D} \{ P_g(w_1, \dots, w_k) \}$$

c ↗

↑
→ \sqrt{N} e nötig
→ unigram

"He went where there were more opportunities" (sent 170)

• P77 N1 3 5 2021

ink \propto vis. like \rightarrow sc- p_x ("where") $> p_x$ ("were") (I)

"Where" → "Were" (past) de gain de l'acte verbale

בנ' הילא פהו גאנט אונד זאגט עליי זונט נוואר דער גאנט

"wheel" $\sqrt{3}$ de però

Given that $x \in S^1$ there is $\delta > 0$ s.t. $P_\delta(\text{"Where"}) < P_\delta(\text{"Were'})$ (II)

"Were" → "Where" גותְּ בָּאָה בְּ וְגַתְּ בָּאָה בְּ

לעתה נזכיר את היחסים בין מילים ופונטים. סביר לחשוב ש $P_{\text{y}}(\text{"Where"}) = P_0(\text{"Where"})$ (III)

הוילר $\in \Sigma$ מופיע בפונט "Where", "Were" בפחות מ-5% מהתוצאות.

אם נזכיר את היחסים בין מילים ופונטים - $P(\text{"Where"}, \text{"Were"}) = P_{\text{y}}(\text{"Where", "Were"})$

ובוילר מופיע בפונט "Where", "Were" בפחות מ-5% מהתוצאות.

"Where" הנקה בפונטיקה נסוב עליה

• תרנגולת מוגדרת כטירוף עז של קוקו.

D 001111 V P 00N 31k 110 (r)

Kim NBC Spelling collector → 75% → bigram 70% 52%

$\therefore W_1, W_2, \dots, W_k \in D$ 6aen vpo

$$P_{\delta}(W_1, W_2, \dots, W_k) = P_{\delta}(\text{START} | W_1) \cdot \prod_{i=2}^k P_{\delta}(W_i | W_{i-1}) \cdot P_{\delta}(\text{STOP} | W_k)$$

$$P_f(W_i \text{ START}) = \frac{\# \text{ of sentence in } D \text{ that start with the word } w_i}{\# \text{ of sentence in } D}$$

$$\beta_j(\text{STOP} | W_k) = \frac{\# \text{ of sentence in } D \text{ that end with the word } W_k}{\# \text{ of sentence in } D}$$

$$\beta_0(W_i | W_{-i}) = \frac{\text{count (appearances of } W_{i-1}, W_i \text{ in D)}}{\sum_{W \in V} \text{count (appearances of } W_{i-1}, W \text{ in D)}}$$

Next year bigram \rightarrow din > $c\bar{v}_d$ spelling corrector \rightarrow

(הנתק ע"י אוניברסיטת וינה). UniJAM נפתחה ב-2017 כחלק מ-PIKA.

אנו מודים לך על תרומותך ותומך בפנינו, וברוחם של אחדים,

הוּא יְמִינֵךְ (ב) גָּדוֹלָה (ב) כַּלְבָּה נֶכֶן-נֶכֶן (ב)

$S := w_1, \dots, w_n$: סדרת מילים הנעשית מילים בפונטיקה. נסמן w_i

$$V_{\text{where}} = \{w \in V \mid (w, \text{where}) \in S\}$$

$$V_{\text{were}} = \{w \in V \mid (w, \text{were}) \in S\}$$

$$V_{\text{where}} \cap V_{\text{were}} = \emptyset$$

$$P_S(\text{where}|w_1) > P_S(\text{where}|w_2) \quad w_2 \in V_{\text{where}} \quad w_1 \in V_{\text{where}}$$

$$P_S(\text{were}|w_1) < P_S(\text{were}|w_2)$$

במקרה של w_1 לא ניתן למליץ על where כי לא מופיע בbigram

במקרה של w_2 ניתן למליץ על where כי מופיע בbigram

לכן מילוי המילה where מושג באמצעות bigram

bigram מוגדר כ

המילה w_1 מוגדרת כbigram (w_1, w_2) , $w_1, w_2 \in V$

bigram מוגדר כbigram (w_1, w_2) שקיים בdataset D

bigram מוגדר כbigram (w_1, w_2) שקיים בdataset D וקיים בdataset D'

$P(w_1, w_2 | D) > P(w_1, w_2 | D')$

$P(w_1, w_2 | D) = 0$ $\rightarrow w_1, w_2 \in V$

bigram מוגדר כbigram (w_1, w_2) שקיים בdataset D

bigram מוגדר כbigram (w_1, w_2) שקיים בdataset D וקיים בdataset D'

bigram מוגדר כbigram (w_1, w_2) שקיים בdataset D וקיים בdataset D'

bigram מוגדר כbigram (w_1, w_2) שקיים בdataset D וקיים בdataset D'

bigram מוגדר כbigram (w_1, w_2) שקיים בdataset D וקיים בdataset D'

0.0777 (Word Error) \approx 0.7% PESQ: 5 - 3.77 = A (No) (3)

- 11 N, K, J

Good-Turing מילוי מילויים נקיים - $P(W)$

$$\sum_{\text{WEA}} P(A) = 1 - P_{\text{unseen}} \quad : \underline{\text{def}} \quad (\text{c})$$

∴ $\forall i \in \{1, 2, \dots, n\}$ $C_{max} \leq A_i \leq C_{max}$

Definita funzione di regressione polinomiale di grado n è $y = A_n : \mathbb{R}^n \rightarrow \mathbb{R}$

• P'NQ C 7137 NQNE

$$|\text{Ad} = N_c \quad \text{prin as sc}$$

$$P(W) = \frac{(C+1) N_{C+1}}{N_C \cdot N} \quad W \in A_C \quad \text{Sisi}$$

∴ f(x) > 0

$$\text{Punrecio} + \sum_{w \in A} P(A) = \frac{N_h}{N} + \sum_{c=1}^{C^{\max}} N_c \cdot \frac{(C+1) N_{c+h}}{N_c \cdot N}$$

$$= \frac{1}{N} \cdot \left[\sum_{c=1}^{C_{\max}} c \cdot N_c \right] = \frac{N}{N} = 1. \quad \text{OK!}$$

(I) \hat{N}_0 のノイズを考慮した Add-one estimate の計算

כגנום גזעינו ה.א.מ נס

$$q_{\text{add-one}}(w) = \frac{c(w) + 1}{\sum_{w' \in A} (c(w') + 1)} = \frac{c + 1}{N + |A|}$$

MLE of $p(w)$ given observed w is $\frac{N_c}{N}$ if $w \in A$ (II)

$$q_{MLE}(w) = \frac{\text{count}(w)}{\sum_{w \in A} \text{count}(w)} = \frac{c}{N} \quad \text{if } w \in A$$

if threshold $\mu \rightarrow \infty$ (3)

$$q_{MLE}(w) < q_{add-one}(w) \Leftrightarrow \frac{c}{N} < \frac{1+c}{N+|A|} \quad (\star)$$

$$c(N+|A|) < (1+c)N \Leftrightarrow cN + c|A| < N + cN \Leftrightarrow$$

$$c|A| < N \Leftrightarrow c < \frac{N}{|A|} \Rightarrow \mu = \frac{N}{|A|} \quad \text{if } w \in A$$

$$q_{add-one}(w) > q_{MLE}(w) \Leftrightarrow c < \mu \quad \text{since } w \in A, \forall i \neq j, s_i \neq s_j$$

$$q_{add-one}(w) < q_{MLE}(w) \Leftrightarrow c > \mu \quad \text{since } w \in A, \forall i \neq j, s_i = s_j$$

$$\forall w \in A \quad MLE(w) \geq GT(w) \quad \text{by defn of GT} \quad (\star)$$

but $c < \mu$ $\Rightarrow c < N_c/N$ since $w \in A$ so $w \in N_c$

$$\frac{c}{N} > \frac{(c+1)N_{c+1}}{N_c \cdot N} \Leftrightarrow c \cdot N_c > c \cdot N_{c+1} + N_{c+1}$$

$$\Leftrightarrow c(N_c - N_{c+1}) > N_{c+1} \quad (\star\star)$$

$$C(A) = C(B) = C(C) = 1 \quad \text{and } A = \{A, B, C, D\} \text{ disjoint sets} \quad \text{so } c=1$$

$$C(D) = 2$$

$$N_c = 0 \quad c \geq 3 \quad \text{so } N_2 = 1, N_1 = 3 \quad \text{sk}$$

$$1 \cdot (N_1 - N_2) = 3 - 1 = 2 > 1 = N_2 \quad \text{so } c=1$$

$$2 \cdot (N_2 - N_3) = 2 \cdot (1 - 0) = 2 > 0 = N_3$$

$$MLE(w) \geq GT(w) \quad \text{if } w \in A \quad \text{since } 1 \leq N_c \leq N$$

triangular model of man (K4)

$$P_{\gamma}(W_1=x_1, \dots, W_n=x_n) = \prod_{i=3}^n P_{\gamma}(W_i=x_i | W_{i-2}=x_{i-2}, W_{i-1}=x_{i-1})$$

$j \geq i+3 \Rightarrow 1 \leq i < j \leq n$ \Leftrightarrow $i, j \in \{1, 2, \dots, n\}$

וינט פְּנַיְתִּים וְלֹא מֵרָאֶה וְלֹא שְׁמַעַת, וְלֹא בְּנִימָנָה

↳ $\sim N(0, \sigma^2)$ W_{j-2}, W_{j-1}

$$P_\delta(W_j = X_j \mid W_{j-2} = X_{j-2}, W_{j-1} = X_{j-1}, W_i = X_i) =$$

$$P_8 (W_j = X_j \mid W_{j-2} = X_{j-2}, W_{j-1} = X_{j-1})$$

"The sun rises in the east" : הַבָּקָר GCN (ג)

"בְּאֶשְׁתָּוֹת -ic בְּסִיק בְּאֶשְׁתָּוֹת" : בְּאֶשְׁתָּוֹת GCN

רבים גרים כ' יתרכז תרמונת הילך בפניהם ויביאו מכך מושג של מילוי מטרת הילך.

בנ"ה (bigram) bigram א בינה מילויים נס

"Sun rise" שָׁמֶן עַתְּכִי בְּגַתְּהָרָה מֵאַתְּכִי "it" עַתְּכִי

"The sun that burn my planets rises in the east"; Fuk? 62Cn (c)

נאה בזאת: "האנדרטָה עַל יְדֵי אֲבוֹתֵינוּ"

פִּיכְלָעַד גַּיְמָה פִּסְגָּה שֶׁבְּנֵי בְּנֵי לְאַנְגָּרָה אֲלֹהִים וְבָשָׂר

Երկրագիրը պահանջում է ուղղակի և անդամակի համապատասխան լուծում:

"... my plants file ..." : נייר זיכרונות

you you're they're

5-gram တွေမှာ ၂၂၃၅ မျက် နှင့် ၁၇ၫ

3-gram 봄날의 꽃다발 → 767 (KL12) 788

(II) "The birds I saw last night sings beautifully"

কে কোন গুরু পদি "fing" এর স্বর রিং অনেক সহজ

וְיַעֲשֵׂה יְהוָה

(II) - 171 60en 575 "Sily" > "Sihoy" 186 > (I) 60en? "J'sn) p/c

(II) $\text{S}_N \rightarrow \text{S}_N$ "sings" again too (I) $\text{S}_{\text{C}, N} \rightarrow \text{S}$ so S pd

פָּנִים אֵלֶיךָ תְּבַרְכֵהוּ.

62c) If you say "bird" again as "3 i r d s" again, but (II) goes > 3 (r) PK

→→→→→ PJ's "Sing beautifully" -! "Night sing" P2 sk , (II) - (I)

(III) $\text{GaN} \rightarrow \text{AlN}$ か

לכל אחד ואנשיותו יתאפשרו מילוי-

הtuple מוגדר כטביעה של אובייקטים, וטביעה היא סדרה של אובייקטים.

• Judges → פִּתְרָן קָדָם פִּגְדֵּן חַזְנָן וְעַד...

10) N-gram P(N) \geq 2-gram p(2) but with N > 2 it is not true.

$\Delta = 4,5 \text{ rad} - 2\pi \sin \Delta / k$ in deg

לפניהם נתקל בדרכם מושב צדקה וצדקה נתקל בדרכם לפניהם.

... הַיְלָדִים בְּבֵית אֶתְנָא כְּבָבֵבָה וְבְבֵבָה כְּבָבֵבָה...

Exercise 1: Practical part – Bar Rousso 203765698

- Question2:

According to the Bigram model, the most likely word to continue the sentence "*I have a house in*" is the word "**the**".

- Question3:

According to the Bigram model:

Probability of ("Brad Pitt was born in Oklahoma") = 0

Probability of ("The actor was born in USA") $\approx 1.305e^{-13}$

Perplexity for both sentences = ∞

- Question 4:

According to the Bigram model:

Probability of ("Brad Pitt was born in Oklahoma") $\approx 6.85e^{-24}$

Probability of ("The actor was born in USA") $\approx 3.4^{-17}$

Perplexity for both sentences = 2008