$c \simeq v_j \rightarrow \alpha \simeq 1 \rightarrow \frac{e^{k_i q}}{\sum_{j=1}^{n} e^{k_j q}} \simeq 1 \rightarrow k_j q \gg k_i q$ for $j \neq i$

(1a)

$c \simeq \frac{1}{2}(v_a + v_b) \rightarrow \alpha_a = \alpha_b = [\frac{1}{2}, \frac{1}{2}, \cdots]$

$\alpha_i \; i \neq b, i \neq a = [0, 0, \cdots]$

(1b)

$k_i \Rightarrow$ one hot vectors

$\hookrightarrow q = \beta(k_a + k_b), \beta > 0$

$\alpha_b = \alpha_a = \frac{e^{\beta}}{2e^{\beta}} = \frac{1}{2}$ $\quad \alpha_i \; i \neq b, i \neq a = 0$

$\hookrightarrow c = 0v_1 + 0v_2 + \cdots 0v_i \cdots +$

$\frac{1}{2}v_a + \frac{1}{2}v_b + \cdots$

$0v_n$

$\rightarrow c = \frac{1}{2}(v_a + v_b)$

(1 c i)

$$q = \beta(u_a + u_b), \beta \gg 0$$

since $\Sigma_i = \alpha I$ and α is close to zero → then we can say x_i is unperfect one hot because u_i is one hot. So with large value of β this problem is approximately the same as previous one.

1 c ii

$$k_a \sim \mathcal{N}(\mu_a, \Sigma_a)$$

$$\Sigma_a = \frac{1}{2}(\mu_a \mu_a^T) + \alpha I$$

$$\alpha \simeq 0$$

$$\rightarrow k_a \simeq \gamma \mu_a$$

$$\gamma \sim \mathcal{N}(1, \frac{1}{2})$$

ex: $\gamma = [0.3, 0.7, 0.5, \ldots]$

$$q = \beta (u_a + u_b)$$

$$k_a \cdot q = \gamma_a \beta$$

$$k_b \cdot q = \gamma_b \beta$$

$$k_i \cdot q = 0$$

$$i \neq a, \; i \neq b$$

$$\rightarrow \alpha_a = \frac{e^{\gamma_a \beta}}{e^{\gamma_a \beta} + e^{\gamma_b \beta}}$$

$$\alpha_b = \frac{e^{\gamma_b \beta}}{e^{\gamma_a \beta} + e^{\gamma_b \beta}}$$

$$c \simeq \frac{e^{\gamma_a \beta}}{e^{\gamma_a \beta} + e^{\gamma_b \beta}} v_a + \frac{e^{\gamma_b \beta}}{e^{\gamma_a \beta} + e^{\gamma_b \beta}} v_b$$

if $|\gamma_a| > |\gamma_b|$

↳ c looks like v_a

if $|\gamma_b| > |\gamma_a|$

↳ c looks like v_b

(1 d i)

$$c = \frac{1}{2}(c_a + c_b)$$

$$c \simeq \frac{1}{2}(v_a + v_b)$$

$$\Rightarrow v_a + v_b \simeq c_a + c_b$$

$$\Rightarrow \eta_a = \beta_a \nu_a \qquad \eta_b = \beta_b \nu_b$$

$$\beta_a, \beta_b > 0$$

$$\alpha_a = \frac{e^{\beta_a \mu_a}}{0 + 0 + e^{\beta_a \mu_a}} = 1$$

$$\hookrightarrow c_a = v_a$$

$$\alpha_b = \frac{e^{\beta_b \mu_b}}{0 + 0 + e^{\beta_b \mu_b}} = 1$$

$$\hookrightarrow c_b = v_b$$

$$c = \frac{1}{2}(v_a + v_b)$$

Id ii (like Ic ii)

$$K_a = \gamma_a \mu_a$$

$$K_b = \gamma_b \mu_b$$

$$q_a = \beta_a \mu_a \quad q_b = \beta_b \mu_b$$

$$\beta_a \cdot \beta_b > 0$$

$$K_a \cdot q_a = \gamma_a \beta_a \rightarrow \alpha_a \equiv 1$$

$$K_b \cdot q_b = \gamma_b \beta_b \rightarrow \alpha_b \equiv 1$$

$$c_a \equiv v_a \qquad c_b \equiv v_b$$

$$c \equiv \tfrac{1}{2}(v_a + v_b)$$

(e i)

$$\alpha_2 = \frac{e^{|v_a|}}{e^0 + e^0 + e^{|v_a|}} \simeq 1$$

since β is very large and $|v_n| = 1/\beta$

$$\alpha_{i,i\neq 2} = \frac{e^0}{e^0 + e^0 + e^{|v_a|}} \simeq 0$$

$$c_2 \simeq \alpha_2 v_2 \simeq v_2 \simeq x_2 = v_a$$

$$\hookrightarrow c_2 \simeq v_a$$

no it cannot oppose v_b and M_2

unlike v_a, v_b ~~does~~ do not cancel out and solely

so for example by adding $w_j x_i$ weight increases but this does so by increasing v_j and v_b. the same thing happens for w_c.

$$V = (|u_b| - |u_c|) \cdot \frac{1}{\beta^2}$$

$$\alpha = (u_d \cdot u_a + u_c u_d) \cdot \frac{1}{\beta^2}$$

$$K = \bar{\bar{I}}$$

$$\rightarrow v_1 = u_b \quad \alpha_1 = u_c$$

$$v_2 = 0 \quad \alpha_2 = u_d$$

$$v_3 = u_b - u_c$$

$$\alpha_3 = 0$$

$$\alpha_1 = 0, 0, 1 \qquad \alpha_2 = 1, 0, 0$$

$$c_1 \simeq v_3 = v_b - v_c \rightarrow c_2 = v_1 = v_b$$

2d

dev 1.0 r.

gendon 5 i.

---

28 20 i.

---

29 13.6 i.

29 (ii)

Synthesizer self attention is not able to understand the context of different positions.

CSS in part e provided more knowledge the the model. (3a)

---

(1) misinformation.

(2) biased output (3b)

It may use some notion of similarity between name to purpose the birth place. (3c)
This causes bias (racism, sexism and ...)