Assignment 5.

1. (a) i. $\forall i, \alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. and α_i represents the portion of $e^{k_i^T q}$ in $\sum_j e^{k_j^T q}$.

ii. If there exists a $e^{k_j^T q} \gg \sum_{i \neq j} e^{k_i^T q}$, then $\alpha_j \gg \sum_{i \neq j} \alpha_i$. Maybe k_j and q are very alike.

iii. The output c will be very similar to v_j.

iv. If the query vector resembles certain k_j, consequently its output will be nearly identical to the correspond value v_j.

(b) i. Suppose $M = \sum_{i=1}^{m} \lambda_i a_i a_i^T$. $v_a = \sum_{i=1}^{m} c_i a_i$.

$$Ms = \sum_{i=1}^{m} \lambda_i a_i a_i^T (v_a + v_b) = \sum_{i=1}^{m} \lambda_i c_i a_i a_i^T a_i = \sum_{i=1}^{m} c_i a_i$$

Since $\{a_1, \cdots, a_m\}$ are orthogonal, $\lambda_i c_i a_i^T a_i = c_i$ for any $i \in [1, m]$.

$$\lambda_i = \frac{1}{a_i^T a_i} \quad \text{then} \quad M = \sum_{i=1}^{m} \frac{a_i a_i^T}{a_i^T a_i}.$$

ii. $\alpha_a = \alpha_b$ then $k_a^T q = k_b^T q$, q and $(k_a - k_b)$ are orthogonal.

$$(k_a^T - k_b^T)(c_1 k_a + c_2 k_b) = 0. \qquad \frac{c_1}{c_2} = \frac{k_b^T k_b}{k_a^T k_a}$$

A possible $q = k_b^T k_b k_a + k_a^T k_a k_b$.

(c) i. Set $q = 10(\mu_b^T \mu_b \mu_a + \mu_a^T \mu_a \mu_b) = 10(\mu_a + \mu_b)$ suppose $k_i = \mu_i + \Delta_i$, $\Delta_i \to 0$.

$$\exp(k_i^T q) = \begin{cases} \exp(\Delta_i^T q) \simeq 1 + \Delta_i^T q \approx 1, & i \neq a, b \\ \exp(\mu_a^T q + \Delta_a^T q) = \left(\exp\left(1 + \frac{\Delta_a^T q}{\mu_a^T q}\right)\right)^{\mu_a^T q} \approx e^{10 \mu_a^T \mu_a \mu_b^T \mu_b}(1 + \Delta_a^T q) \approx e^{10}, & i = a \\ \exp(\mu_b^T q + \Delta_b^T q) \approx e^{10 \mu_a^T \mu_a \mu_b^T \mu_b}(1 + \Delta_b^T q) \approx e^{10}, & i = b. \end{cases}$$

$$\alpha_i = \begin{cases} \frac{e^{10}}{2e^{10} + n - 2} \approx \frac{1}{2}, & i = a, b \\ \frac{1}{2e^{10} + n - 2} \approx 0, & i \neq a, b \end{cases}$$

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

ii. If $\Sigma_a = \beta I + \frac{1}{2}(\mu_a \mu_a^T)$. for vanishingly small β.

and $q = 10(\mu_a + \mu_b)$

set $k_a = \eta \cdot \mu_a + \Delta_a$, where $\eta \sim N(1, \frac{1}{2})$.

Then
$$k_i^T q = \begin{cases} 0, & i \neq a,b \\ 10\eta, & i=a \\ 10, & i=b. \end{cases} \qquad \alpha_i = \begin{cases} 0, & i \neq a,b \\ \frac{1}{1+e^{10(1-\eta)}}, & i=a \\ \frac{1}{1+e^{10(\eta-1)}}, & i=b \end{cases}$$

$$c \approx \frac{1}{1+e^{10(\eta-1)}} \left(e^{10(\eta-1)} \cdot v_a + v_b \right)$$

The value of η has a great effect on c. i.e. If $\eta > 1.1$, $c \approx v_a$.

Since $t = \frac{1}{1+e^{10(\eta-1)}} \in [0,1]$. $\quad Var[c] = v_a \cdot Var[t] + v_b \cdot Var[1-t]$
$$= (v_a + v_b) \cdot Var[t]$$

$$Var[t] \in (0, \tfrac{1}{4}]$$

(d) i. Let $q_1 = q_2 = 10(\mu_a + \mu_b)$. $c \approx \frac{1}{2}(v_a + v_b)$

ii. Assume $k_a = \eta_a \cdot \mu_a + \Delta_a$, $k_b = \eta_b \cdot \mu_a + \Delta_b$.

$$k_i^T q = \begin{cases} 0, & i \neq a,b \\ 10\eta_a, & i=a \\ 10\eta_b, & i=b. \end{cases}$$

$$c = \frac{e^{10\eta_a} v_a + e^{10\eta_b} v_b}{e^{10\eta_a} + e^{10\eta_b}} = \frac{e^{10(\eta_a-\eta_b)} v_a + v_b}{e^{10(\eta_a-\eta_b)} + 1}$$

Set $t = \frac{e^{10(\eta_a-\eta_b)}}{e^{10(\eta_a-\eta_b)}+1} \in [0,1]$. $Var[c] = v_a \cdot Var[t] + v_b \cdot Var[1-t]$
$$= (v_a + v_b) \cdot Var[t].$$

2. (d) acc: $9/500 = 1.8\%$

(f) acc: $59/500 = 11.8\%$

(g) i. acc: $14/500 = 2.8\%$ (not sure what has gone wrong)

ii. $$\frac{\text{Complexity (Perceiver)}}{\text{Complexity (Vanilla)}} = \frac{(L-2)m^2 + 2\ell m}{L \cdot \ell^2}$$

3. (a) The model has learned useful representations from a large and diverse pretrain dataset, which enables it to apply its knowledge to a downstream task with less training set and time.

(b) i. The wrong answers might cause problems for the users.

ii. Considering it's nearly impossible to distinguish correct and incorrect outputs. the whole system is not reliable

(c) Find a most similar name from the train dataset. then predict its hometown.

The effect of this strategy is limited, and it harms the trustworthiness of the model.