

1. W modelu IBM 1 przy ustalonym dopasowaniu a logarytm funkcji wiarygodności L_a ma postać

$$L_a = \log\left(\prod p_{w,v}^{c_{w,v}}\right)$$

gdzie $p_{w,v} = P(w|v)$ to prawdopodobieństwo że słowo w jest tłumaczeniem słowa v zaś $c_{w,v}$ to ilość wystąpień w jako tłumaczenie v (tzn. takich gdzie a dopasowuje w do v). Uzasadnij to. Używając ten wzór uzasadnij że maksimum L_a jest osiągane gdy

$$p_{w,v} = \frac{c_{w,v}}{\sum_u c_{u,v}}$$

(użyj mnożników Lagrange'a i warunek $\sum_u p_{u,v} = 1$). W algorytmie EM maksymalizujemy funkcję Q postaci

$$\sum L_a q_a$$

gdzie q_a jest (zależnym od danych) bierzącym oszacowaniem prawdopodobieństwa dopasowania a . Znajdź i uzasadnij wzór na $p_{w,v}$ maksymalizujące Q . Uwaga: w algorytmie EM Q maksymalizujemy traktując q_a jako stałe.