Question :-

Bigram -

what is the probability of word w given before word v

$P(w | v) \longrightarrow P(w_2 | w_1)$

→ sometimes bigram won't occur leading to 0 counts

→ for this we use interpolation

unigram + bigram

→ $P(w|v) = \gamma f(w|v) + (1-\gamma) f(w)$

→ Directly modify counts in training data

3 cases

i) $C^*(v,w) = C(v,w) + \theta$

ii) $C^*(v,w) = C(v,w) + \theta\, C(w)$

iii) $C^*(v,w) = C(v,w) + \theta\, C(v)\, f(w)$

so

smoothed bigram prob

$$p^*(w|v) = \frac{C^*(v,w)}{\sum_{w' \in V} C^*(v,w')}$$

$N(v,w)$ - held out data

1) θ that maximizes the log-probability

$$p(H) = \sum_{v \in V} \sum_{w \in V} N(v, w) \log p^*(w|v)$$

Case 1 :-

$$P(w_i | w_{i-1}) = \frac{\text{Count}(w_{i-1}, w_i)}{\text{Count}(w_{i-1})}$$

$$p^*(w|v) = \frac{C^*(v, w)}{C^*(v)}$$

$$p(H) = \sum_{v \in V} \sum_{w \in V} N(v, w) \log \frac{C^*(v, w)}{C^*(v)}$$

$$\frac{d}{d\theta} p(H) = \sum_{v \in V} \sum_{w \in V} N(v, w) \frac{d}{d\theta} \log \frac{C^*(v, w) + \theta}{C^*(v) + \theta}$$

$$= \sum_{v \in V} \sum_{w \in V} N(v,w) \frac{1}{C^*(v,w) + \theta} - \frac{1}{C^*(v) + \theta}$$

set the derivative equal to 0 & solve for θ to find critical points

$$\sum_{v \in V} \sum_{w \in V} N(v,w) \left(\frac{c^*(v) + \cancel{(\theta)} - C^*(v,w) - \cancel{(\theta)}}{C^*(v,w) \cdot C^*(v) + C^*(v,w) \cdot \theta + \theta\, C^*(v) + \theta^2} \right) = 0$$

$$\sum_{v \in V} \sum_{w \in V} N(v,w) \left(\frac{(C^*(v) - C^*(v,w))}{C^*(v,w) \cdot C^*(v) + \theta(C^*(v,w) + C^*(v)) + \theta^2} \right) = 0$$

Now compute the second derivative of $P(H)$ at the critical point

if negative then critical point corresponds to local maximum

$$\frac{d}{d\theta} \sum_{v \in V} \sum_{w \in V} N(v,w) \left(\frac{(C^*(v) - C^*(v,w))}{C^*(v,w) \cdot C^*(v) + \theta(C^*(v,w) + C^*(v)) + \theta^2} \right)$$

$$= - \sum_{v \in V} \sum_{w \in V} N(v,w) \left(\frac{(C^*(v) - C^*(v,w))}{C^*(v,w) \cdot C^*(v) + \theta(C^*(v,w) + C^*(v)) + \theta^2} \right) \times \left[2\theta + \left(C^*(v,w) + C^*(v) \right) \right]$$

In the numerator we have 2 values which will definitely be ≥ 0

all the values are a positive number

so it is Maximum

Case 2 :-

$$C^*(v,w) = C(v,w) + \theta\, C(w)$$

$$\sum_{w' \in V} C^*(v, w')$$

$$= \sum_{w' \in V} C^*(v, w') + \theta\, C(w')$$

$$= C^*(v) + \theta \sum_{w' \in V} C(w')$$

$$= C^*(v) + N\theta$$

N is the length of vocabulary

$$P(H) = \sum_{v \in V} \sum_{w \in V} N(v, w) \log \left[\frac{C^*(v,w) + \theta C(w)}{C^*(v) + N\theta} \right]$$

differentiate & equate to 0

$$\frac{d\,P(H)}{d\theta} = \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C(w)}{C^*(v,w) + \theta C(w)} - \frac{N}{C^*(v) + N\theta} \right] = 0$$

$$= \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(w)\, C^*(v) + \cancel{\theta N C^*(w)} - N\, C^*(v,w) - \cancel{N\theta C^*(w)}}{C^*(v,w) . C^*(v) + N\, C^*(v,w)\theta + N\theta^2 C(w) + \theta C(w) C(v)} \right]$$

$$= \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(w)\, C^*(v) - N\, C^*(v,w)}{C^*(v,w) . C^*(v) + N\, C^*(v,w)\theta + N\theta^2 C(w) + \theta C(w) C(v)} \right]$$

Second differentiation

$$= \frac{d}{d\theta} \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(w)\, C^*(v) - N\, C^*(v,w)}{C^*(v,w) . C^*(v) + N\, C^*(v,w)\theta + N\theta^2 C(w) + \theta C(w) C(v)} \right]$$

$$= - \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(w)\, C^*(v) - N\, C^*(v,w)}{N\theta^2 C^*(w) + (C^*(v,w) . C^*(v))^2 + \theta (C^*(v,w) N + C^*(w) C(v))} \right]$$

$$\left[2 N\theta\, C^*(w) + C^*(v,w) N + C^*(w) C^*(v) \right]$$

= it is - (number) so maximum

Case 3:-

$$C^*(v,w) = C^*(v,w) + \theta\, C^*(v)\, f(w)$$

$$p^*(w|v) = \frac{C^*(v,w)}{\sum_{w' \in V} C^*(v,w')}$$

$$\sum_{w' \in V} C^*(v,w') = \sum_{w' \in V} C(v,w') + \theta\, C^*(v) \frac{C^*(w)}{N}$$

$$= C^*(v)\left[1 + \theta \sum_{w' \in V} \frac{C^*(w)}{N}\right]$$

$$= C^*(v)\left[1 + \theta \frac{N}{N}\right]$$

$$= C^*(v)\left[1 + \theta\right]$$

Now Substitute

$$P(H) = \sum_{v \in V} \sum_{w \in V} N(v,w) \log P^*(w|v)$$

$$= \sum_{v \in V} \sum_{w \in V} N(v,w) \log \frac{C(v,w) + \theta C(v) f(w)}{C(v)[1+\theta]}$$

differentiate and equate to 0

$$\frac{d}{d\theta} P(H) = \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(v) f(w)}{C^*(v,w) + \theta C^*(v) f(w)} - \frac{C^*(v)}{C^*(v)[1+\theta]} \right]$$

$$= \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(v)^2 f(w)[1+\theta] - C^*(v) C^*(v,w) - C^*(v)^2 f(w) \theta}{C^*(v) C^*(v,w) + C^*(v) C^*(v,w)\theta + \theta C^*(v)^2 f(w) + \theta^2 C^*(v)^2 f(w)} \right]$$

$$= \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(v)^2 f(w) - C^*(v) C^*(v,w)}{\theta^2 C^*(v)^2 f(w) + \theta C^*(v)[C^*(v,w) + C^*(v) f(w)] + C^*(w) C^*(v,w)} \right]$$

differentiate again

differentiate again

$$= -\sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(v)^2 f(w) - C^*(v) C^*(v,w)}{\theta^2 C^*(v)^2 f(w) + \theta C^*(v)[C^*(v,w) + C^*(v) f(w)] + C^*(v) C^*(v)^2} \right]$$

$$\cdot \left[2\theta C^*(v)^2 f(w) + C^*(v)[C^*(v,w) + C^*(v) f(w)] \right]$$

its -ve and the value inside is +ve

so maximum

2) Case 1 :-

$$c^*(v,w) = c^*(v,w) + \theta$$

here $N(v,w) = c(v,w)$

so substitute here

$$p^*(w|v) = \frac{c(v,w) + \theta}{\sum_{w' \in V} [c(v,w') + \theta]}$$

$$= \frac{c(v,w) + \theta}{\sum_{w' \in V} c(v,w') + \theta |V|}$$

differentiate

its constant

$$\sum_{w' \in V} \frac{c(v,w') - c(v,w)|V|}{\left(\sum_{w' \in V} c(v,w') + \theta |V|\right)^2}$$

its +ve for any +ve θ

the θ does not change the probabilities

Therefore θ doesn't affect $P(H)$ & optimal value is $\theta = 0$

$$- \sum_{v \in V} \sum_{w \in V} N(v,w) \left(\frac{(C^*(v) - C^*(v,w))}{C^*(v,w) \cdot C^*(v) + \theta(C^*(v,w) + C^*(v)) + \theta^2} \right)$$

$$\times \left[2\theta + \left(C^*(v,w) + C^*(v) \right) \right]$$

from the above equation in question 1, after double differentiating w r t to θ evaluated at θ is negative, it means that $\theta = 0$ is indeed a local maximum.

function is concave at that point

Given $\theta = 0$ emerges as a local maxima, it reinforces the conclusion that $\theta = 0$ is goot when $N(v,w) = C(v,w)$

Case 2:-

$$C^*(v,w) = C(v,w) + \theta C(w)$$

$$P^*(w|v) = \frac{C(v,w) + \theta C(w)}{\sum_{w' \in V} C(v,w') + \theta C(w')}$$

Based on the equation derived above

$$= - \sum_{v \in V} \sum_{w \in V} N(v,w) \left(\frac{(C^*(v) - C^*(v,w))}{C^*(v,w) \cdot C^*(v) + \theta(C^*(v,w) + C^*(v)) + \theta^2} \right)$$
$$\times \left[2\theta + \left(C^*(v,w) + C^*(v) \right) \right]$$

we can see that
the θ does not change the probabilities
Therefore θ doesn't affect $P(H)$ & optimal value is $\theta = 0$

Case 3:-

$$C^*(v,w) = C(v,w) + \theta C(v) f(w)$$

Based on above equation

$$- \sum_{v \in V} \sum_{w \in V} N(v,w) \left[\frac{C^*(v)^2 f(w) - C^*(v) C^*(v,w)}{\theta^2 C^*(v)^2 f(w) + \theta C^*(v) [C^*(v,w) + C^*(v) f(w)] + C^*(v) C^*(v)^2} \right]$$
$$\cdot \left[2\theta C^*(v)^2 f(w) + C^*(v) [C^*(v,w) + C^*(v) f(w)] \right]$$

we can see that
the θ does not change the probabilities
Therefore θ doesn't affect $P(H)$ & optimal value is $\theta = 0$

3) Case 1 :-

$$c^*(v,w) = c(v,w) + \theta$$

$$p^*(w|v) = \frac{c(v,w) + \theta}{\sum_{w' \in V} [c(v,w') + \theta]}$$

$$= \frac{c(v,w) + \theta}{c(v) + \theta}$$

$$= \frac{c(v,w)}{c(v) + \theta} + \frac{\theta}{c(v) + \theta}$$

$$f_2(w|v) = \frac{c(w,v)}{c(v)}$$

$$f_1(w) = \frac{c(w)}{N}$$

$$\gamma f(w|v) + (1-\gamma) f(w) = \frac{C(v,w)}{C(v)+\theta} + \frac{\theta}{C(v)+\theta}$$

$$= f(w|v)\left[\frac{C(v)}{C(v)+\theta}\right] + \frac{\theta}{C(v)+\theta}$$

$$f(w|v) = \frac{C(w,v)}{C(v)} \qquad \gamma = \frac{C(v)}{C(v)+\theta}$$

$$f(w) = 1$$

Case 2 :-

$$p^*(w|v) = \frac{C(v,w) + \theta\, C(w)}{C(v) + N\theta}$$

$$p^*(w|v) = \gamma f_2(w|v) + (1-\gamma) f_1(w)$$

$$\frac{c(v,w) + \theta c(w)}{c(v) + N\theta} = \gamma f_2(w|v) + (1-\gamma) f_1(w)$$

$$f(w|v) = \frac{c(w,v)}{c(v)} + f(w) = \frac{c(w)}{N}$$

$$\frac{c(v,w)}{c(v) + N(\theta)} + \frac{\theta c(w)}{c(v) + N\theta}$$

$$= \gamma f(w|v) + (1-\gamma) f(w)$$

$$= c(w,v) = f(w|v) \cdot c(v)$$

$$= c(w) = f(w) \cdot n$$

$$= f(w|v) \left[\frac{c(v)}{c(v) + N\theta} \right] + \frac{\theta n f(w)}{c(v) + N\theta}$$

$$= \gamma\, f(w|v) + (1-\gamma)\, f(w)$$

$$\gamma = \frac{c(v)}{c(v) + N\theta}$$

$$f(w|v) = \frac{c(w,v)}{c(v)}$$

$$f(w) = \frac{c(w)}{n}$$

Case 3:-

$$p^*(w|v) = \frac{c(v,w) + \theta\, c(v)\, f(w)}{c(v)\,[1+\theta]}$$

$$= \frac{c(v,w)}{c(v)\,[1+\theta]} + \frac{\theta\, c(v)\, f(w)}{c(v)\,[1+\theta]}$$

$$= \frac{c(v)\ f(w|v)}{c(v)\ [1+\theta]} + \frac{\theta\ c(v)\ f(w)}{c(v)[1+\theta]}$$

$$\gamma = \frac{1}{1+\theta}$$

$$f(w) = f(w)$$

$$f(w|v) = \frac{c\ (v,w)}{c(v)}$$