



# **SNS COLLEGE OF ENGINEERING**



**Kurumbapalayam(Po), Coimbatore – 641 97**

**Accredited by NAAC-UGC with 'A' Grade**

**Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai**

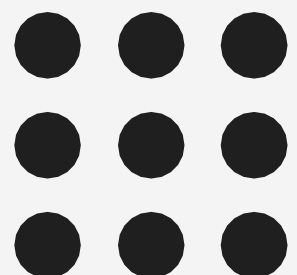
## **Department of Artificial Intelligence and Data Science**

**Course Name – 19AD601 – Natural Language  
Processing**

**III Year / VI Semester**

**Unit 2 – WORD LEVEL ANALYSIS**

**Topic 4- Interpolation and Backoff**





# Interpolation and Backoff

In backoff, we use the trigram if the evidence is sufficient, otherwise we use the bigram, otherwise the unigram.

In other words, we only “back off” to a lower-order n-gram if we have zero evidence for a higher-order n-gram.

By contrast, in interpolation, we always mix the probability estimates from all the n-gram estimators, weighting and combining the trigram, bigram, and unigram counts.

In simple linear interpolation, we combine different order n-grams by linearly interpolating them.

Thus, we estimate the trigram probability  $P(w_n | w_{n-2}w_{n-1})$  by mixing together the unigram, bigram, and trigram probabilities, each weighted by a



# Interpolation and Backoff



$\lambda$ :

$$\begin{aligned} \hat{P}(w_n|w_{n-2}w_{n-1}) &= \lambda_1 P(w_n) \\ &\quad + \lambda_2 P(w_n|w_{n-1}) \\ &\quad + \lambda_3 P(w_n|w_{n-2}w_{n-1}) \end{aligned}$$



# Interpolation and Backoff

In a slightly more sophisticated version of linear interpolation, each  $\lambda$  weight is computed by conditioning on the context. Interpolation with context-conditioned weights:

$$\begin{aligned}\hat{P}(w_n | w_{n-2}w_{n-1}) &= \lambda_1(w_{n-2:n-1})P(w_n) \\ &\quad + \lambda_2(w_{n-2:n-1})P(w_n | w_{n-1}) \\ &\quad + \lambda_3(w_{n-2:n-1})P(w_n | w_{n-2}w_{n-1})\end{aligned}$$



**THANK YOU**