Welcome back!

everyone in their online classes now

credit: Funny Texts on FB

CS 378 N-gram recap

Today: recap n-gram LMs (brief recap Tues)

LM: $P(\bar{w}) = P(w_1)\,P(w_2|w_1)\,P(w_3|w_1 w_2)\cdots$

dist over sentences

n-gram LM: assume word w_i depends only on previous $n-1$ words

$P(\text{I want to go to Maui}) =$ 3-gram

$P(\text{I} \mid \langle s\rangle\ \langle s\rangle)\ P(\text{want} \mid \langle s\rangle\ \text{I})$

$P(\text{to} \mid \text{I want})\ P(\text{go} \mid \text{want to}) \cdots$

Parameters

I want to go to Maui
You should go to Denmark
John will go to Maui

$P(w \mid c = \text{go to}) =$

2/3	Maui
1/3	Denmark
0	go
0	to
0	:
:	

call this $\hat{P}(w \mid \text{go to})$ empirical distribution

Smoothing empirical dist has lots of zeroes

Naive smoothing: $P_s(w \mid \text{go to})$

$$= \lambda \hat{P}(w \mid \text{go to}) + (1-\lambda)\hat{P}(w)$$

unigram

$> 0 \ \forall w$

$$P_S(w \mid \text{go to}) = \lambda_1 \hat{P}(w \mid \text{go to}) + (1-\lambda_1) P_S(w \mid \text{to})$$

$$P_S(w \mid \text{to}) = \lambda_2 \hat{P}(w \mid \text{to}) + (1-\lambda_2) \hat{P}(w)$$

back off

Absolute Discounting (AD)

empirical dist. w/ "discounts"

$$P_{AD}(w \mid \text{go to}) = \frac{\max(0, \text{count}(w \mid \text{go to}) - k)}{\text{count}(\text{go to})} + \lambda P_{AD}(w \mid \text{to})$$

"k" / count(go to) · num words w/ nonzero count after "go to" (assume $k < 1$)

counts		probs
2	Mavi	1.6/3
1	Den.	0.6/3
0	⋮	0
0	⟹	0
0	$k=0.4$	⋮
0		

$$+ \frac{0.8}{3} P_{AD}(w \mid \text{to})$$

Kneser-Ney: AD + one trick