Reading #1: MaxEnt

LING 572

Advanced Statistical Methods for NLP
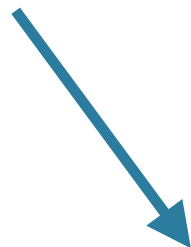
Papers for reading assignment

- <u>Ratnaparkhi 1997</u>

  - Sections 1-3

  - Sections 4-8

  - You can skip the proof

- <u>Berger et al 1996</u>

  - Sections 1-3.3

Notation

We use this one

	Input	Output
Berger et al 1996	x	y
Ratnaparkhi 1997	b	a
Ratnaparkhi 1996	h	t
Klein and Manning 2003	d	c

Questions

(Q1) Let $P(X = i)$ be the probability of getting an *i* when rolling a dice (e.g. possible outcomes are *i=1,2,…,6*). What is the value of $P(X = i)$ with the maximum entropy if the following is true?

(a) $P(X = 1) + P(X = 2) = 1/2$

(b) $P(X = 1) + P(X = 2) = 1/2$ and $P(X = 6) = 1/3$

Questions

(Q2) In the text classification task, |V| is the number of features and |C| is the number of classes. How many *feature functions* are there?

(Q3) What are the similarities and differences between MaxEnt and Naive Bayes with respect to modeling, training, and decoding?

Due: 11am on January 28 2020