

# Reading #1: MaxEnt

LING572

Fei Xia

# Papers for the reading assignment

- [Ratnaparkhi \(1997\)](#)
  - Sections 1-3
  - Section 4-8
  - You can skip the proof
- [Berger et al. \(1996\)](#)
  - Sections 1-3.3

# Notation

	Input	Output
--	-------	--------

(Berger et. al., 1996)	x	y
------------------------	---	---

(Ratnaparkhi, 1997)	b	a
---------------------	---	---

(Ratnaparkhi, 1996)	h	t
---------------------	---	---

(Klein and Manning, 2003)	d	c
---------------------------	---	---

We following the notation in (Berger et al., 1996)

# Questions

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

(a)  $P(X=1) + P(X=2) = \frac{1}{2}$

(b)  $P(X=1) + P(X=2) = \frac{1}{2}$  and  $P(X=6) = \frac{1}{3}$

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

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

Due: 11am next Thurs (1/26/2017), 25 points