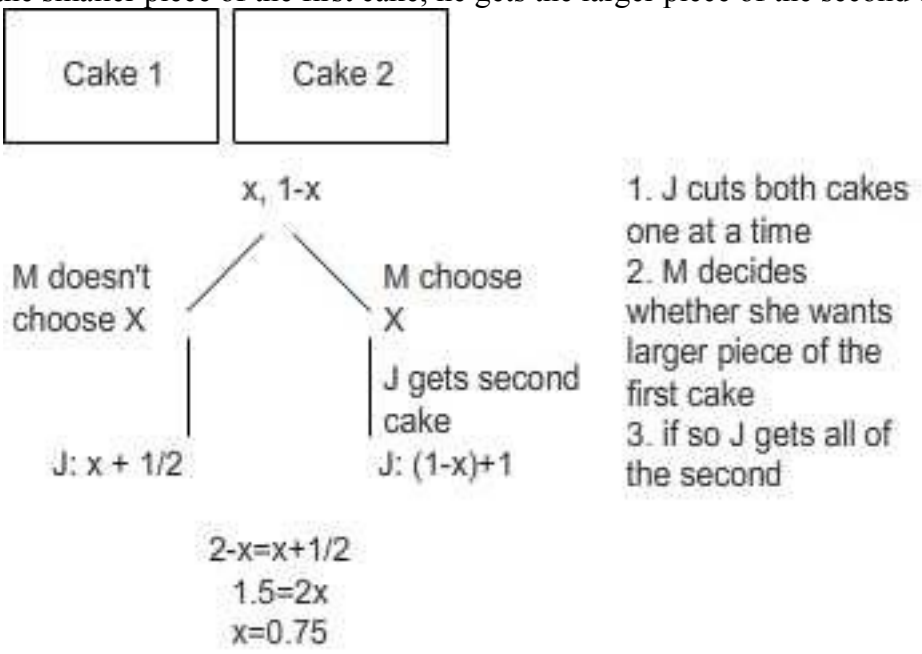
Computational Thought
Professor Dennis Shasha
Tengchao Zhou scribes
November 18th 2009

DECISION TREE

This class starts with a quiz about the decision tree.
Shasha reviews the cutting cake case.

Jeremie cuts both cakes. Marie gets to choose which piece she wants of first cake. If she does choose, then she gets the larger piece of the first cake; J gets all of the second cake. If she doesn't , then she gets the smaller piece of the first cake; he gets the larger piece of the second cake.



KIT CATS

While collecting the quzi, Shasha tells us a short history of kit cats:
Kit cats were used to reward Shasha' students in previous years. Shasha derives this idea from rewarding dolphins with fish. So dolphins == Shasha' students in previous years.

CODING

Shasha shows us a PowerPoint designed by a professor of Carnegie Mellon University, whose name is Andrew M. Moore. The PowerPoint is about Information Gain.

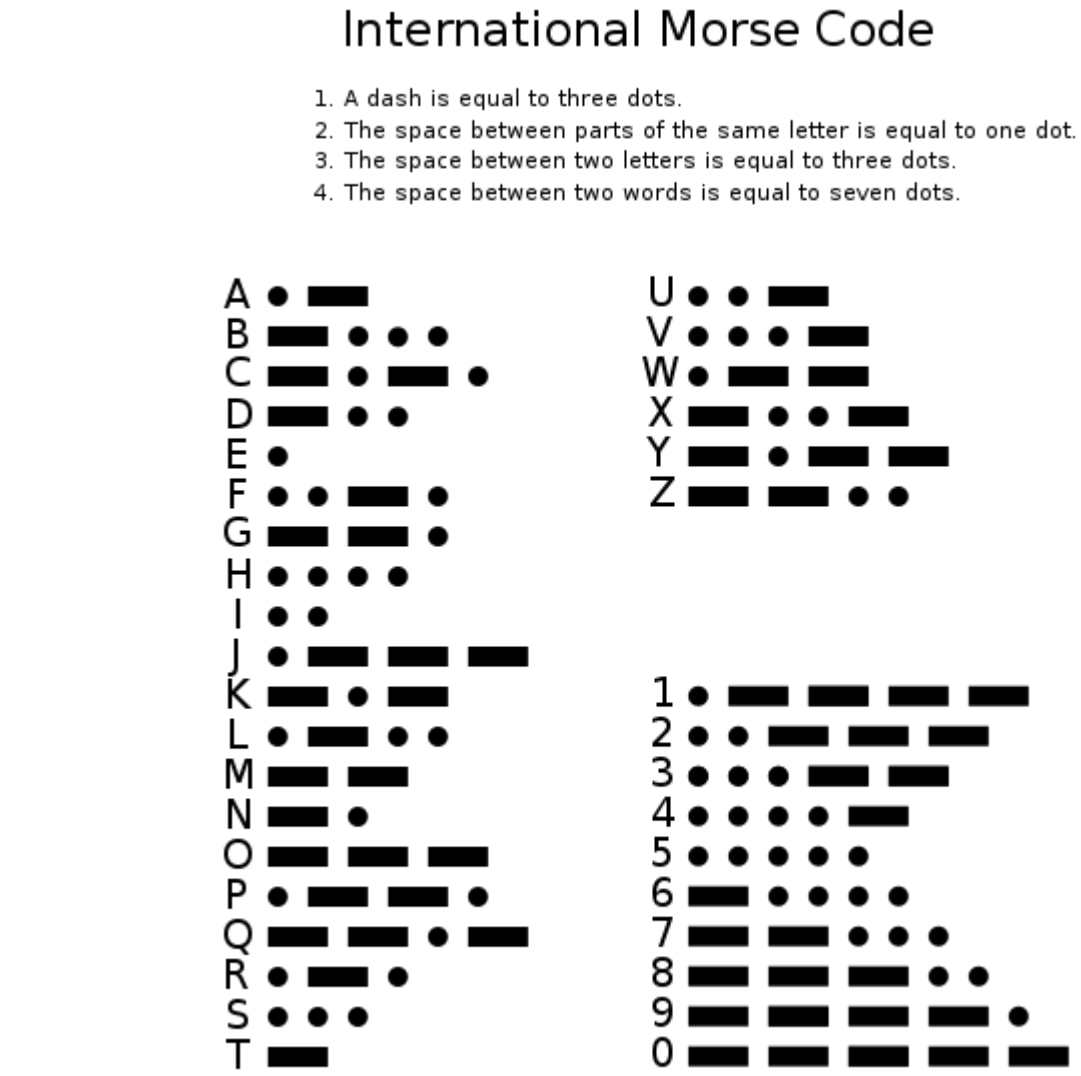
In a series of random letters ADCDDADDADCCBBB..., the possibility of the letter A, B, C, D are equal, all ¼.
P(X=A) = ¼ P(X=B) = ¼ P(X=C) = ¼ P(X=D) = ¼
We can use 00, 01, 10 and 11 to represent them respectively.

A	B	C	D
00	01	10	11

So we can rewrite ADCDDADDADCCBBB... as 00111011110011110011101001010 1...
Each time we read two digits, we can revert them into the original letters.

Shasha introduces us the international Morse Code.
Based one the frequency that a letter appear in English words, the more frequent letter will be assigned shorter code. For example, the most frequent letter is e, so e is assigned with a dot. The second frequent letter is t, so it's assigned with a dash.

e	.
t	-



Rhey T. Snodgrass & Victor F. Camp, 1922

SOS is the commonly used description for the international Morse code distress signal (· · · — — — · · ·).

When the possibility of appearing the letters varied, we can adjust the way we present them. We can assign them with different length of digit.

Letter	Frequency	The Number of Digits	Representation
A	P1=1/2	$-\log_2 P1=1$	0
B	P2=1/4	$-\log_2 P2=2$	10
C	P3=1/8	$-\log_2 P3=3$	110
D	P4=1/8	$-\log_2 P4=3$	111

$$-P1\cdot\log_2 P1-P2\cdot\log_2 P2-P3\cdot\log_2 P3-\ldots-Pn\cdot\log_2 Pn=-\sum_{j=1}^{n} Pn\cdot\log_2 Pn$$

1*1/2+2*1/4+3*1/8+3*1/8
=1/2+1/2+3/8+3/8
=1+3/4

ENTROPY

Then Shasha talks about High Entropy and Low Entropy.

High Entropy means something is uniform(boring).
Low Entropy means something is varied(peaks and valleys).

Let's see the energy.
When energy is contained in the gasoline, it's Low Entropy.
When energy transformed into noise, motion power and so on, it's High Entropy.

Shasha shows another interesting example to understand the Low and High Entropy.
When all the soups are in the bow, it's Low Entropy.
When the soups go everywhere, it's High Entropy.

LINGUISTIC

In the last few minutes, Shasha talks about some linguistic stuff.
Let's compare the letter t in two words.
top: The t in top is word-initial aspiration.
stop: The t in stop is unvoiced.

The tot and the dot are minimal pairs.
Voice is phonemic in English.

In the last second, Shasha shows us a pair of similar German words:
HÖLLE means hell (the word's pronunciation is like hello)
HÖHLE means hole