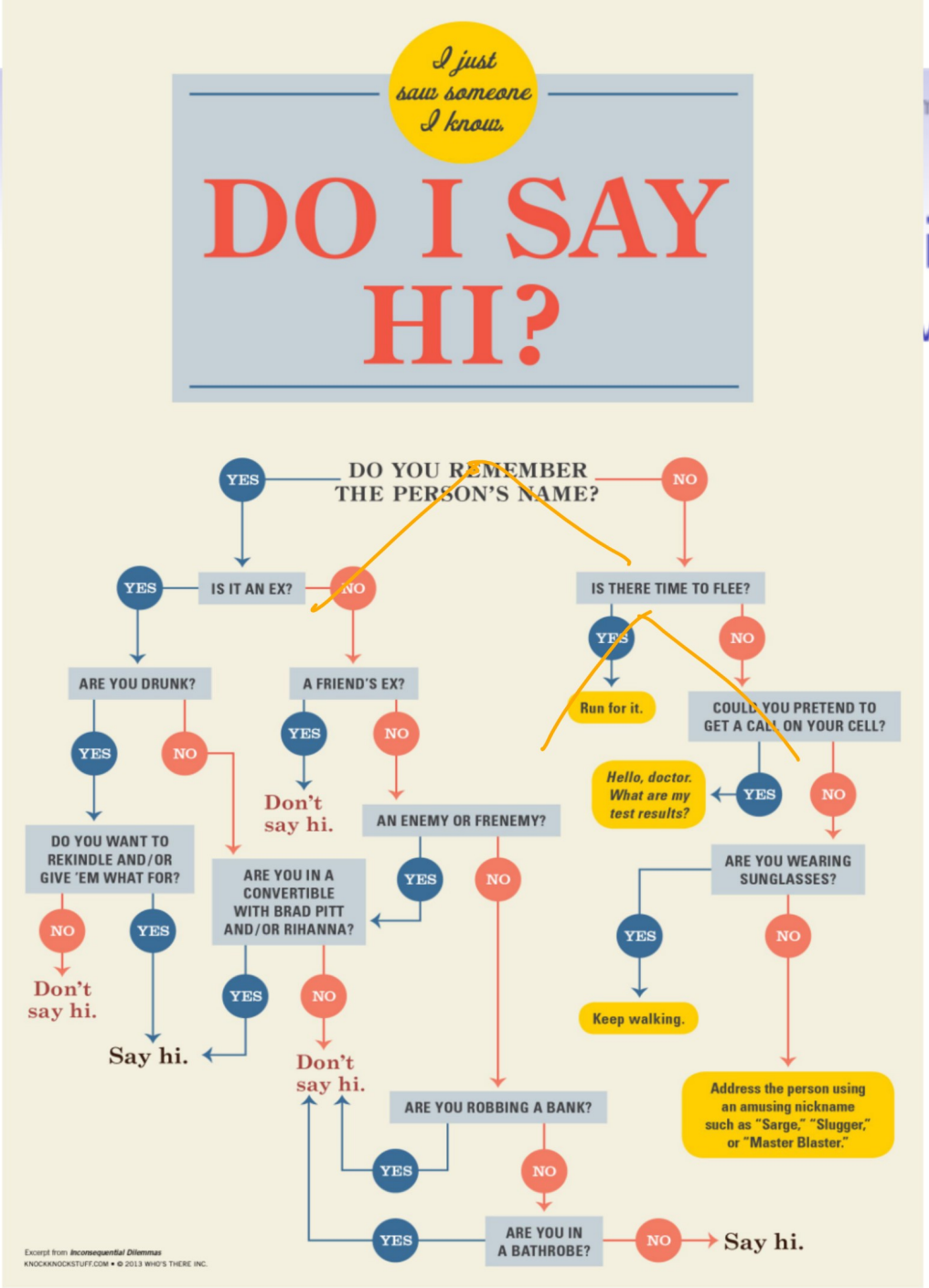
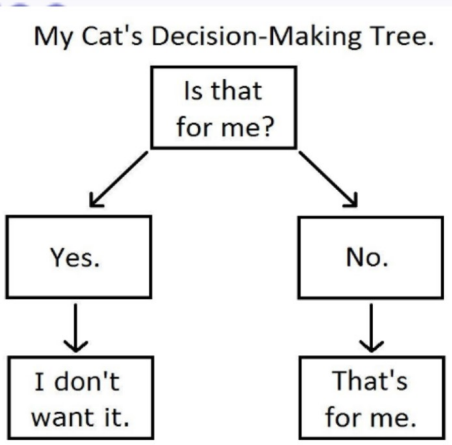
CS540 Introduction to Artificial Intelligence
Lecture 6

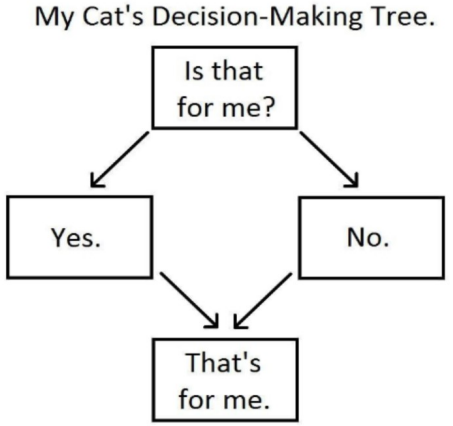
Young Wu
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

June 10, 2020

ion T
vation

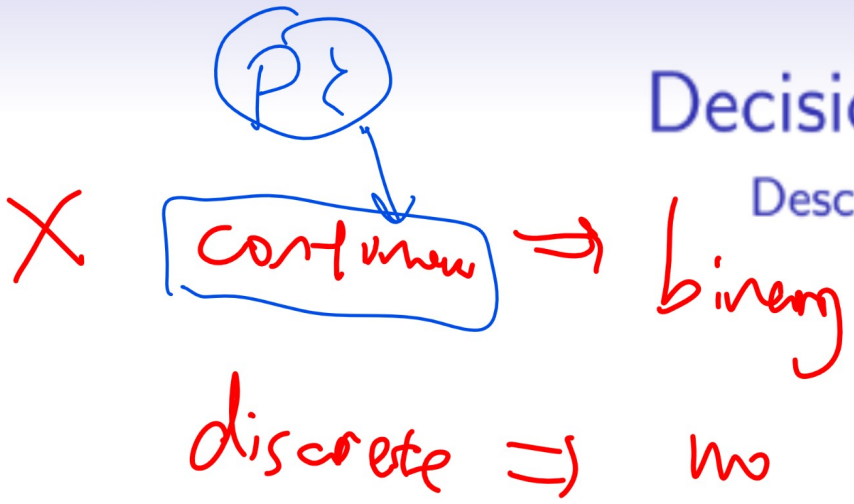

I just saw someone I know.
DO I SAY HI?
DO YOU REMEMBER THE PERSON'S NAME?
YES
NO
IS IT AN EX?
ARE YOU DRUNK?
A FRIEND'S EX?
IS THERE TIME TO FLEE?
Run for it.
COULD YOU PRETEND TO GET A CALL ON YOUR CELL?
Hello, doctor. What are my test results?
DO YOU WANT TO REKINDLE AND/OR GIVE 'EM WHAT FOR?
Don't say hi.
AN ENEMY OR FRENEMY?
ARE YOU IN A CONVERTIBLE WITH BRAD PITT AND/OR RIHANNA?
ARE YOU WEARING SUNGLASSES?
Keep walking.
Say hi.
ARE YOU ROBBING A BANK?
Address the person using an amusing nickname such as "Sarge," "Slugger," or "Master Blaster."
ARE YOU IN A BATHROBE?
Excerpt from Inconsequential Dilemmas
KNOCKKNOCKSTUFF.COM • © 2013 WHO'S THERE INC.

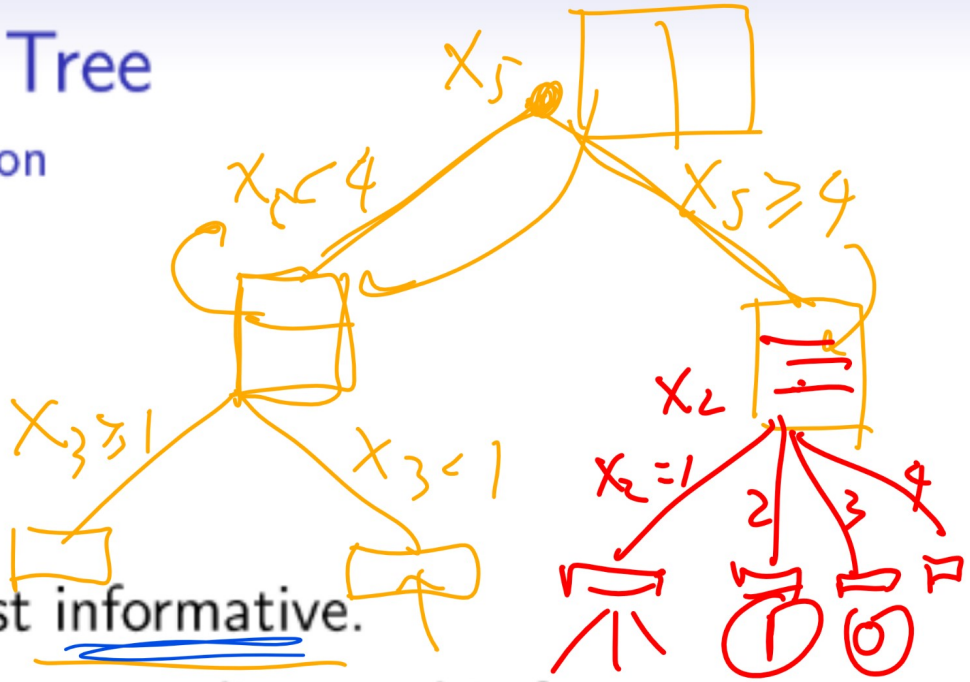

My Cat's Decision-Making Tree.
Is that for me?
Yes.
No.
I don't want it.
That's for me.


My Cat's Decision-Making Tree.
Is that for me?
Yes.
No.
That's for me.

Decision Tree

Description





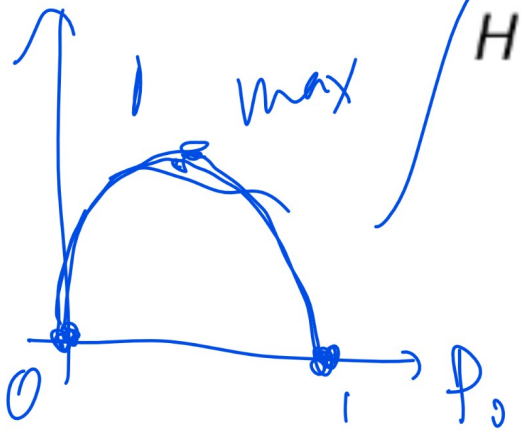
- Find the feature that is the most informative.
- Split the training set into subsets according to this feature.
- Repeat on the subsets until all the labels in the subset are the same.

Binary Entropy

Definition

- Entropy is the measure of uncertainty.
- The value of something uncertain is more informative than the value of something certain.
- For binary labels, $y_i \in \{0, 1\}$, suppose p_0 fraction of labels are 0 and $1 - p_0 = p_1$ fraction of the training set labels are 1, the entropy is:

$$H(Y) = p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right)$$
$$= -p_0 \log_2(p_0) - p_1 \log_2(p_1)$$



Entropy

Definition

- If there are K classes and p_y fraction of the training set labels are in class y, with $y \in \{1, 2, ..., K\}$, the entropy is:

$$H(Y) = \sum_{y=1}^{K} p_y \log_2 \left(\frac{1}{p_y}\right)$$

$$= -\sum_{y=1}^{K} p_y \log_2 (p_y)$$

Conditional Entropy

Definition

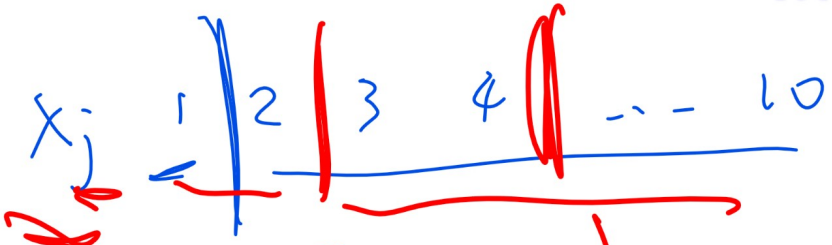
- Conditional entropy is the entropy of the conditional distribution. Let K_X be the possible values of a feature X and K_Y be the possible labels Y. Define p_x as the fraction of the instances that is x, and $p_{y|x}$ as the fraction of the labels that are y among the ones with instance x.

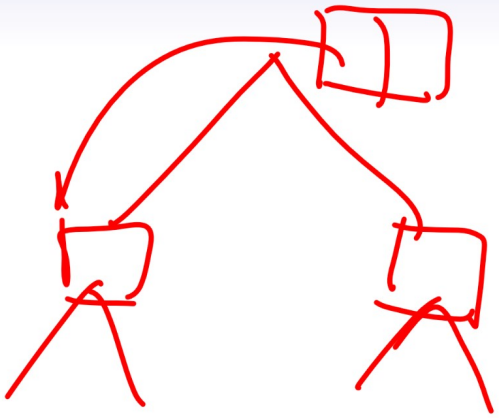
$$H(Y|X=x) = -\sum_{y=1}^{K_Y} p_{y|x} \log_2 (p_{y|x})$$

$$H(Y|X) = \sum_{x=1}^{K_X} p_x H(Y|X=x)$$

Information Gain

Definition





- The information gain is defined as the difference between the entropy and the conditional entropy.

$$I(Y|X) = H(Y) - H(Y|X).$$

- The larger than information gain, the larger the reduction in uncertainty, and the better predictor the feature is.

Splitting Discrete Features

Definition

- The most informative feature is the one with the largest information gain.

$$\arg\max_j I\left(Y|X_j\right)$$

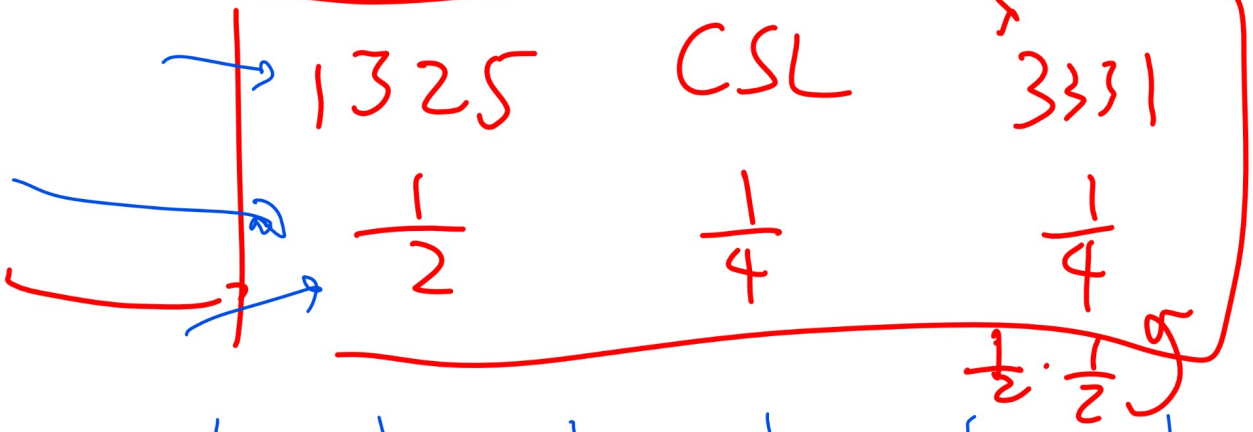
- Splitting means dividing the training set into K_{X_j} subsets.

$$\left\{(x_i, y_i) : x_{ij} = 1\right\}, \left\{(x_i, y_i) : x_{ij} = 2\right\}, ..., \left\{(x_i, y_i) : x_{ij} = K_{X_j}\right\}$$

Entropy

Quiz

- Fall 2010 Final Q10
- Running from You-Know-Who, Harry enters the CS building on the 1st floor. He flips a fair coin: if it is heads he hides in room 1325; otherwise, he climbs to the 2nd floor. In that case he flips the coin again: if it is heads he hides in CSL; otherwise, he climbs to the 3rd floor and hides in 3331. What is the entropy of Harry's location?
- A: 0.75
- B: 1
- C: 1.5
- D: 1.75
- E: None of the above.



$\frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1.5$

$-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{4}\log_2\frac{1}{4}$

Entropy Math

Quiz

Entropy 2

Quiz

~~Q5~~ last. Q6

- A bag contains a red ball, a green ball, a blue ball, and a black ball. Randomly draw a ball from the bag with equal probability. What is the entropy of the outcome?
- A: 1
- B: $\log_2(3)$
- C: 1.5
- D: 2
- E: 4

$Pr\{Y=1\}$ $Pr\{\hat{Y}=1\}$

$y_i \log q_i$

$+ (1-y_i) \log(1-q_i)$

$Pr\{Y=0\}$ $Pr\{\hat{Y}=0\}$

$-\left(\frac{1}{4}\cdot\log_2\frac{1}{4} + \frac{1}{4}\cdot\log_2\frac{1}{4} \dots\right) = -\log_2\frac{1}{4} = 2$

Pruning Diagram

Disucssion

Baigging and Boosting Diagram

Discussion

K Nearest Neighbor

Description

- Given a new instance, find the K instances in the training set that are the closest.
- Predict the label of the new instance by the majority of the labels of the K instances.

Distance Function

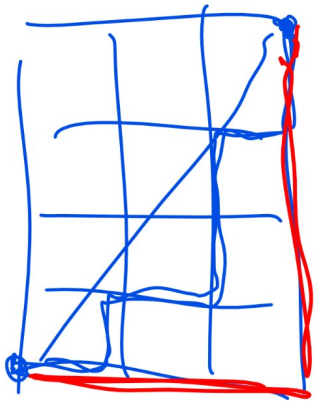
Definition

- Many distance functions can be used in place of the Euclidean distance.

$$\rho(x, x') = \|x - x'\|_2 = \sqrt{\sum_{j=1}^{m} (x_j - x'_j)^2}$$

- An example is Manhattan distance.

$$\rho(x, x') = \sum_{j=1}^{m} |x_j - x'_j|$$



1 Nearest Neighbor

Quiz

Q1

- Spring 2018 Midterm Q7
- Find the 1 Nearest Neighbor label for $\begin{bmatrix} 3 \\ 6 \end{bmatrix}$ using Manhattan distance.

x_1	1	1	3	5	2
x_2	1	7	3	4	5
y	0	1	1	0	0

dist 7 3 3 4 2

- A: 0
- B: 1

3 Nearest Neighbor

Quiz

new point

- Find the 3 Nearest Neighbor label for $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$ using Manhattan distance.

training set

x_1	1	1	3	5	2
x_2	1	7	3	4	5
y	0	1	1	0	0

dis = 4 6 | 0 3 3

- A: 0
- B: 1

Cross Validation

Quiz

accuracy K-fold

- Given the following training data. What is the 2 fold cross validation accuracy if 1 nearest neighbor classifier with Manhattan distance is used? The first fold is the first five data points.

fold 2 fold 1

x	1	1	2	2	3	3	4	4	5	5
y	1	2	3	3	2	2	3	3	2	1

label accuracy

2 2 2 2 2 2 2 2 2 2

2/5 2/5

2CV accuracy = 4/10 = 40%

Cross Validation Diagram

Quiz

Cross Validation 2

Quiz

Q2

- Given the following training data. What is the 10 fold cross validation accuracy if 1 nearest neighbor classifier with Manhattan distance is used?

x	1	1	2	2	3	3	4	4	5	5
y	1	2	3	3	2	2	3	3	2	1

- A: 20 percent, B: 40, C: 60, D: 80, E: 100

2 1 3 3 2 2 3 3 1 2

6/10 Correct.