CS540 Introduction to Artificial Intelligence
Lecture 6

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

June 1, 2020

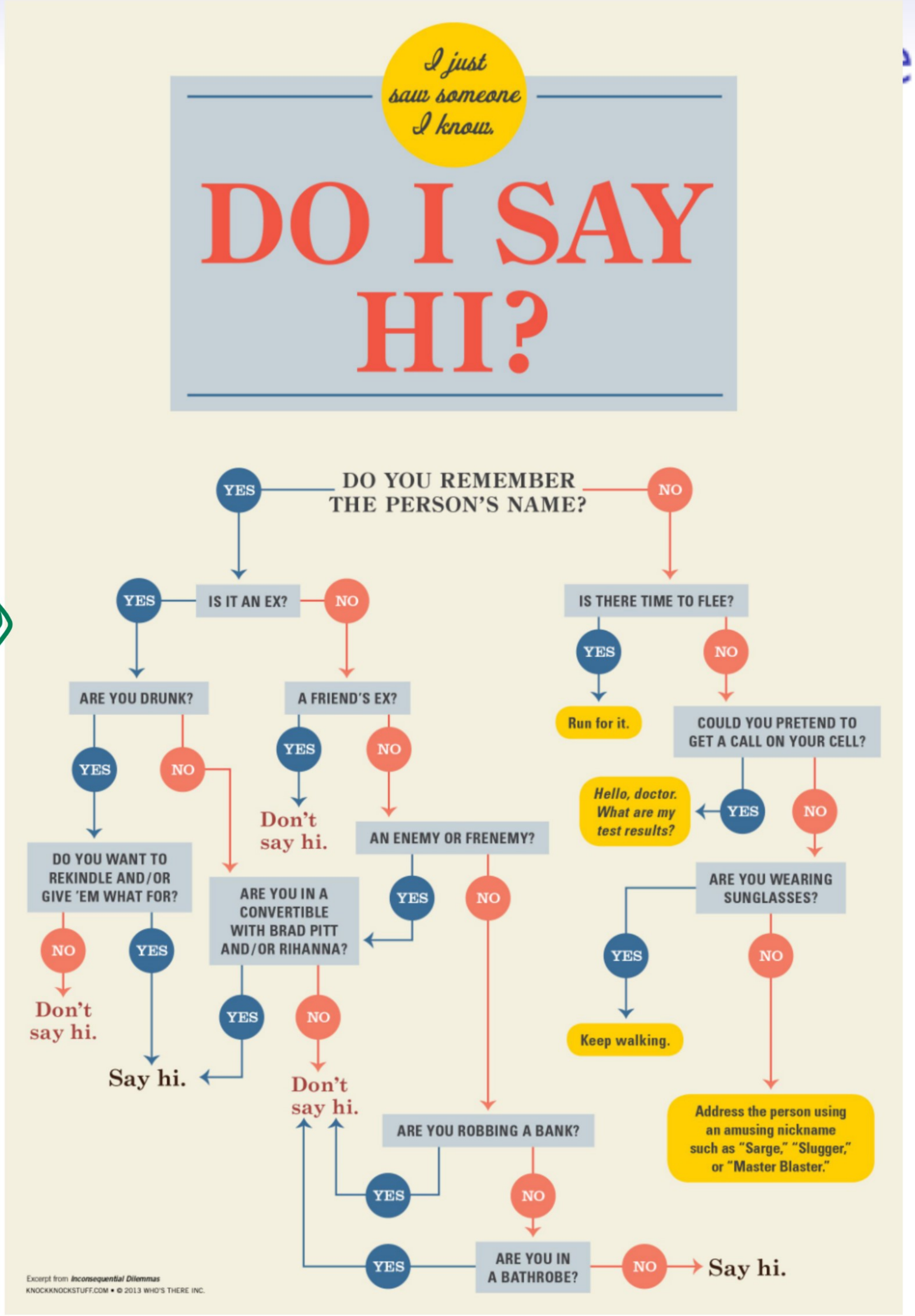
Hat Game

Quiz (Participation)

- 5 kids are wearing either green or red hats in a party: they can see every other kid's hat but not their own.
- Dad said to everyone: at least one of you is wearing green hat.
- Dad asked everyone: do you know the color of your hat?
- Everyone said no. ← at least 2 green hats,
- Dad asked again: do you know the color of your hat?
- Everyone said no. ← at lease 3 green hats,
- Dad asked again: do you know the color of your hat?
- Some kids (at least one) said yes. ← only 2 other green all others red
- No one lied. How many kids are wearing green hats?
- A: 1... B: 2... [✓ C: 3...] D: 4... E: 5

Hat Game Diagram

Discussion



I just saw someone I know.

DO I SAY HI?

DO YOU REMEMBER THE PERSON'S NAME?

YES

IS IT AN EX?

YES

ARE YOU DRUNK?

YES

DO YOU WANT TO REKINDLE AND/OR GIVE 'EM WHAT FOR?

NO

Don't say hi.

YES

Say hi.

NO

ARE YOU IN A CONVERTIBLE WITH BRAD PITT AND/OR RIHANNA?

YES

Say hi.

NO

Don't say hi.

NO

A FRIEND'S EX?

YES

Don't say hi.

NO

AN ENEMY OR FRENEMY?

YES

ARE YOU IN A CONVERTIBLE WITH BRAD PITT AND/OR RIHANNA?

NO

ARE YOU ROBBING A BANK?

YES

Don't say hi.

NO

ARE YOU IN A BATHROBE?

YES

Don't say hi.

NO

Say hi.

NO

IS THERE TIME TO FLEE?

YES

Run for it.

NO

COULD YOU PRETEND TO GET A CALL ON YOUR CELL?

YES

Hello, doctor. What are my test results?

NO

ARE YOU WEARING SUNGLASSES?

YES

Keep walking.

NO

Address the person using an amusing nickname such as "Sarge," "Slugger," or "Master Blaster."

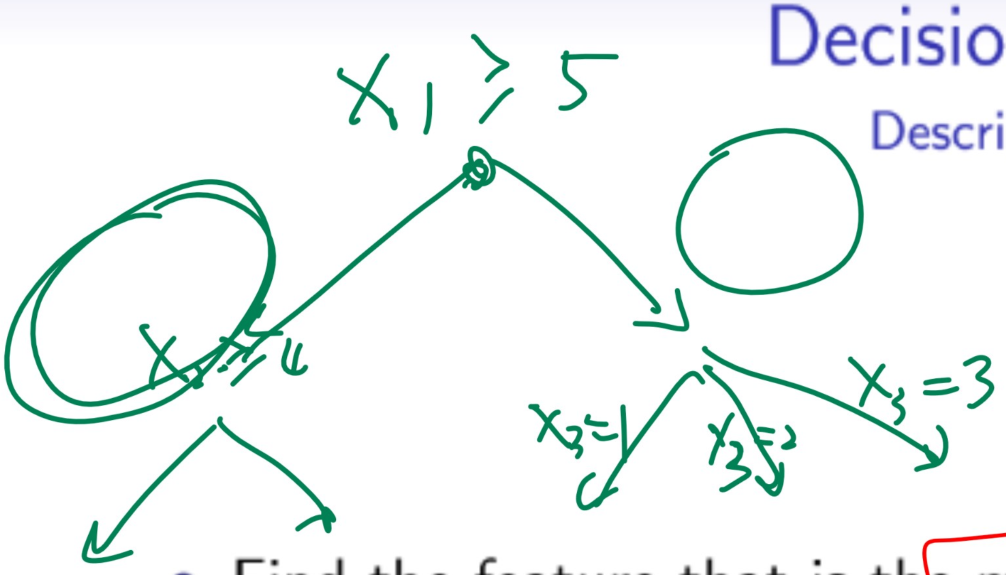
Excerpt from *Inconsequential Dilemmas*

KNOCKKNOCKSTUFF.COM • © 2013 WHO'S THERE INC.

w^T

Decision Tree

Description



y x_1 x_2 x_3

0

1

- 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

informative

- 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:

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

Entropy

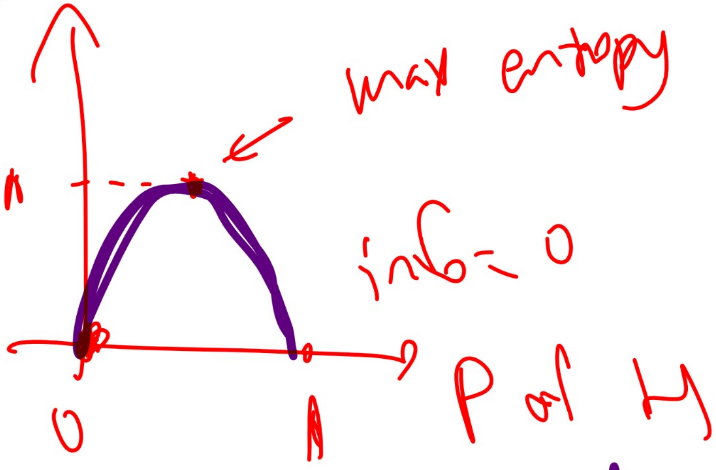
Definition

- 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:

0 √ 1

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

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



sum of entropy of indep = entropy of sum.

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.

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

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

Information Gain

Definition

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

$$\boxed{I(Y|X)} = \underline{H(Y)} - \underline{H(Y|\boxed{X})}.$$

X_1

X_2

- 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.

$$\boxed{\arg\max_{j} I\left(Y|X_j\right)}$$

- 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.

1325 CSL 3331

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{4}$$

$$\frac{1}{2} = 2^{-1}$$

$$\frac{1}{4} = 2^{-2}$$

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

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

Entropy Math

Quiz

Entropy 2

Quiz

Q8

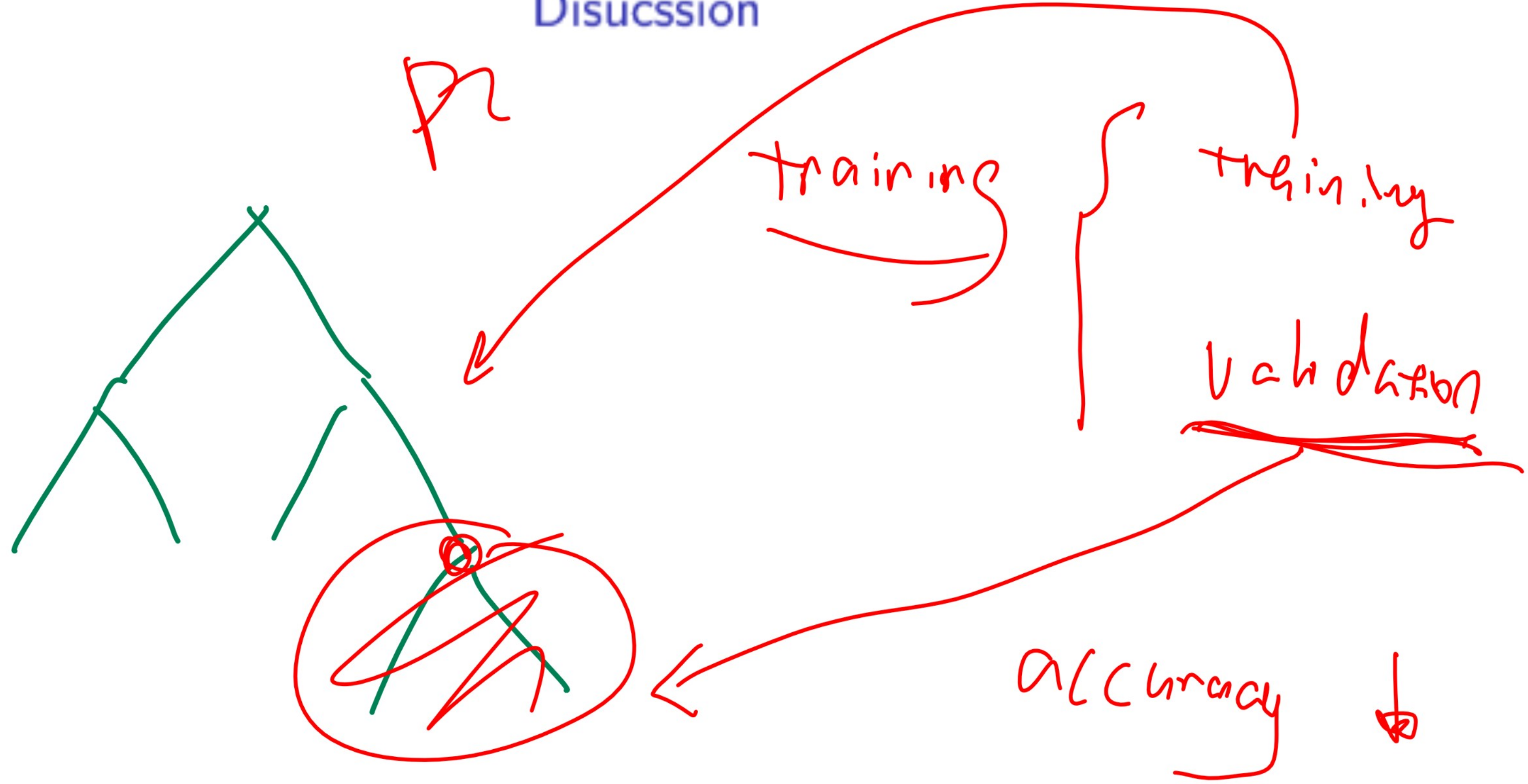
- A bag contains a red ball, a green ball, a blue ball, and a black ball. Randomly draw a ball from the bag with equal probability. What is the entropy of the outcome?
- A: 1
- B: $\log_2(3)$
- C: 1.5
- [✓ D: 2]
- E: 4

	red	green	blue	black
→ 4	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$$-\underbrace{4\cdot\frac{1}{4}}_{1}\cdot\underbrace{\log\frac{1}{4}}_{-2} = 2$$

Pruning Diagram

Disucssion



Ba~~i~~gging 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

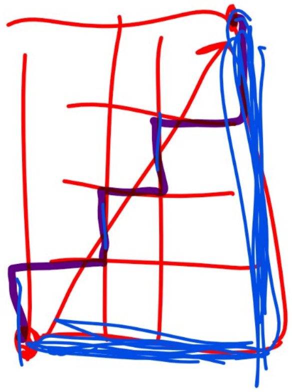
Definition

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

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

- An example is Manhattan distance.

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



1 Nearest Neighbor

Quiz

- 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

d 7 3 3 4 2

- [✓ A: 0]
- B: 1

3 Nearest Neighbor

Quiz

kNN

Q9

1

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

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

d 4 6 0 3 3

1 2 3

majority

- [✓ A: 0]
- B: 1

if tie use first example for cs540.

W2 →

Cross Validation

Quiz

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 1 fold 2

accuracy 40%

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

train on 1, test on 2 → 2 ~~2~~ ~~2~~ 2 ~~2~~

train on 2 test on 1 ~~2~~ 2 ~~2~~ ~~2~~ 2

CV → calculate accuracy on training set.

Cross Validation Diagram

Quiz

Cross Validation 2

Quiz

x_1

x_2

y

Q10.

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

train KNN

or other 9 points

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

label

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

60%

fold 1 ~~2~~ ~~1~~ 3 3 2 2 3 3 ~~1~~ ~~2~~

✓ ✓ ✓ ✓ ✓ ✓

$[x_{i1} = 1 \quad x_{i'1} = 2$

$x_i - x_{i'} = |2-1| = 1$

$$\left[\begin{array}{ll} x_{i1}=1 & x_{i'1}=2 \\ x_{i2}=1 & x_{i'2}=3 \\ \begin{pmatrix}1\\1\end{pmatrix} & \begin{pmatrix}2\\3\end{pmatrix} \end{array}\right. \longrightarrow$$

$$\begin{aligned} x_i - x_{i'}| &= |1-2| \\ &+ |1-3| \\ &= 3 \end{aligned}$$