CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2019

HAVERFORD
COLLEGE

Admin

- Lab 2 **TODAY**
- Next office hours **Friday 3-5pm**
- **TA hours** (in H204)
  - Sunday 7-8pm (Pablo)
  - Monday 7-8pm (Charlie)

Outline for September 12

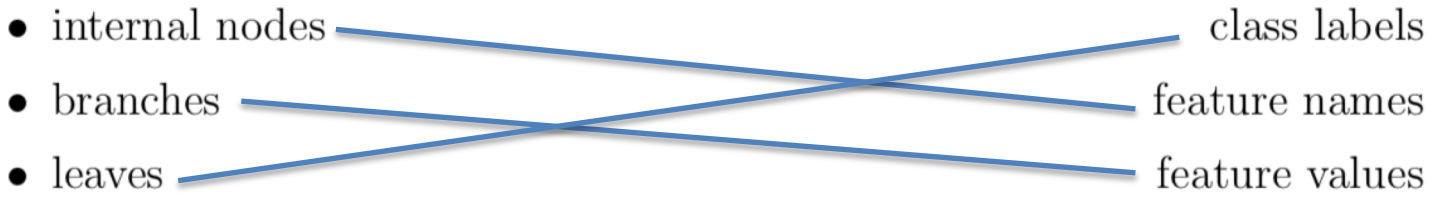
- Reading quiz
- Continue entropy and information gain
- Continuous features
- Lab 2 implementation suggestions
- Learning problem so far + terminology

Outline for September 12

- Reading quiz
- Continue entropy and information gain
- Continuous features
- Lab 2 implementation suggestions
- Learning problem so far + terminology

Reading Quiz #2

1)



Reading Quiz #2

1)

- internal nodes
- branches
- leaves

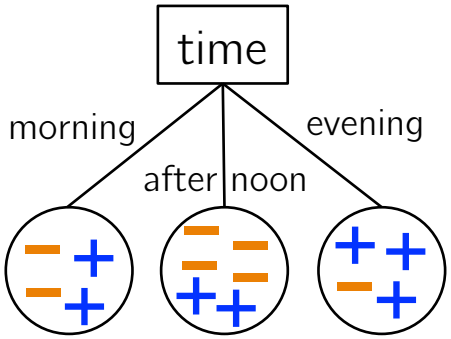
class labels

feature names

feature values

2) (a) +

(b) 5/14



Reading Quiz #2

1)

- internal nodes
- branches
- leaves

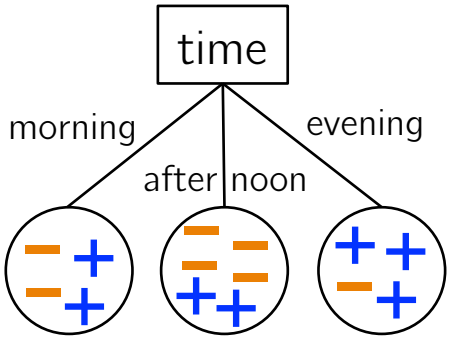
class labels

feature names

feature values

2) (a) +

(b) 5/14



3) high

Reading Quiz #2

1)

- internal nodes
- branches
- leaves

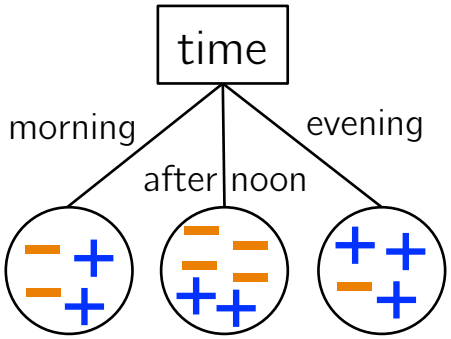
class labels

feature names

feature values

2) (a) +

(b) 5/14

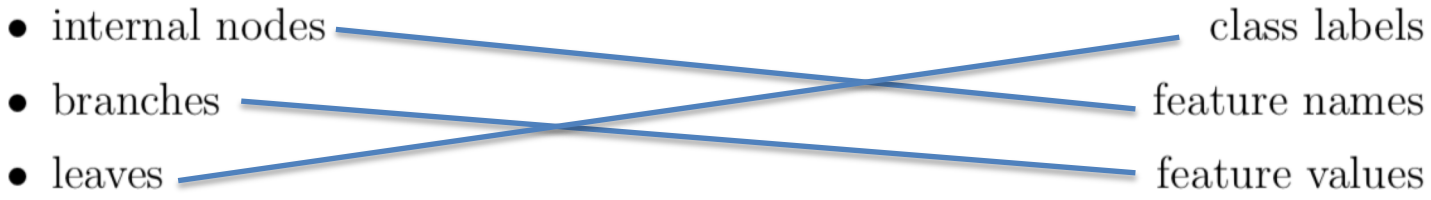


3) high

4) zero/one loss

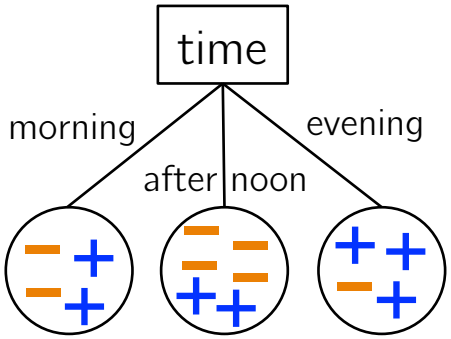
Reading Quiz #2

1)



2) (a) +

(b) 5/14



3) high

4) zero/one loss

5) Case when no training examples have the feature value

Outline for September 12

- Reading quiz
- Continue entropy and information gain
- Continuous features
- Lab 2 implementation suggestions
- Learning problem so far + terminology

Real-World Examples

- Medical diagnostics

- Credit risk analysis

- Modeling calendar scheduling preferences

Decision Trees in Chemistry reactions

- Example of decision trees in practice
- Use decision trees to interpret another ML algorithm (SVMs)

Machine-learning-assisted materials discovery using failed experiments

Paul Raccuglia, Katherine C. Elbert, Philip D. F. Adler, Casey Falk, Malia B. Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A. Friedler ✉, Joshua Schrier ✉ & Alexander J. Norquist ✉

Nature **533**, 73–76 (05 May 2016) | Download Citation ⬇

Optional Reading!

Note: we can put the negative sign inside or outside the ceiling function, since the ceiling function rounds up in an absolute value sense. i.e. ceil(2.25) = 3 and ceil(-2.25) = -3.

Y year	prob (p)	cummulative prob	in binary	$\lceil \log_2(p) \rceil$	code
senior	0.5	0	0.0000...	1	0
junior	0.25	0.5	0.100...	2	10
soph	0.125	0.75	0.110...	3	110
first	0.125	0.875	0.1110...	3	111

ceiling (round up)

inside or outside

#digits

Shannon encoding.

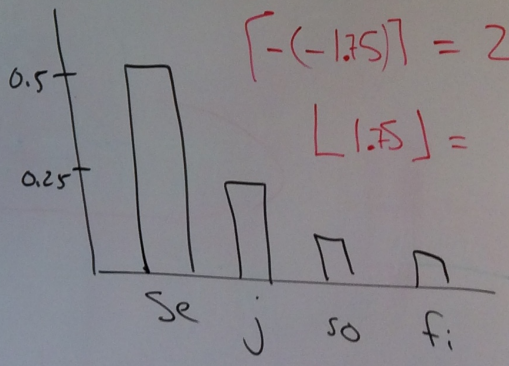
1948

entropy: avg # of bits needed to transmit one example.

110 110 100 0 111 0

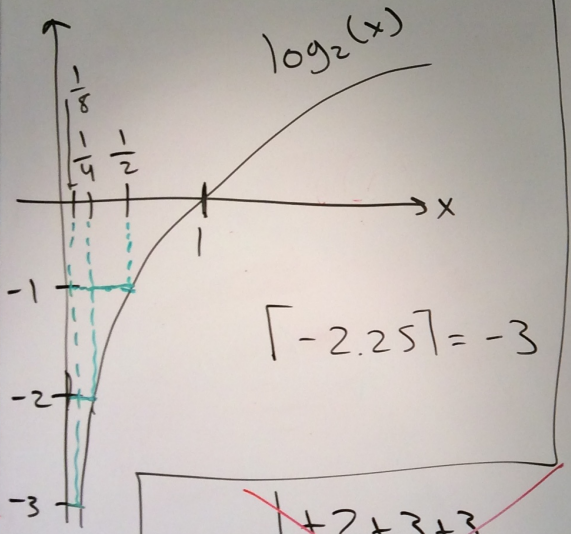
$\lceil -(-1.75) \rceil = 2$

$\lfloor 1.75 \rfloor = 1$



binary: $\square \cdot 2^2 + \square \cdot 2^1 + \square \cdot 2^0 + \square \cdot 2^{-1} + \square \cdot 2^{-2} \ldots$

decimal $\frac{1}{2}$ $\frac{1}{4}$



$\lceil -2.25 \rceil = -3$

$\frac{1+2+3+3}{4} = 2.25$

entropy

$$H(Y) = -\sum_{c \in vals(Y)} P(Y=c) \cdot \log_2(P(Y=c))$$

function2 · function1 · # of bits

$$H(year) = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \left(\frac{1}{8} \cdot 3\right) \cdot 2$$

$$\boxed{H(year) = 1.75}$$

Conditional Entropy

class · one feature · function 3

$$H(Y|X=v) = -\sum_{c \in vals(Y)} P(Y=c|X=v) \log_2\left(P(Y=c|X=v)\right)$$

$$P(\underbrace{Y=yes}_{A} | \underbrace{X=sunny}_{B}) = \frac{P(\underbrace{yes \text{ AND } sun}_{A \text{ & } B})}{P(\underbrace{sun}_{B})} = \frac{2}{5}$$

$$= \frac{2/14}{5/14}$$

Practice Problem

$H(Y|X) = \sum_{v \in vals(X)} P(X=v)\, H(Y|X=v)$

function 4

Information Gain

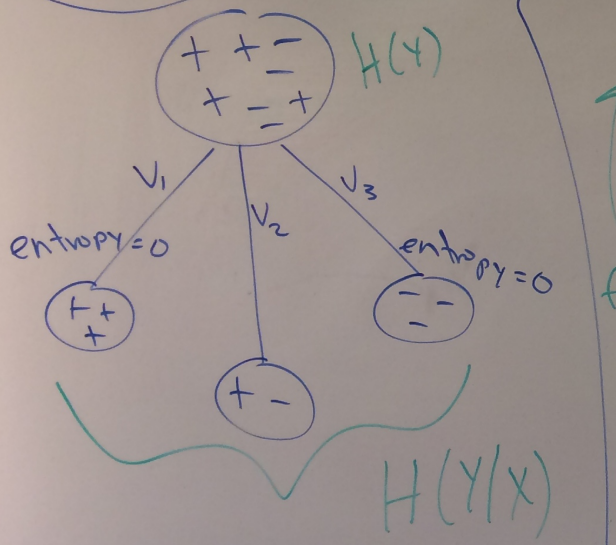
$Gain(Y,X) = H(Y) - H(Y|X)$

want to be high!

starts high

want low!

function 5



start at root

Goal: $H(Li|FA)$

famous actors

start

$P(Li = Yes | FA = No)$

$P(Li = No | FA = No)$

0.85

Handout 4: work with your group!

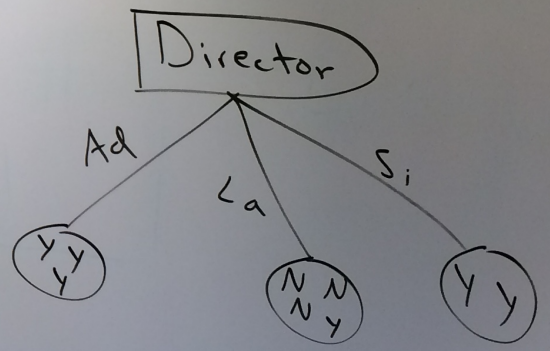
Handout 4

Movie	Type	Len	Director	Famous	Liked?
m_1	Com	S	Ad	N	Y
m_2	Ani	S	La	N	N
m_3	Dra	M	Ad	N	Y
m_4	Ani	L	La	Y	N
m_5	Com	L	La	Y	N
m_6	Dra	M	Si	Y	Y
m_7	Ani	S	Si	N	Y
m_8	Com	L	Ad	Y	Y
m_9	Dra	M	La	N	Y

$$H(Y) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)$$

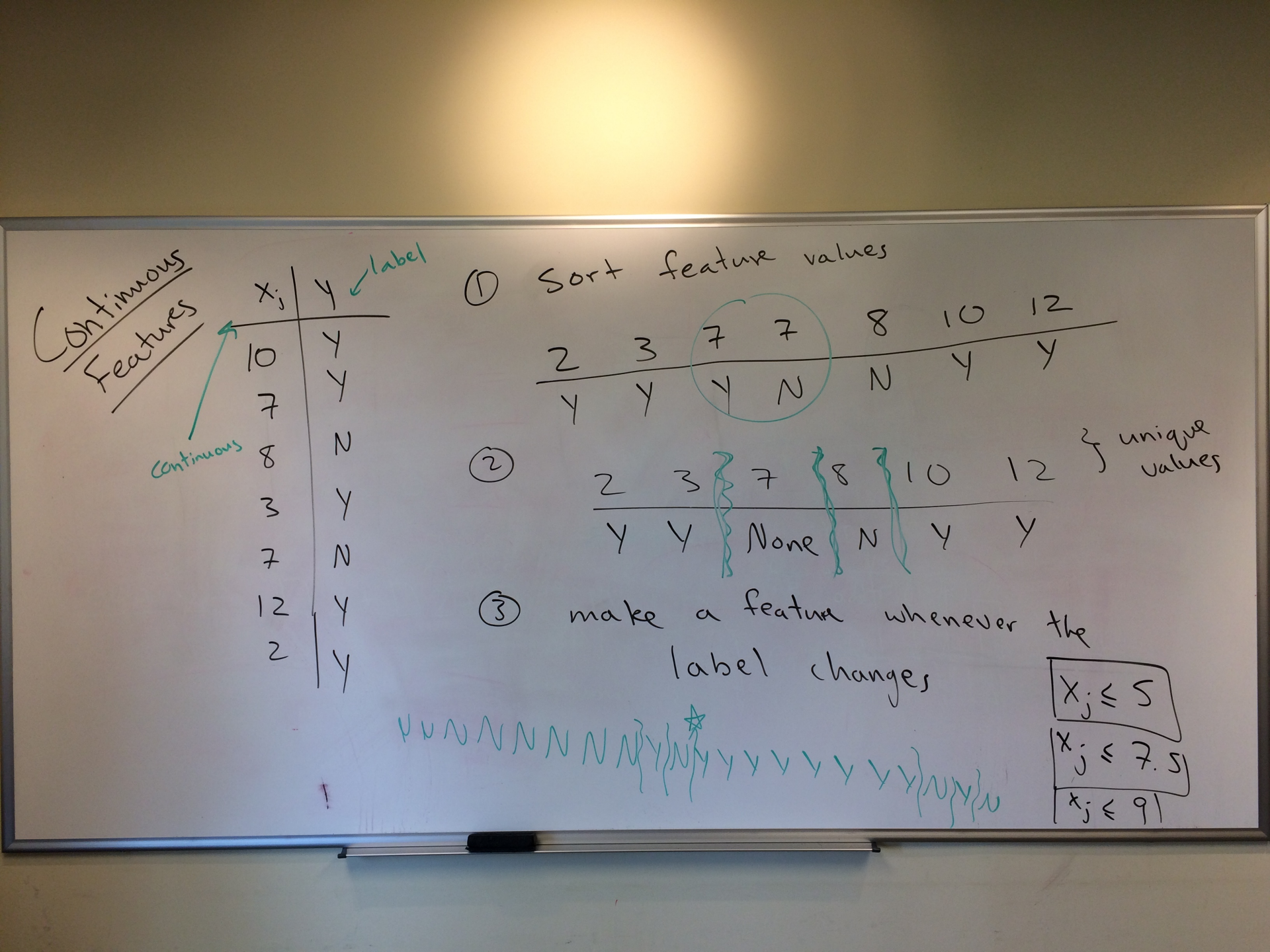
$$\approx 0.92$$

$$Gain(\tilde{Y}_i, T) = 0.92 - 0.61 = 0.31$$



Outline for September 12

- Reading quiz
- Continue entropy and information gain
- Continuous features
- Lab 2 implementation suggestions
- Learning problem so far + terminology


Continuous Features
x_j
y
label
continuous
10 Y
7 Y
8 N
3 Y
7 N
12 Y
2 Y
① Sort feature values
2 3 7 7 8 10 12
Y Y Y N N Y Y
②
2 3 7 8 10 12
Y Y None N Y Y
unique values
③ make a feature whenever the label changes
$x_j \leq 5$
$x_j \leq 7.5$
$x_j \leq 9$

Outline for September 12

- Reading quiz
- Continue entropy and information gain
- Continuous features
- Lab 2 implementation suggestions
- Learning problem so far + terminology

Implementation Suggestions

- Start slow with entropy! Build up function by function
- Think back to trees in data structures
- Distinguish between data (X,y) and options for data (values for each feature, classes for y)

Outline for September 12

- Reading quiz
- Continue entropy and information gain
- Continuous features
- Lab 2 implementation suggestions
- Learning problem so far + terminology

Next time!