Warm-up as you login

Search internet to find:

1. Machine learning classification dataset
   - Discrete/unordered output
2. Machine learning regression dataset
   - Continuous output

Are the input values discrete/continuous?

Are the input values a single value or multiple values?

What assumptions might we make with these datasets?

Announcements

Help us help you

- Student survey (see Piazza)
- Name pronunciation (via Canvas)

Assignments:

- HW1
  - Out today
  - Due Thu, 9/10, 11:59 pm (all times will be Pittsburgh time)

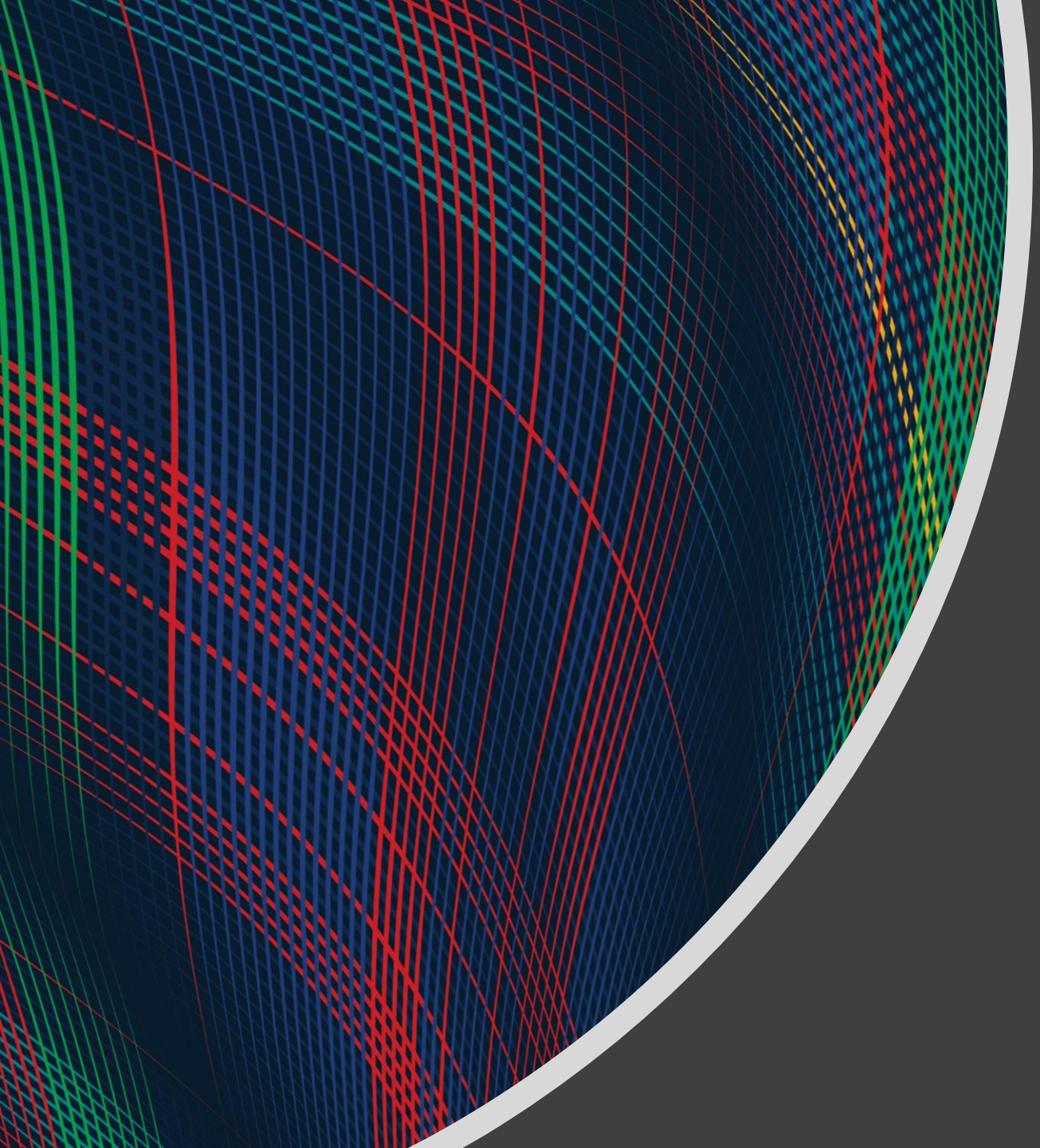
No class Monday: Labor Day

Q&A

Q: In Lecture 1, why did we use the term **experience** instead of just **data**?

A: Because our concern isn't just the data itself, but also where the data comes from (e.g. an agent interacting with the world vs. knowledge from a book).

As well, the word *experience* better aligns with the notion of what humans require in order to learn.



Introduction to Machine Learning

Decision Trees

Instructor: Pat Virtue

Plan

Today

- Problem formulation (notation)
- Algorithm 0: Memorization
- Algorithm 1: Majority vote
- Algorithm 2: Decision Stump

Monday

- Decision trees

Well-Posed Learning Problems

Three components *<T,P,E>*:

1. Task, *T*
2. Performance measure, *P*
3. Experience, *E*

Definition of learning:

A computer program **learns** if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

Definition from (Mitchell, 1997)

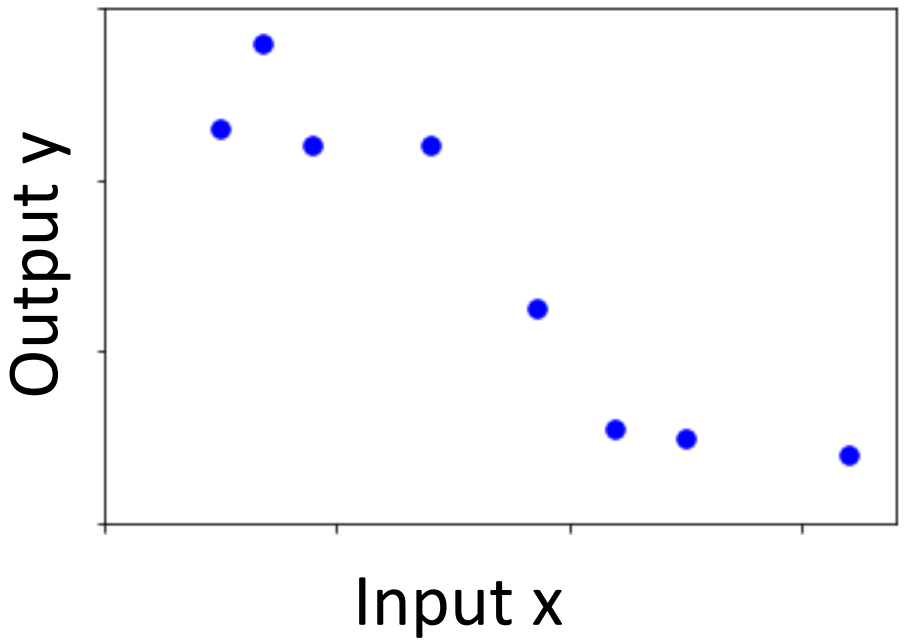
Problem Formulation

Experience

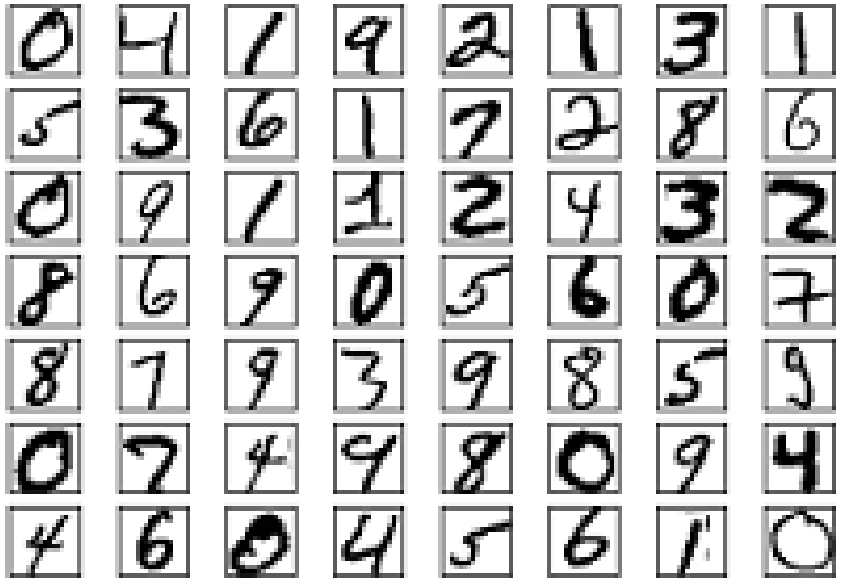
Hypothesis

Performance measure

Problem Formulation



Problem Formulation



Problem Formulation

Medical Prediction

Outcome	Fetal Position	Fetal Distress	Previous C-sec
Natural	Vertex	N	N
C-section	Breech	N	N
Natural	Vertex	Y	Y
C-section	Vertex	N	Y
Natural	Abnormal	N	N

Problem Formulation

Medical Prediction

Y	X_1	X_2	X_3
Outcome	**Fetal Position**	**Fetal Distress**	**Previous C-sec**
Natural	Vertex	N	N
C-section	Breech	N	N
Natural	Vertex	Y	Y
C-section	Vertex	N	Y
Natural	Abnormal	N	N

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [x_1, x_2, x_3]^T$$

$x_1 \in \{Vertex, Breech, Abn\}$
$x_2 \in \{Y, N\}$
$x_3 \in \{Y, N\}$

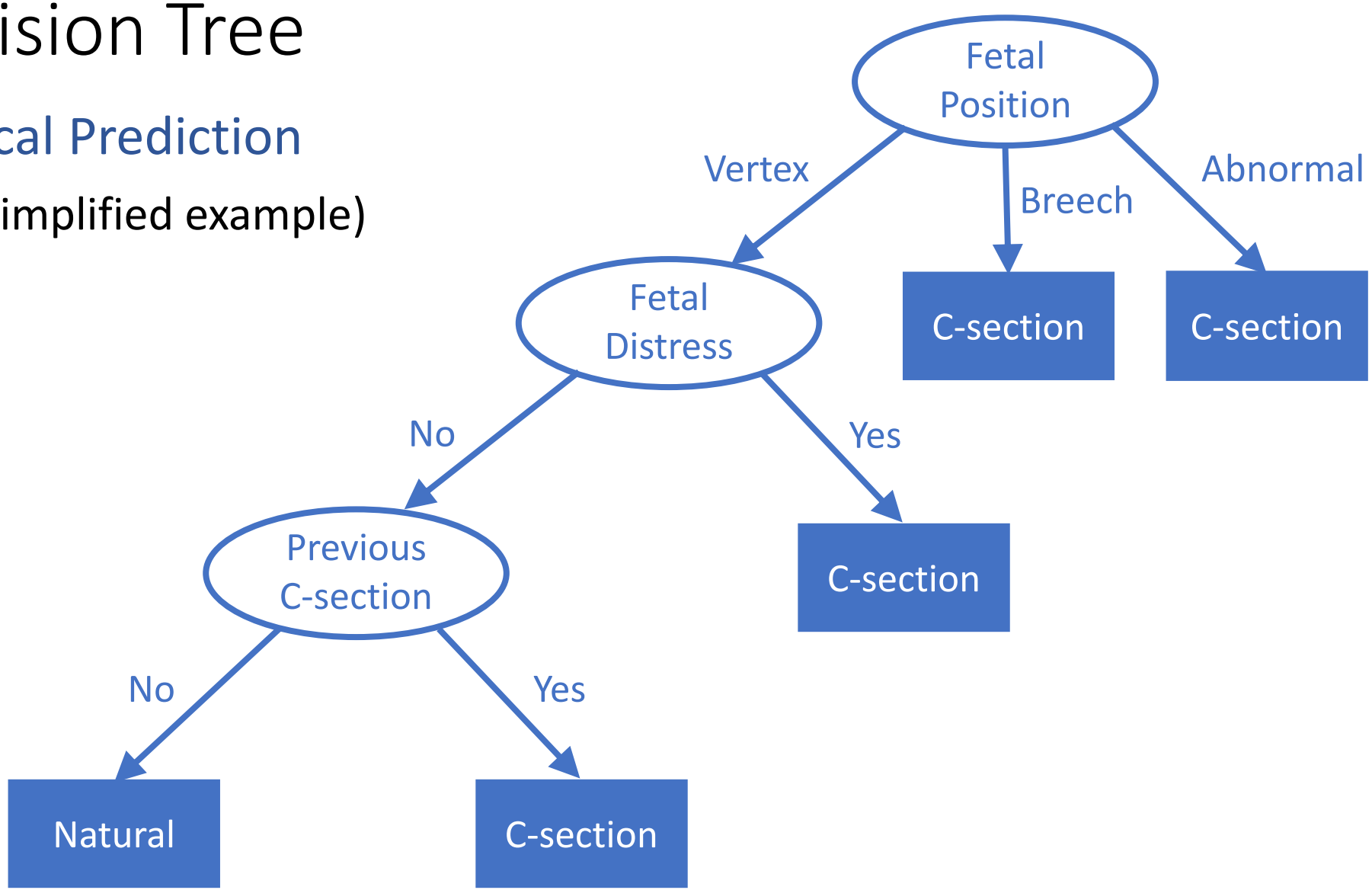
$y \in \{Csection, Natural\}$

$\hat{y} = h(\boldsymbol{x})$

Decision Tree

Medical Prediction

(Oversimplified example)



How could we implement training and prediction?

Algorithm 0: Memorization algorithm

Slide credit: CMU MLD Matt Gormley

Piazza Poll 1

Does the memorization algorithm learn?

A. Yes
B. No
C. I have no clue

Slide credit: CMU MLD Matt Gormley

How could we implement training and prediction?

Algorithm 1: Majority vote algorithm

Slide credit: CMU MLD Matt Gormley

Piazza Poll 2

What does the majority vote algorithm return on this training data?

A. A
B. B
C. C
D. 0
E. 1
F. +
G. -

Dataset:

Output Y, Attributes A, B, C

Y	A	B	C
-	1	0	0
-	1	0	1
-	1	0	0
+	0	0	1
+	1	1	0
+	1	1	1
+	1	1	0
+	1	1	1

Slide credit: CMU MLD Matt Gormley

Decision Stumps

Split data based on a single attribute

Majority vote at leaves

Dataset:

Output Y, Attributes A, B, C

Y	A	B	C
-	1	0	0
-	1	0	1
-	1	0	0
+	0	0	1
+	1	1	0
+	1	1	1
+	1	1	0
+	1	1	1

Slide credit: CMU MLD Matt Gormley

How could we implement training and prediction?

Algorithm 2: Decision stump algorithm

Slide credit: CMU MLD Matt Gormley

Piazza Poll 3

Splitting on which attribute {A, B, C} creates a decision stump with the lowest training error?

Dataset:

Output Y, Attributes A, B, C

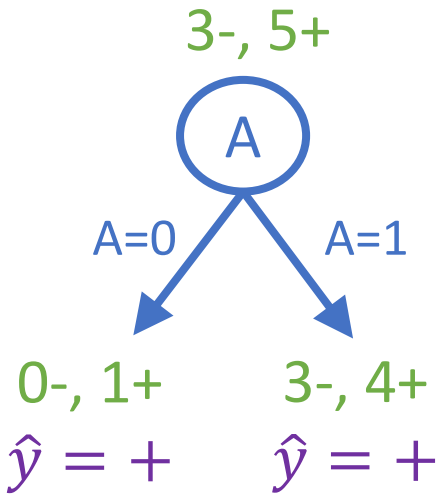
Y	A	B	C
-	1	0	0
-	1	0	1
-	1	0	0
+	0	0	1
+	1	1	0
+	1	1	1
+	1	1	0
+	1	1	1

Slide credit: CMU MLD Matt Gormley

Piazza Poll 3

Splitting on which attribute {A, B, C} creates a decision stump with the lowest training error?

Answer: B



Error rate: (0 + 3) / 8
= 3/8

Dataset:

Output Y, Attributes A, B, C

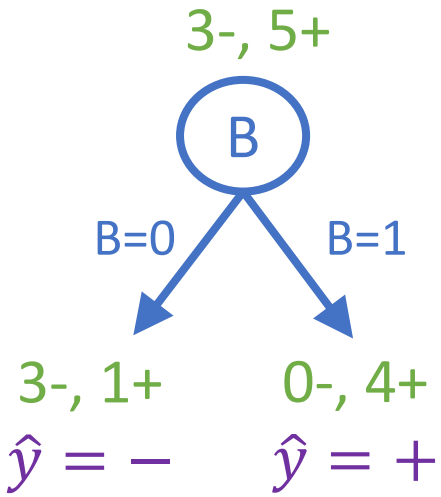
Y	A	B	C
-	1	0	0
-	1	0	1
-	1	0	0
+	0	0	1
+	1	1	0
+	1	1	1
+	1	1	0
+	1	1	1

Slide credit: CMU MLD Matt Gormley

Piazza Poll 3

Splitting on which attribute {A, B, C} creates a decision stump with the lowest training error?

Answer: B



Error rate: (1 + 0) / 8
= 1/8

Dataset:

Output Y, Attributes A, B, C

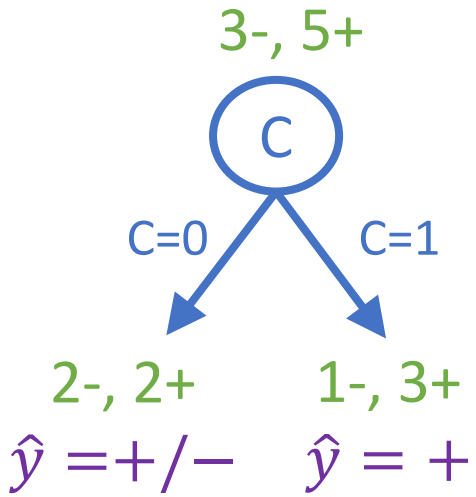
Y	A	B	C
-	1	0	0
-	1	0	1
-	1	0	0
+	0	0	1
+	1	1	0
+	1	1	1
+	1	1	0
+	1	1	1

Slide credit: CMU MLD Matt Gormley

Piazza Poll 3

Splitting on which attribute {A, B, C} creates a decision stump with the lowest training error?

Answer: B



Error rate: (2 + 1) / 8
= 3/8

Dataset:

Output Y, Attributes A, B, C

Y	A	B	C
-	1	0	0
-	1	0	1
-	1	0	0
+	0	0	1
+	1	1	0
+	1	1	1
+	1	1	0
+	1	1	1

Slide credit: CMU MLD Matt Gormley