10-315
Introduction to ML

Decision Trees

Instructor: Pat Virtue

Today

Autoencoders

More course info

Announcements

ML data and notation

Using data to learn

Decision trees

Autoencoders

Previous slides

Today

Autoencoders

More course info

Announcements

ML data and notation

Using data to learn

Decision trees

Machine Learning Systems

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)

Input/Output Tasks

Task

numerical data → Iris classification → category

hour/time → Traffic prediction → num

img → Image classification → category

img → Image denoising → img

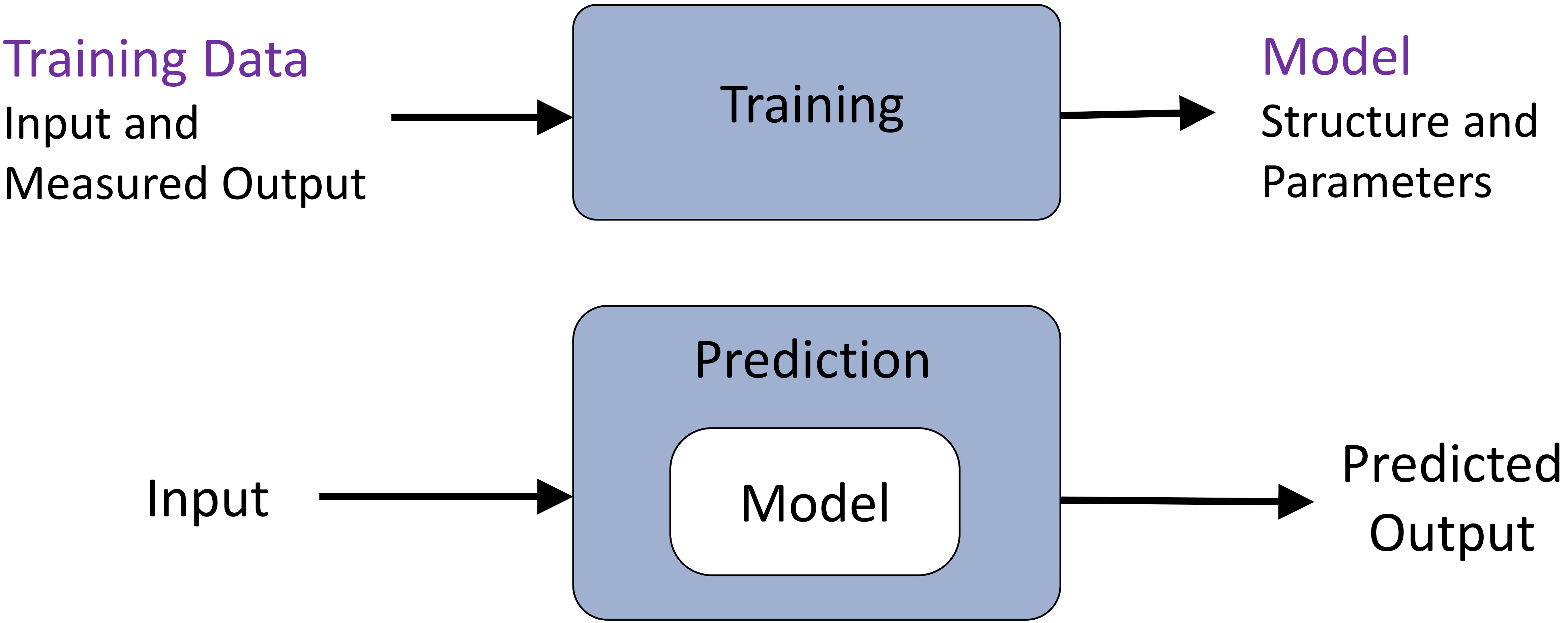
img → Medical image recon → img

text → Text to image generation → img

rand num → Face generation → img

Machine Learning

Using (training) data to learn a model that we'll later use for prediction



Example Dataset: Fisher Iris Dataset

Fisher (1936) used 150 measurements of flowers from 3 different species: Iris setosa (0), Iris virginica (1), Iris versicolor (2) collected by Anderson (1936)

Species	Sepal Length	Sepal Width	Petal Length	Petal Width
0	4.3	3.0	1.1	0.1
0	4.9	3.6	1.4	0.1
0	5.3	3.7	1.5	0.2
1	4.9	2.4	3.3	1.0
1	5.7	2.8	4.1	1.3
1	6.3	3.3	4.7	1.6
2	5.9	3.0	5.1	1.8

Images and full dataset: https://en.wikipedia.org/wiki/Iris_flower_data_set

Example Dataset: Fisher Iris Dataset

```python
from sklearn import datasets

iris = datasets.load_iris()
X = iris.data
y = iris.target
```

Assume samples in data are i.i.d.

Dataset notation

$$\mathcal{D} = \left\{\left(y^{(i)}, \mathbf{x}^{(i)}\right)\right\}_{i=1}^{N}$$

$$= \left\{\left(y^{(i)}, x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\right)\right\}_{i=1}^{N}$$

Linear algebra can represent all data

$\mathbf{y} \in \{0,1,2\}^N$

$X \in \mathbb{R}^{N\times 4}$ (design matrix)

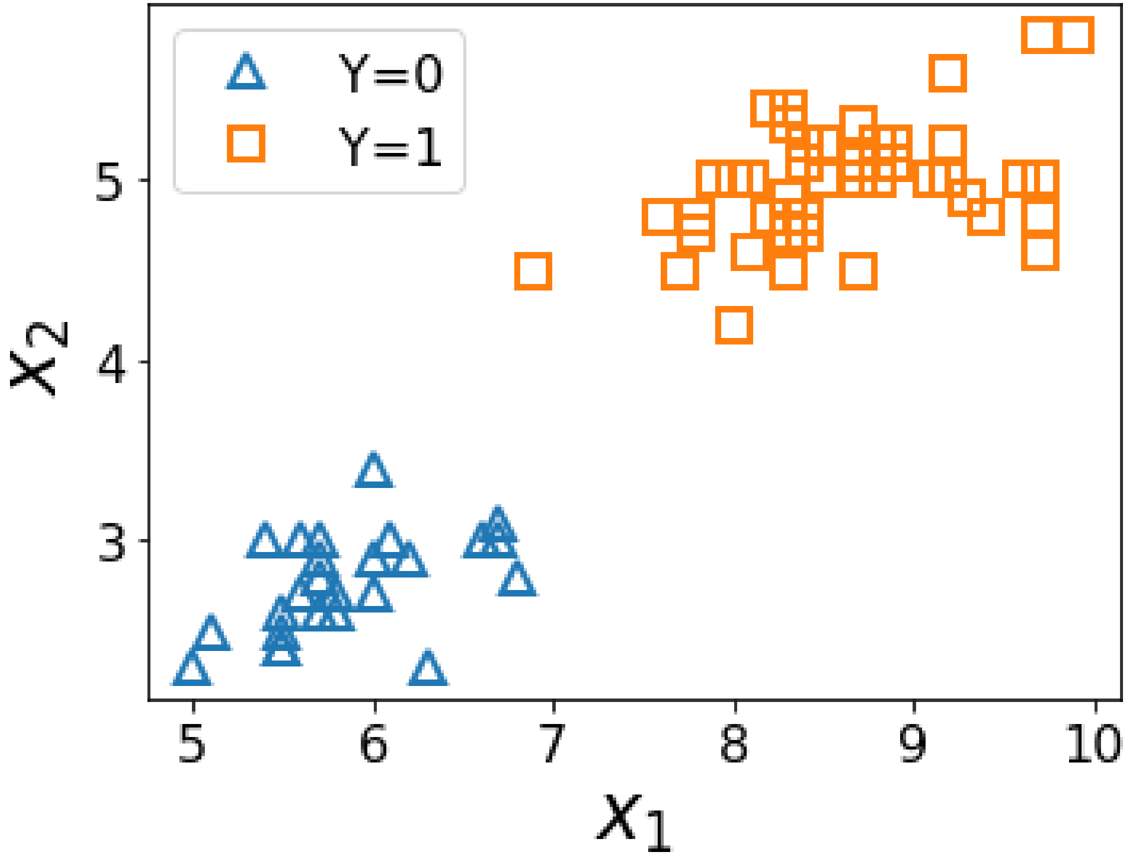
Species	Sepal Length	Sepal Width	Petal Length	Petal Width
0	4.3	3.0	1.1	0.1
0	4.9	3.6	1.4	0.1
0	5.3	3.7	1.5	0.2
1	4.9	2.4	3.3	1.0
1	5.7	2.8	4.1	1.3
1	6.3	3.3	4.7	1.6
2	5.9	3.0	5.1	1.8

Images and full dataset: https://en.wikipedia.org/wiki/Iris_flower_data_set

ML Task: Classification

Predict species label from first two input measurements

$h(\mathbf{x}) \rightarrow \hat{y}$



Species	Sepal Length	Sepal Width
0	4.3	3.0
0	4.9	3.6
0	5.3	3.7
1	4.9	2.4
1	5.7	2.8
1	6.3	3.3
2	5.9	3.0

Images and full dataset: https://en.wikipedia.org/wiki/Iris_flower_data_set

ML Tasks

Supervised learning

$$\mathcal{D} = \left\{\left(y^{(i)}, \mathbf{x}^{(i)}\right)\right\}_{i=1}^{N} \qquad h(\mathbf{x}) \to \hat{y}$$

Classification

- Output labels
- $y \in \mathcal{Y}$, where $\mathcal{Y}$ is discrete and order of values has no meaning

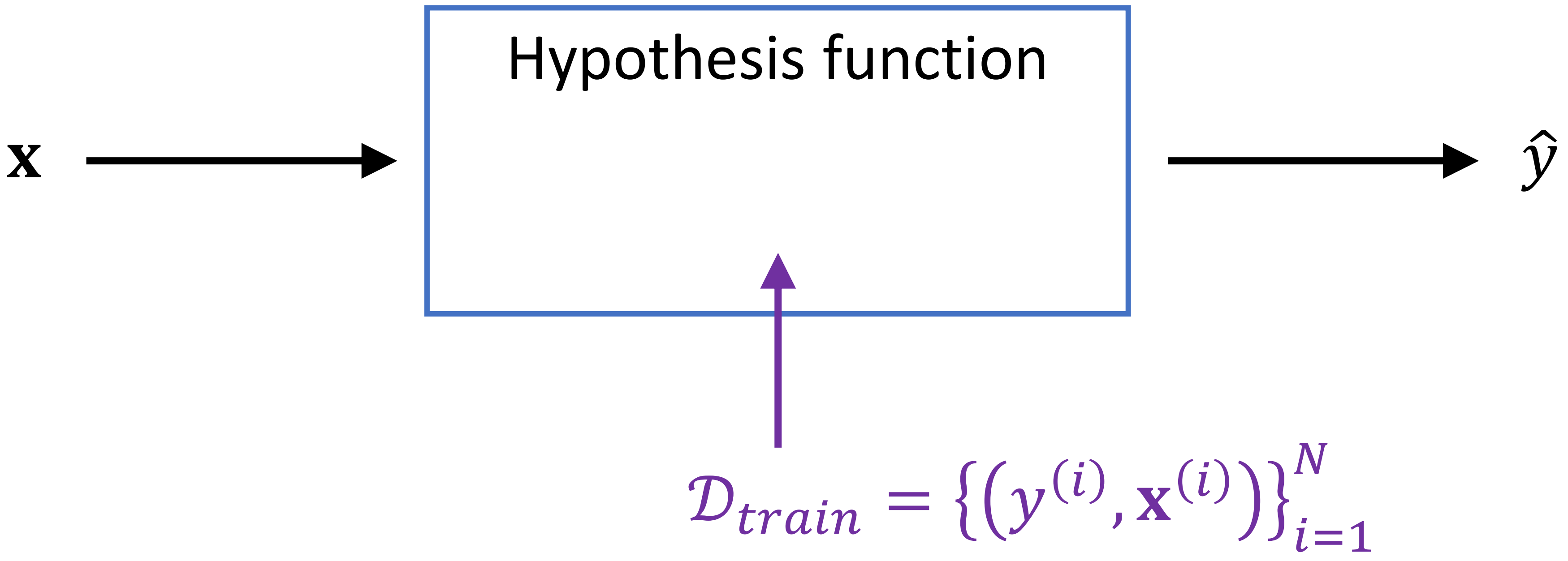
Regression

- Output values
- $y \in \mathcal{Y}$, where $\mathcal{Y}$ is usually continuous, order of values has meaning

Supervised Learning

Use training dataset to learn "best" hypothesis function

Predict output from input features

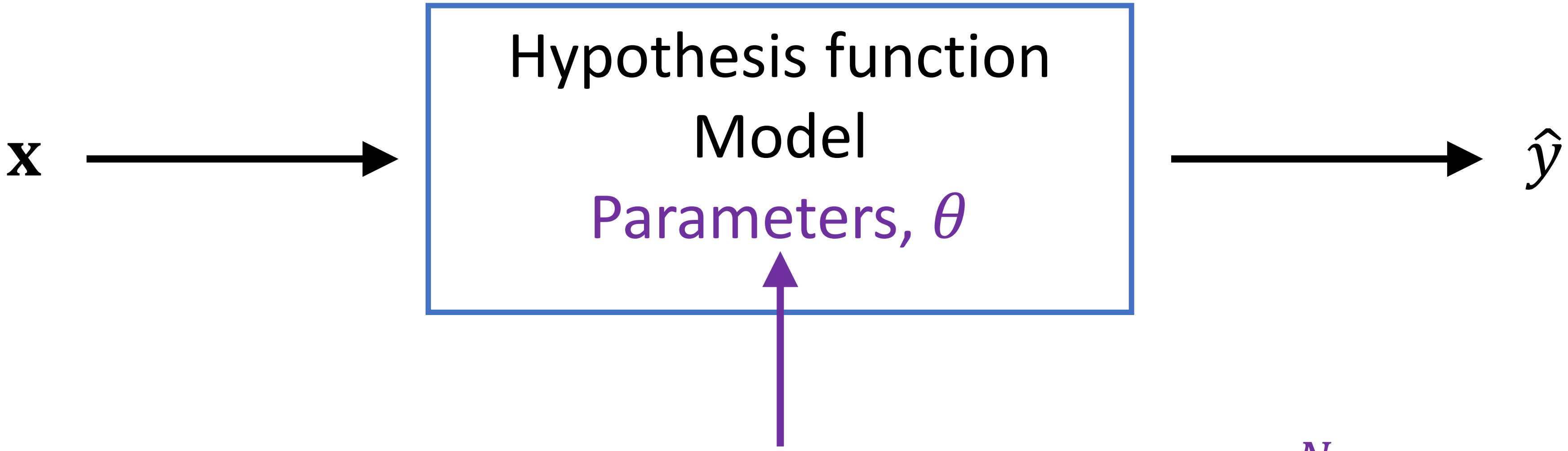


Need performance measure to determine meaning of "best" function

Supervised Learning

Use training dataset to learn "best" hypothesis function

Predict output from input features



$$\mathcal{D}_{train} = \left\{\left(y^{(i)}, \mathbf{x}^{(i)}\right)\right\}_{i=1}^{N}$$

Need performance measure to determine meaning of "best" function

Classification

Iris data example

$$\mathcal{D} = \left\{\left(y^{(i)}, \mathbf{x}^{(i)}\right)\right\}_{i=1}^{N}, \text{ where } \mathbf{x}^{(i)} \in \mathbb{R}^4, y^{(i)} \in \{0, 1, 2\}$$

Predict species label from input measurements

$$h(\mathbf{x}) \rightarrow \hat{y}$$

Performance measure?

Classification error rate

- Fraction of times $y \neq \hat{y}$ in a given dataset

Species	Sepal Length	Sepal Width	Petal Length	Petal Width
0	4.3	3.0	1.1	0.1
0	4.9	3.6	1.4	0.1
0	5.3	3.7	1.5	0.2
1	4.9	2.4	3.3	1.0
1	5.7	2.8	4.1	1.3
1	6.3	3.3	4.7	1.6
2	5.9	3.0	5.1	1.8

Images and full dataset: https://en.wikipedia.org/wiki/Iris_flower_data_set

Classification

Iris data example

$\mathcal{D} = \left\{\left(y^{(i)}, \mathbf{x}^{(i)}\right)\right\}_{i=1}^{N}$, where $\mathbf{x}^{(i)} \in \mathbb{R}^4$, $y^{(i)} \in \{0, 1, 2\}$

Predict species label from input measurements

$$h(\mathbf{x}) \rightarrow \hat{y}$$

Performance measure?

Classification error rate

- Fraction of times $y \neq \hat{y}$ in a given dataset
- $\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(y^{(i)} \neq \hat{y}^{(i)}\right)$

Notation alert: Indicator function

$$\mathbb{I}(z) = \mathbf{1}(z) = \begin{cases} 1 & \text{if } z \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Species	Sepal Length	Sepal Width	Petal Length	Petal Width
0	4.3	3.0	1.1	0.1
0	4.9	3.6	1.4	0.1
0	5.3	3.7	1.5	0.2
1	4.9	2.4	3.3	1.0
1	5.7	2.8	4.1	1.3
1	6.3	3.3	4.7	1.6
2	5.9	3.0	5.1	1.8

Images and full dataset: https://en.wikipedia.org/wiki/Iris_flower_data_set

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 Outcome	X_1 Fetal Position	X_2 Fetal Distress	X_3 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

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

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

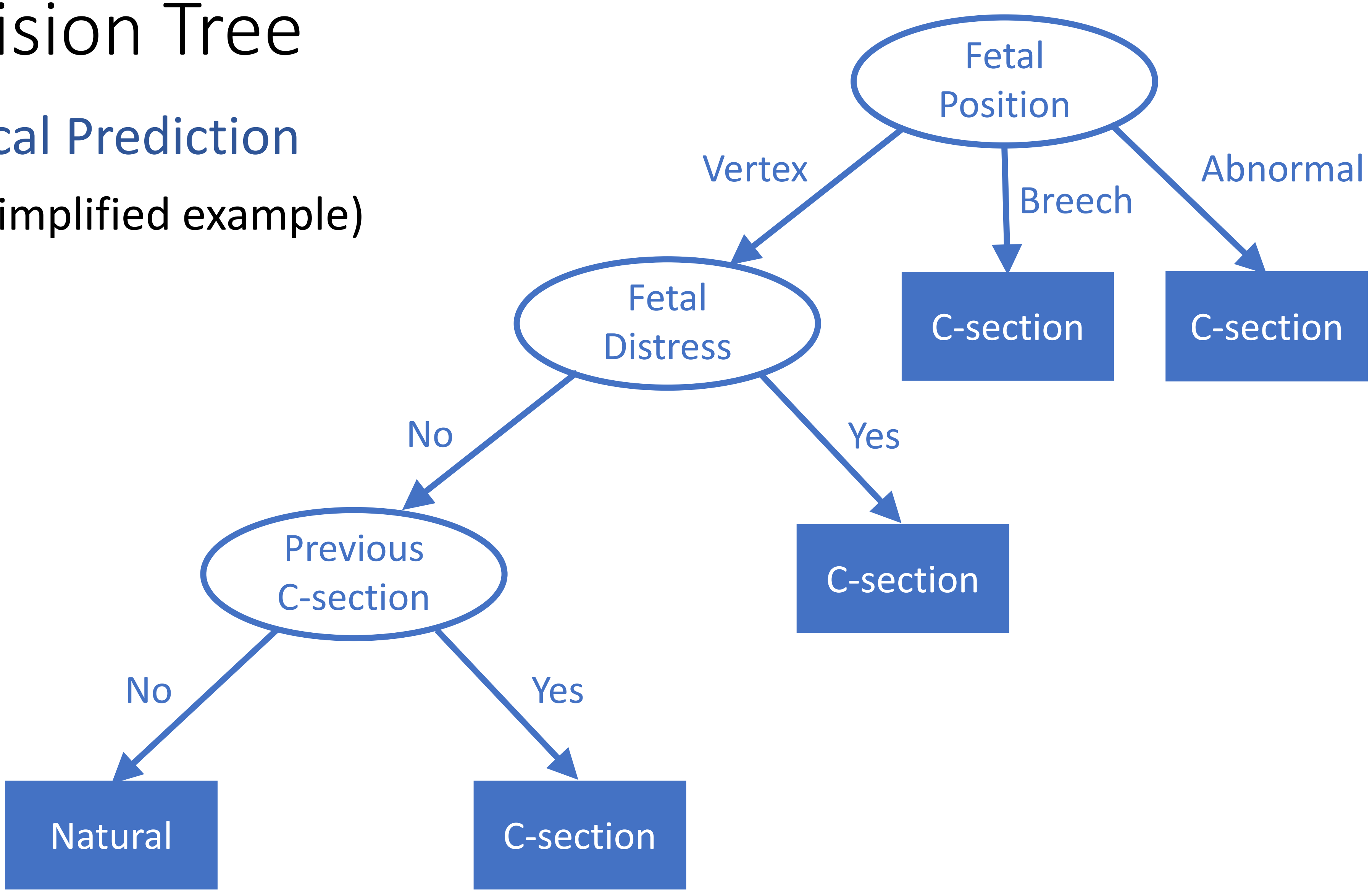
$$y \in \{Csection, Natural\}$$

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

Decision Tree

Medical Prediction

(Oversimplified example)



How could we implement training and prediction?

Algorithm 0: Memorization algorithm

Piazza Poll 1

Does the memorization algorithm learn?

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