Mahidol University
Faculty of Medicine Ramathibodi Hospital
Section for Clinical Epidemiology and Biostatistics

Introduction to Machine Learning

Ratchainant Thammasudjarit, Ph.D.

Wisdom of the Land

Class Policies

- Quiz (10%)
- Assignment (90%)

Topics

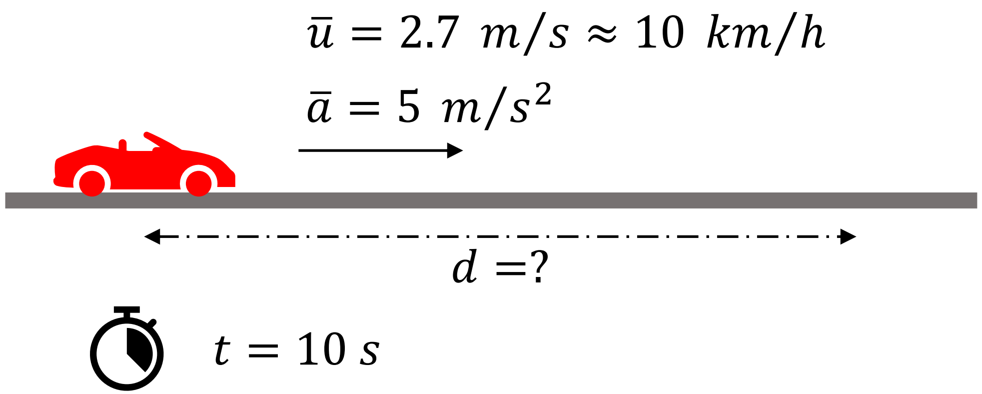
- Introduction to Machine Learning
- Naïve Bayes Classifier
- Decision Tree
- Logistic Regression
- Perceptron and Neural Network
- Model Deployment

Today Class

- AL, ML and DL
- Taxonomies and Use Cases

AI, ML and DL

Why do we need them



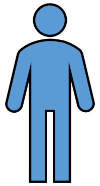
Solutions:

$$
\begin{aligned}
d &= \bar{u}t + \frac{1}{2}\bar{a}t^2 \\
&= 2.7(10) + \frac{1}{2}(5)(10^2) \\
&= 277 \ m
\end{aligned}
$$

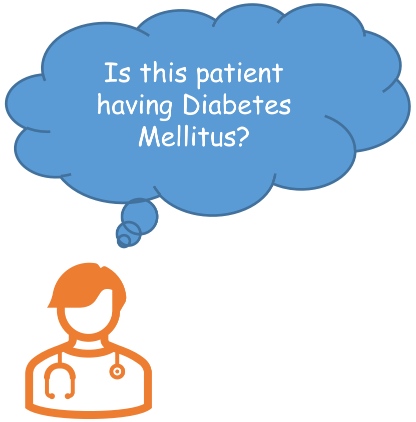
AI, ML and DL

Why do we need them

- Many real-world problems cannot be defined equations



BMI	Sex	LDL
35	Male	133



AI, ML and DL

Why do we need them

Detecting Spam Email

- Which email is spam mail
- How to define equation to detect spam mail

Sender	xxx@truedigitalacademy.com
Subject	Notification Acceptance

I am pleased to inform you that your paper has been found acceptable for publication pending minor revision. I anticipate that you will easily be able to answer the criticisms of the reviewers in a satisfactory manner.

- Many real-world problems cannot be defined equations

Sender	fifi12v@outlook.com
Subject	re

My name is [redacted]. I am a senior government official in the Department of Public works and housing of the Republic of South Africa and the Chairman of the board in charge of the contracts award execution and supervision of all contracts in the Ministry of works and Housing. In 2020, I personally monitored and Supervised the awarded contracts for the supply of building materials and construction of Ultra modern shopping malls, low-cost housing units, airports maintenance and Stadiums and hospitals /laboratories and Medical Equipment's for COVID-19 preparations in the Eastern Cape, Western Cape, Natal and Gauteng Provinces here in South Africa.

AI, ML and DL

What are AI, ML, and DL

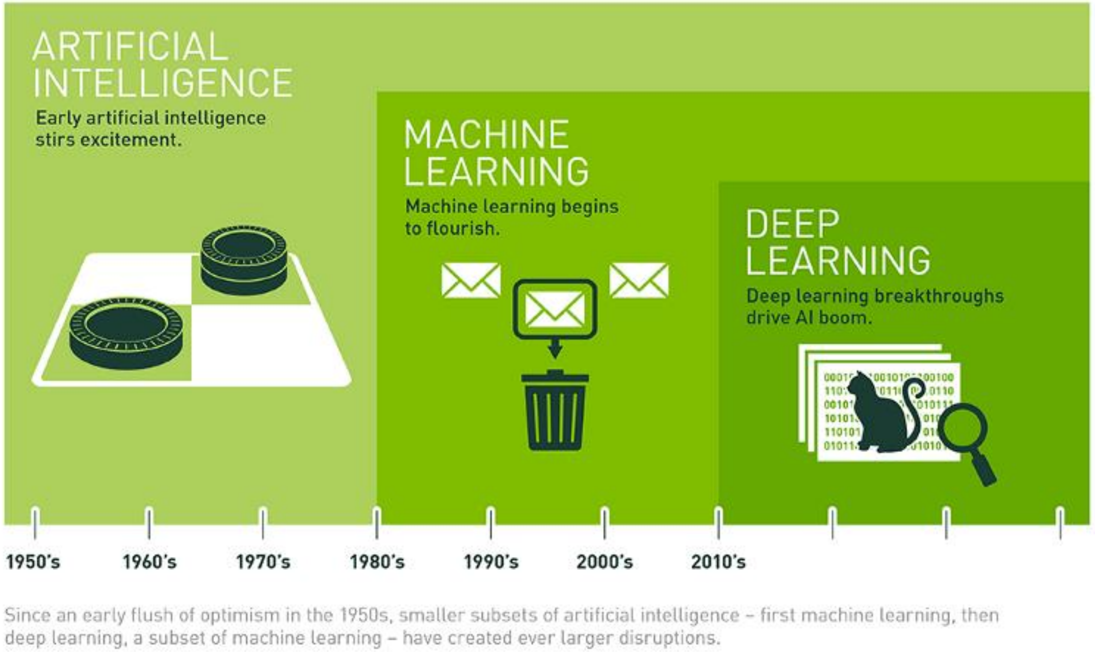


Image Source: Nvidia

- From AI to DL
  - Problem Complexity
  - Computational Power

ML and Statistics

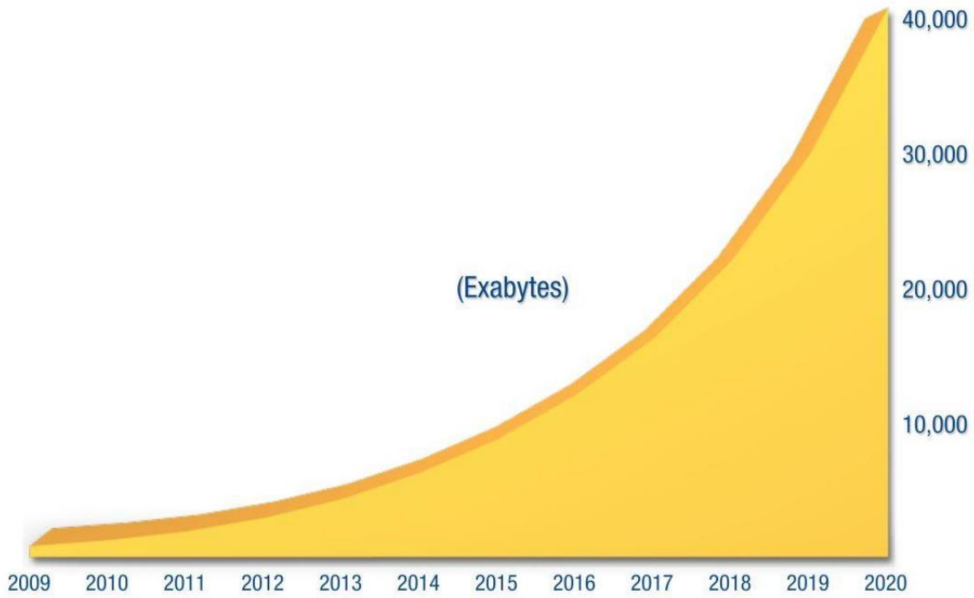
Commons and differences

- Common
  - Learning from data
- Difference
  - ML: Emphasize on model performance
  - Statistics: Emphasize on statistical inference

Why healthcare needs ML

How much data is being generated daily

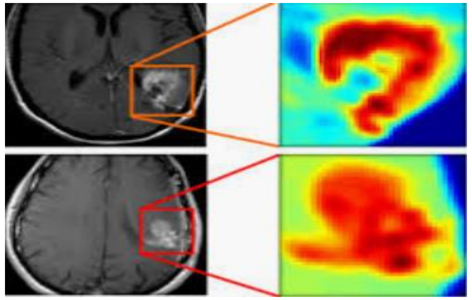
- The growth of data



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

Why healthcare needs ML

World is changing



- Task complexity
  - Predefining relationship between variables are difficult in many real-world cases

Benefits of ML in Healthcare

Better life

Types of Machine Learning

Supervised Learning

- Learning from **known samples**
- Prediction
  - Classification
  - Regression

Types of Machine Learning

Supervised Learning

Bacteria Keratitis Infection

Fungal Keratitis Infection

- Classification

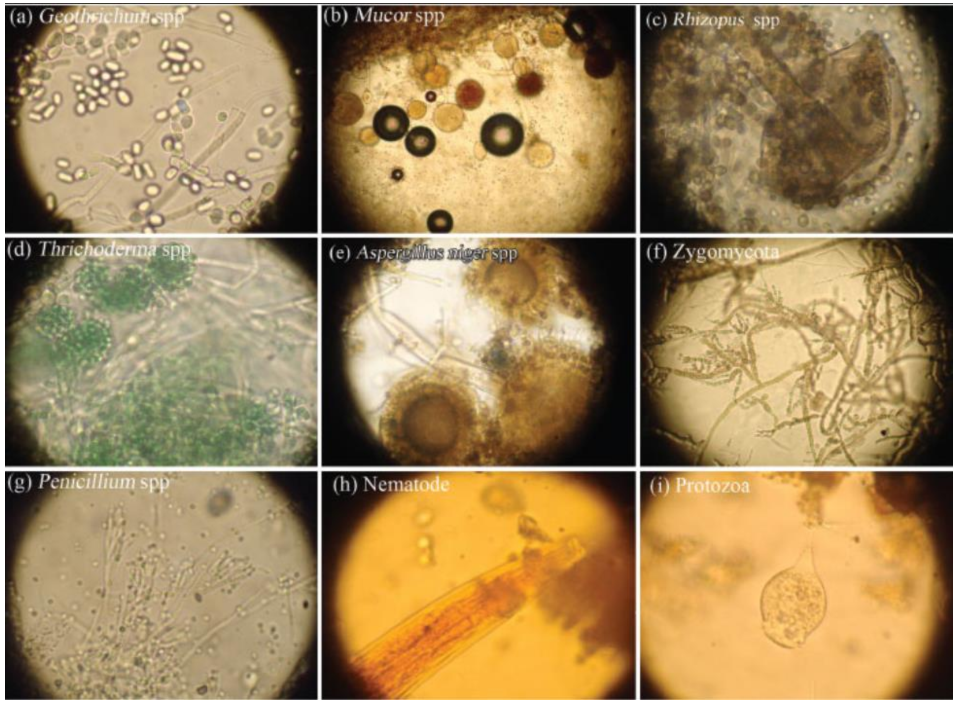
Microbial Keratitis Classification

Is a given image either bacteria or fungal infection?

- Predicting two possible class is called **binary classification problem**

Types of Machine Learning

Supervised Learning



- Classification

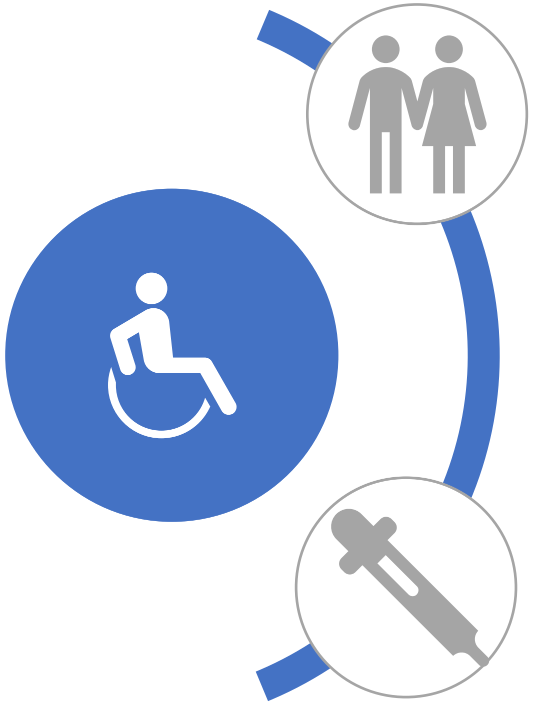
Microbial Classification

What is the microbial type of a given image?

- Predicting more than two possible class is called **Multi-classification problem**

Types of Machine Learning

Supervised Learning



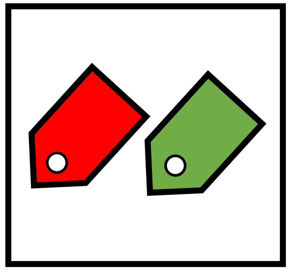
- Regression

Disease Monitoring

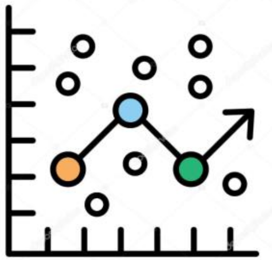
Estimating R_0 of Covid-19

Types of Machine Learning

Supervised Learning



Classification



Regression

- Discussions
  - What is the difference between classification and regression

Types of Machine Learning

Unsupervised Learning

- Learning from **samples**
- Tasks
  - Structure discovery
    - Clustering
    - Dimensionality Reduction

Types of Machine Learning

Unsupervised Learning

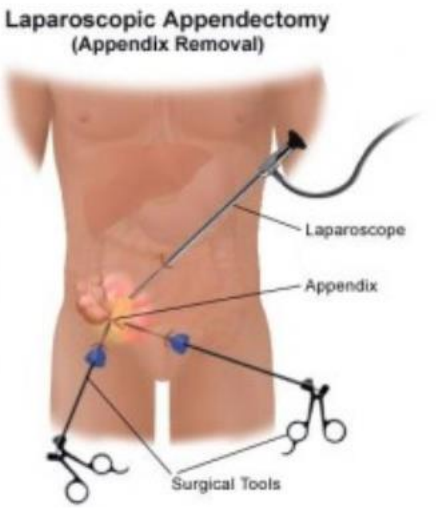
- Clustering

Diagnosis Related Group

What is the DRG of a given procedure?

Types of Machine Learning

Unsupervised Learning



- Each procedure consumes resources
- Procedure clustering helps hospital manage financial better

Types of Machine Learning

Reinforcement Learning

- Self-learning
- Tasks
  - Optimization
    - Decision Making

Types of Machine Learning

Reinforcement Learning

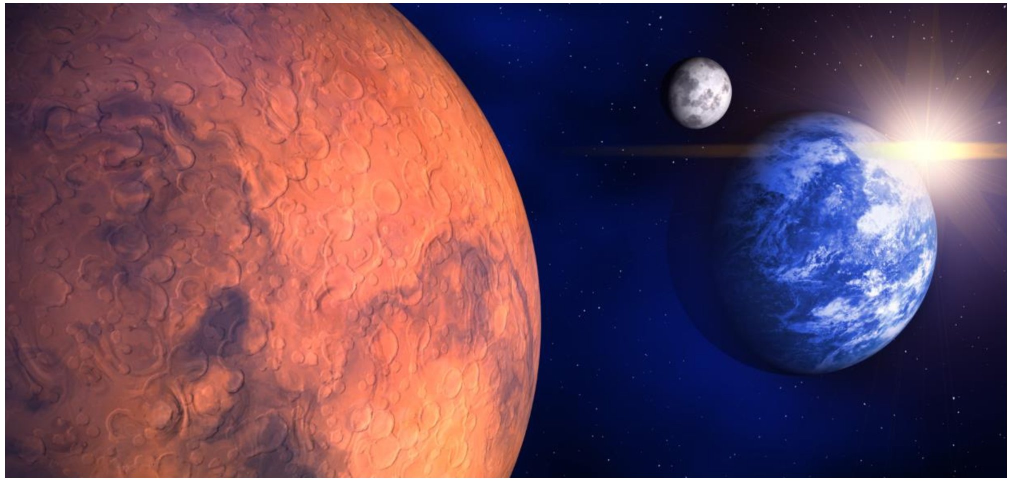
- Decision Making

Robotic Navigation

Exploring landscape and self-decide to perform action, e.g. moving forward, stop, turn left, turn right

Types of Machine Learning

Reinforcement Learning

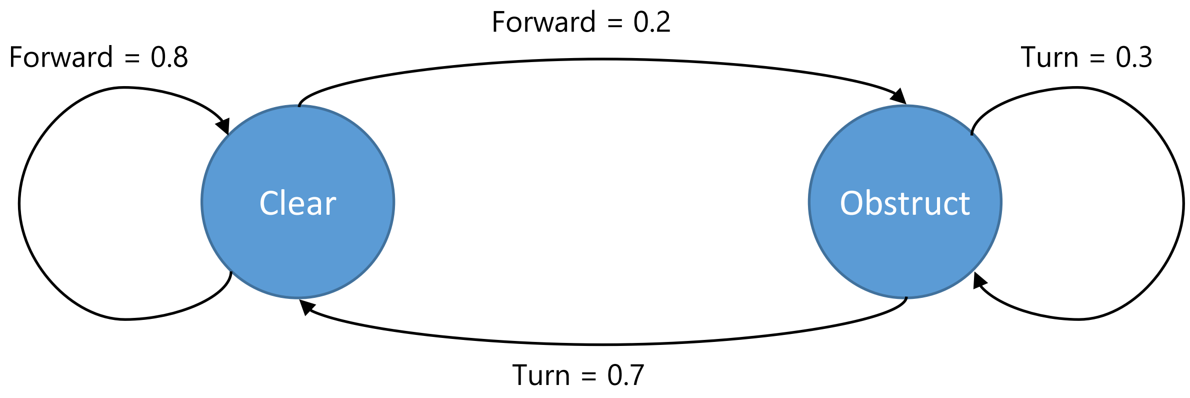


- Distance from the Earth to the Mars is 225 million kilometers
- Light speed is 300,000 km per second
- Light takes about 14 minutes traveling from the Earth to the Mars
- What if we would like to control the rover to explore the Mars

Types of Machine Learning

Reinforcement Learning

- State is a consequence of action
- Optimize objective function
  - Maximize explored area



Types of Machine Learning

Summary

Supervised Learning

- Prediction
  - Classification
  - Regression

Unsupervised Learning

- Structure Discovery
  - Clustering
  - Dimensionality Reduction

Reinforcement Learning

- Optimization
  - Decision Making

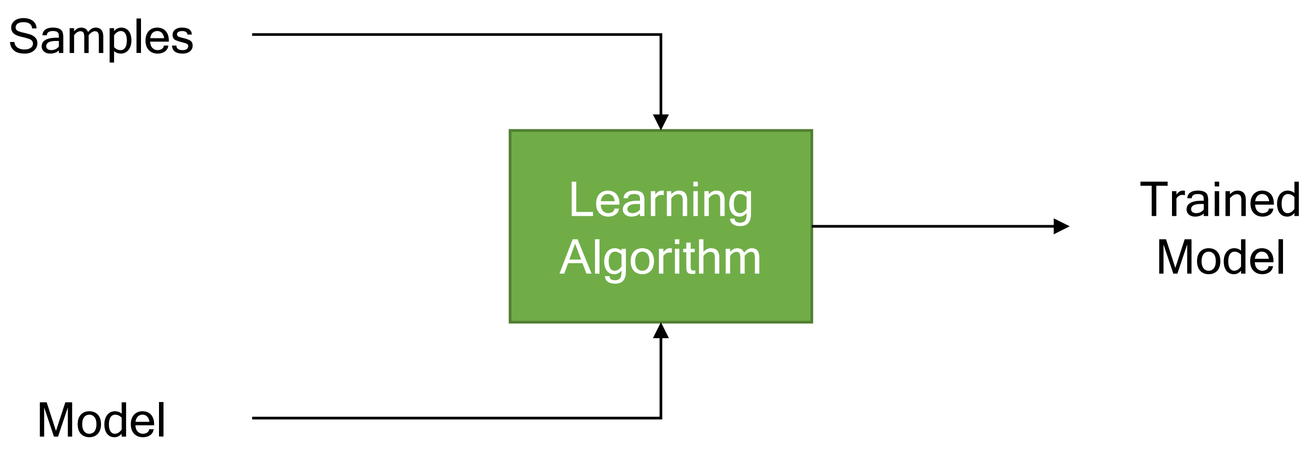
Machine Learning Components

Mathematically

$$\boldsymbol{X} \longrightarrow \boxed{F(\boldsymbol{X})} \longrightarrow \mathbf{y}$$

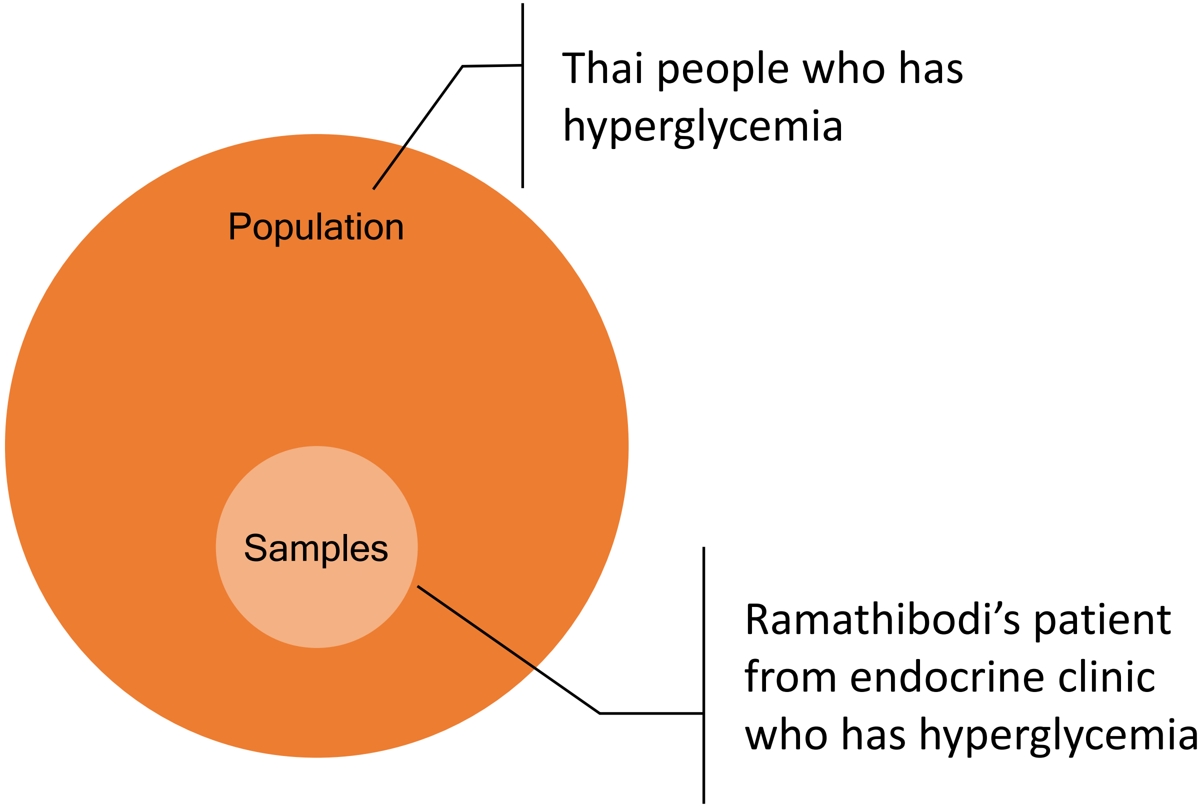
Machine Learning Components

Overall



Machine Learning Components

Samples



Machine Learning Components

Sample Terminologies and Notations

Example of diabetes prediction

Age	Sex	Diabetes
55	Male	yes
37	Female	no
48	Male	yes
60	Female	no
42	Male	yes

- Features are predictors or independent variables
- Standard notation is $\mathbf{x}$ as feature and x as value
  - $\mathbf{x}_{age} = \langle x \in I^{+} \mid 0 \leq x \leq 100 \rangle$
  - $\mathbf{x}_{sex} = \langle x \mid x \in \{Male, Female\} \rangle$
- Multiple features are represented by $\mathbf{X}$

Machine Learning Components

Sample Terminologies and Notations

Example of diabetes prediction

Age	Sex	Diabetes
55	Male	yes
37	Female	no
48	Male	yes
60	Female	no
42	Male	yes

- Target class is a variable to be predicted
  - Only required for supervised learning
- Standard notation is $\mathbf{y}$ as target class and y as class label
  - $\mathbf{y} = \langle y \mid y \in \{yes, no\} \rangle$

Machine Learning Components

Sample Terminologies and Notations

Example of diabetes prediction

Age	Sex	Diabetes
55	Male	yes
37	Female	no
48	Male	yes
60	Female	no
42	Male	yes

$(\mathbf{X}^{(2)}, y^{(2)})$

- Supervised learning
  - The sample i is a pair of feature vector and its class label
  - The standard notation is $(\mathbf{X}^{(i)}, y^{(i)})$
- Others
  - The sample i is a feature vector
  - The standard notation is $\mathbf{X}^{(i)}$

Machine Learning Components

Sample Terminologies and Notations

Example of diabetes prediction

Age	Sex	Diabetes
55	Male	yes
37	Female	no
48	Male	yes
60	Female	no
42	Male	yes

- Supervised Learning
  - Dataset is a collection of m pairs of feature vector and its class label
  - The standard notation is $(\mathbf{X}, \mathbf{y})$

Machine Learning Components

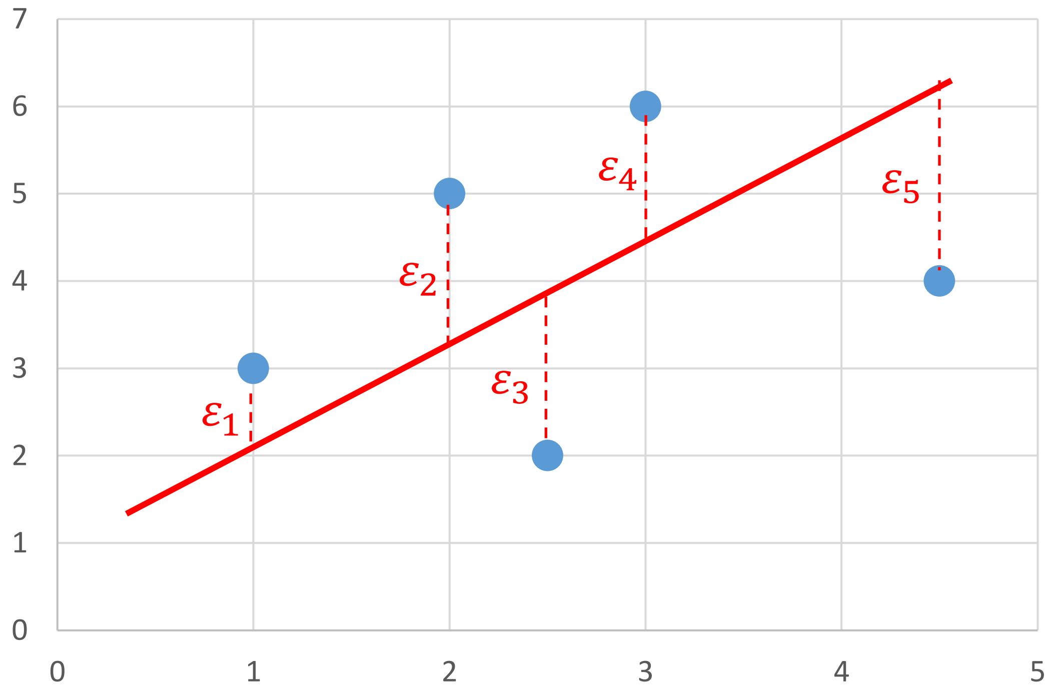
Sample Terminologies and Notations

Age	Sex	Diabetes
55	Male	yes
37	Female	no
48	Male	yes
60	Female	no
42	Male	yes

- Summary
  - Italic lowercase represents a scalar
  - Non-italic lowercase represents a vector
  - Non-italic uppercase represents a matrix

Machine Learning Components

Model



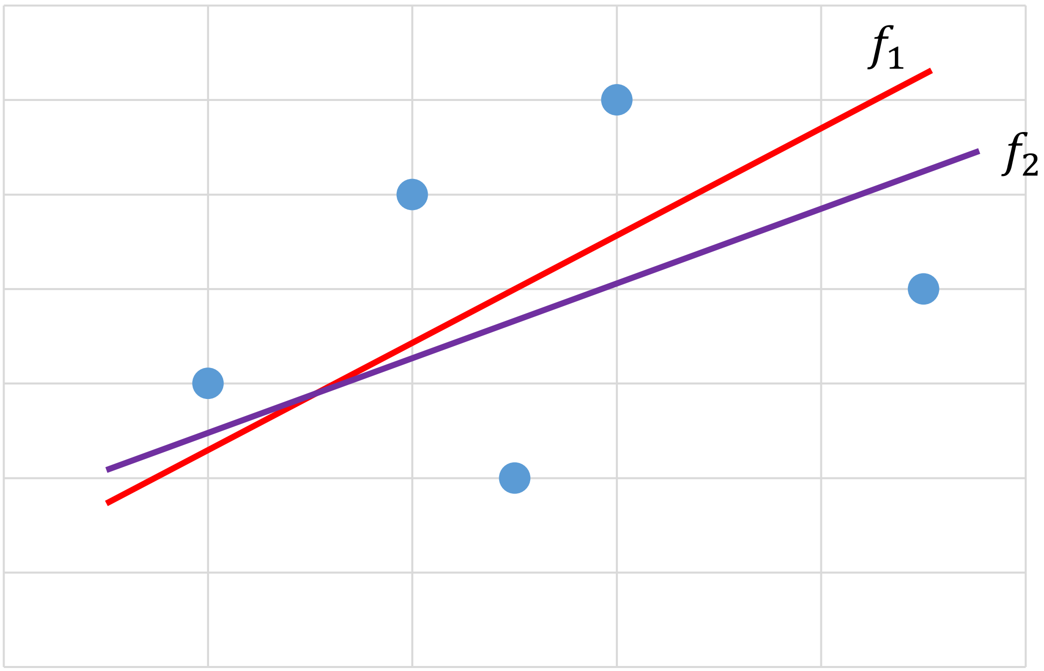
- Expression of relationships between variables
  - Example: Linear relationship

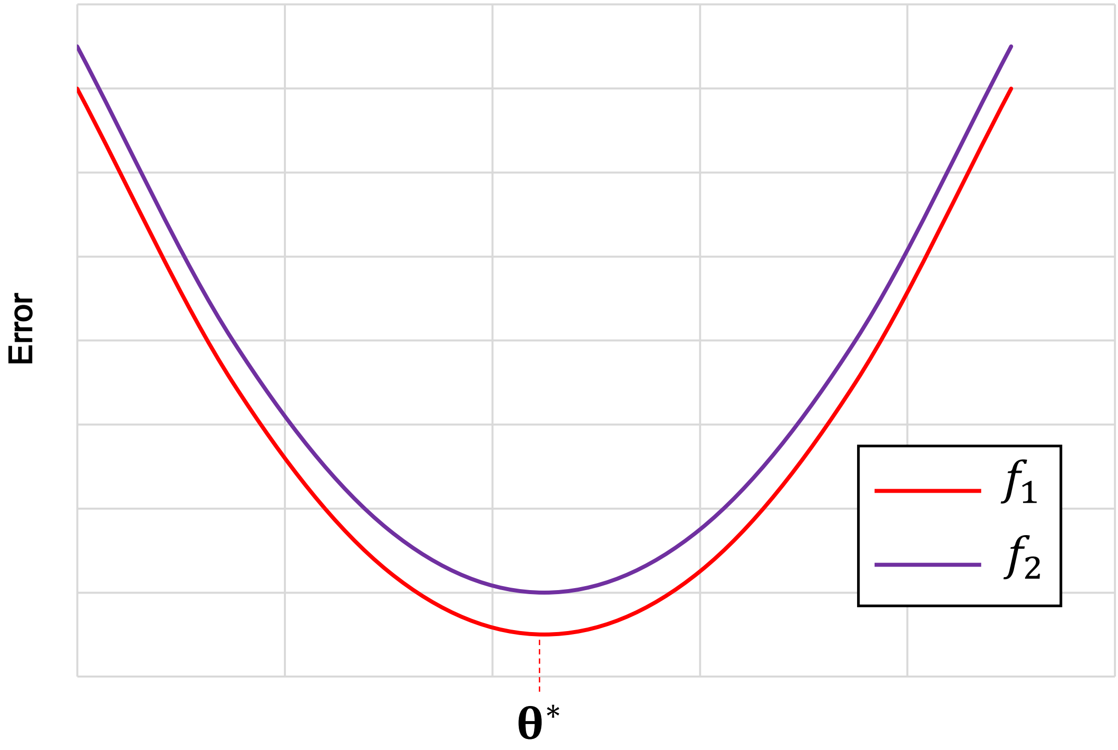
$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}$$

- Goal
  - Learn the **optimal parameters $\boldsymbol{\theta}^* = \langle \beta_0, \beta_1 \rangle$** from dataset $(\mathbf{X}, \mathbf{y})$ that minimize error

Machine Learning Components

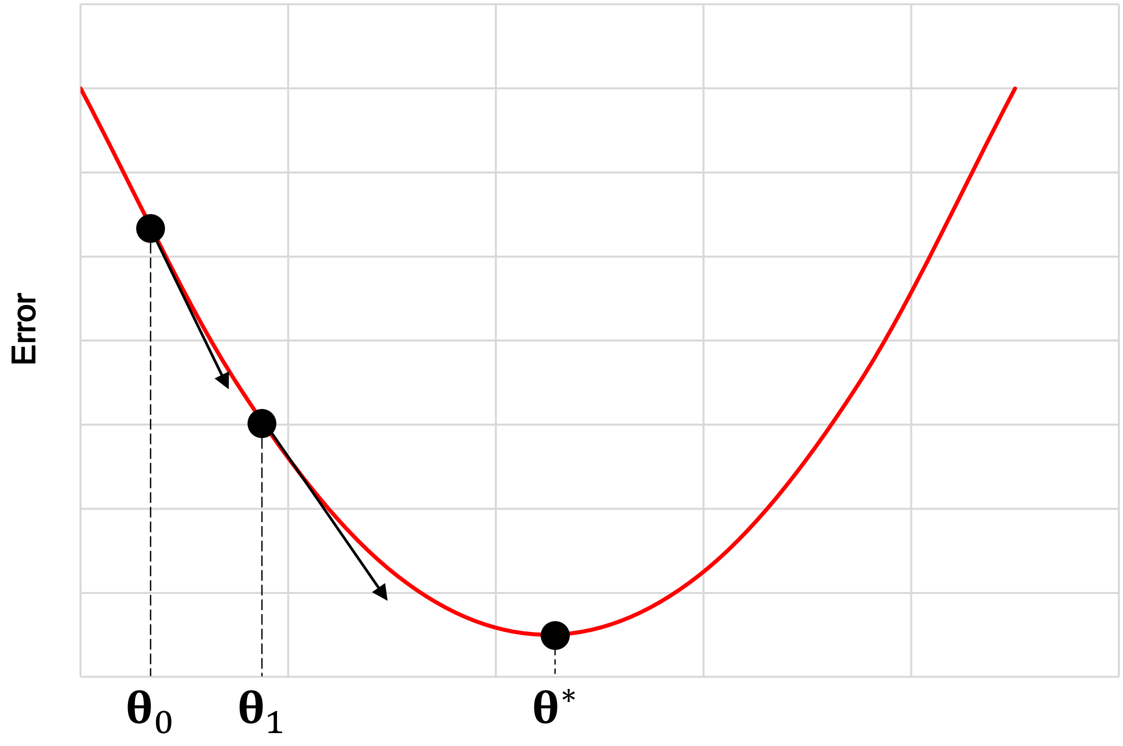
Model





Machine Learning Components

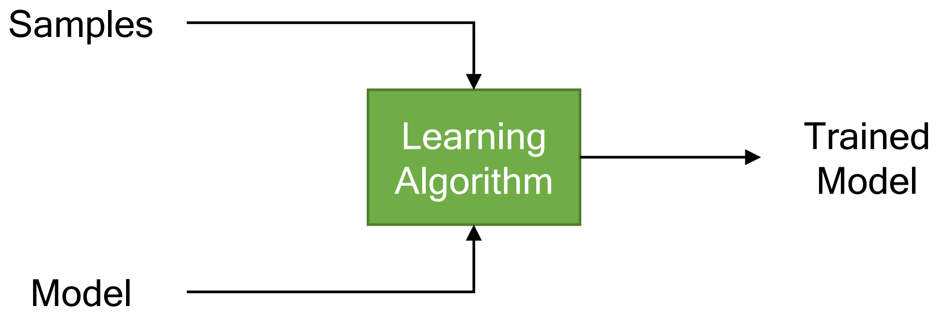
Learning Algorithm



- A procedure to search for the optimum parameters $\boldsymbol{\theta}^*$

Machine Learning Components

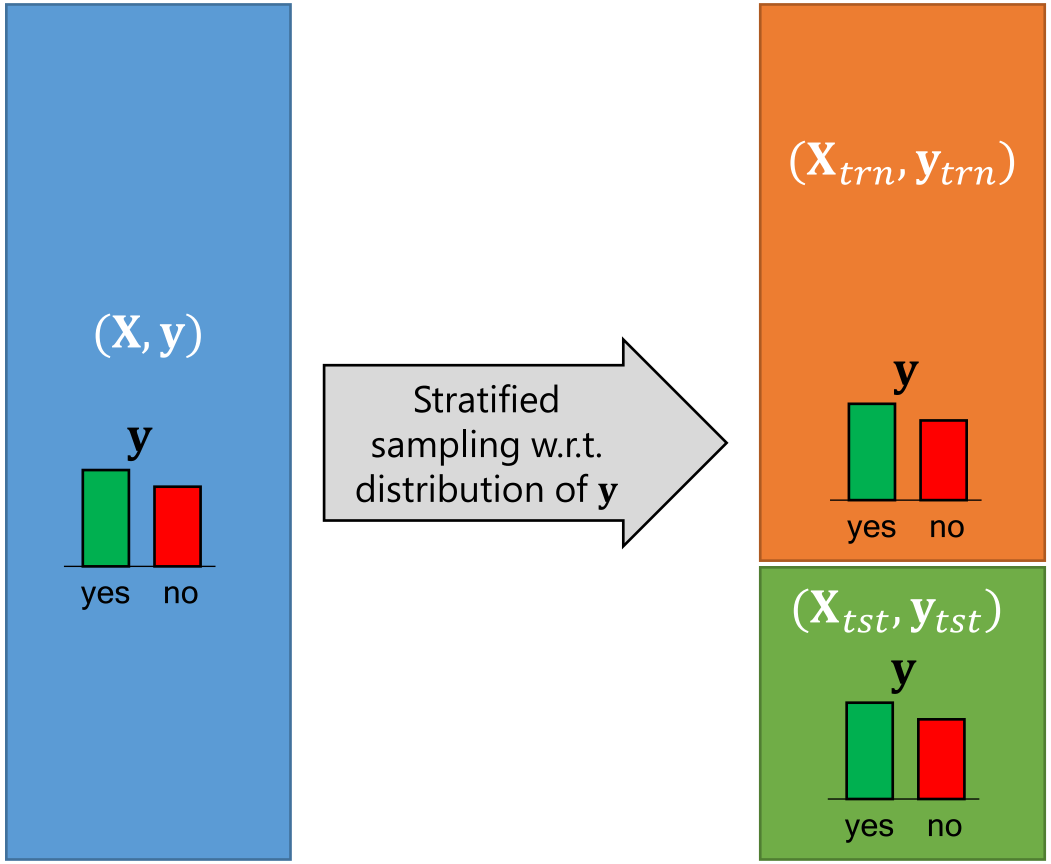
Model and Learning Algorithm



- Summary
  - Input
    - Data
    - Model
  - Process
    - Learning Algorithm
  - Output
    - Model parameters (Trained Model)

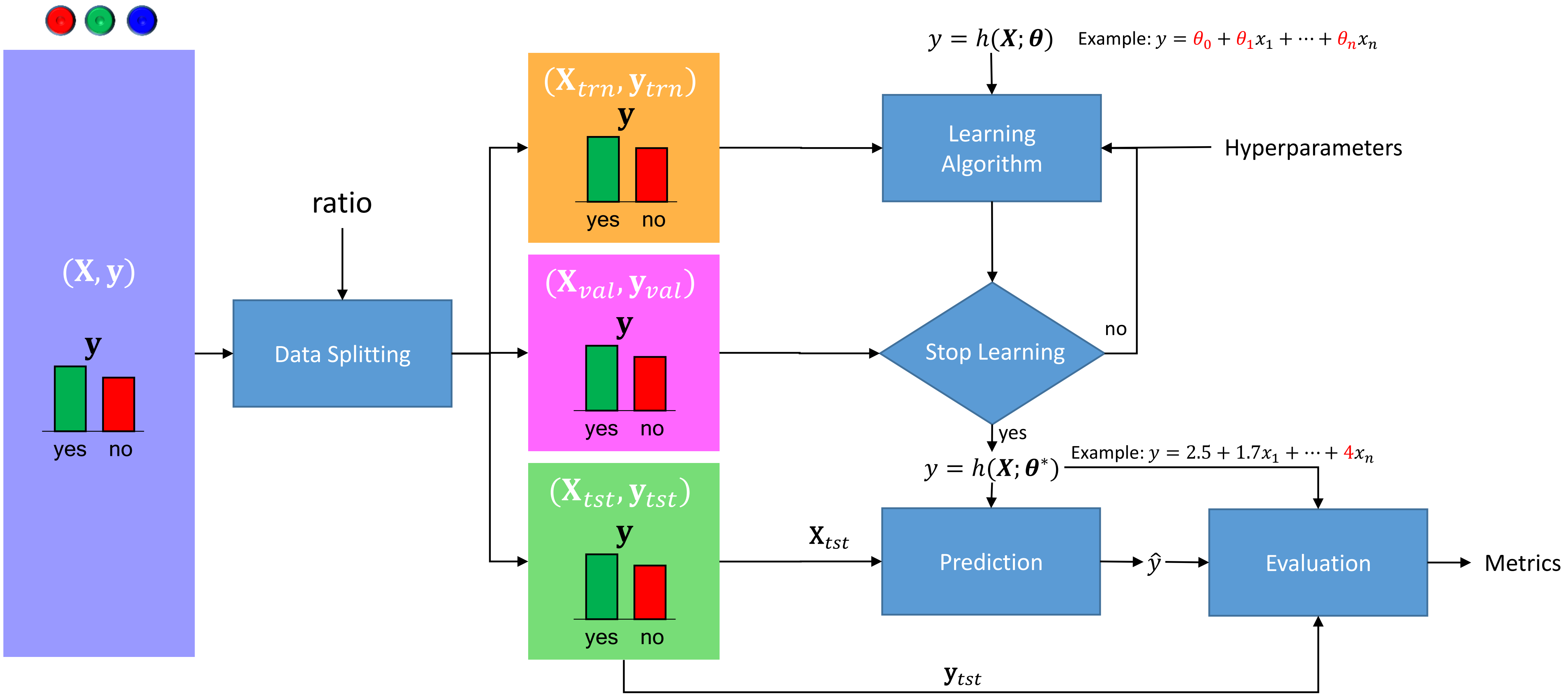
ML Methodology

Data Splitting

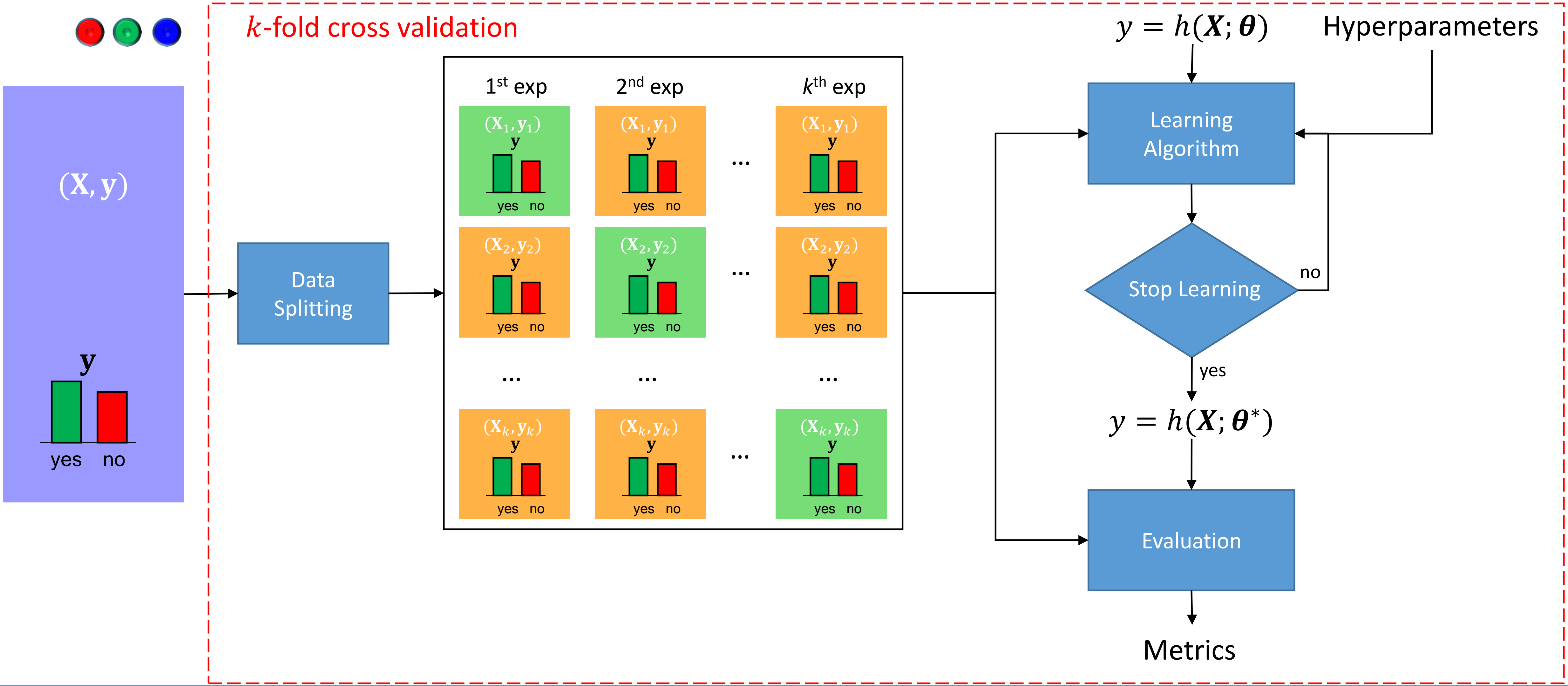


- Classical Machine Learning
  - Number of samples is not very large
  - Training samples: Test samples
    - 80:20
    - 70:30

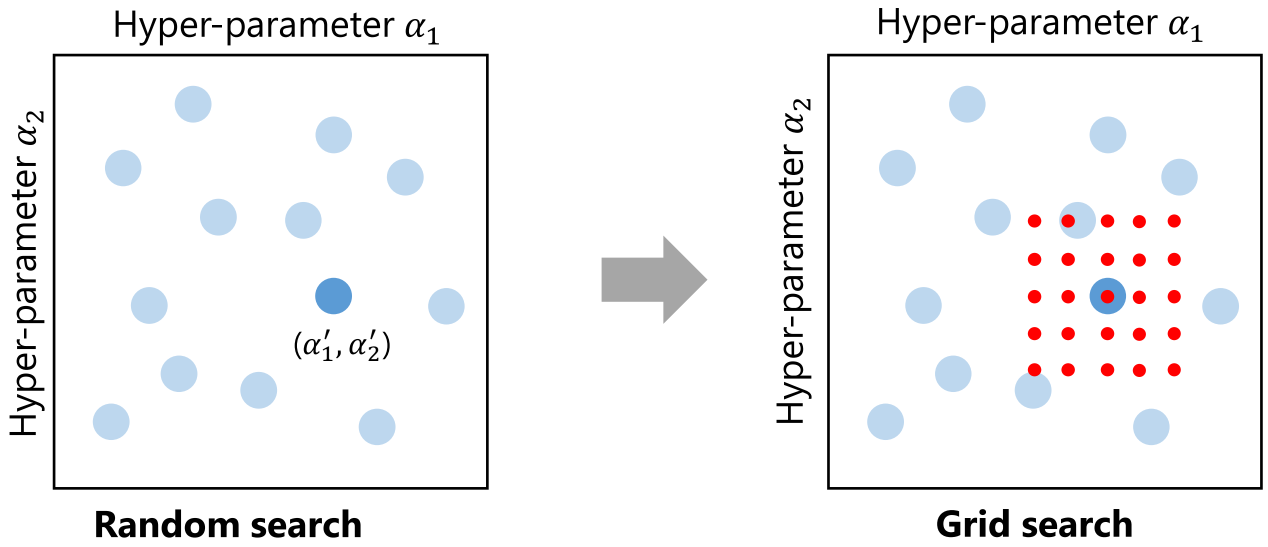
ML Methodology (Classification)



ML Methodology (Classification)



Hyperparameter Tuning



Model Evaluation

Confusion Matrix

Age	Sex	y	ŷ	Error-Type
20	Male	$\neg dm$	$\neg dm$	tn
55	Female	dm	$\neg dm$	fn
42	Male	dm	dm	tp
48	Male	$\neg dm$	dm	fp
45	Male	dm	dm	tp
58	Male	$\neg dm$	dm	fp
60	Male	dm	dm	tp
28	Female	$\neg dm$	$\neg dm$	tn
50	Male	dm	dm	tp
50	Female	$\neg dm$	$\neg dm$	tn

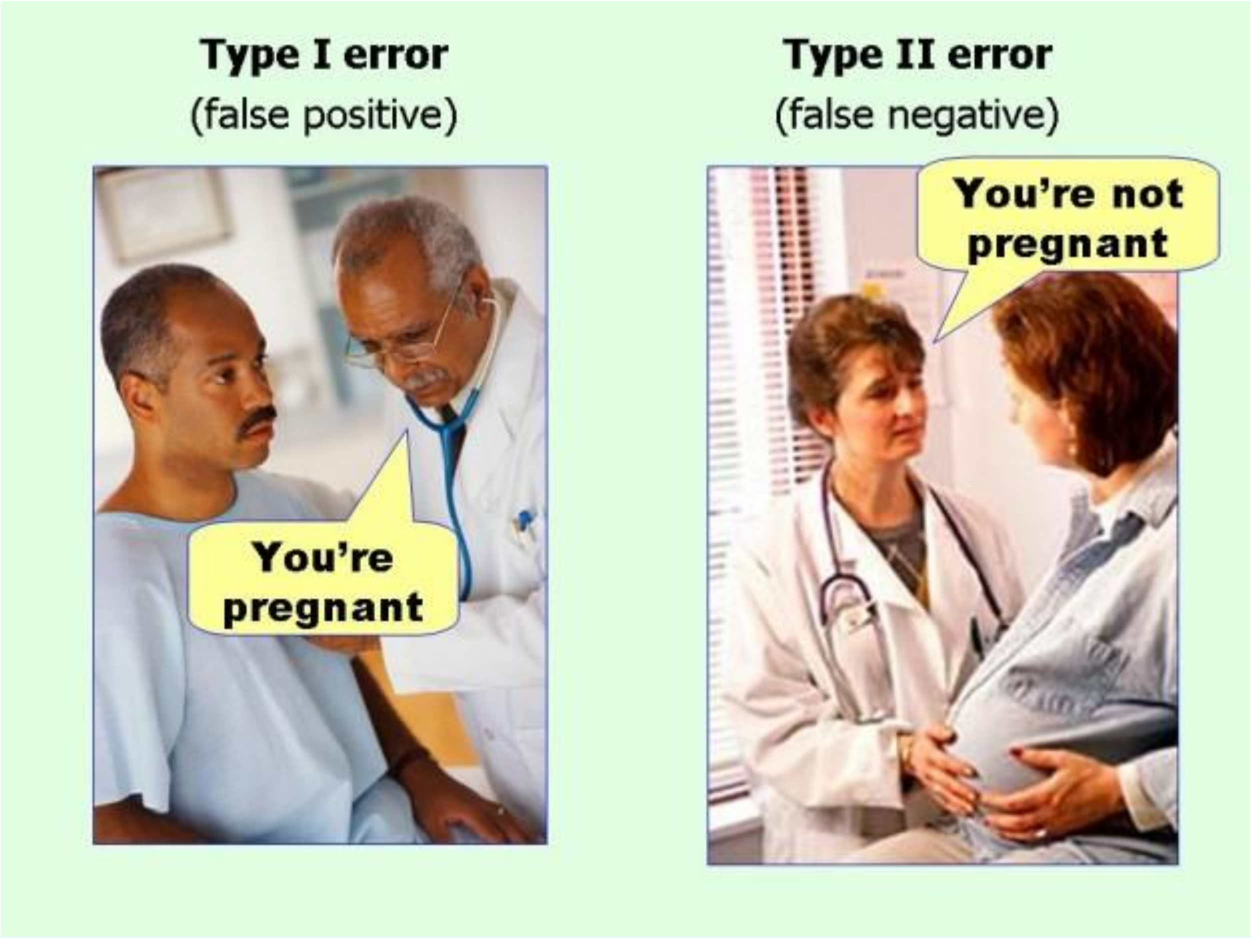
- Type of Errors
  - fp: Type-I Error
  - fn: Type-II Error

Predict \ Actual	y	$\sim y$
y	tp	fp
$\sim y$	fn	tn

Predict \ Actual	y	$\sim y$
y	4	2
$\sim y$	1	3

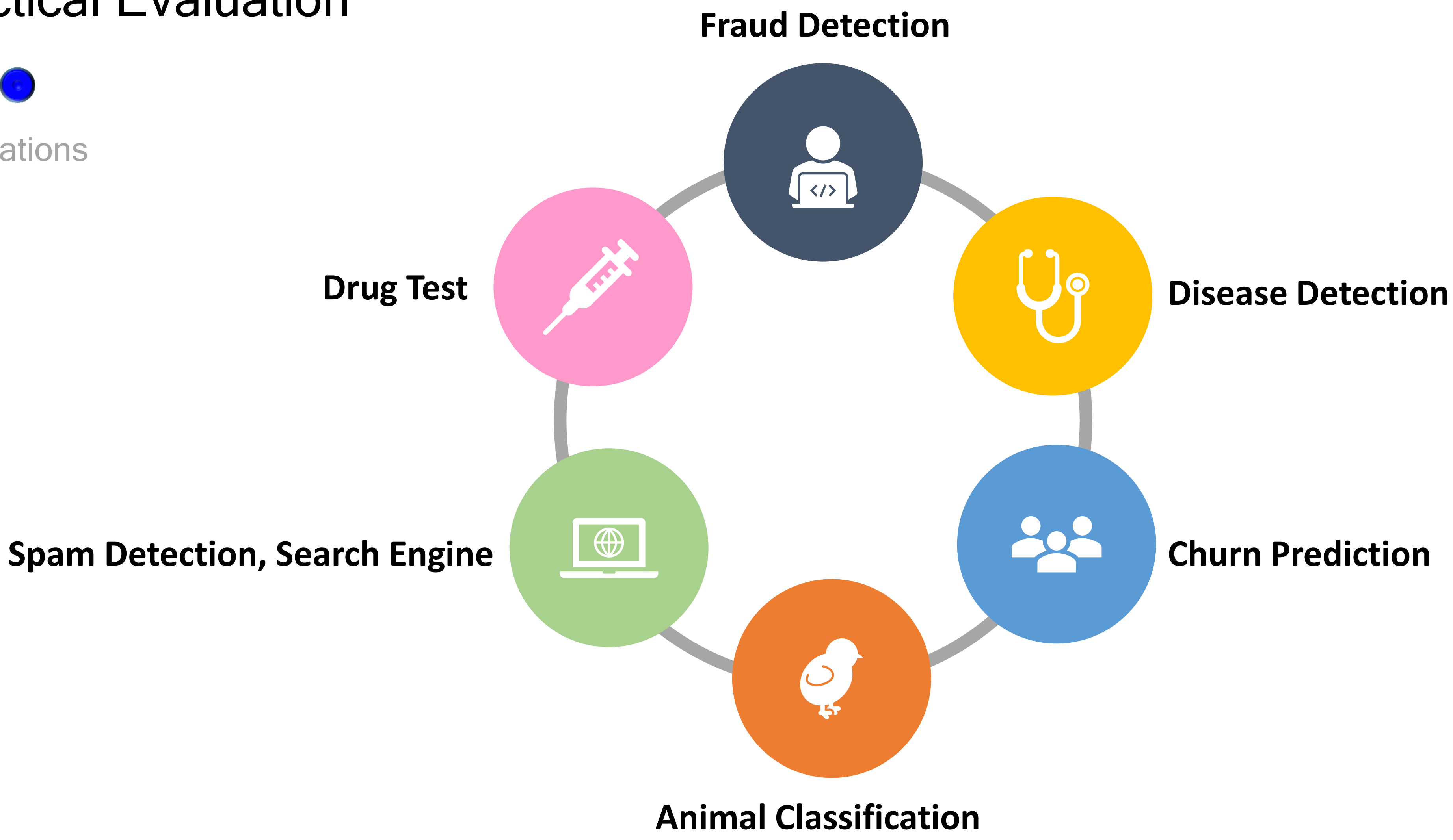
Model Evaluation

Type of Errors



Practical Evaluation

Applications



Practical Evaluation

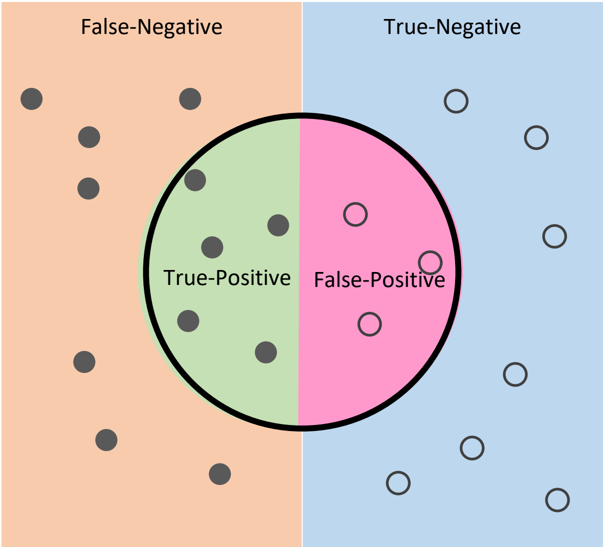
FAQ

Accuracy: 0.90

Q: Is my model good enough?

A: Depends on application, impact, and expectations

Choose the right measures



Accuracy answers the following question:

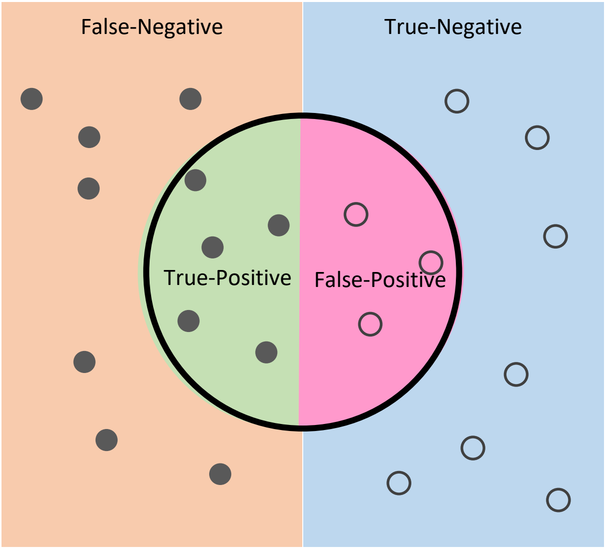
How many samples are correctly labeled out of all samples?

Accuracy is a good measure when impact of FP and FN are similar and balanced class distribution

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Ex. Distinguishing Male and Female Chick

Choose the right measures



Precision answers the following question:

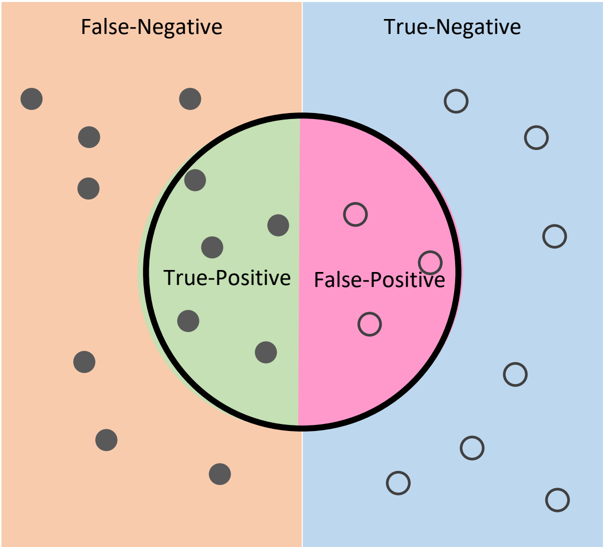
How many samples labeled as positive are actually positive samples?

Precision is a good measure when impact of FP must be minimized

$$precision = \frac{TP}{TP + FP}$$

Ex. Spam Mail Detection

Choose the right measures



Recall (a.k.a. Sensitivity) answers the following question:

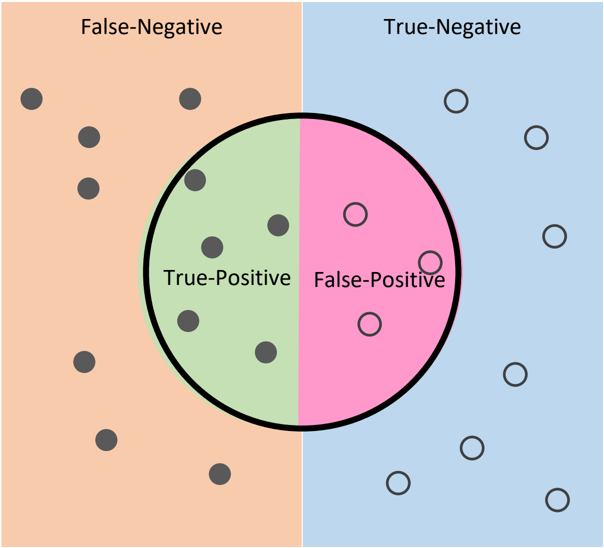
How many samples from all positive samples are correctly predicted?

Recall is a good measure when impact of FN must be minimized

$$recall = \frac{TP}{TP + FN}$$

Ex. Disease detection, Fraud Detection, Churn Prediction

Choose the right measures



Specificity answers the following question:

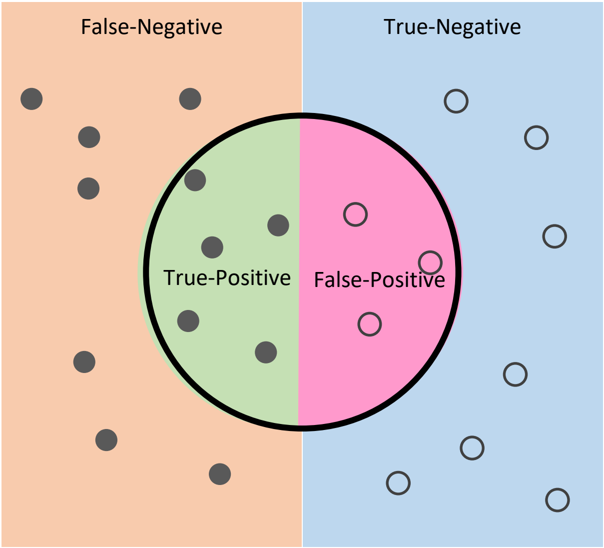
How many samples from all negative samples are correctly predicted?

Specificity is a good measure when we want to cover all TN and do not want FP

$$specificity = \frac{TN}{TN + FP}$$

Ex. Drug Test, Alcohol Test

Choose the right measures



F1-Score (a.k.a. F-Score) balances between precision and recall:

F1-Score is a good measure when impact of FP and FN are different and imbalanced class distribution

$$F1 = \frac{2PR}{P + R}$$

Ex. Search Engine

Model selection

Classifiers	Accuracy	Runtime (ms)
A_1	90%	80
A_2	92%	90
A_3	95%	1,500

Which classifier is the best?

- Given the following models

  Linear Weight Combination (Bad Idea)

  Optimizing subjected to satisfying (Better Idea)

  Ex. Accuracy represents optimizing while runtime represents satisfying

  Selection: Maximize accuracy subjected to runtime < 100 ms

Model selection

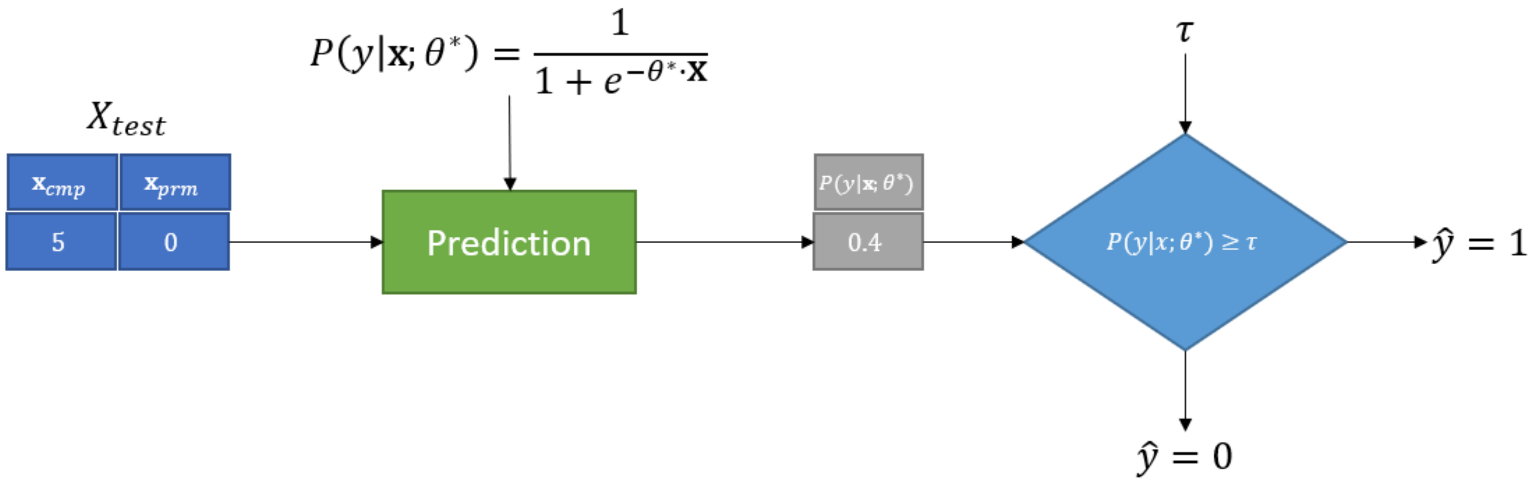
Classifiers	Metric 1	...	Metric k
A_1			
...			
A_N			

- Summary

Maximize one objective subjected to at least one constraint

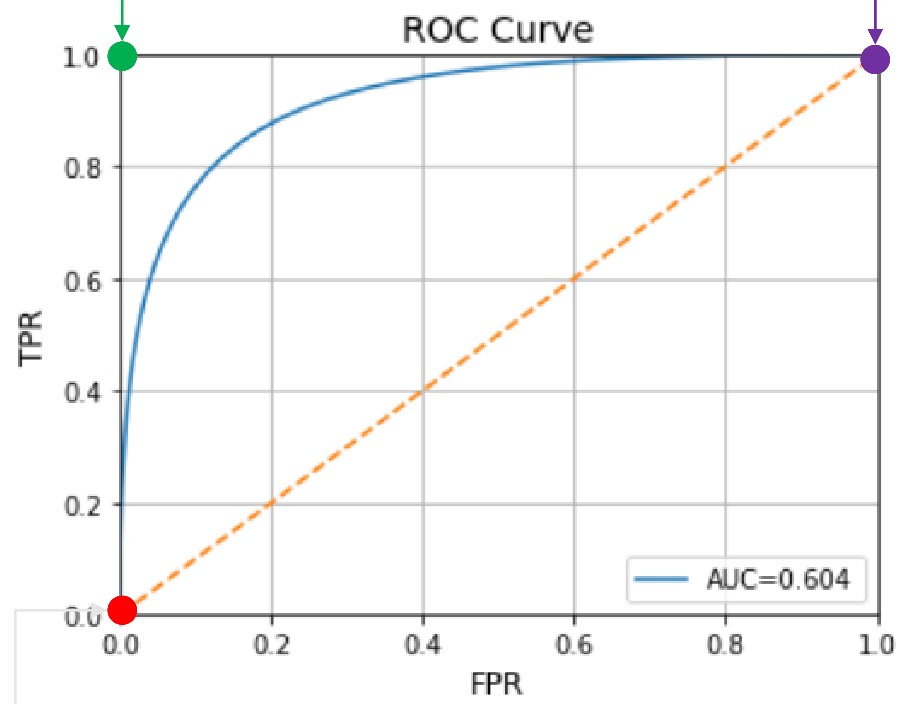
Model Evaluation

Receiver Operating Characteristics (ROC)



The closer the better model

$\tau = 0$, any sample was classified as the positive sample



$\tau = 1$, any sample was classified as the negative sample

Multiclass Classification Evaluation

Actual (y) \ Predict ($\hat{y}$)	no	wait	yes
no	16	0	5
wait	0	14	0
yes	1	1	6

Class: no

$TP_{no} = 16$

$TN_{no} = 14 + 0 + 1 + 6 = 21$

$FP_{no} = 0 + 1 = 1$

$FN_{no} = 0 + 5 = 5$

Actual (y) \ Predict ($\hat{y}$)	no	wait	yes
no	16	0	5
wait	0	14	0
yes	1	1	6

Class: wait

$TP_{wait} = 14$

$TN_{wait} = 16 + 6 + 1 + 6 = 29$

$FP_{wait} = 0 + 1 = 1$

$FN_{wait} = 0 + 0 = 0$

Actual (y) \ Predict ($\hat{y}$)	no	wait	yes
no	16	0	5
wait	0	14	0
yes	1	1	6

Class: yes

$TP_{yes} = 6$

$TN_{yes} = 16 + 0 + 0 + 14 = 30$

$FP_{yes} = 5 + 0 = 5$

$FN_{yes} = 1 + 1 = 2$

Micro Average:

$$\text{precision} = \frac{TP_{no} + TP_{wait} + TP_{yes}}{TP_{no} + TP_{wait} + TP_{yes} + FP_{no} + FP_{wait} + FP_{yes}}$$

$$= \frac{16 + 14 + 6}{16 + 14 + 6 + 1 + 1 + 5} = \frac{36}{43} \approx 0.837$$

$$\text{recall} = \frac{TP_{no} + TP_{wait} + TP_{yes}}{TP_{no} + TP_{wait} + TP_{yes} + FN_{no} + FN_{wait} + FN_{yes}}$$

$$= \frac{16 + 14 + 6}{16 + 14 + 6 + 5 + 0 + 2} = \frac{36}{43} \approx 0.837$$

$$\text{F1} = \frac{2PR}{P + R} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

$$= \frac{TP_{no} + TP_{wait} + TP_{yes}}{TP_{no} + TP_{wait} + TP_{yes} + \frac{1}{2}(FP_{no} + FP_{wait} + FP_{yes} + FN_{no} + FN_{wait} + FN_{yes})}$$

$$= \frac{16 + 14 + 6}{16 + 14 + 6 + \frac{1}{2}(1 + 1 + 5 + 5 + 0 + 2)} = \frac{36}{43} \approx 0.837$$

There is no micro average for accuracy

$$\text{accuracy} = \frac{\#correctly\ classified\ samples}{\#samples}$$

$$= \frac{16 + 14 + 6}{16 + 14 + 6 + 0 + 5 + 0 + 0 + 1 + 1} = \frac{36}{43} \approx 0.837$$

Multiclass Classification Evaluation

Predict ($\hat{y}$) / **Actual** (y)

	no	wait	yes
no	16	0	5
wait	0	14	0
yes	1	1	6

Class: no

$TP_{no} = 16$

$TN_{no} = 14 + 0 + 1 + 6 = 21$

$FP_{no} = 0 + 1 = 1$

$FN_{no} = 0 + 5 = 5$

Predict ($\hat{y}$) / **Actual** (y)

	no	wait	yes
no	16	0	5
wait	0	14	0
yes	1	1	6

Class: wait

$TP_{wait} = 14$

$TN_{wait} = 16 + 6 + 1 + 6 = 29$

$FP_{wait} = 0 + 1 = 1$

$FN_{wait} = 0 + 0 = 0$

Predict ($\hat{y}$) / **Actual** (y)

	no	wait	yes
no	16	0	5
wait	0	14	0
yes	1	1	6

Class: yes

$TP_{yes} = 6$

$TN_{yes} = 16 + 0 + 0 + 14 = 30$

$FP_{yes} = 5 + 0 = 5$

$FN_{yes} = 1 + 1 = 2$

Macro Average:

$$\text{precision} = \frac{\text{precision}_{no} + \text{precision}_{wait} + \text{precision}_{yes}}{3}$$

$$= \frac{\frac{16}{16+1} + \frac{14}{14+1} + \frac{6}{6+5}}{3} \approx 0.807$$

$$\text{recall} = \frac{\text{recall}_{no} + \text{recall}_{wait} + \text{recall}_{yes}}{3}$$

$$= \frac{\frac{16}{16+5} + \frac{14}{14+0} + \frac{6}{6+2}}{3} \approx 0.837$$

$$F1 = \frac{F1_{no} + F1_{wait} + F1_{yes}}{3}$$

$$= \frac{\frac{16}{16+\frac{1}{2}(1+5)} + \frac{14}{14+\frac{1}{2}(1+0)} + \frac{6}{6+\frac{1}{2}(5+2)}}{3} \approx 0.813$$

There is no micro average for accuracy

$$\text{accuracy} = \frac{\#correctly\ classified\ samples}{\#samples}$$

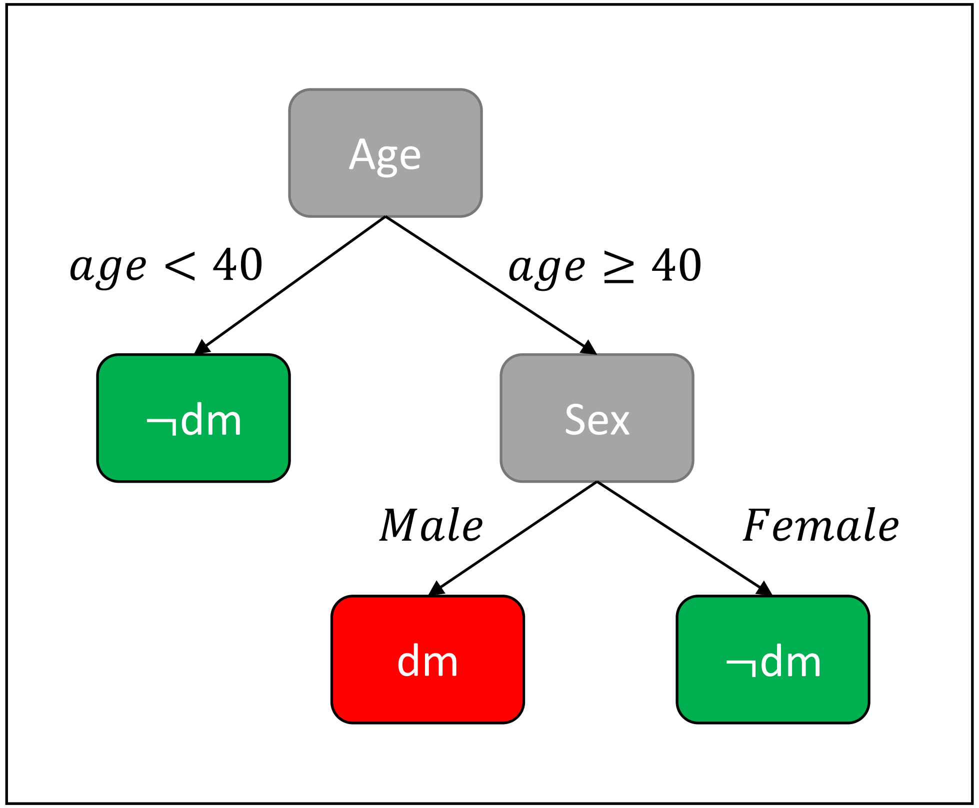
$$= \frac{16+14+6}{16+14+6+0+5+0+0+1+1} = \frac{36}{43} \approx 0.837$$

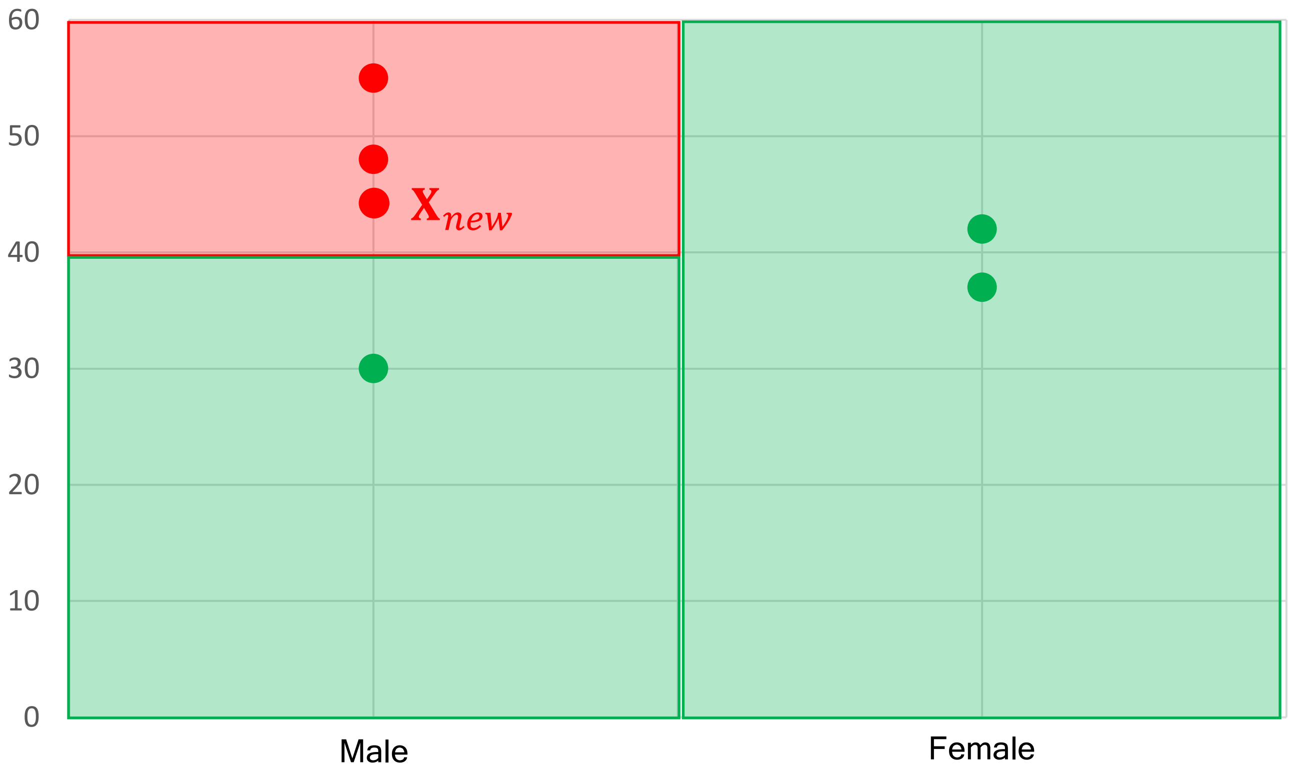
Using Trained Model

Applying a trained model to an unseen sample

Unseen sample $\mathbf{X}_{new}$

Age	Sex	Diabetes
45	Male	?





ML Tools

Most frequently used tools

	Programming Language		IDE		GUI
Anaconda Good for production	Python	R	Spyder	Jupyter	Orange
			Good for library development		
Colab Good for training	Python	R		Jupyter	
			Good for writing tutorial		