

# CoderSchool: Machine Learning

Week 1: Welcome to Class!

WiFi: CoderSchool

Chatroom: <https://gitter.im/coderschool/ml-may18>

**Kunal Jathal**  
Musician

**Zein Tawil**  
Scientist



# Kunal

Introducing your Instructor

**Kunal** has been involved in music technology and machine learning for a few years now, and has worked previously @ Microsoft, MusixMatch, and currently does audio engineering in Virtual Reality, at Verizon Labs in New York City.

He has been in Vietnam for 2 weeks now and LOVES it!

Especially the food ...

A portrait of a young man with curly brown hair and a beard, wearing a brown button-down shirt. He is smiling and looking towards the camera.

# Zein

Introducing your TA

**Zein** is a data scientist at Primer AI, a San Francisco based NLP-AI company. Before Primer, he also TA'd at an SF based data science school (Galvanize), and was a student there before that. He recently moved to Vietnam because his girlfriend, Francisca, got a job here! Zein is excited to get to know the city, country, and region. I like exploring new places and watching soccer--can't wait for the World Cup.

Welcome to  
CoderSchool!



# Attendance

Use your phone to log yourself  
in for class. Thanks!

Scan me!



# Class Structure

- 5 Week Class
- Monday Lecture: slides and code walkthrough.
- Wednesday Lab: hands-on pair programming.
- Homework due every Sunday evening at 22:00.

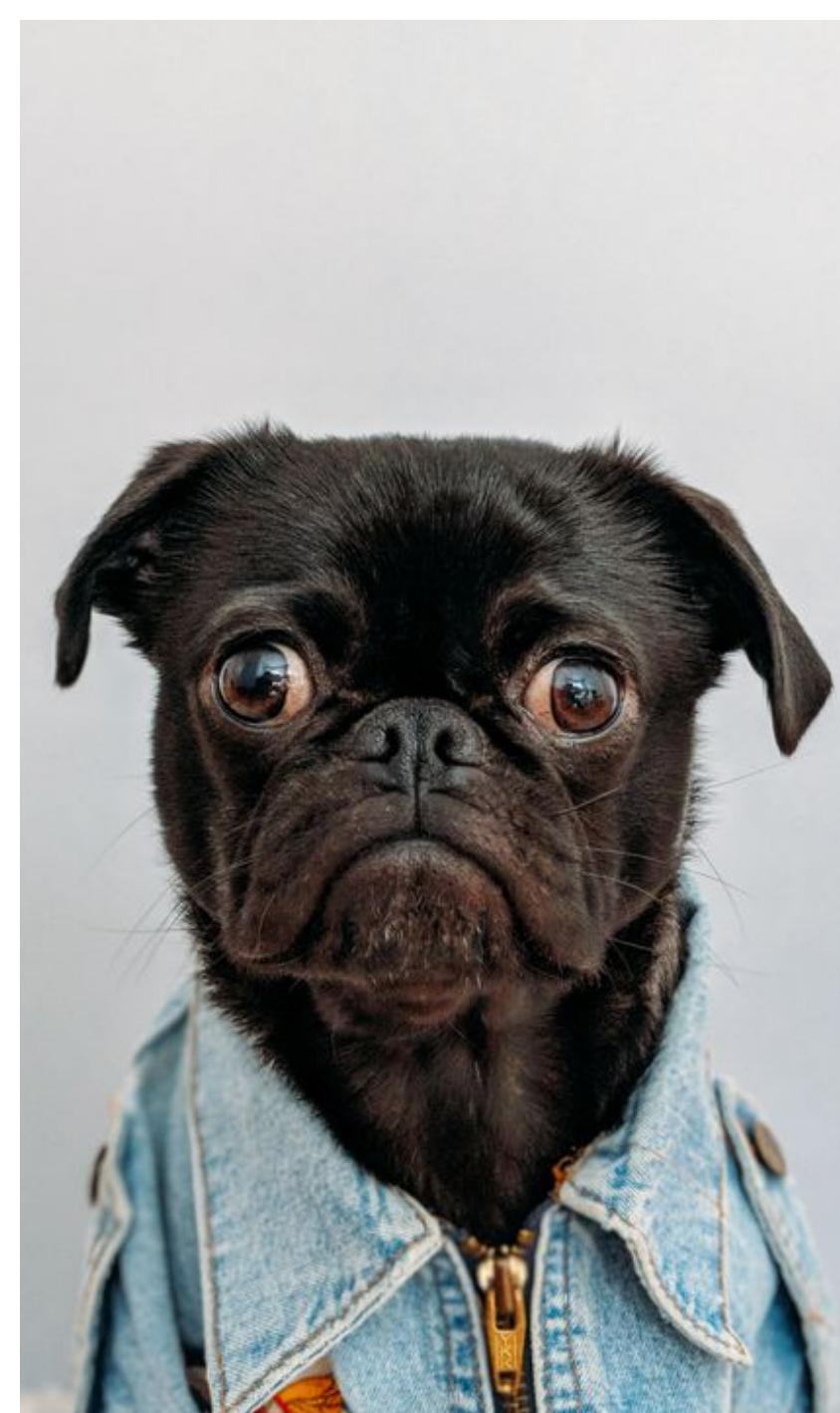
**Learn** by doing

**Learn** by asking

**Learn** by helping

# Class Rules

- Come to class. Maximum 1 Excused Absence.  
Absence MUST be excused: send a clear email as early as possible.
- Come to class on time.
- Submit homework no later than 22:00 Sunday.  
You are allowed one 48-hour homework extension. Only one.
- You must be honest.



You will be asked  
to leave the course  
if you don't follow  
these rules.

# How to get help

## Gitter

<https://gitter.im/coderschool/ml-may18>

## Email

ml@coderschool.vn

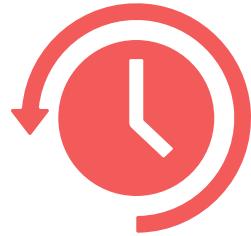
## In Person

12 Ton Dan, we're usually around.

## Learning Portal

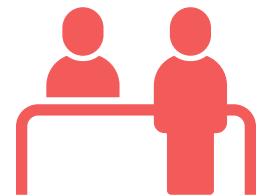
[http://learning.coderschool.vn/courses/intro\\_to\\_machine\\_learning/](http://learning.coderschool.vn/courses/intro_to_machine_learning/)

# Tips for success



## Start Homework Early

The earlier you start, the more time you have to ask questions and think clearly.



## Ask Questions!

If you're stuck for more than 30 minutes on something, please try to ask your teachers or classmates. Sometimes the simple act of asking a question will help you solve the problem yourself.



## Push yourself

This is the best time for you to learn; you have a lot of support. Try to do all of the assignments as well as you can, and you will learn a lot.

Week #1

# Discussion

# Discussion Questions



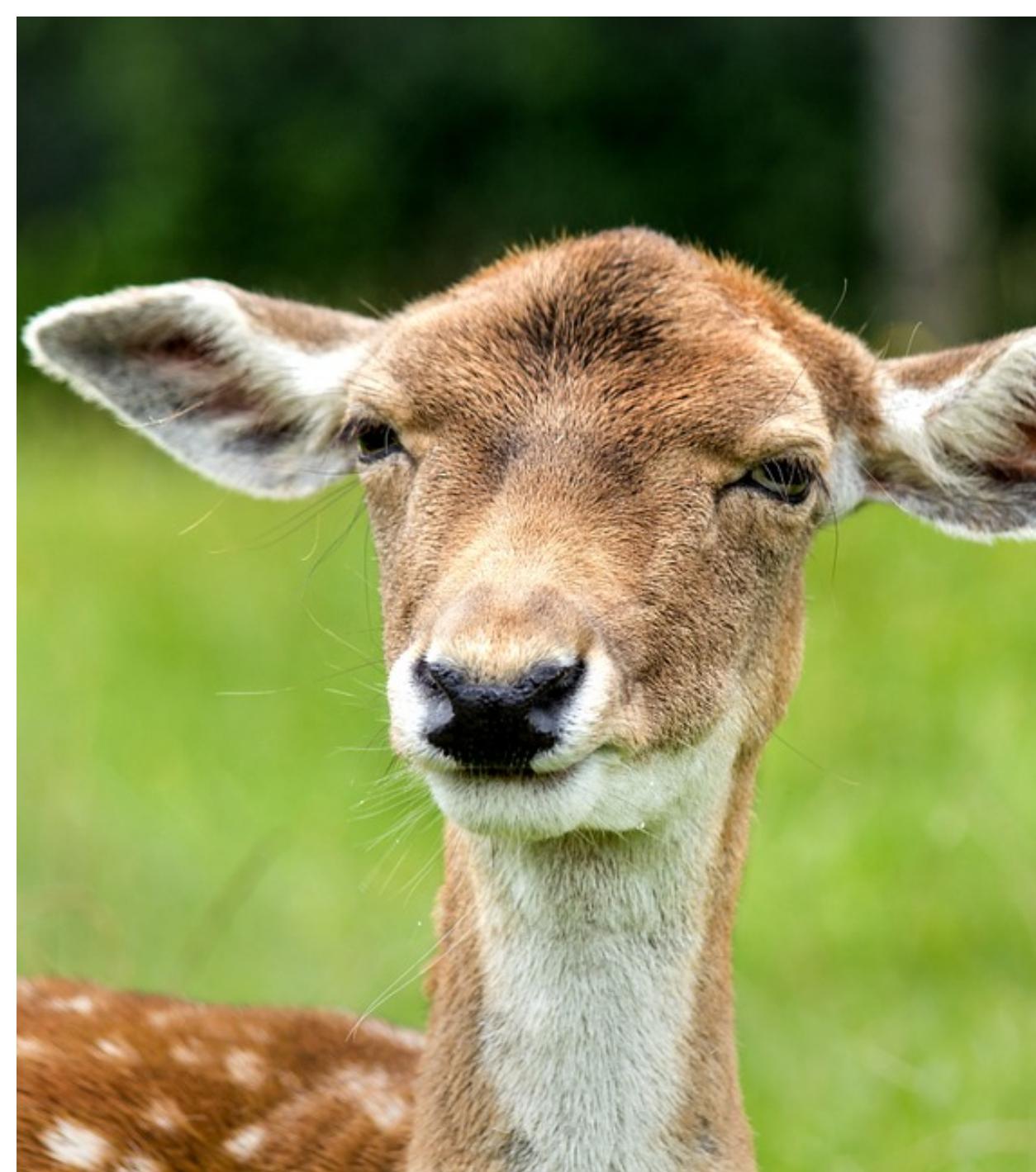
- What's your name?
- Why did you decide to take this course?
- What do you hope to achieve with machine learning?



Best  
Preworks!

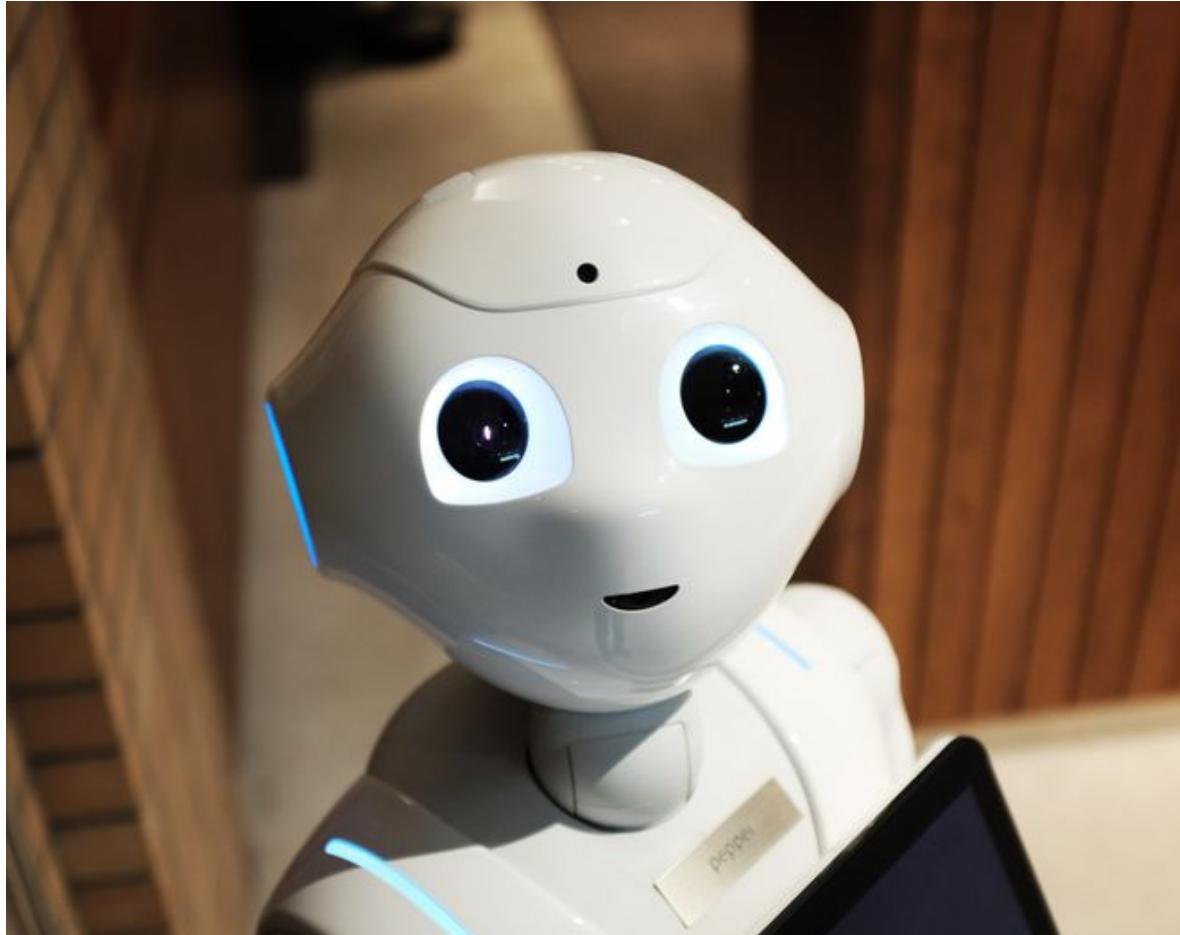
Our Goal

You will all be **able to build  
machine learning applications.**



But wait...what  
is Machine  
Learning?

# What is Machine Learning?



- Machine learning is a method of data analysis that automates analytical model building
- Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

# What is it used for?

1 | Fraud Detection

2 | Web Search Results

3 | Targeted Advertising

4 | Credit Scoring

5 | Facial and Image Recognition

6 | Financial Modeling

7 | Self Driving Cars

8 | Recommendation Engines

9 | Email Spam Filtering

10 | Text Sentiment Analysis

# Types of Machine Learning

- **Supervised Learning**

Trained with labeled data.

The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors.

Commonly used in applications where historical data predicts likely future

- **Unsupervised Learning**

Data has no historical labels.

The system is not told the “right answer.” The algorithm must figure out what is being shown.

The goal is to explore the data and find some structure within.

Segment text topics, recommend items and identify data outliers the data and find some structure within.

- **Reinforcement Learning**

Deep Learning

# Layout of this Course

Introduction to  
Machine  
Learning

Building Simple Classifiers

Music Information  
Retrieval

Exploring supervised learning, using Music Information Retrieval  
from the Web

Natural Language Processing

Sentiment Analysis on Song Lyrics

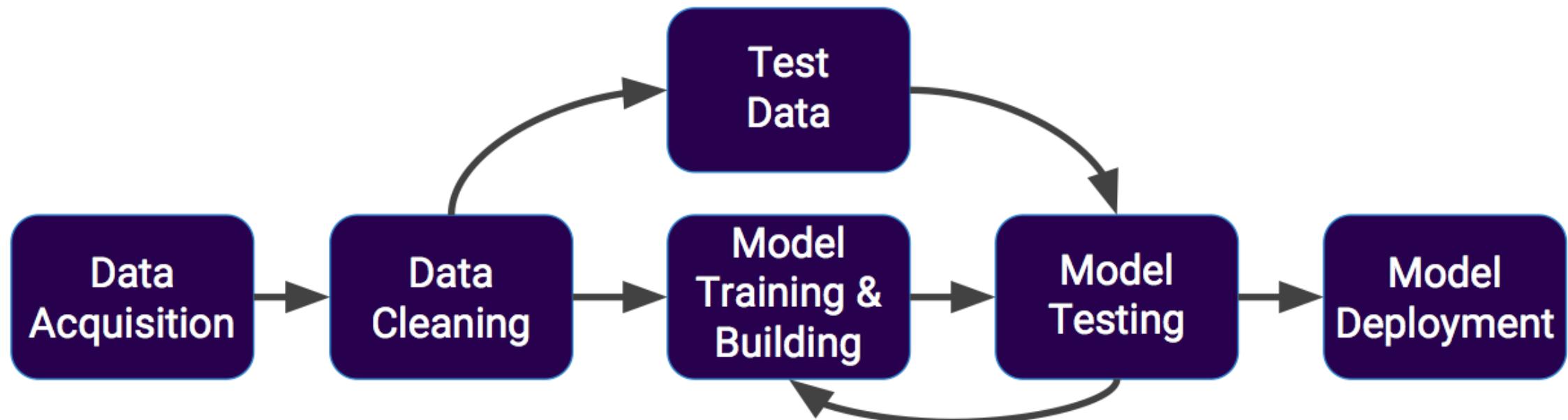
Music Recommendation System

Building a complete music recommendation system

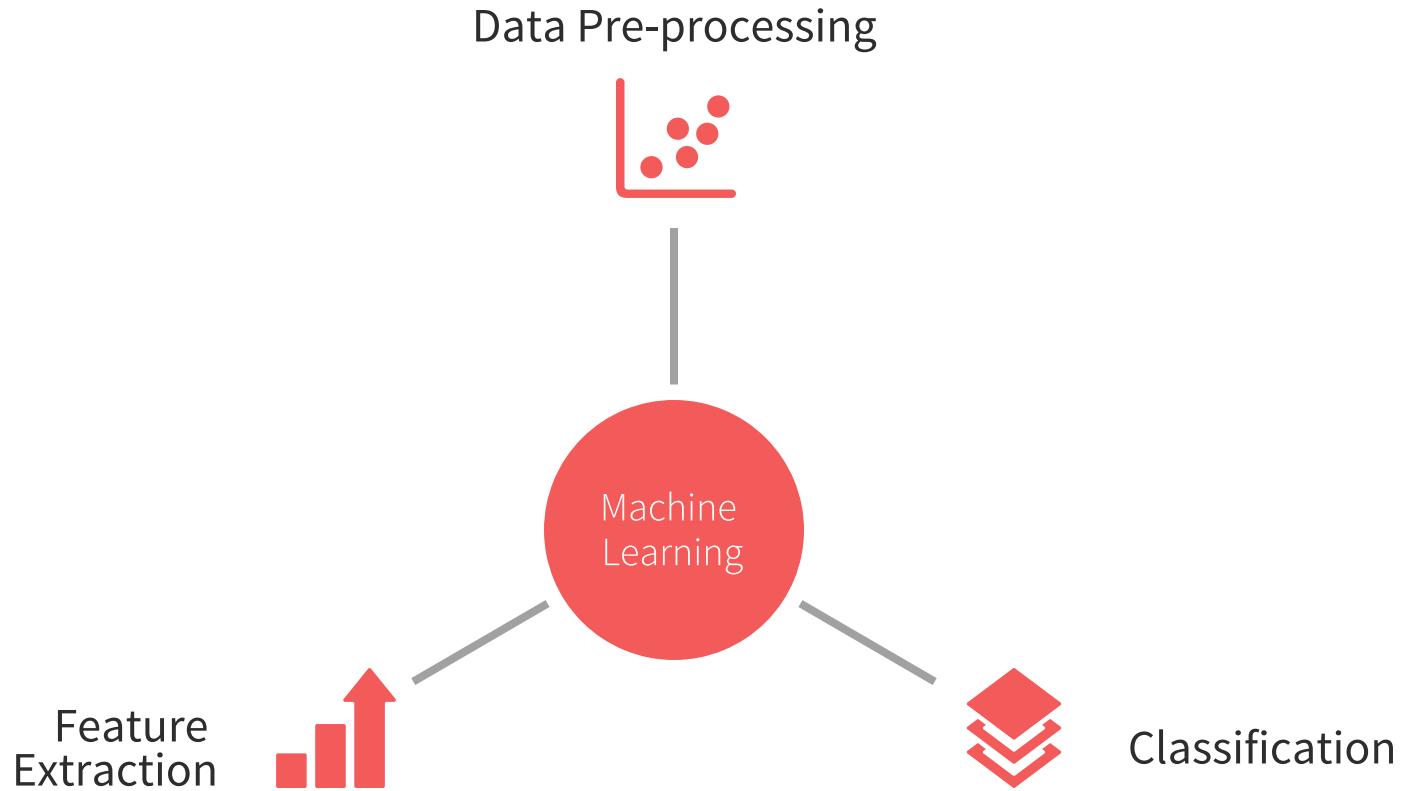
Module 1

# Introduction to Machine Learning Concepts

# Steps in Machine Learning



# Three Main Areas



# Feature Extraction

# Feature Extraction: What is a feature?

- A feature is some piece of information

It describes our data in a way that is useful and relevant.  
Egs: if our data was cats, a feature might be the number of
- Features need to be numbers, or converted to numbers

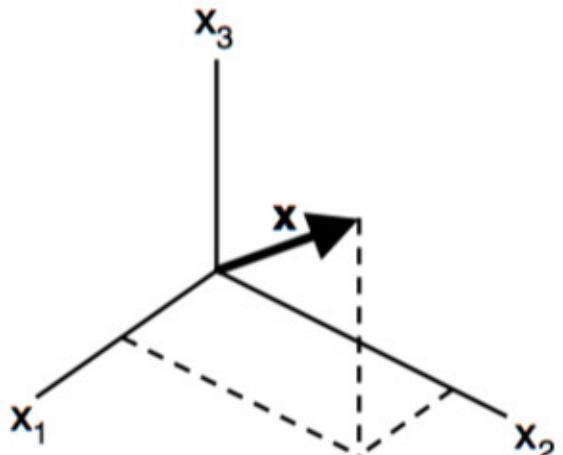
This is because computers process numbers, and eventually do math operations on features.
- Features are organized in the form of vectors

A single feature vector will contain the numerical values for many features.

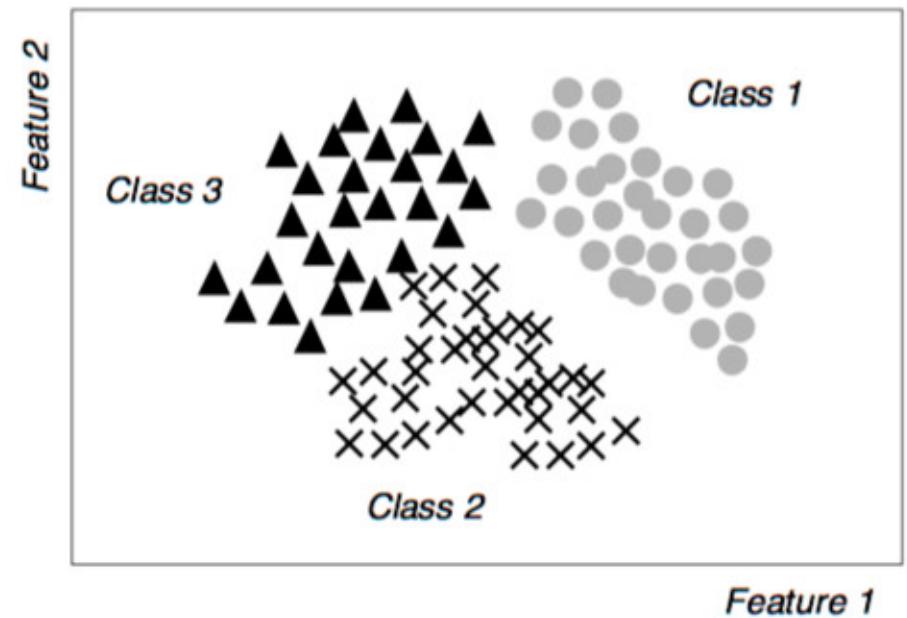
# Features in Space

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

**Feature vector**



**Feature space (3D)**

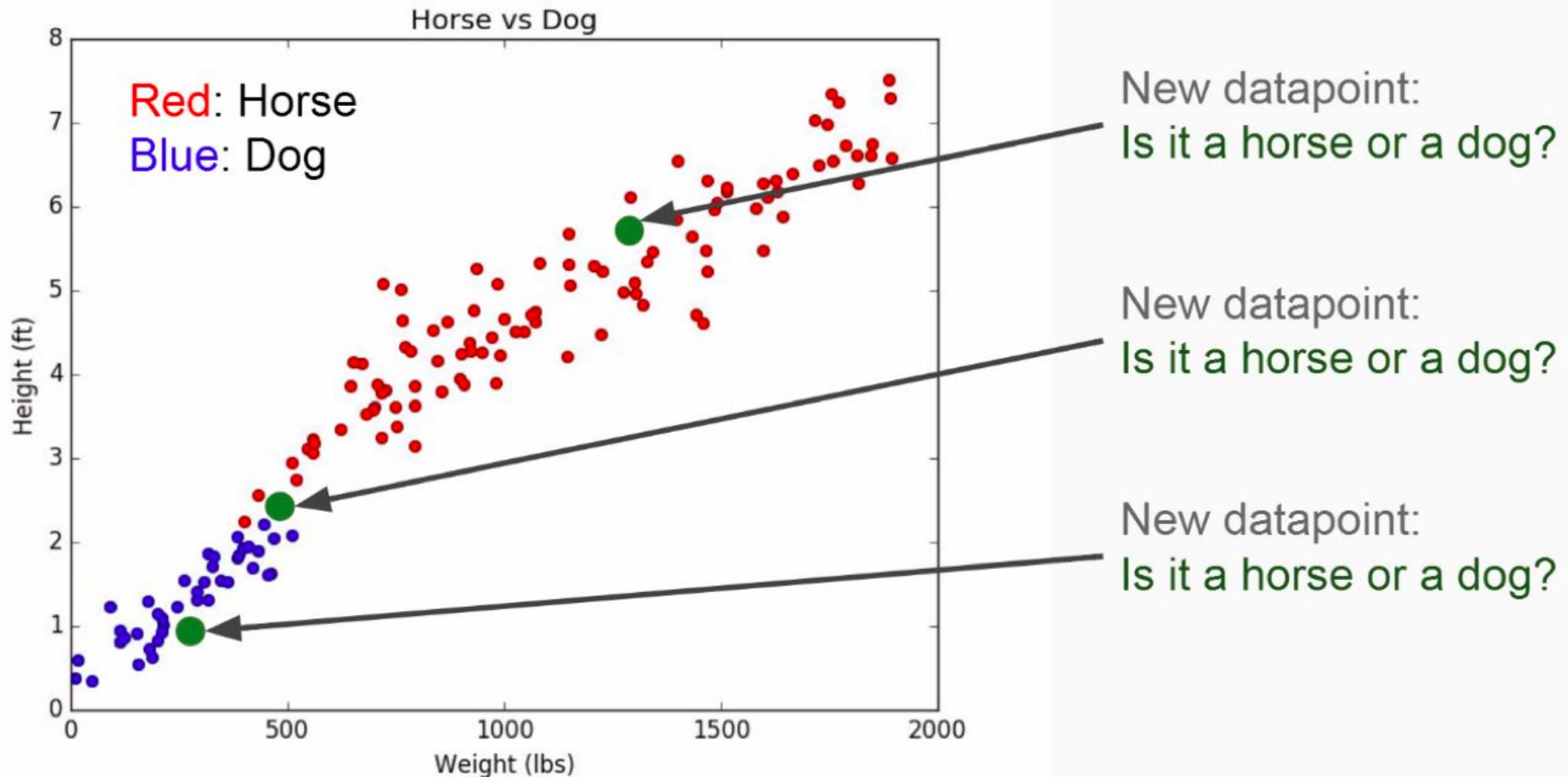


**Scatter plot (2D)**

# Classification

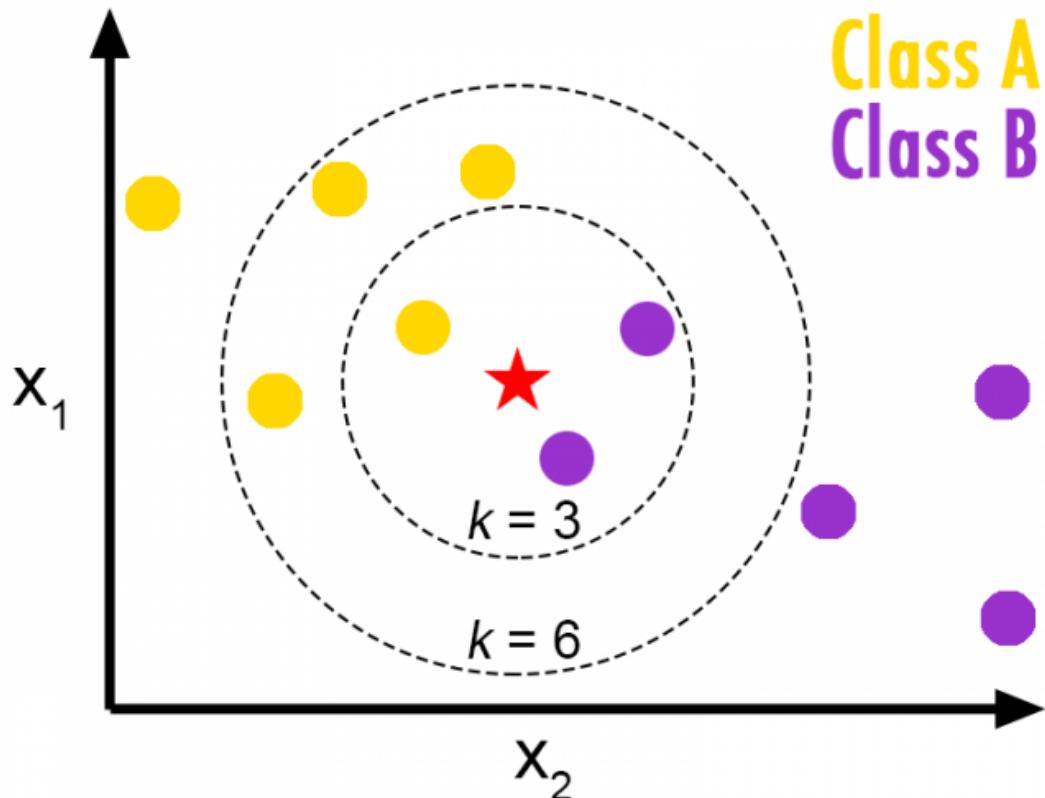
Supervised Learning: k Nearest Neighbors

# Classification: k Nearest Neighbors



# k Nearest Neighbors

Choosing a K will affect what class a new point is assigned to



# Classification

Training and Testing Our Classifier

# Training & Testing: Cross Validation

Data is divided into groups called 'folds'. In this example, there are 5 folds.

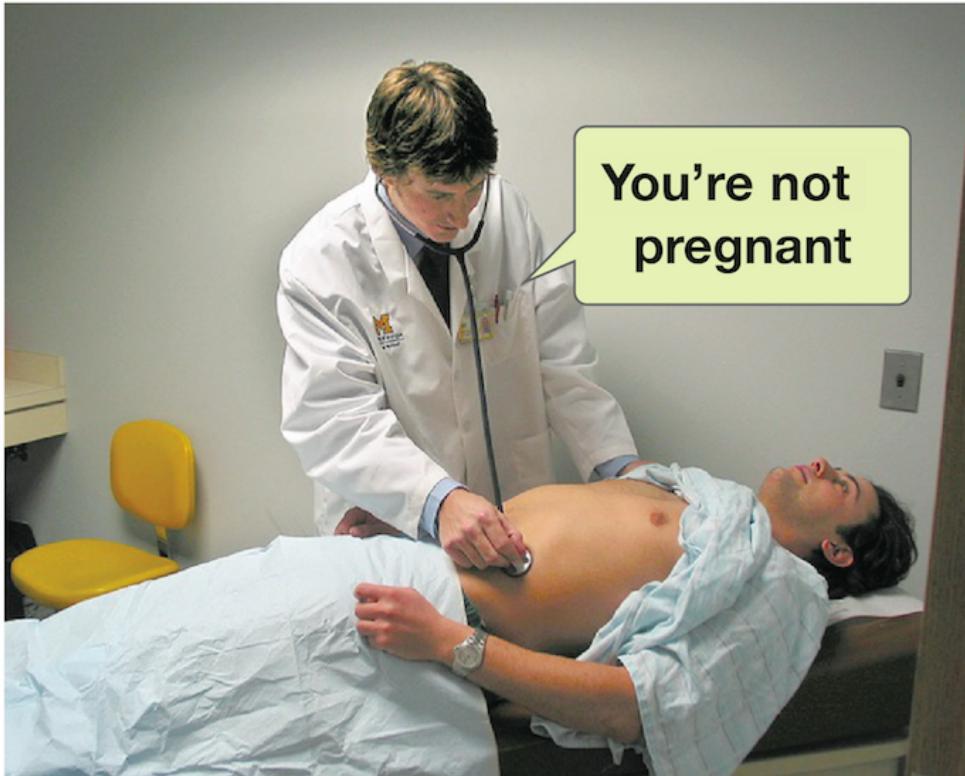


# Confusion Matrix

		Prediction outcome		
		positive	negative	
Actual value	positive	$TP$	$FN$	$TP + FN$
	negative	$FP$	$TN$	$FP + TN$
		$TP + FP$	$FN + TN$	

# Confusion Matrix

true negative

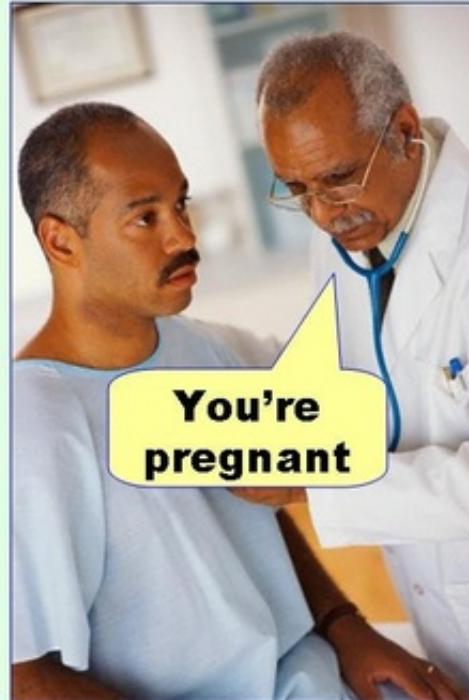


true positive

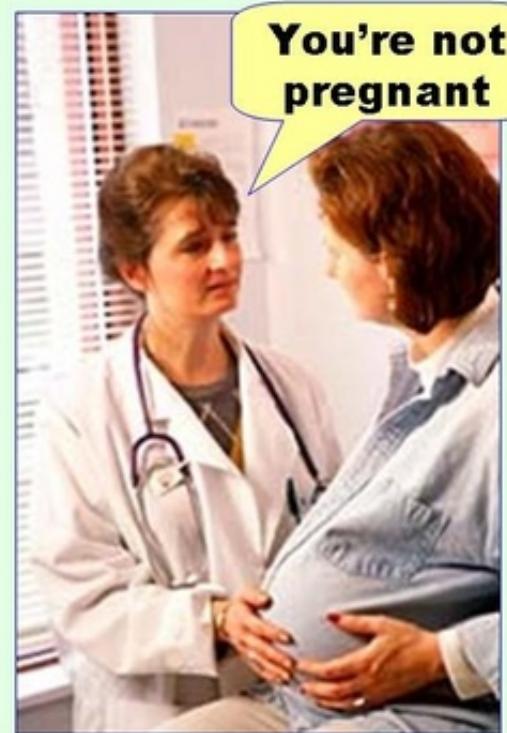


# Confusion Matrix

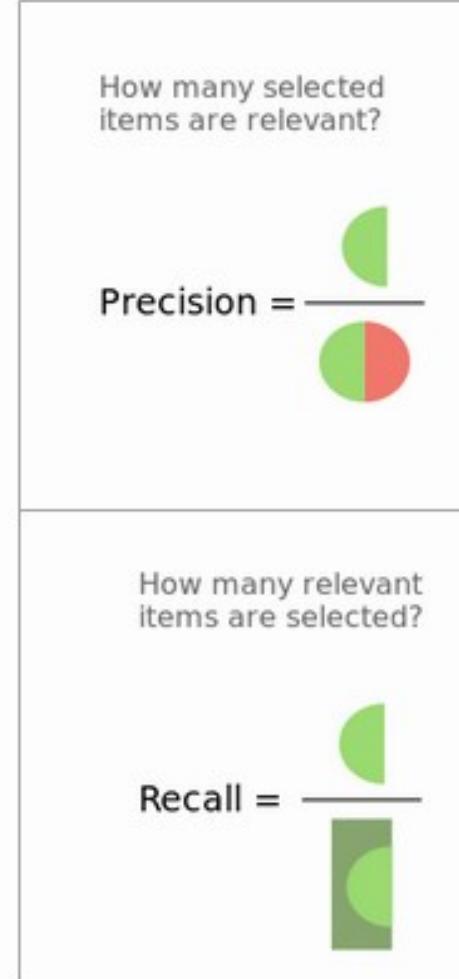
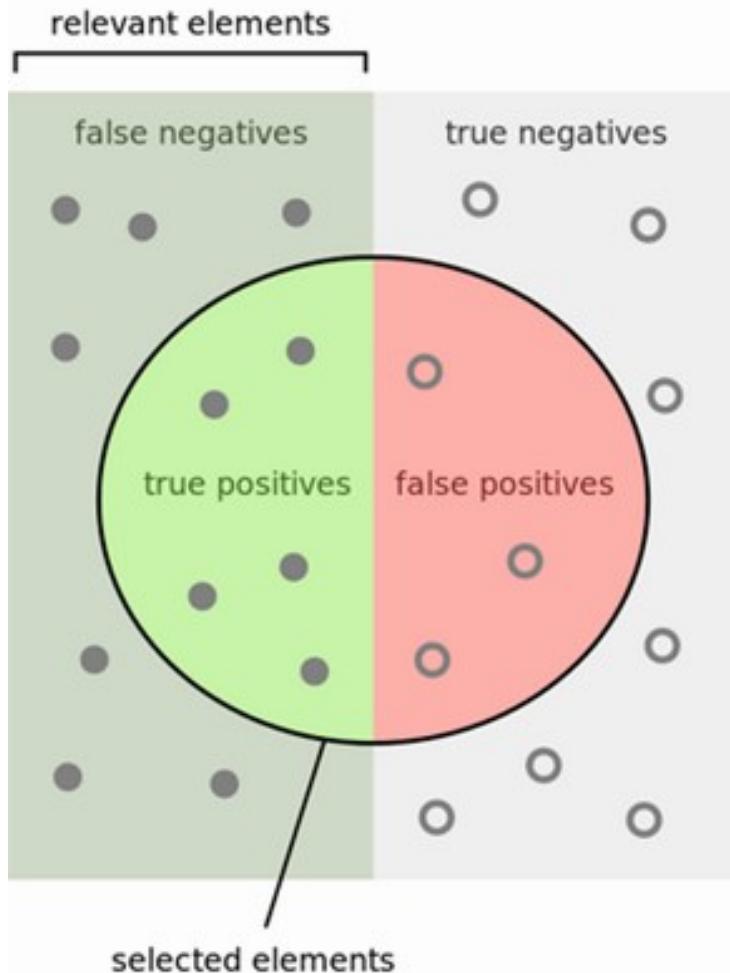
**Type I error**  
(false positive)



**Type II error**  
(false negative)



# Confusion Matrix



# Confusion Matrix

$$Precision = \frac{\#True\ Positives}{\#True\ Positives + \#False\ Positives}$$

$$Recall = \frac{\#True\ Positives}{\#True\ Positives + \#False\ Negatives}$$

# Homework Assignment

Walkthrough

# One Last Thing!

## 1. Join our class Gitter.

<https://gitter.im/coderschool/ml-may18>

## 2. Click on the link provided in the chat

<http://apply.coderschool.vn/cohorts/machinelearning-may18/versions/general/>

## 3. That's it. Thanks!

This helps us track you, grade your homework, etc.