

A person in a puffer jacket and shorts stands on dark, wet rocks at a rocky coastline during sunset. The sun is low on the horizon, creating a bright orange glow and a long, shimmering reflection on the calm water. The sky is filled with dramatic, dark clouds. A diagonal line divides the image, with the left side being darker and the right side showing more of the sunset's light.

# Machine Learning Crash Course

WEEK 9

“

“I believe this artificial intelligence is going to be our partner. If we misuse it, it will be a risk. If we use it right, it can be our partner.” ~ Masayoshi Son

“

“I believe true AI, will be a cybernetic representation of the human brain; which will learn through the very same stages of development as humans (E.g. Infancy to Adolescence). If we raise such a entity in a way as to instill good morals, we will be able to work with the AI. Otherwise there is little to no guarantee that such an entity would collaborate with us. This vision of true AI, will be a reflection of that of our objectives, not that of the creator/s, but as a society. True AI, symbolically represents the end of an old way of thinking and the beginning of anything being possible if done right.” ~ Kyle D. Zeller

# Topics

- Anomaly Detection
  - Density Estimation
  - System Design
- Recommender Systems
  - Problem Formation
  - Content Recommendations
  - Potential Issues
  - Alternate Parameter Learning Choices
- Large Scale Machine Learning
  - ML with Large Data Sets
  - Stochastic Gradient Descent
  - Mini-Batch Gradient Descent
  - Online Learning
  - Map Reduce & Data Parallelism







# 1. Anomaly Detection

## Introduction

- What is it?
- What do we use it for?
- What ways can we represent these types of problems?



$$\hat{f} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

## Density Estimation or Kernel Density Estimation (KDE)

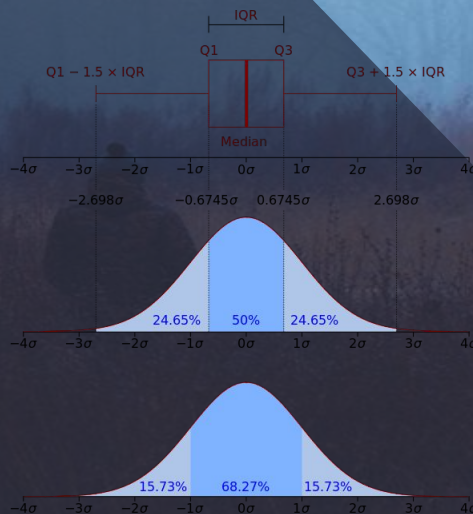
DEMO

### What is it?

“A non-parametric way to estimate the **probability density function** of a random variable.” ~Wiki

### Probability Density Function:

A modeling of the likelihood/probability that the value of some arbitrary random variable equals a particular sample.



### Smoothing Function K

This is used to capture important features in the data, while leaving out noisy artifacts.

### Our Choice of K:

We will consider a gaussian smoothing function for our KDE.

## System Design

## DEMO

$$\hat{f} = \sum_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_j}} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}}$$

- Choose the smoothing parameter  $h > 0$  (also called the **bandwidth**)
- $n$  is the size of your dataset
- Choose features  $x_i$  to be indicative of anomalous examples
- Fit your parameters  $u_1 \dots u_n$  and  $\sigma_1 \dots \sigma_n$
- Recompute the KDE
- Check for anomalies iff  $p(x) < \varepsilon$



A low-angle, upward-looking perspective of several modern skyscrapers with glass facades, creating a sense of height and architectural scale. The buildings are dark and their windows reflect light, set against a pale, overcast sky.

## 2. Recommender Systems



# Introduction

- What is it?
- What do we use it for?
- What ways can we represent these types of problems?

## Problem Formation

- Consider Movie Ratings (Trivial Case)
  - $n$  users /  $m$  movies / [1-5] ratings
  - Based purely on the rating of the movie you could show the user similar rated movies





# Content Recommendations

## DEMO

- Given the example of movies
- Consider category based scoring (e.g. action, comedy, and/or romance), as well as the ratings
- Develop a parameter to learn the preference, based on how a user rated the movie and categorized it





## Potential Issues

- Mislabeled content
- Deliberate alterations ratings or categorization
- No labeling of content




# Alternate Parameter Learning Choices

## DEMO

- Clustering Approach
  - KMeans
  - FuzzyCMeans
- Neural Networks
  - MLP



The background of the slide features a close-up photograph of autumn leaves, likely Japanese maple, in shades of purple, red, and orange. A diagonal line splits the image from the top-left to the bottom-right. The area to the left of this line is a solid dark grey, providing a high-contrast background for the white text. The area to the right shows the detailed texture and vibrant colors of the leaves.

## 3. Large Scale Machine Learning



# ML with Large Data Sets

## Issues

- Computational issues  
e.g. one step of  
gradient descent over  
100 million data  
points

## Data Intuition

More data will triumph over  
a better algorithm.

**DEMO**



# Stochastic Gradient Descent

## DEMO

### Original Descent Algorithm

Batch Gradient Descent, is when you take the sum across “all” the examples.

### Stochastic Gradient Descent

Look at a single training example selected at random.

### ALGORITHM

1. Randomly Shuffle training data
2. Repeat the update step “m” times

# Mini-Batch Gradient Descent

## DEMO

Use  $b$  examples in each iteration.

It is fairly common to randomly select batches of data.

It can outperform stochastic gradient descent in a vectorized implementation.

### ALGORITHM

1. Randomly select batches of  $b$  length
2. Repeat the update step ( $mb$ ) times



# Online Learning

## DEMO

### What is it?

A circumstance, when we want to learn from a continuous stream of information.

### When can we use it?

A large quantity of live data and/or users.

### Examples

Online shopping websites, ad monitoring

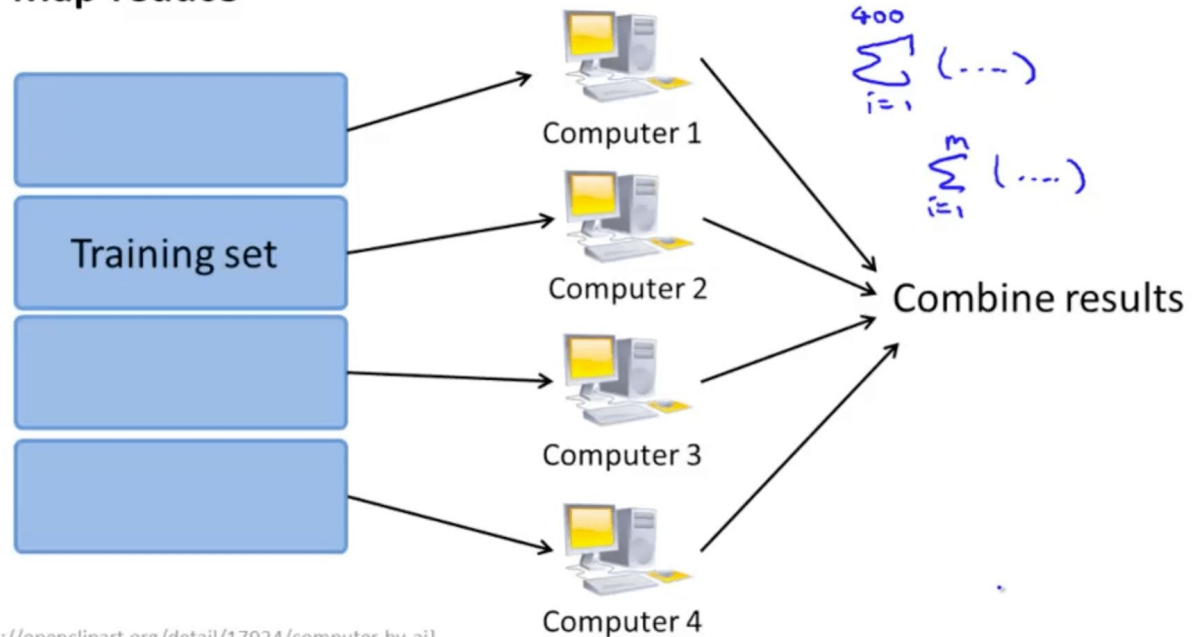
# Map Reduce and Data Parallelism

## DEMO

### Examples

- Multi-Threading
- Cluster Computing
- Running code on the GPU

### Map-reduce







# CODE

A close-up photograph of a hand held palm-up, with a fine, bright pink powder being poured from it. The powder is captured in mid-air, creating a dynamic, textured cloud. The background is dark and out of focus. A diagonal line, separating a dark blue/purple gradient on the left from a dark red/maroon gradient on the right, runs from the top-left towards the bottom-right, passing over the hand and powder.

# THANKS!

Any questions?

You can find me on:

- LinkedIn
- GitHub
  - ECE-Engineer
- Gmail
  - [kzeller@oswego.edu](mailto:kzeller@oswego.edu)
- Twitter

# CREDITS

- [https://en.wikipedia.org/wiki/Probability\\_density\\_function](https://en.wikipedia.org/wiki/Probability_density_function)
- [https://en.wikipedia.org/wiki/Kernel\\_density\\_estimation](https://en.wikipedia.org/wiki/Kernel_density_estimation)
- <https://chemicalstatistician.wordpress.com/2013/06/09/exploratory-data-analysis-kernel-density-estimation-in-r-on-ozone-pollution-data-in-new-york-and-ozonopolis/>
- <https://www.ritchieng.com>
- <http://www.stat.cmu.edu/~larry/=sml/densityestimation.pdf>