

Machine Learning in Practice

Chul Min Yeum

Assistant Professor

Civil and Environmental Engineering

University of Waterloo, Canada

CIVE 497 – CIVE 700: Smart Structure Technology

Last updated: 2020-03-28



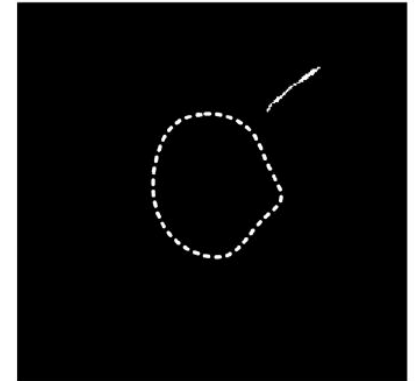
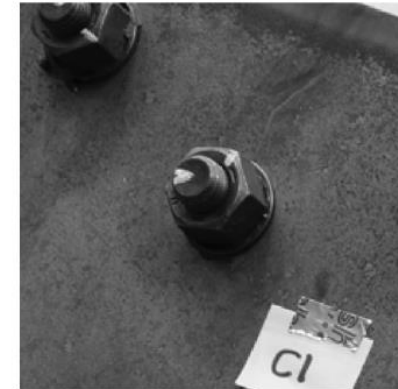
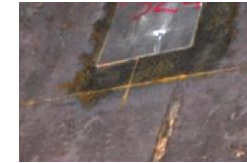
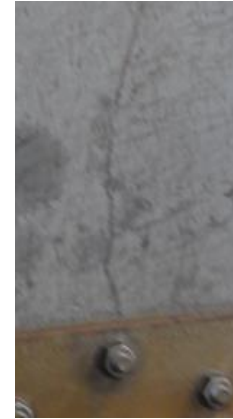
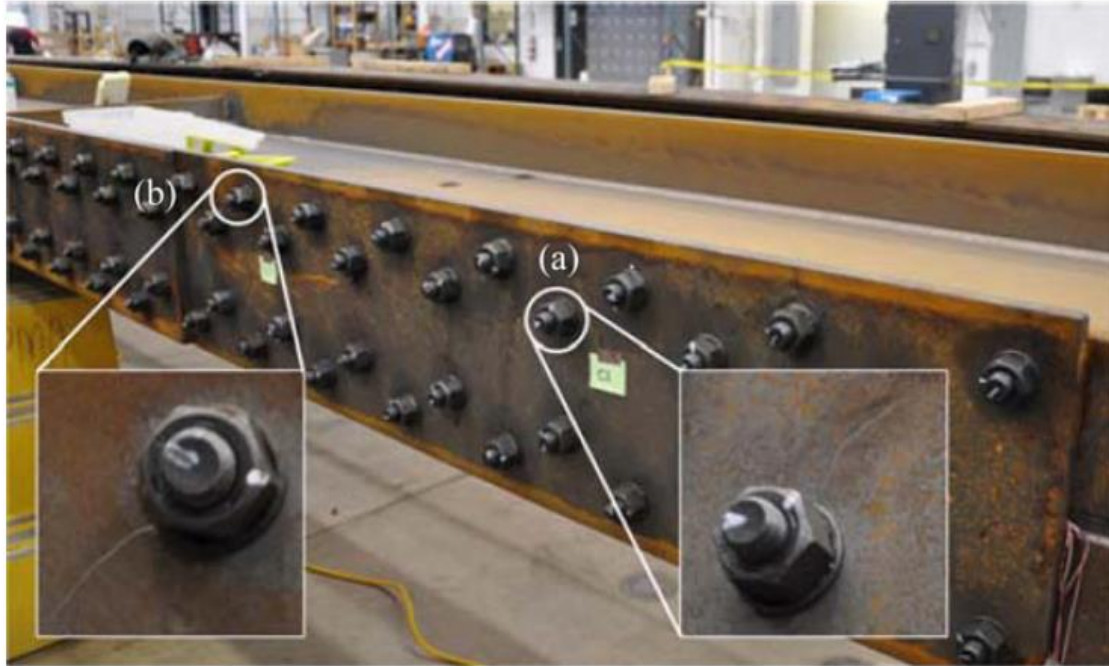
UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

- Step 1: Understand domain, prior knowledge, and goals
- Step 2: Data integration, selection, cleaning, pre-processing, etc.
- Step 3: Learn models
- Step 4: Interpret results
- Step 5: Deploy developed models

Machine Learning in Practice (Step 1: Understand domain, prior knowledge, and goals)

- **Step 1: Understand domain, prior knowledge, and goals**
 - What problems are you going to solve?
 - Do you have to utilize machine learning?
 - What prior knowledge do you involve?
- Step 2: Data integration, selection, cleaning, pre-processing, etc.
- Step 3: Learn models
- Step 4: Interpret results
- Step 5: Deploy developed models

Vision-based Crack Detection

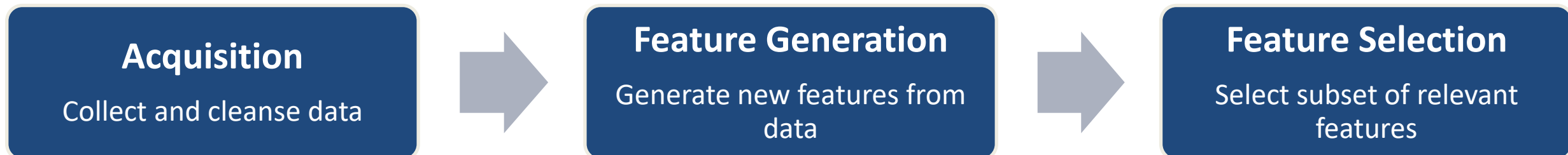


Detection of crack damage

Prior knowledge:

- Bolt holes have an initial flaw.
- Fatigue crack damage is initiated from a bolt hole boundary

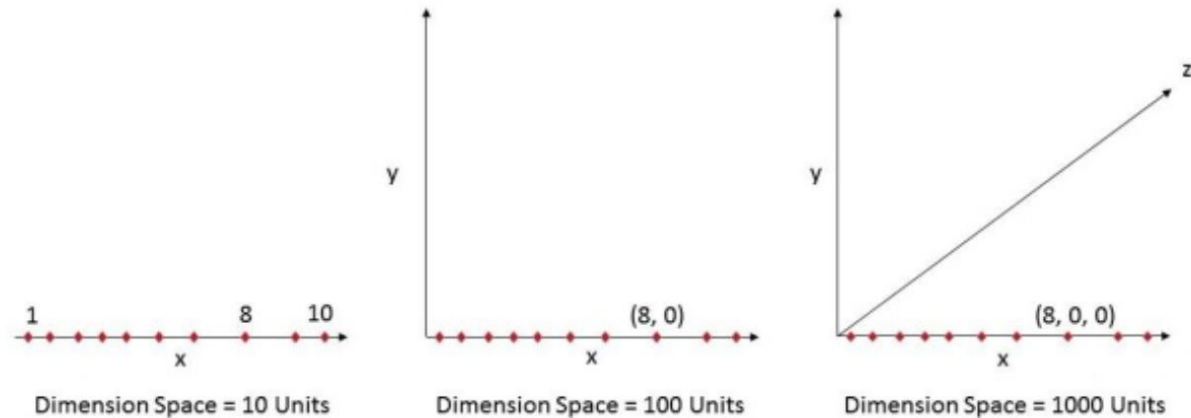
- Step 1: Understand domain, prior knowledge, and goals
- **Step 2: Data integration, selection, cleaning, pre-processing, etc.**
 - How to collect data?
 - How to pre-process the data?
 - What features/data are using for model development?
- Step 3: Learn models
- Step 4: Interpret results
- Step 5: Deploy developed models



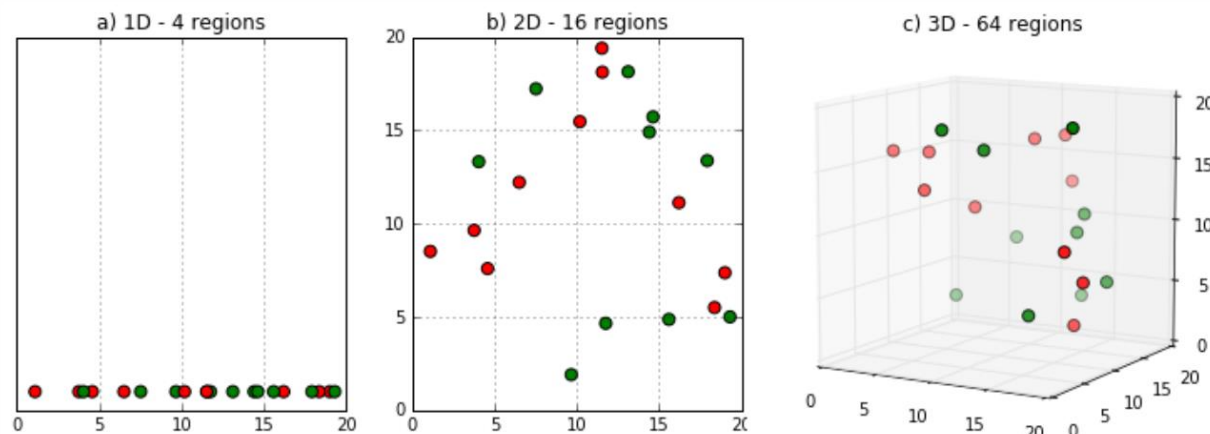
Post Earthquake Disaster Image Classification



Curse of Dimensionality



- As the number of features or dimensions grows, the amount of data we need to generalize accurately grows exponentially
- Think of image recognition problem of high resolution images $1280 \times 720 = 921,600$ pixels i.e. 921600 dimensions.
- That's why it's called **Curse of Dimensionality**. Value added by additional dimension is much smaller compared to overhead it adds to the algorithm.

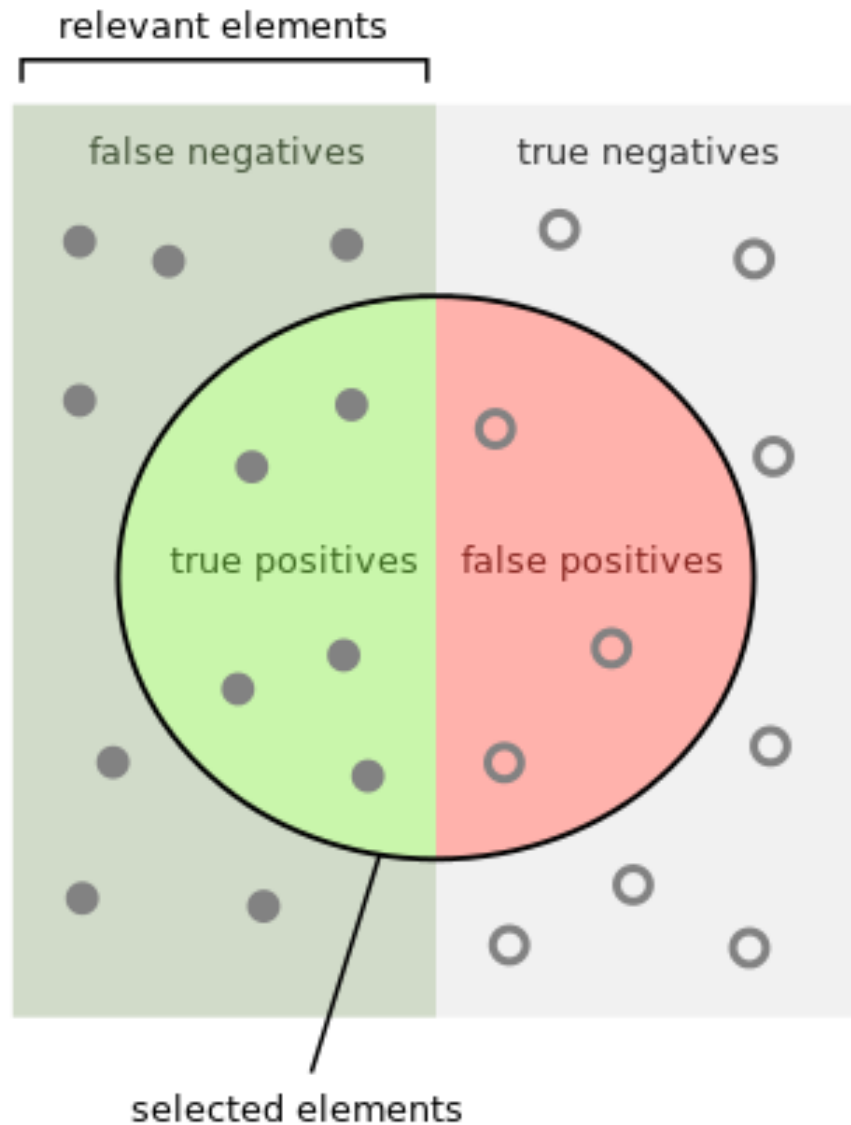


SIFT Descriptor

- Step 1: Understand domain, prior knowledge, and goals
- Step 2: Data integration, selection, cleaning, pre-processing, etc.
- **Step 3: Learn models**
 - How to represent the target function or algorithm
 - How to explore and determine the optimum hyper-parameters
- Step 4: Interpret results
- Step 5: Deploy developed models

- Step 1: Understand domain, prior knowledge, and goals
- Step 2: Data integration, selection, cleaning, pre-processing, etc.
- Step 3: Learn models
- **Step 4: Interpret results**
 - Performance metric
 - Explainability or interpretability
- Step 5: Deploy developed models

Precision and Recall



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

- Step 1: Understand domain, prior knowledge, and goals
- Step 2: Data integration, selection, cleaning, pre-processing, etc.
- Step 3: Learn models
- Step 4: Interpret results
- **Step 5: Deploy developed models**