ML Experimental Design Discussion

Overview of Assignment 1

Winter 2025

UNIVERSITY OF CALGARY

ML Experimental Design Overview

- Literature Review
- Data
- Task Definition
- Anticipated Issues

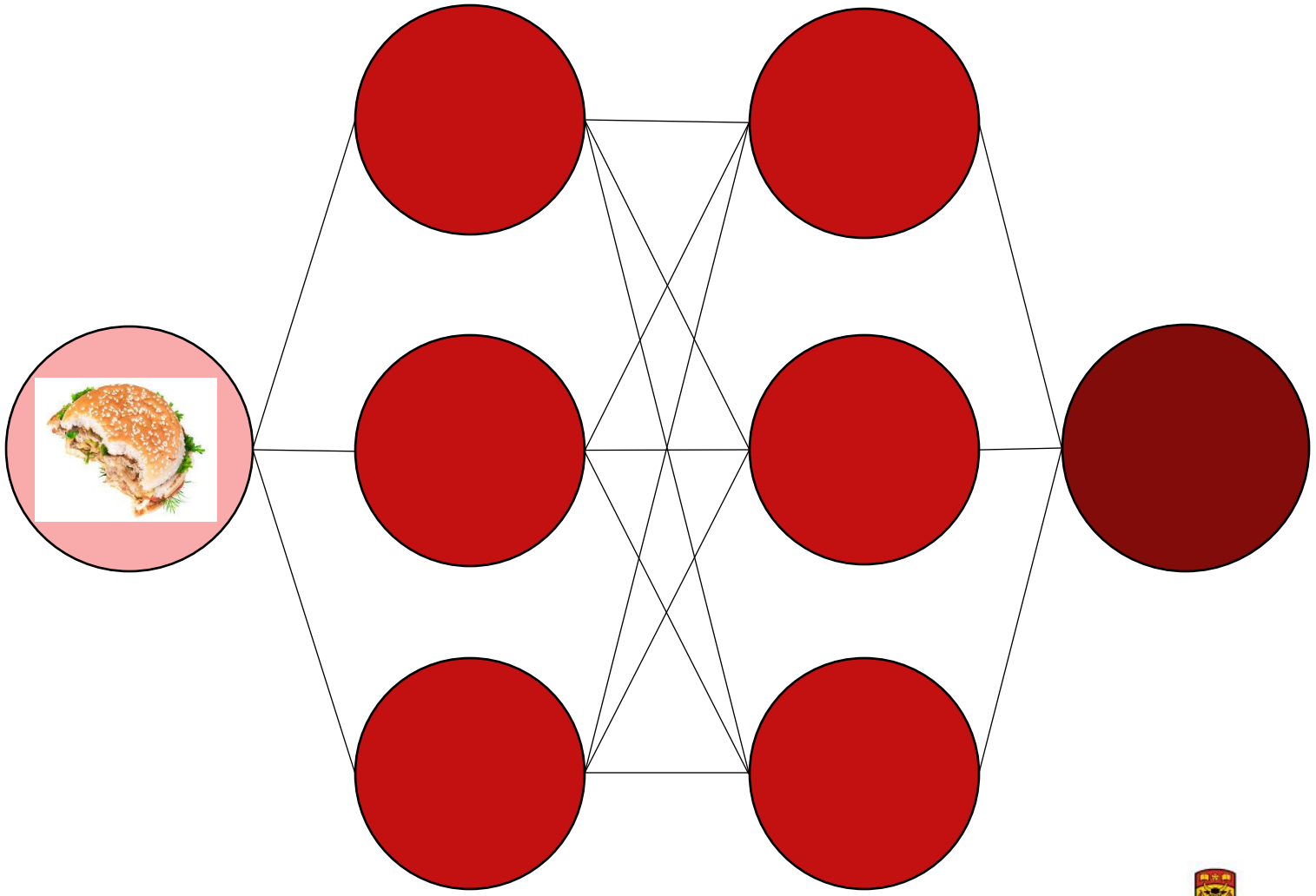
Problem Definition – Assignment 1

The city of Calgary assigned your team to develop a garbage classification system that, given a cellphone picture of an object you want to throw away and a short sentence describing the object, the system tells you whether to throw it in the “green”, “blue”, “black” trash bin or somewhere else. You can see more information about the city of Calgary's garbage collection system here: https://www.calgary.ca/uep/wrs/what-goes-where/default.html.

Whenever we're starting with a new ML problem, a literature review of similar works can help provide a good starting point!

Literature Review

- Model architecture
- Data
- Training procedure



UNIVERSITY OF CALGARY

Experimental Design Considerations

Data

Data Source

- Write a quick description of the following piece of garbage which would help with sorting it?

Data Diversity

- How might the object of interest vary?
- How might images vary?
- How may these characteristics bias the model's performance?

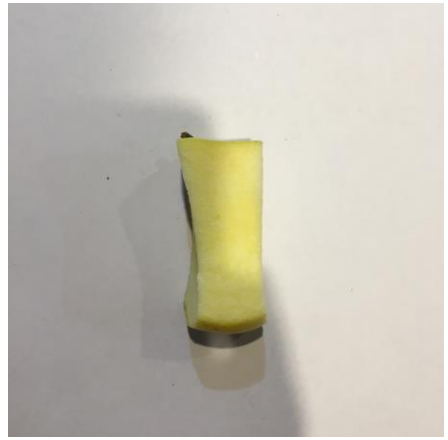
UNIVERSITY OF CALGARY

Data Preprocessing

- Why preprocess our data?
- From an imaging perspective, what are some popular preprocessing steps?

Data Preprocessing

- Object segmentation
- White background
- Pre-filtering
- Convert images to grayscale
- The object material alone does not determine the appropriate trash bin



Experimental Design Considerations

Data

Task Definition

Task Definition

Task Complexity

Multi-object detection

Single object background segmentation

Pros vs. Cons?

UNIVERSITY OF CALGARY

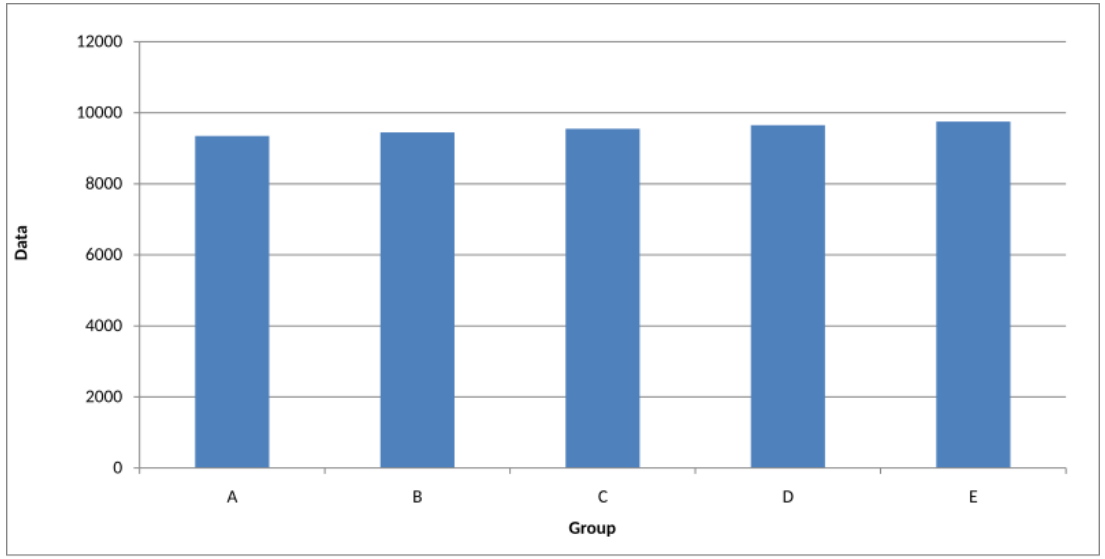
Task Definition

Model Complexity

How is the final model intended to be used?

Task Definition

Selecting A Performance Metric



What are some characteristics of a good performance metric?

UNIVERSITY OF CALGARY

Experimental Design Considerations

Data

Task Definition

Addressing Anticipated Issues

Potential Issues

Rare Occurrences - Outliers

"Another issue that can arise is the inability of the system to classify certain objects due to their rarity and the lack of data. A prime example would be Floppy Disks or VHS tapes that are generally not sold commercially anymore and are hard to find."

Potential Issues

Non-Traditional Classification

Not green, nor black nor blue trash bin?

UNIVERSITY OF CALGARY

Potential Issues

Subjective Classification

Green or blue? What if the box is closed?

"Another potential issue is the class overlap for certain items. For example, a clean pizza box can be disposed of in the blue bin, however if it is greasy/covered in food toppings, it should be disposed of in the green bin. Depending on how the photo is taken, it may be impossible to know for sure which class this item belongs to."

Some Thoughts On Starting The Assignment…

- What considerations are needed when building your dataset?
- How does your dataset fit into the bigger picture (task definition)?

Philosophical question…

Trash?

Thank you!

UNIVERSITY OF CALGARY