# Teaching Gradient Descent Through Analogies, Step by Step

## Evaluating and using analogies to teach concepts in Machine Learning to Computer Science students

Author: Thomas Koppelaar – t.j.koppelaar@student.tudelft.nl
Responsible Professor: Gosia Migut - m.a.migut@tudelft.nl

## 1. Introduction and Background

- Artificial Intelligence (AI) and Machine Learning (ML) are applied in many important sectors of the world.
- Machine Learning is being taught in more and more lecture halls every year.
- Notional Machines have been researched in Computing Education, but not ML specific.
- One exploratory paper on using analogies in ML.

**Aim of this paper: Expand on ML education using analogies.**

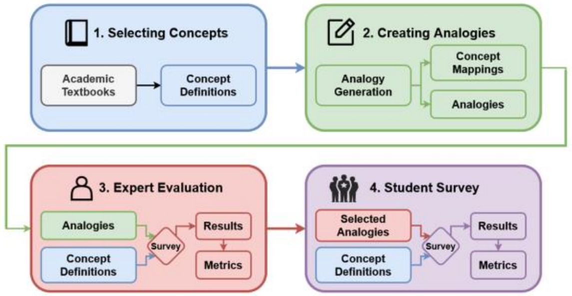


Figure 1: A full overview of the Methodology.

## 2. Research Question(s)

**How does the use of analogies in explaining Gradient Descent affect the learning proficiency for Computer Science students?**

1. How do experts in Machine Learning evaluate different analogies?
2. What knowledge do Computer Science students gain from learning about Gradient Descent using analogies?
3. How do Computer Science students evaluate their engagement with the topic when using analogies to teach Machine Learning?

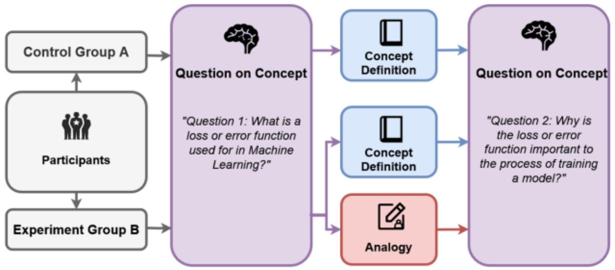


Figure 2: A diagram showcasing the setup for the first part of the student survey.

## 3. Methodology

Research goals:

1. Construct analogies for (concepts related to) Gradient Descent
2. Have experts review analogies
3. Measure learning proficiency through survey. Educational goals are: Remember and Understand
4. Have students evaluate their engagement

## 3.1 Example analogy

| Concept | Definition | Analogy |
|---|---|---|
| Optimization and Error / Loss / Cost function | *Optimization refers to the task of either minimizing or maximizing some function f(x) by altering x. When we are minimizing it, we may also call it the cost function, loss function, or error function.* | *Imagine you're in a radioactive zone. We're using a Geiger counter (function) to measure the radiation in different spots, Optimization refers to the task of either looking for a safe zone (minimization), or looking for high spots of radiation (maximization). When we are looking for a safe spot, we are minimizing the radiation we measure on our geiger counter through measurements and calibration (loss function).* |

Figure 3: An example concept definition, alongside a handmade analogy. All analogies are available at: https://ml-teaching-analogies.github.io/

## 3.2 Measuring effectiveness of analogies

Metrics used in survey:

1. **Target concept coverage** – Does the analogy cover the topics in the description?
2. **Mapping strength** - Are these mappings structurally consistent?
3. **Metaphoricity** - Conceptual distance between the source and the target concept.

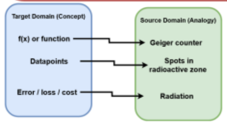


Figure [todo]: An example concept mapping.

For each Analogy:
1. Definition was given
2. Analogy was shown
3. Scoring given: 1 – Low, 2 – Mid, 3 - High

## 4.1 Expert Evaluation Results (SQ1)

N = 15*, 10 BSc students, 3 TA's, 1 MSc student, 1 Lecturer/professor

** 16 responses with 1 invalid*

| Analogy | Average Score | Krippendorff's Alpha |
|---|---|---|
| **Optimization and Error / Loss / Cost function (OLEC)** | **2.667** | -0.064 |
| Gradient | 2.222 | -0.185 |
| **Gradient Descent (GD)** | **2.389** | **0.111** |
| Critical Points | 2 | -0.111 |
| Batch Gradient Descent | 2.111 | -0.233 |
| Stochastic Gradient Descent | 1.833 | -0.064 |

Figure [todo]: The average score and Krippendorff's Alpha for each analogy, given by the experts in the expert evaluation.

## 4.2 Student Survey Results (SQ2&3)

*N = 15*

| Metaphor | | μ | σ | t-test | p |
|---|---|---|---|---|---|
| OLEC | Control | 0.375 | 0.518 | -0.293 | 0.778 |
| | Experiment | **0.429** | 0.535 | | |
| GD | Control | **0.750** | 0.916 | 0.717 | 0.497 |
| | Experiment | 0.571 | 0.378 | | |

Figure [todo]: The mean knowledge gain per concept from the student survey. Values are rounded to 3 decimal places.

**Conclusion: No statistically significant results for SQ2&3.**

| | | μ | σ | t-test | p |
|---|---|---|---|---|---|
| Attention | Control | **2.250** | 1.238 | 0.381 | 0.714 |
| | Experiment | 2.083 | 1.265 | | |
| Relevance | Control | **3.375** | 0.916 | 0.642 | 0.541 |
| | Experiment | 3.167 | 1.304 | | |
| Confidence | Control | **3.500** | 1.604 | 0.176 | 0.865 |
| | Experiment | 3.400 | 1.140 | | |
| Satisfaction | Control | **2.500** | 1.414 | 0.666 | 0.527 |
| | Experiment | 2.167 | 0.548 | | |

Figure [todo]: The mean ARCS metrics from the responses of the student survey. Values are rounded to 3 decimal places.

## 5. Discussion

- General: Low sample size
- Expert Evaluation:
  - Low agreement / Systematic disagreement
  - Single review iteration
  - Equal review weights for all knowledge levels
- Student Survey
  - Better performance on Post-questions
  - One-way information stream

## 6. Conclusion and Future Work

1. Analogies for concepts relevant to ML / Gradient Descent
2. No statistically significant learning proficiency/engagement
3. Exploratory paper, framework/method for evaluation

Suggestions for Future Work:

- Mimic traditional examination setup
- Measure over longer period of time
- Larger sample size

## Links

Link to Paper

Link to Analogies