A Systematic Way of Creating Assessments for Machine Learning Courses

Author: Kerem Cakici Supervisor: Gosia Migut Responsible Professor: Prof. Dr. M. Specht

TU Delft Delft University of Technology

Introduction

- Machine learning growing
- Lack of expertise
- Enhancing machine learning education
- Assessments
- Skill Circuit
- Reliably show progress
- Formative vs Summative
- Learning Outcomes (Objectives)
- Bloom's taxonomy

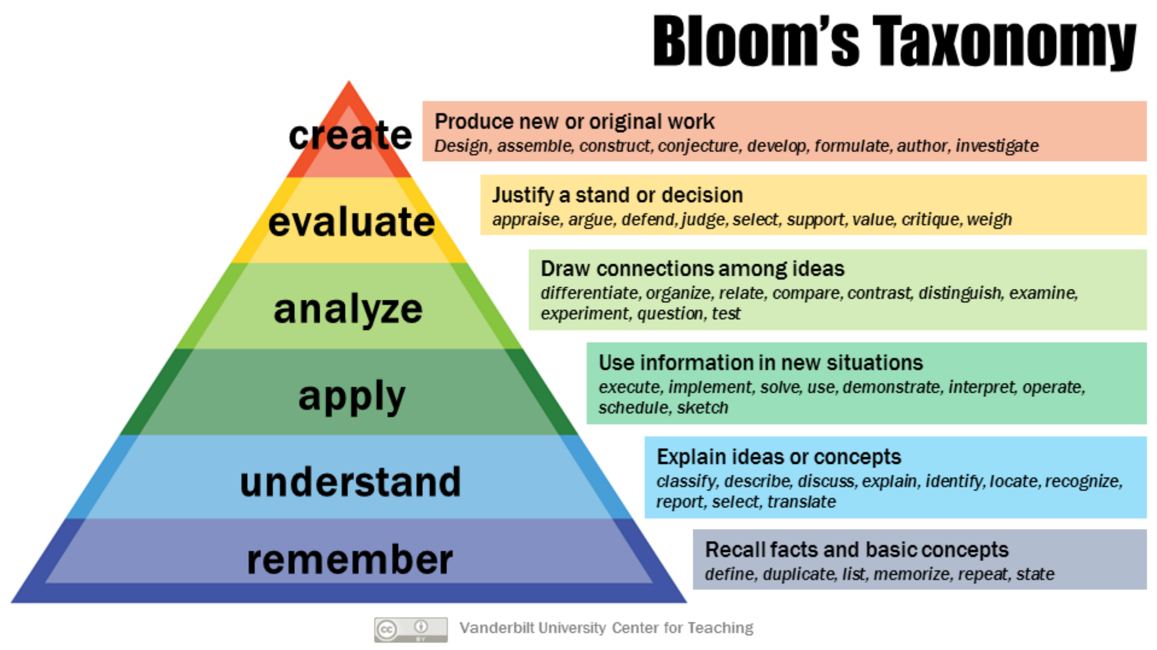


Figure 1: Bloom's revised taxonomy [1]

Research Question

How to design assessments that reliably show progress on a module in machine learning for both students and teachers using Bloom's taxonomy and learning outcomes?

- What is a comprehensive set of learning outcomes of the module?
- What are the corresponding cognitive categories from Bloom's taxonomy of the learning outcomes?
- How can learning outcomes translate into sample assessment questions?
- How does Bloom's taxonomy help to find the appropriate way to assess a learning outcome?
- What are the ways of evaluating assessment questions?

Methodology

- Literature study
- Choose a module (Non-parametric density estimation)
- Identify problems in the initial learning outcomes
- Improve initial learning outcomes
- Revise & Classify learning outcomes
- Create an assessment design guideline
- Choose assessment methodologies for learning outcomes
- Create sample questions
- Evaluate by interviews
- Analyse the interviews by using conventional content analysis

Designing Assessment

Reliability: If an assessment question is reliable, it should produce the same result in the same conditions consistently, and these results should be matching with the "real" level of the test taker [2].

Validity: For an assessment question to be valid, it needs to measure what needs to be measured [2].

Efficient Assessment Methodology: Most feasible assessment methodology that can test the learning outcome reliably.

Guideline:
- Constructive-alignment
- Consider Bloom's classification
- Consider achievability
- Feasibility
- Using verbs instead of Bloom's classification
- "Know", "Find" and "Apply" mapped using past exams

References

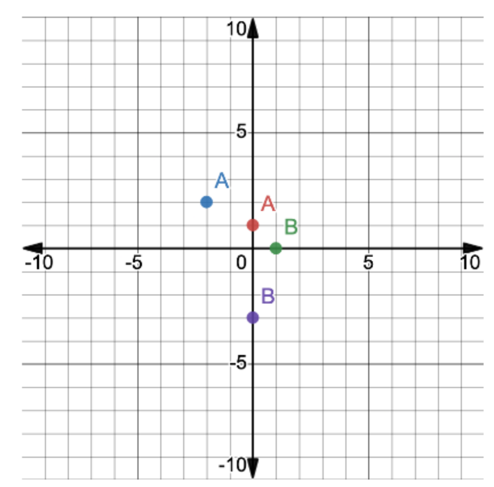
[1] Armstrong, P. (2010). Bloom's Taxonomy. Vanderbilt University Center for Teaching. Retrieved [todaysdate] from https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/.

[2] Dunn Lee et al. The Student Assessment Handbook : New Directions in Traditional and Online Assessment. Routledge, 2004. isbn: 9780415335300.

Revised Learning Outcomes

Learning Outcome	Verb Used	Cognitive Level	Assessment Methodology
Explain the difference between parametric and non-parametric density estimation.	Explain	Analysis	Open-ended Question
Explain Parzen density estimation and the purpose of window functions in detail.	Explain	Understanding	Open-ended Question
Know which parameter needs to be optimized in Parzen density estimation.	Know	Remembering	Multiple-choice Question
Find Parzen probability density function estimates at a given point using (Gaussian, box, tri, triweight) window function.	Find	Application	Multiple-choice Question
Explain k-nearest neighbors, the influence of parameter k, how to optimize parameter k, and how to break ties in detail. **(4th Question)**	Explain	Understanding	Open-ended Question
Compute Euclidean, Manhattan, and Hamming distance. **(1st Question)**	Compute	Application	Multiple-Choice Question
Apply k-nearest neighbors with a specific k parameter, and with one of the above distance metrics.	Apply	Application	Multiple-Choice Question
Implement the k-nearest neighbor classifier in Python. **(2nd Question)**	Implement	Application	Implementation Question
Explain Naïve Bayes, and the effects of choosing Gaussian or Parzen as model per feature in detail.	Explain	Understanding	Open-ended Question
Explain the advantages and disadvantages of the above-mentioned methods. **(3rd Question)**	Explain	Analysis	Open-ended Question

Assessment Questions



1. **(Assesses learning outcome no. 6)** Mark wants to choose which distance metric to use for his k-nearest neighbors algorithm with k = 2 . To do this he first decides to plot a part of the data as the data is two-dimensional. He writes the labels of the points next to them. After he plots a partition of data, he wants to choose the distance metric that will classify a point in the origin as B. Which distance metrics should he choose?
   - A. Manhattan Distance
   - B. Euclidean Distance
   - C. Either of the metrics classifies a point in the origin as A
   - D. Either of the metrics classifies a point in the origin as B
2. **(Assesses learning outcome no. 8)** Implement the following methods for k-nearest neighbors with k = 1:

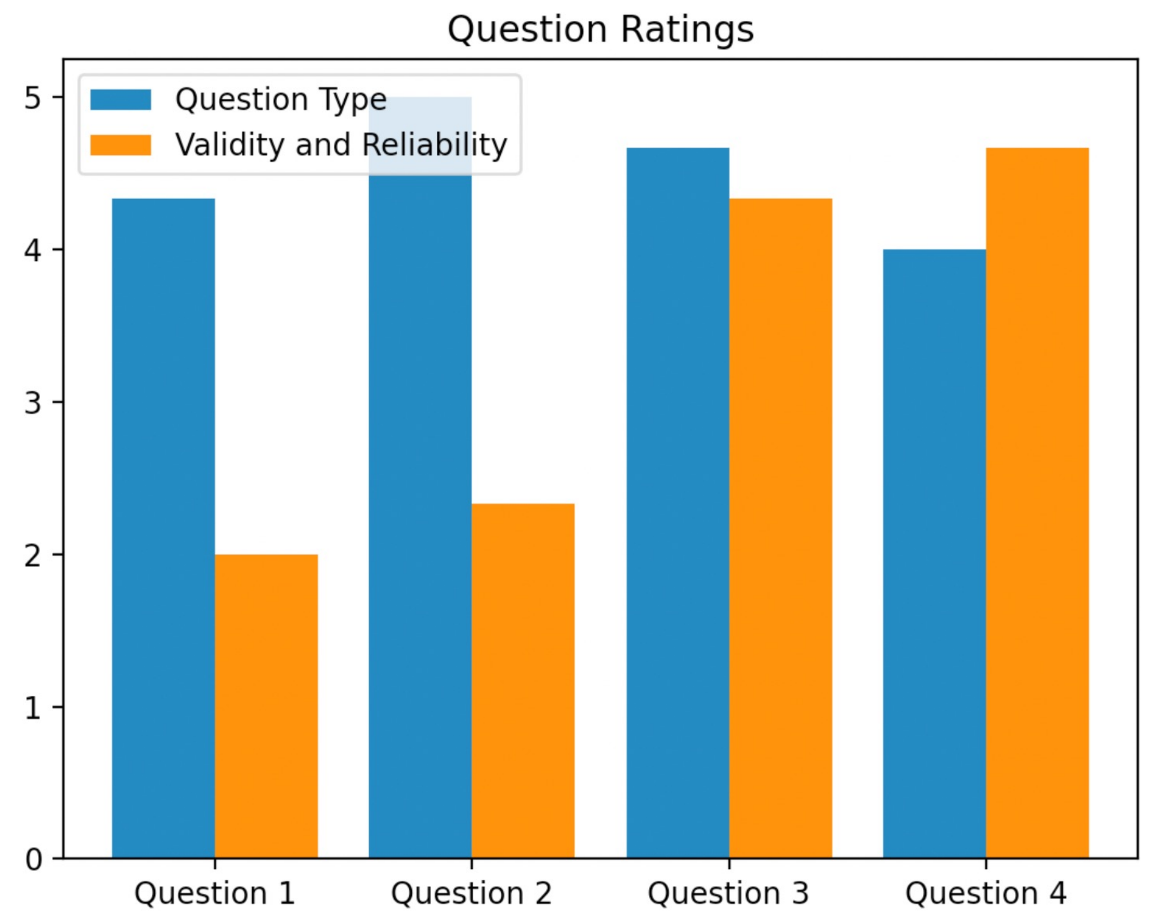
```
# Points are stored in a list of tuples: For example x=1 y=4 label=B as (1, 4, 'B')
# This method should return the classified label for the new point in k nearest neighbor algorithm with k=1
def classify_using_closest_point(points: list, new_point: (int, int)):
    # implement here
    return

# This method should return euclidean distance between two two-dimensional points
def calculate_euclidian(x1, y1, x2, y2):
    # implement here
    return
```

3. **(Assesses learning outcome no. 10)** Can you explain one advantage and one disadvantage k-nearest neighbor algorithm?
4. **(Assesses learning outcome no. 5)** Can you explain two different ways to break ties in k-nearest neighbor algorithm?

Assessment Evaluation

Pre-Assessment evaluation:
- Peer-review process
- Advice and checklist
- Teaching assistant (Feasible)
- Student perspective

Post-Assessment evaluation:
- More students → More data
- "Easy" to see reliability and difficulty
- Editing Rubric for open-ended questions
- Start-over
- Questionnaires and focus group



Evaluation of the Questions

Question 1:
- Isolate the learning outcome (Validity)

Question 2:
- No evaluation or training or testing steps
- Data Structures may be confusing to students

Question 3 & 4:
- Well-made questions
- Too open-ended

Future Work

- Creating a taxonomy specifically for Machine Learning
- Revising learning outcomes for all modules
- Creating Assessment questions and improving the table
- Using statistical tests