

Insight Report

Is k-Nearest Neighbors (kNN) Truly a Supervised Learning Algorithm?

Overview

The k-Nearest Neighbors (kNN) algorithm is widely taught as a classic example of a supervised learning method. However, a closer inspection raises important philosophical and technical questions: Can an algorithm that does not learn model parameters truly be said to “learn”? This report explores the nuance behind this classification.

Key Observations

1. Supervised Learning, Broadly Defined

In machine learning, supervised learning generally refers to any method that:

- Uses labeled training data
- Attempts to generalize a mapping from inputs to outputs
- Makes predictions based on labels seen during training

By this definition, kNN qualifies as supervised learning because it:

- Requires labeled training examples
- Uses those labels during prediction
- Computes prediction performance against known ground truth

2. kNN as a "Lazy Learner"

kNN does not involve a training phase where parameters are adjusted. Instead, it:

- Stores the labeled dataset
- Uses a distance metric to find the nearest neighbors during prediction
- Aggregates labels of the nearest neighbors to decide the output

This is why kNN is often referred to as an instance-based or lazy learner—all computation is deferred until prediction time.

3. Lack of Parameter Updating

Unlike logistic regression or neural networks, kNN does not adjust internal parameters based on prediction errors. There is no gradient descent, backpropagation, or iterative improvement. The "model" is simply the stored dataset.

Philosophical Insight: Is That Really "Learning"?

In a stricter, more traditional sense of learning that involves model evolution or performance improvement with experience, kNN does not learn.

This brings us to Tom Mitchell's well-known definition of machine learning in his book, "Machine Learning," 1992 (<https://www.cs.cmu.edu/~tom/mlbook.html>):

Insight Report

Is k-Nearest Neighbors (kNN) Truly a Supervised Learning Algorithm?

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ."

kNN does not improve performance with more usage or feedback—it merely performs a label lookup using a fixed dataset. In that light, it might be better described as a supervised predictor, rather than a supervised "learner."

Learning vs. Memorization: An Analogy

If a student memorizes formulas and “merely” regurgitates them correctly on a test, we conventionally say they have “learned” something—even if deeper understanding or adaptation is lacking.

Similarly, kNN stores labeled examples and recalls them at prediction time based on similarity—a kind of memorization-based learning that fits the conventional (if debatable) definition of supervised learning.

Therefore, we may summarize:

- ✓ kNN is a supervised learning algorithm (by label-based prediction)—by convention.
- ✗ kNN is not a learning algorithm (in the parametric, iterative improvement sense).

Conclusion

While kNN does not learn in the traditional sense of model adaptation or parameter optimization, **it is nevertheless conventionally classified as a supervised learning algorithm** due to its use of labeled data in making predictions using recall of “memorized” data.

Disclosure

This report was co-created by the instructor with the assistance of the ChatGPT-4o language model (OpenAI). The content was generated through structured prompts, followed by selective editing, original additions, and refinement by the instructor to ensure conceptual accuracy and academic rigor. Students should treat this document as a curated and pedagogically guided resource.