

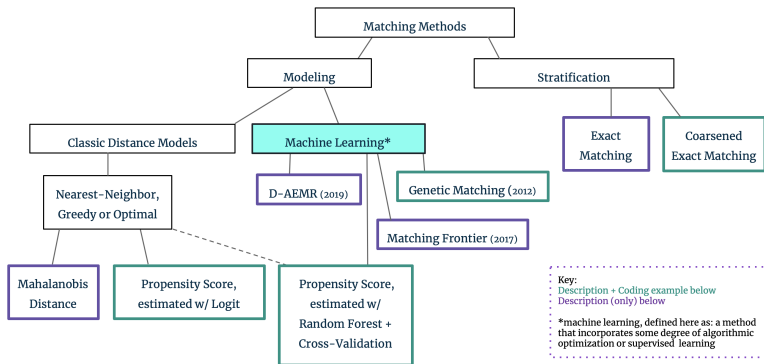


## Cannabis Data Science #79

August 17<sup>th</sup>, 2022



# Matching Models



Authors: Samantha Sizemore and Raiber Alkurdi

*Matching Methods For Causal Inference: A Machine Learning Update*

[https://humboldt-wi.github.io/blog/research/applied\\_predictive\\_modeling\\_19/matching\\_methods/](https://humboldt-wi.github.io/blog/research/applied_predictive_modeling_19/matching_methods/)

# The History of Nearest Neighbor Search

- **Literate Programming**

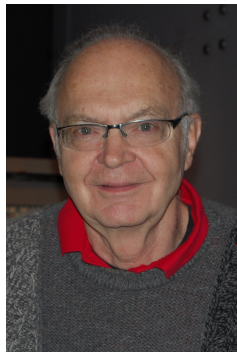
*"Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to **human beings** what we want a computer to do."*

- Donald Knuth

- **Analysis of algorithms**

The study of the complexity of algorithms.

- Creator of the **T<sub>E</sub>X!**



**Donald Knuth**

Computer History Museum's  
Revolution Exhibition (2011)

Author: Alex Handy

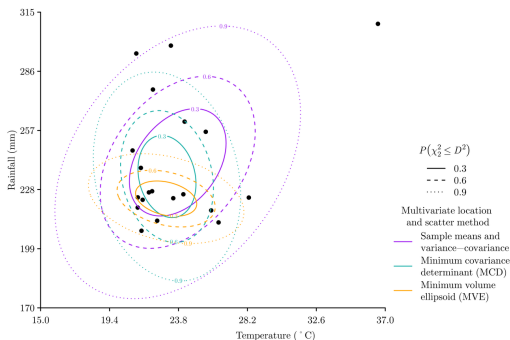
CC BY-SA 2.0

<https://creativecommons.org/licenses/by-sa/2.0>

# Modern Search Algorithms

- **$k$ -NN search** - Find the  $k$  closest points to point  $P$ .
  - ▶ PCA into feature space, followed by  $k$ -NN classification.
- **Mahalanobis distance** - a measure of the distance between a point  $P$  and a distribution  $D$ .

Hypothetical 2D example of Mahalanobis distance.



Author:  
Tretherington  
CC BY-SA 4.0  
<https://creativecommons.org/licenses/by-sa/4.0>

# Supervised learning

**Supervised learning** is task of learning a function that maps an input to an output based on example input-output pairs.

- ❶ Infer a function from labeled training data.
  - ▶ Use training examples to get training data.
  - ▶ Each example is a pair of an input object and a desired output value, e.g. ("Δ-9 THC", "delta\_9\_thc").
- ❷ Map new examples.
  - ▶ Generalize from the training data to unseen situations.
- ❸ Re-train the model with new training examples.

**Goal:** Correctly determine the class labels for unseen instances.

# Cannabis Data Science Application

## Cannabis Data Science Application

- Can we use **matching models** to find cannabis products similar to a given product?



Thank you for coming.

### Insights of the Day

- Small, seemingly inconsequential actions can have profound long-term effects because of compounding.

What is on your mind for next week?