Cultural Analytics

ENGL 64.05

Fall 2019

Prof. James E. Dobson

Nov 6, 2019

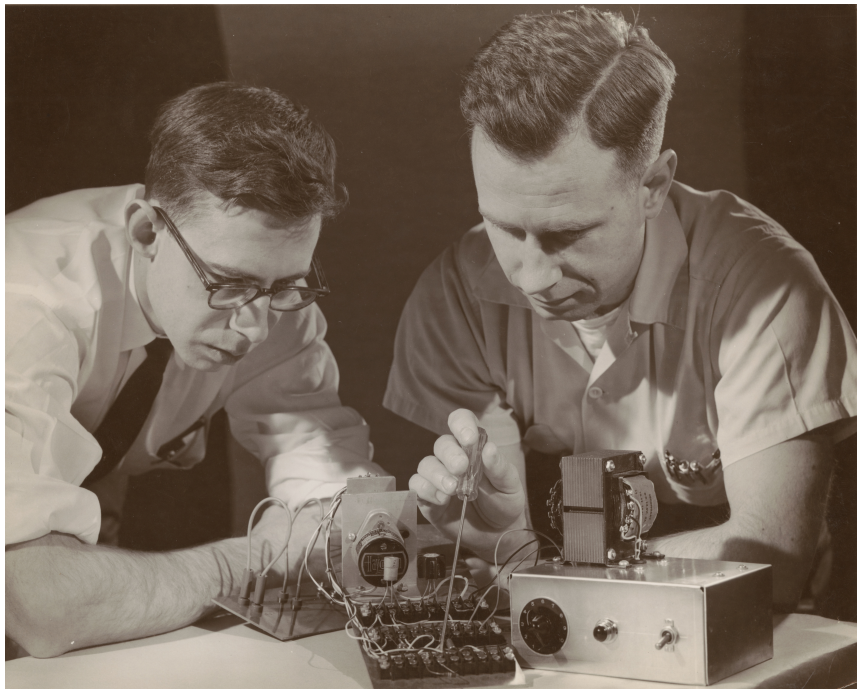
Machine Learning

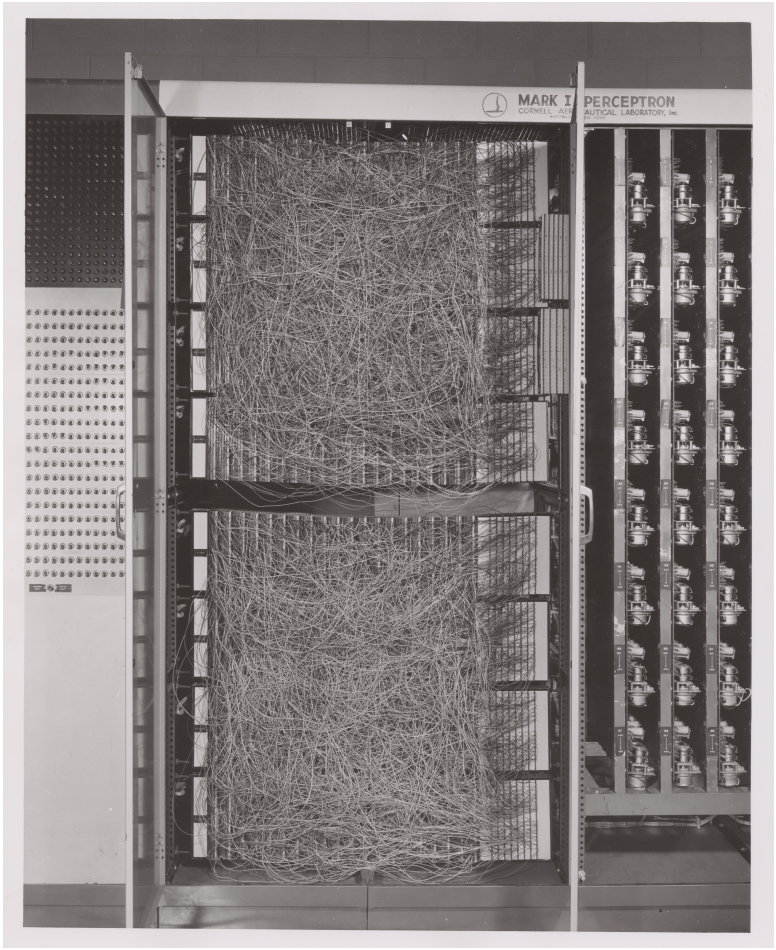
- Discursive formation that has taken different forms at different times.
- Sometimes linked with artificial intelligence.
- Meredith Broussard cites Tom Mitchell's definition:

  "We say that a machine learns with respect to a particular task T, performance metric P, and type of experience E, if the system reliability improves its performance P at task T, following experience E. Depending on how we specify T, P, and E, the learning task might also be called by names such as data mining, autonomous discovery, database updating, programming by example, etc" (92).

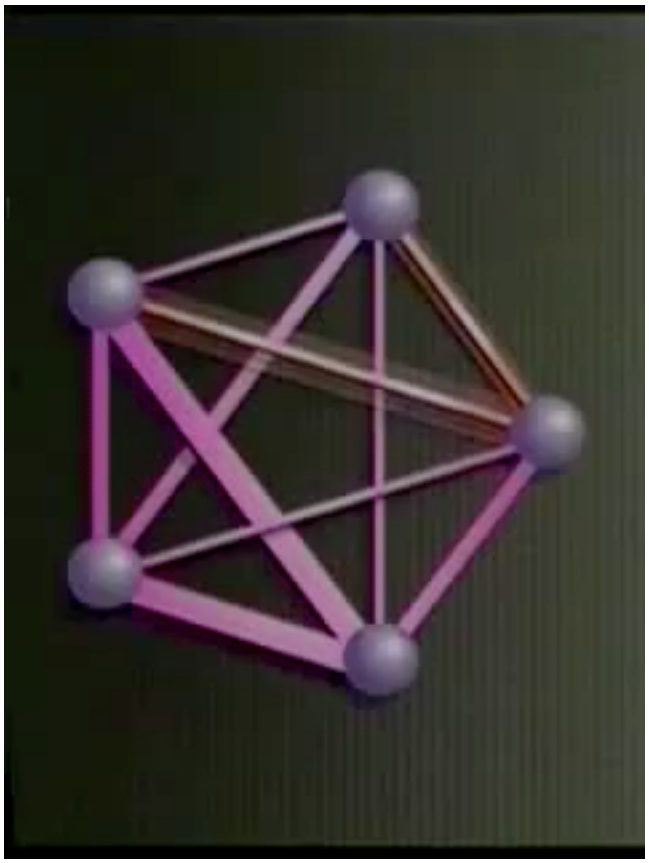
Cultural Analytics

The Machines of Machine Learning





The Mark I Perceptron (1958)



Popular Machine Learning Algorithms

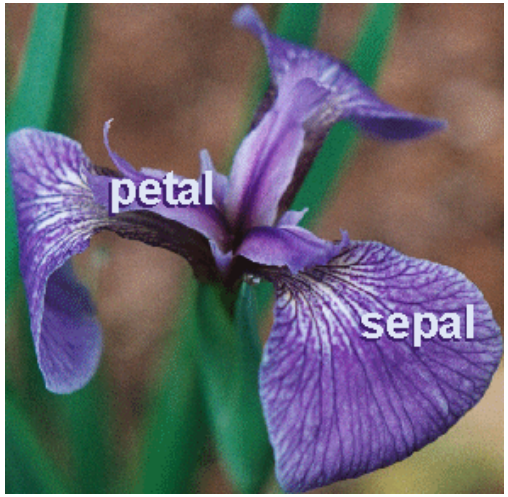
- **LDA**: latent Dirichlet Allocation
- **NB**: Naïve Bayes classification
- **perceptron**: Frank Rosenblatt's neural network model
- **kNN**: k-Nearest Neighbors
- **SVM**: Support Vector Machine
- **Tree**: Decision Tree Learning
- **Random Forest**: builds on decision trees
- **CNN**: convolutional neural network; deep learning

Classification Requirements

- Known and established (testable) classes.
- Features that can be extracted reliably from data.
- Enough data to split into test and training datasets.
- Enough representative samples to extract usable features.

Iris Dataset

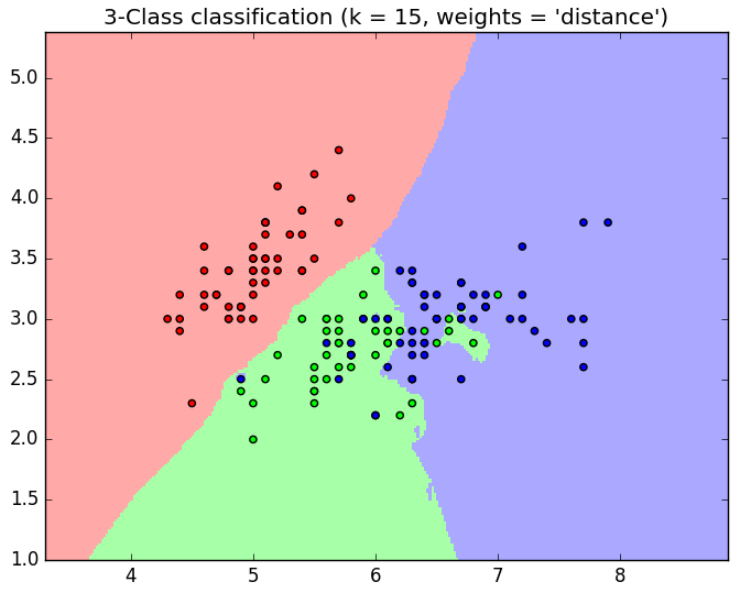
- **Feature Selection:**
  - sepal length in cm
  - sepal width in cm
  - petal length in cm
  - petal width in cm
- **Classes**:
  - Iris Setosa
  - Iris Versicolour
  - Iris Virginica

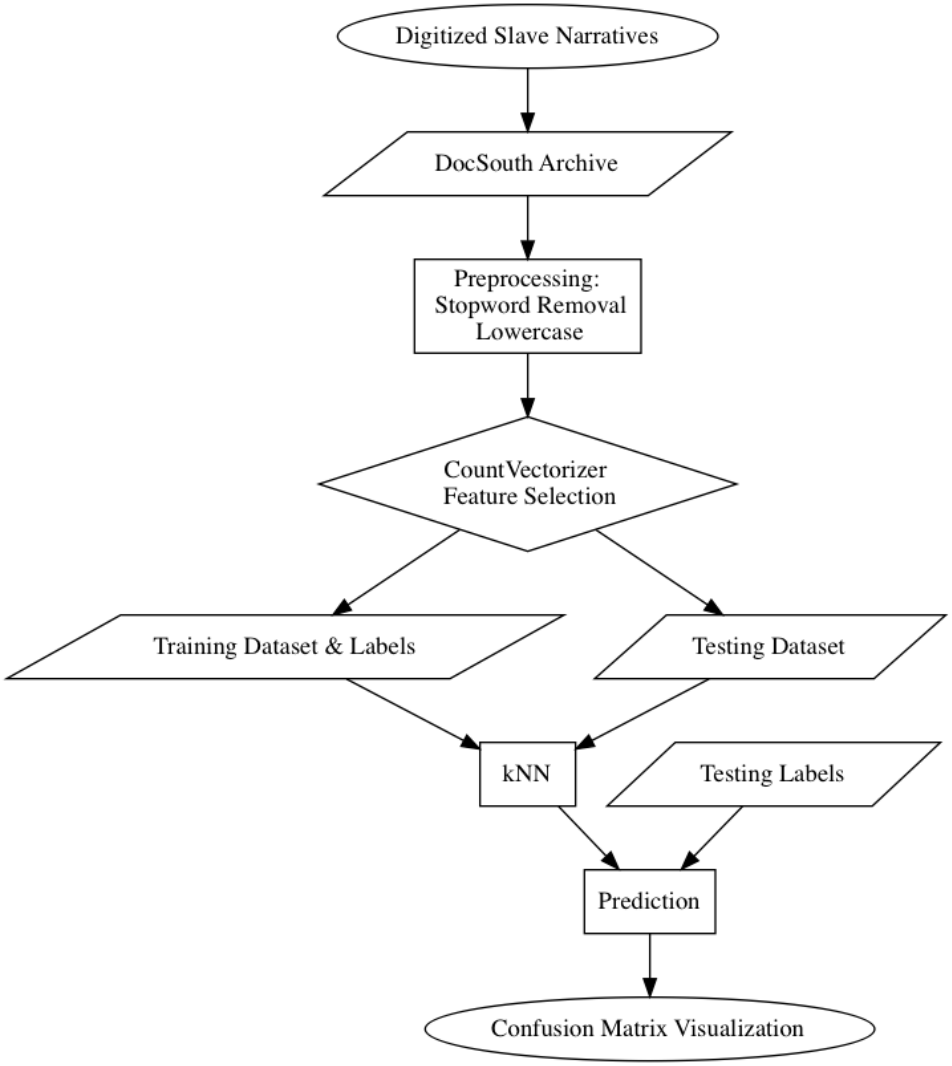


Ronald A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics* 7, no. 2 (1936): 179–88.

Data and Classification

```
[4.9, 3.0, 1.4, 0.2],
[4.7, 3.2, 1.3, 0.2],
[4.6, 3.1, 1.5, 0.2],
[5.0, 3.6, 1.4, 0.2],
[5.4, 3.9, 1.7, 0.4],
[4.6, 3.4, 1.4, 0.3],
[5.0, 3.4, 1.5, 0.2],
[4.4, 2.9, 1.4, 0.2],
[4.9, 3.1, 1.5, 0.1],
[5.4, 3.7, 1.5, 0.2],
[4.8, 3.4, 1.6, 0.2],
[4.8, 3.0, 1.4, 0.1],
[4.3, 3.0, 1.1, 0.1],
[5.8, 4.0, 1.2, 0.2],
[5.7, 4.4, 1.5, 0.4],
[5.4, 3.9, 1.3, 0.4],
[5.1, 3.5, 1.4, 0.3],
[5.7, 3.8, 1.7, 0.3],
[5.1, 3.8, 1.5, 0.3],
[5.4, 3.4, 1.7, 0.2],
```



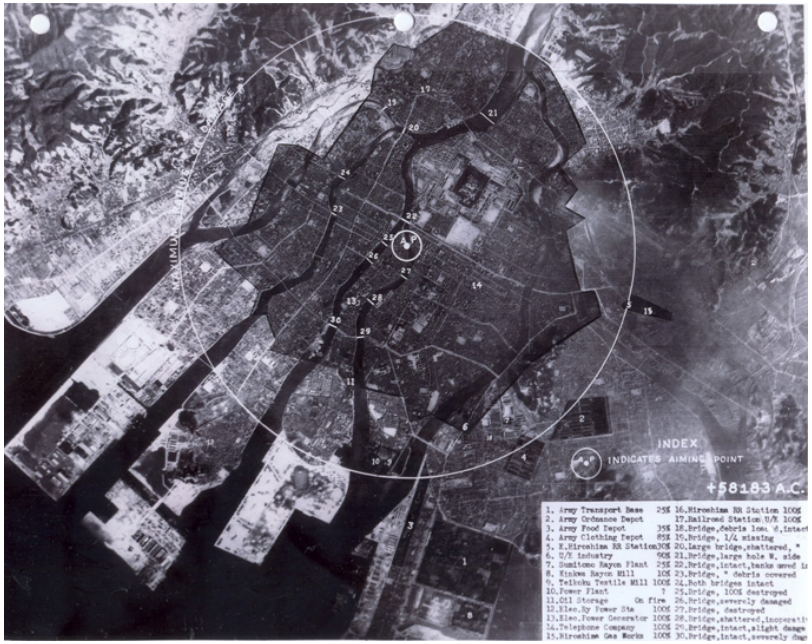


Periodization of a Dataset (DocSouth)

kNN: k-nearest neighbors

- Very simple procedure for learning from data to classify new data.
- All learning takes place at the time of testing data (no proper "training" period).
- I use as a case study for understanding machine learning and classification and the cultural history of the methods.

Hodges and Lehmann in Guam



- Modern academic statistics begin with the UC Berkeley Math department.
- Erich Lehmann recalls that "America's entry into World War II in 1941 put all further academic development on hold. Jerzy Neyman took on war work, and for the next several years this became the laboratory's central and all-consuming activity."
- "The nearest-neighbor rule, as with almost all things computational, is a direct product of military research" (123).
- "Encouraged by Jerzy Neyman, their adviser, a number of Berkeley mathematicians, including Erich Lehmann and a fellow graduate student by the name of Joseph Hodges, signed up to become statistical advisers to the US military. The air force asked them to 'track and analyze bombing accuracy' (28), among other tasks" (123).

Area Bombing

Cultural Analytics

- What was called “area bombing” depended on the close proximity of the targets, a large number of the fire-producing bombs, and low-flying aircraft. Area bombing was enabled by the interpretation of aerial reconnaissance photographs and the marking of dense groupings of buildings.
- While I wouldn’t want to suggest Lehmann and Hodges’s report and the subsequent area bombing strategy employed by LeMay led directly to the development of classification algorithms used in machine-learning technologies, they share, in disturbing ways, a set of spatial logics.

Aerial photograph, Tokyo, March 11, 1945.
Courtesy National Archives, photo no. 342-FH-3A3851-56542ac

Evelyn Fix and Joseph Hodges, Jr.

AD A800 276

DISCRIMINATORY ANALYSIS

NONPARAMETRIC DISCRIMINATION: CONSISTENCY PROPERTIES

Evelyn Fix, Ph.D.
J.L. Hodges, Jr., Ph.D.
University of California, Berkeley

PROJECT NUMBER 21-49-004
REPORT NUMBER 4*

*(Prepared at the University of California
under Contract No. AF41(128)-31)

USAF SCHOOL OF AVIATION MEDICINE
RANDOLPH FIELD, TEXAS
FEBRUARY 1951

Thomas M. Cover and Peter Hart "Nearest Neighbor Pattern Classification" (1967)

- Cover and Hart extend the nearest neighbor rule into a method for *learning* from prior classifications, now called k-nearest neighbors rule.
- "The nearest neighbor decision rule assigns to an unclassified sample point the classification of the nearest of a set of previously classified points."
- "It is reasonable to assume that observations which are close together (in some appropriate metric) will have the same classification, or at least will have almost the same posterior probability distributions on their respective classifications" (21).
- "[S]amples which are close together have categories which are close together" (22).

[T]here are two families, *Blue Squares and Red Triangles*. We call each family [a] **Class**. Their houses are shown in their town map which we call *feature space*. Now a new member comes into the town and creates a new home, which is shown as [a] green circle. He should be added to one of these Blue/Red families…One method is to check who is his nearest neighbour. From the image, it is clear it is the Red Triangle family. So he is also added into Red Triangle. This method is called simply **Nearest Neighbour**, because classification depends only on the nearest neighbour. But there is a problem with that. Red Triangle may be the nearest. But what if there are lot of Blue Squares near to him? Then Blue Squares have more strength in that locality than Red Triangle. So just checking nearest one is not sufficient. Instead we check some k nearest families. Then whoever is majority in them, the new guy belongs to that family.