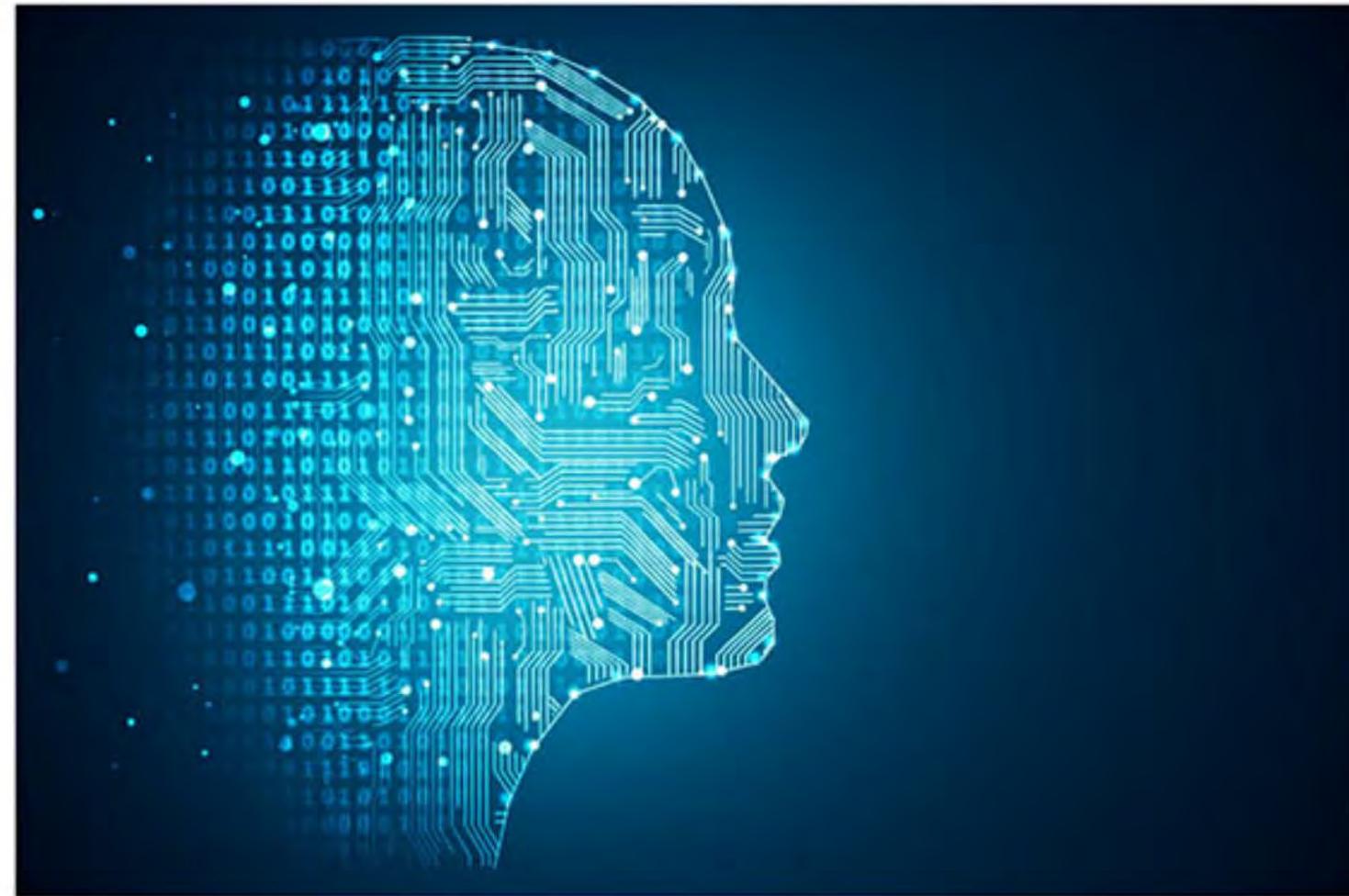




CENTER FOR
STATISTICS AND
MACHINE LEARNING



A Hands-On Introduction to Machine Learning



Julian Gold
Jonathan Hanke

January 13–16, 2026



With materials from:

Brian Arnold, Gage DeZoort, Julian Gold, Jonathan Halverson, Jonathan Hanke, Christina Peters
Jake Snell, Savannah Thias, Amy Winecoff



Mini-Course Outline

Date	Topic	Instructor
Tue. 1/13	Machine Learning Overview and Simple Models	Julian Gold
Wed. 1/14	Model Evaluation and Improving Performance	Julian Gold
Thu. 1/15	Introduction to Neural Networks	Jonathan Hanke
Fri. 1/16	Survey of Neural Network Architectures	Jonathan Hanke

Agenda

- K-nearest Neighbors
 - Regression and classification
- Clustering with K-means
- Evaluation paradigms and improving performance

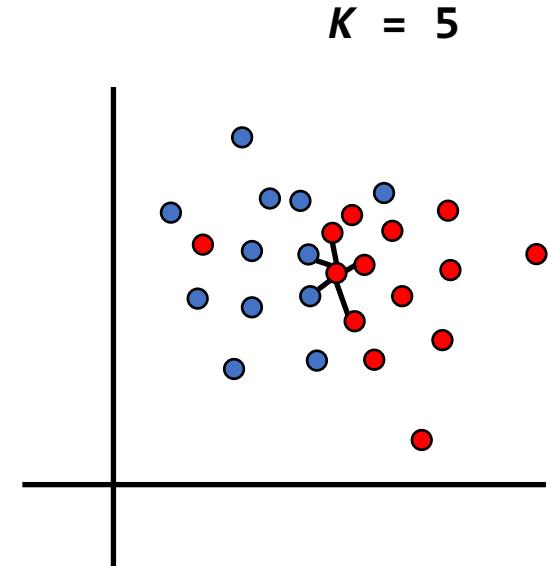
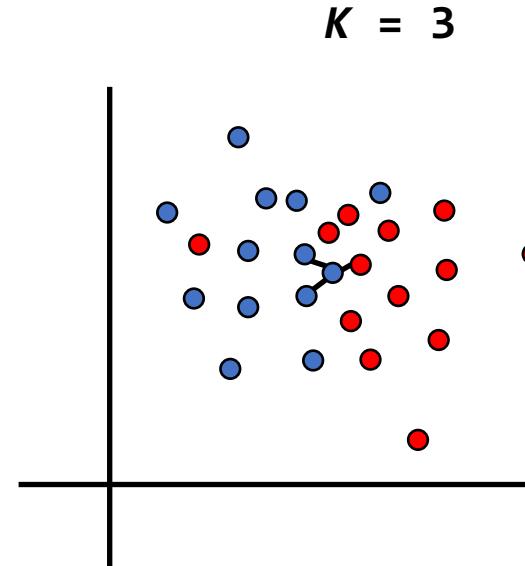
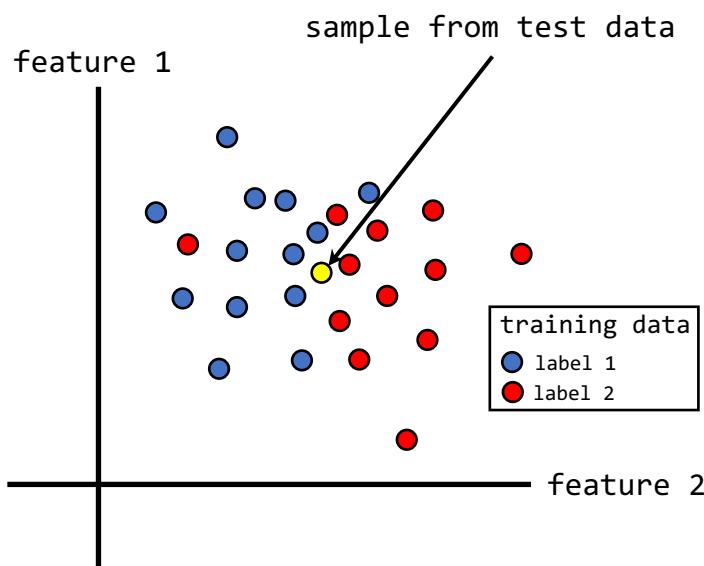


Intro to K -nearest neighbors (KNN)

- simple but powerful
- can be used for classification *or* regression!
- algorithm
 1. for a given test sample (yellow dot), find the K nearest training samples in feature space
 - 2a. for **classification**, assign label by majority vote
 - 2b. for **regression**, assign value by mean of neighbors

K is a tunable parameter!

- choose value that gives better predictions on test data



Coding in Python!

https://github.com/PrincetonUniversity/intro_machine_learning/tree/main/day2

<https://jdh4.github.io/intro-ml>

Google Colab

1. Open notebook (.ipynb):

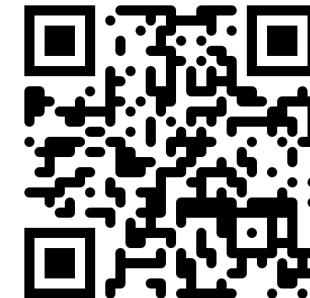


2. Click “Open in Colab”:



JupyterLite

1. Download notebook (.ipynb) from left.
2. Open JupyterLite:



3. Upload the notebook and open:



K-Means Clustering Visualization

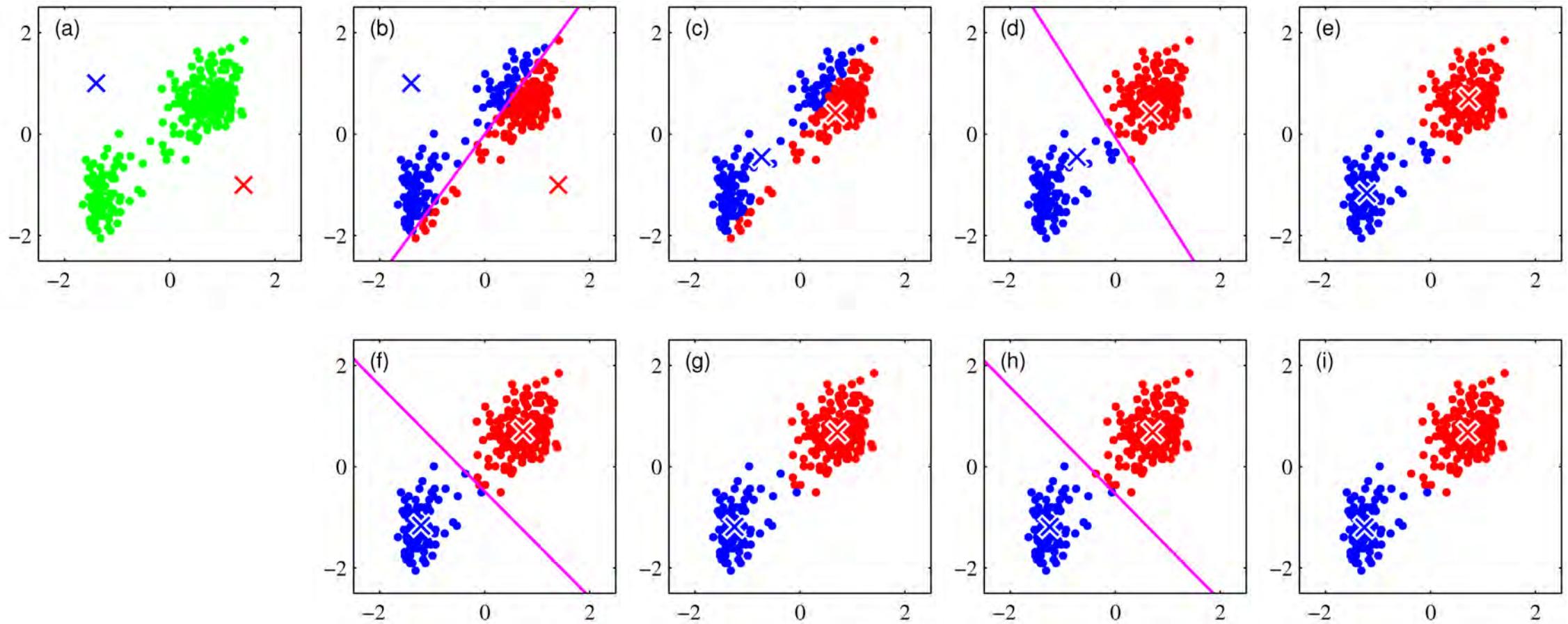


Figure credit: Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*.

Overview of Machine Learning Process

1. Define the problem to be solved.
Datasets? Input features?
Targets? Evaluation metrics?
2. Split the data into train / validation / test.
3. Run the validation loop:
 - a. Choose a set of models.
 - b. Train each model by optimizing its parameters on the training set.
 - c. Evaluate the performance of each model on the validation set.
 - d. Repeat until performance is satisfactory.
4. Evaluate final performance on the test set.

