INTRODUCING UNSUPERVISED LEARNING

Scott O'Hara
Metrowest Developers Machine Learning Group
02/17/2021

REFERENCES

The material for this talk is primarily drawn from the notes, slides and lectures of the courses below:

Applied Machine Learning in Python

University of Michigan, Prof. Kevin Collins Thompson

https://www.coursera.org/learn/python-machine-learning/home/welcome

Machine Learning: Clustering & Retrieval

University of Washington, Profs. Emily Fox & Carlos Guestrin

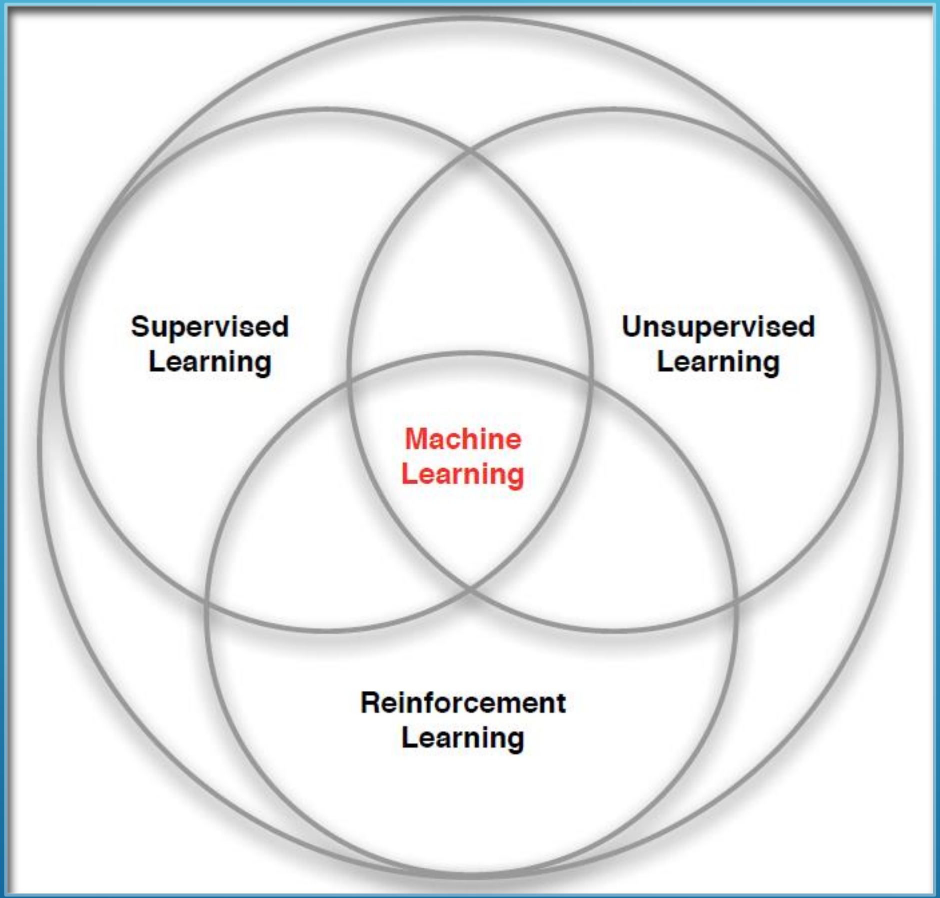
https://www.coursera.org/learn/ml-regression/home/welcome

The Hundred-Page Machine Learning Book (Ch. 9)

Andriy Burkov

http://themlbook.com/

3 TYPES OF MACHINE LEARNING



Supervised Learning – Learn a function from labeled data that maps input attributes to an output label e.g., linear regression, decision trees, SVMs.

Unsupervised Learning – Learn patterns in unlabeled data e.g., principle component analysis or clustering algorithms such as K-means, HAC, or Gaussian mixture models.

Reinforcement Learning – An agent learns to maximize rewards while acting in an uncertain environment.

Credit: image borrowed from lecture slides David Silver, DeepMind, "Introduction to Reinforcement Learning

WHAT IS UNSUPERVISED LEARNING?

- Unsupervised learning involves tasks that operate on datasets without labeled responses or target values.
- The goal is to discover interesting structure or information in the dataset.

Reference: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

APPLICATIONS OF UNSUPERVISED LEARNING

- Visualize structure of a complex dataset.
- Density estimation to predict probabilities of events.
- Compress and summarize the data.
- Extract features for supervised learning.
- Discover important clusters or outliers.

Reference: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

FOUR KINDS OF UNSUPERVISED LEARNING

1. **Density Estimation**
   - Model the probability density function of the unknown probability distribution from which the dataset has been drawn.
2. **Dimensionality Reduction**
   - Finds an approximate version of a dataset using fewer features while retaining some meaningful properties of the original data.
3. **Outlier Detection**
   - Detect the examples in the dataset that are very different from what a typical example in the dataset looks like.
4. **Clustering**
   - The task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more like each other than to those in other groups (clusters).

References:

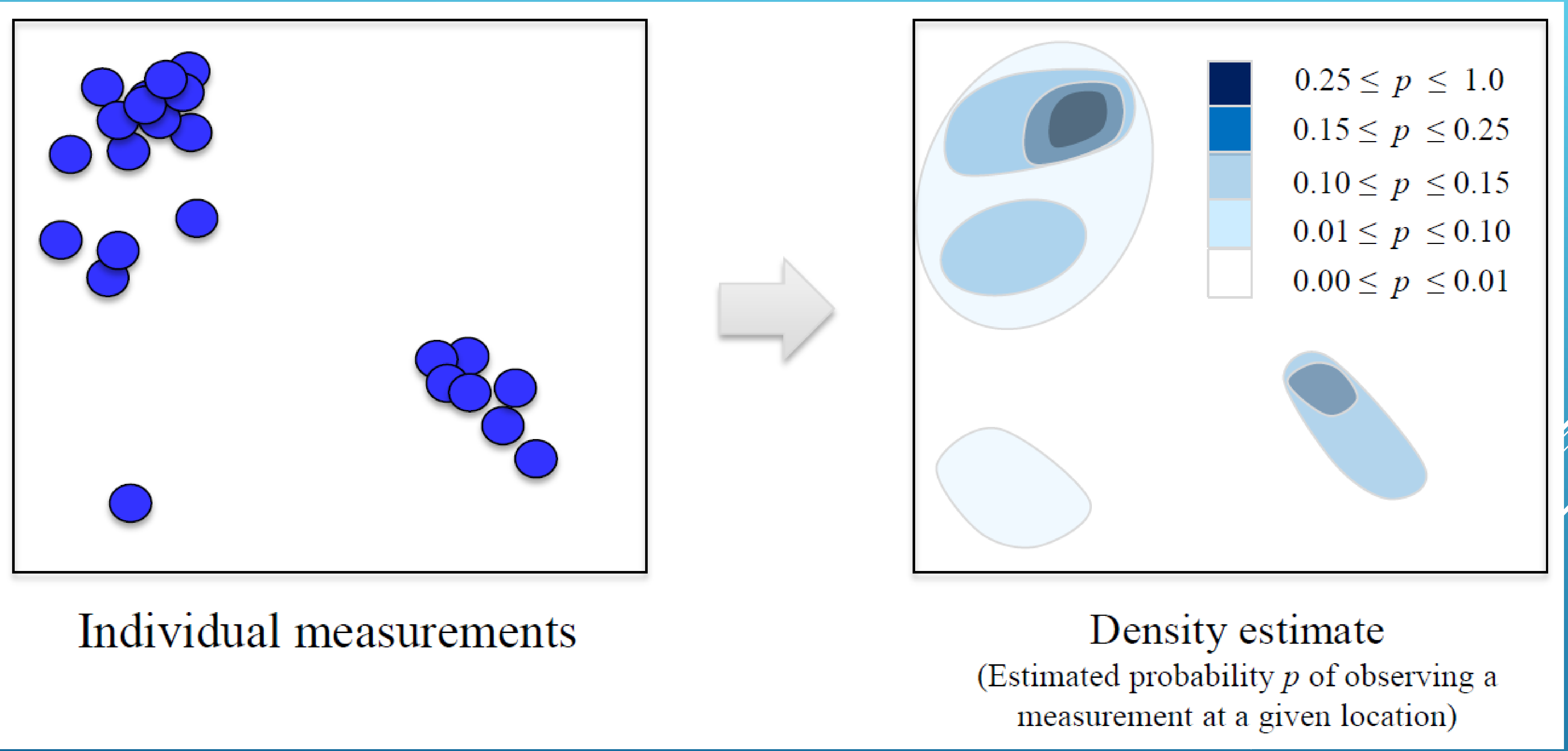
The Hundred-Page Machine Learning Book. Andriy Burkov.

Applied Machine Learning in Python. Coursera. University of Michigan, Prof. Kevin Collins Thompson

Cluster analysis, https://en.wikipedia.org/w/index.php?title=Cluster_analysis&oldid=1002271612 (last visited Jan. 27, 2021).

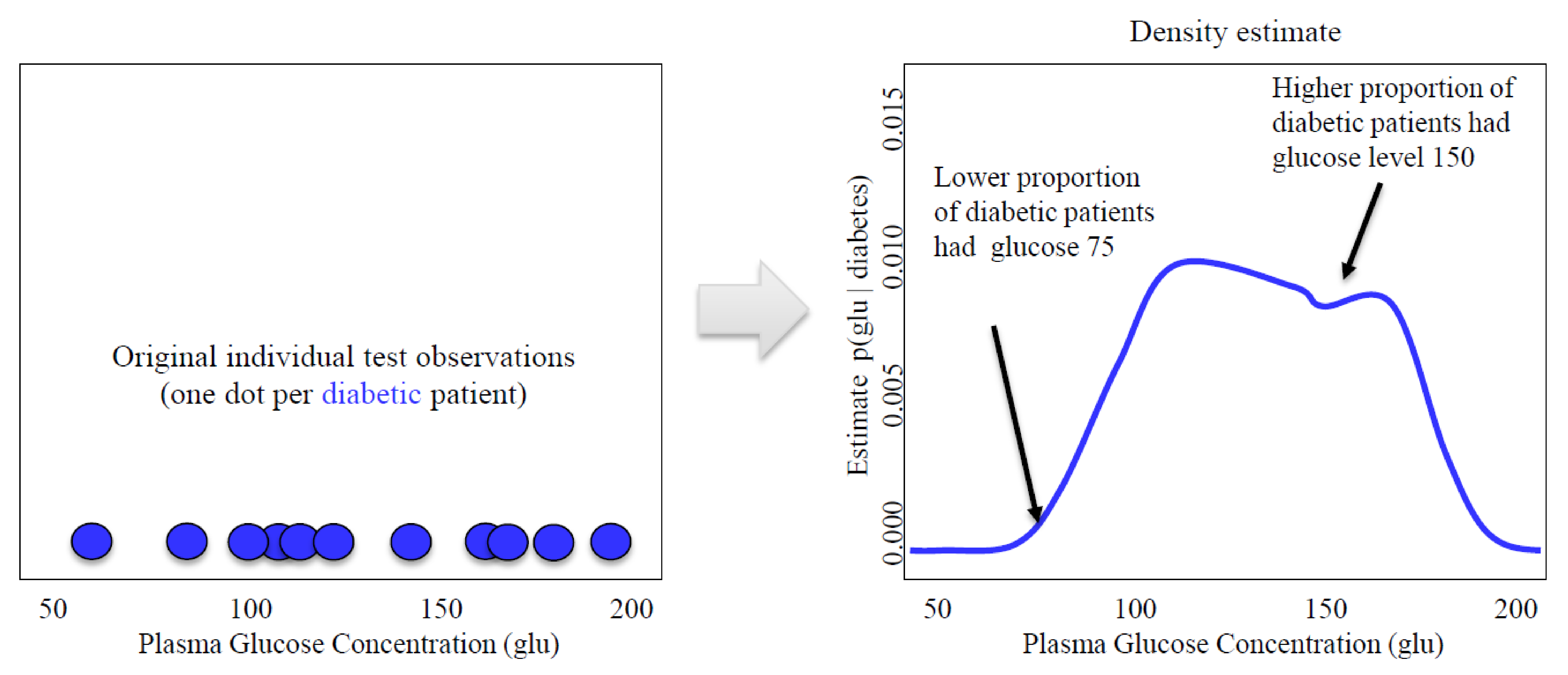
Dimensionality reduction, https://en.wikipedia.org/w/index.php?title=Dimensionality_reduction&oldid=1002754996 (last visited Jan. 27, 2021).

DENSITY ESTIMATION



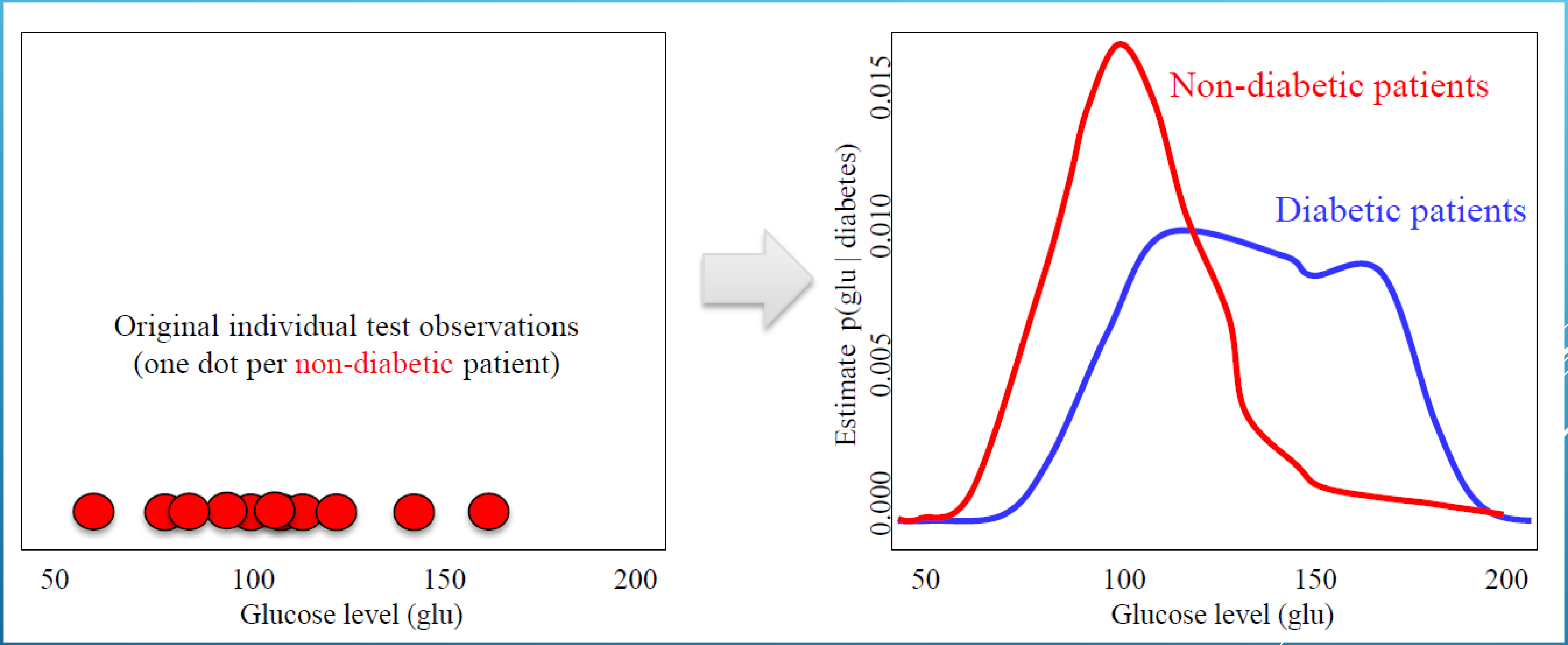
Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

DENSITY ESTIMATION EXAMPLE



Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

DENSITY ESTIMATION EXAMPLE



Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

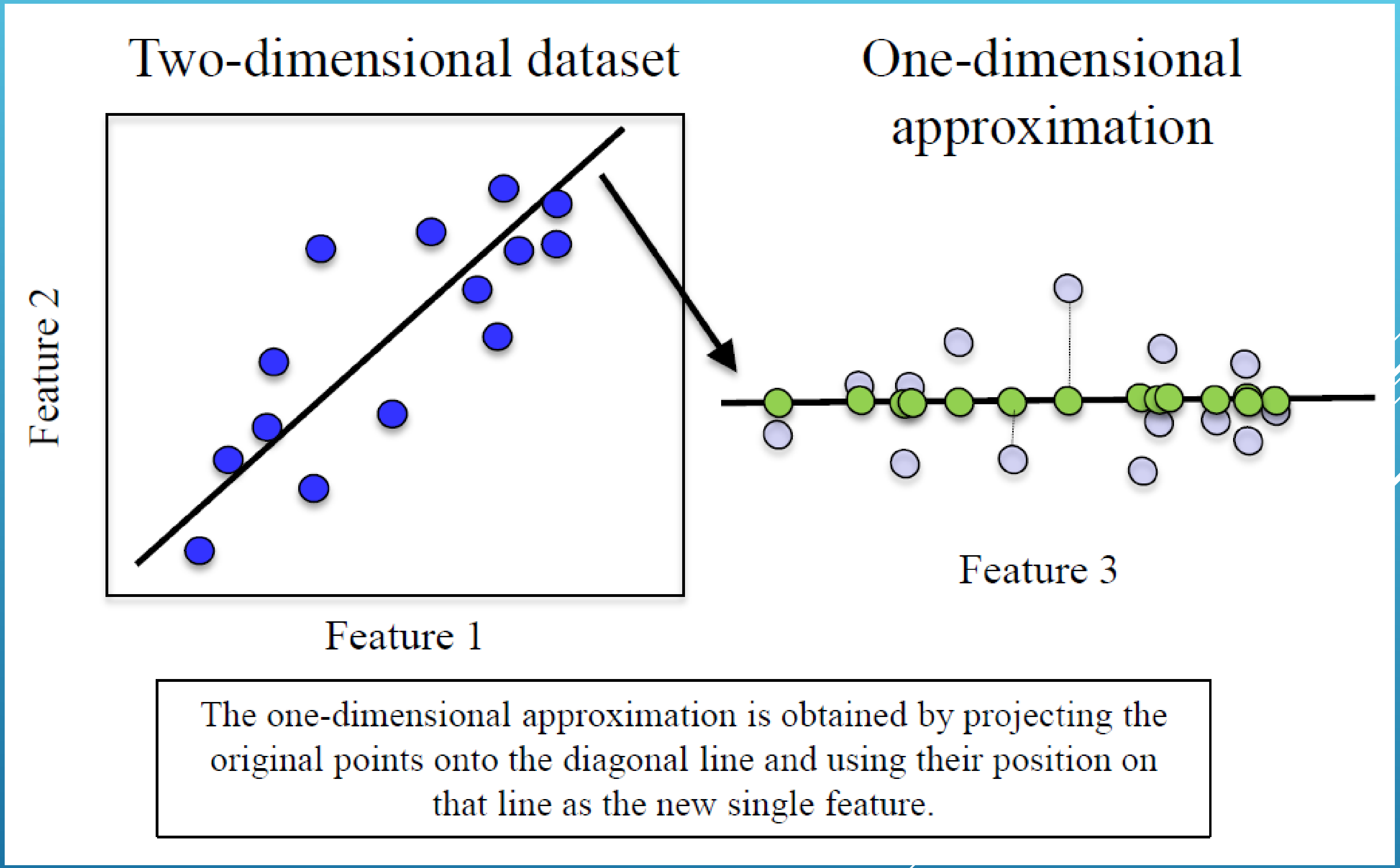
DENSITY ESTIMATION TECHNIQUES

- Histograms
- Kernel Density Estimation
- etc.

Source: *Density estimation, https://en.wikipedia.org/w/index.php?title=Density_estimation&oldid=991862330 (last visited Jan. 27, 2021).*

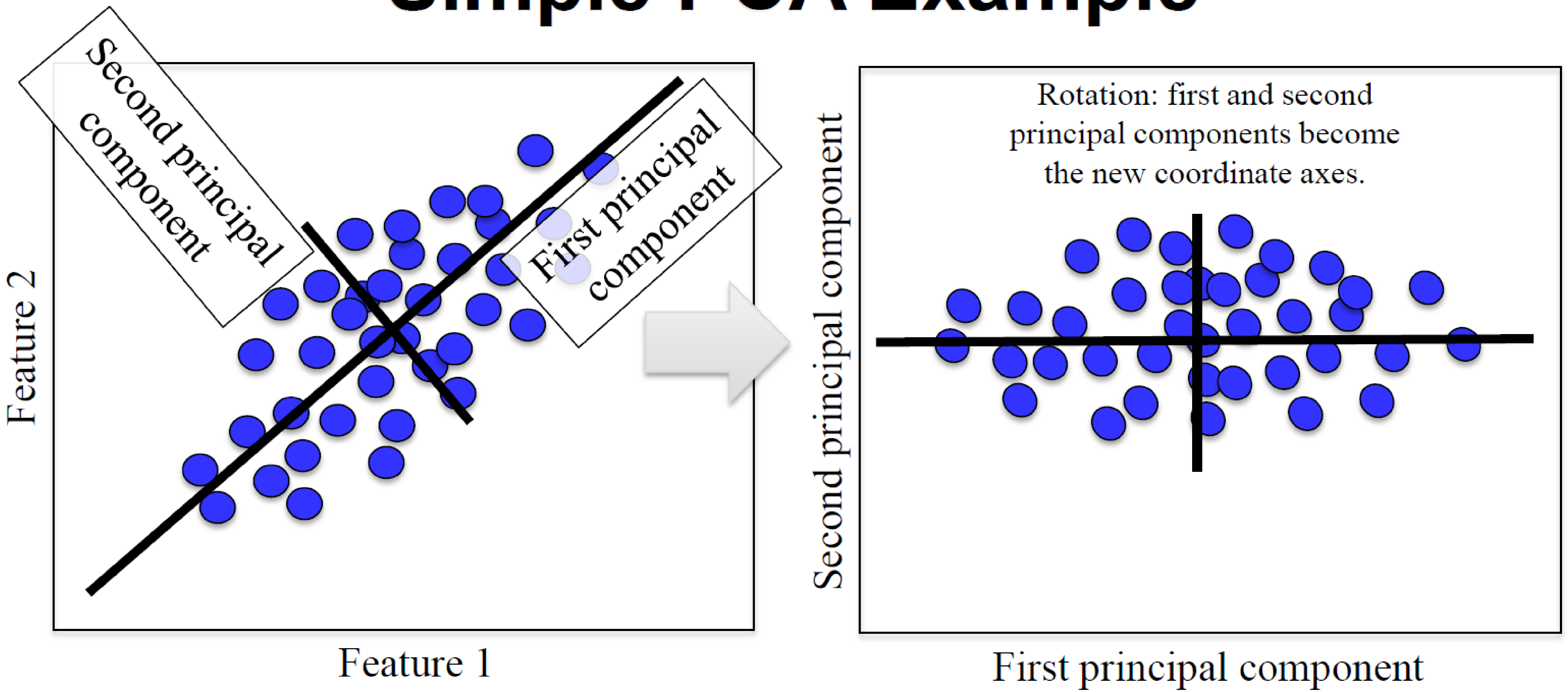
DIMENSIONALITY REDUCTION

- Finds an approximate version of your dataset using fewer features.
- Used for exploring and visualizing a dataset to understand grouping or relationships
- Often visualized using a 2-dimensional scatterplot
- Also used for compression, finding features for supervised learning



Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

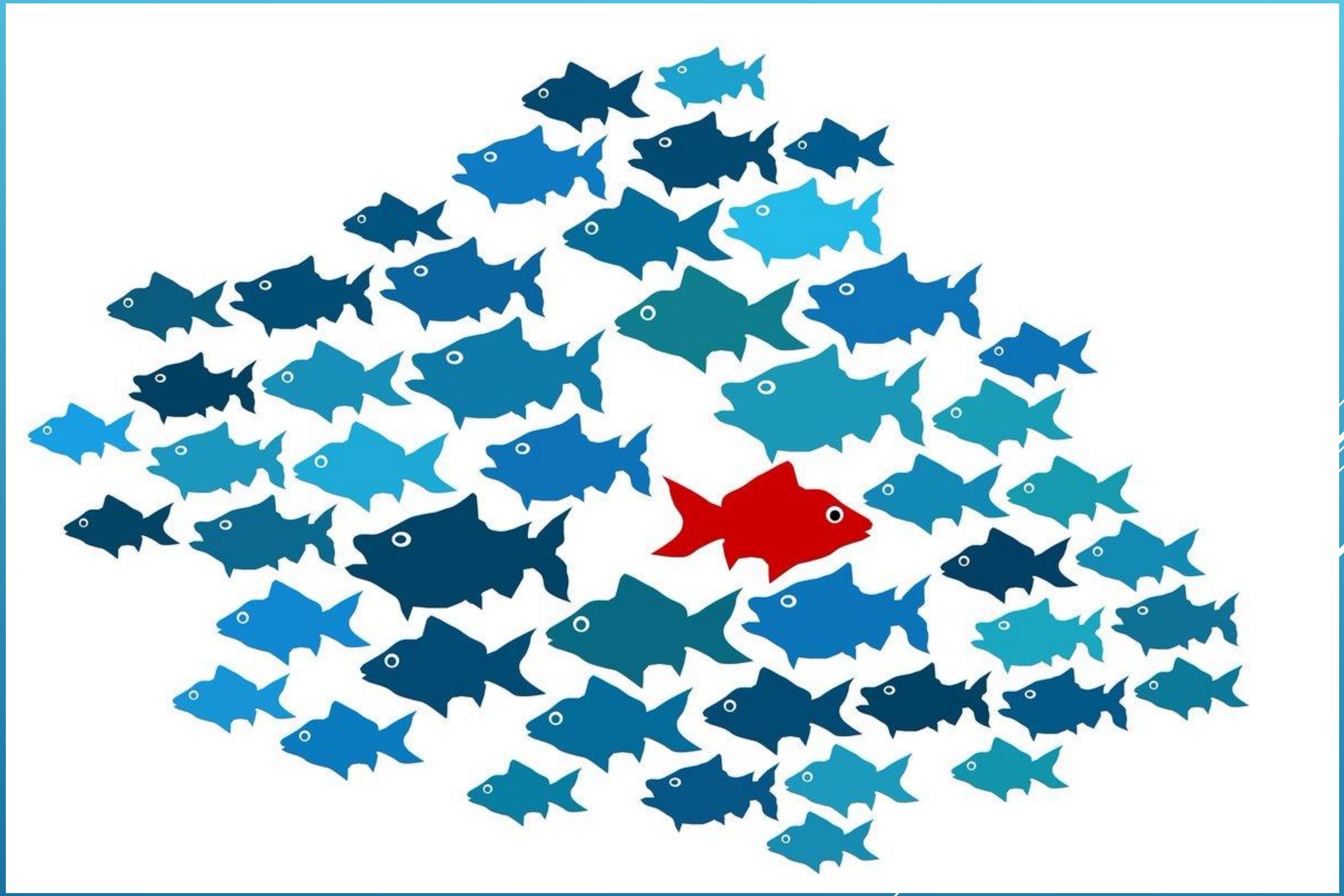
Simple PCA Example



Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

OUTLIER DETECTION

Outlier detection is the problem of detecting the examples in the dataset that are very different from what a typical example in the dataset looks like.



Reference: The Hundred-Page Machine Learning Book. Ch 9. Andriy Burkov.
Image from: A Brief Overview of Outlier Detection Techniques, https://towardsdatascience.com/a-brief-overview-of-outlier-detection-techniques-1e0b2c19e561 (last visited Jan. 27, 2021).

OUTLIER DETECTION ALGORITHMS

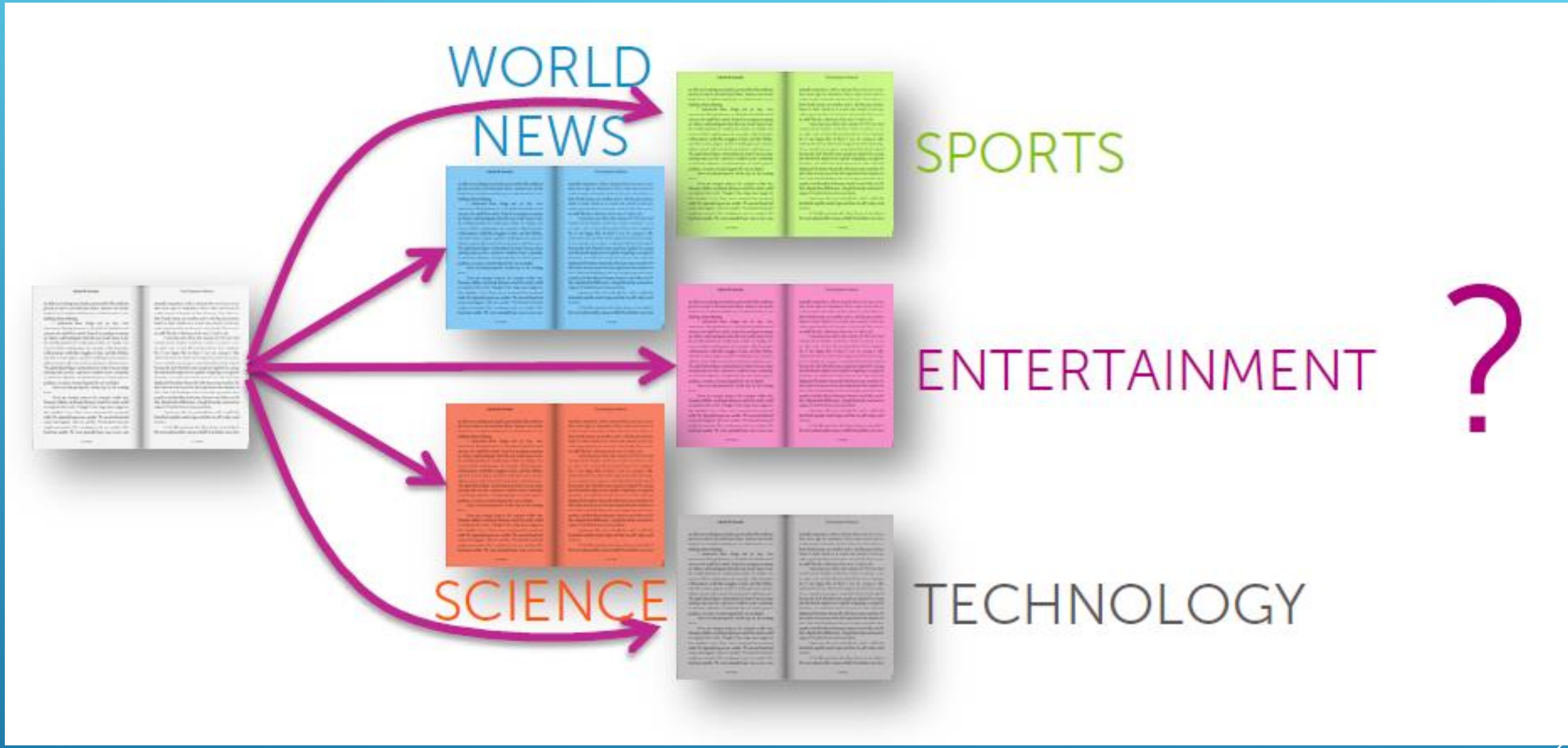
- Many outlier detection algorithms are adapted from more general clustering algorithms.

- One-Class Algorithms:
  - One-class gaussian
  - One-class mixture of gaussians.
  - One-class K-means
  - One-class KNN
  - One-class SVM

- Autoencoder

Source: *Density estimation, https://en.wikipedia.org/w/index.php?title=Density_estimation&oldid=991862330 (last visited Jan. 27, 2021).*

A SUPERVISED LEARNING TASK



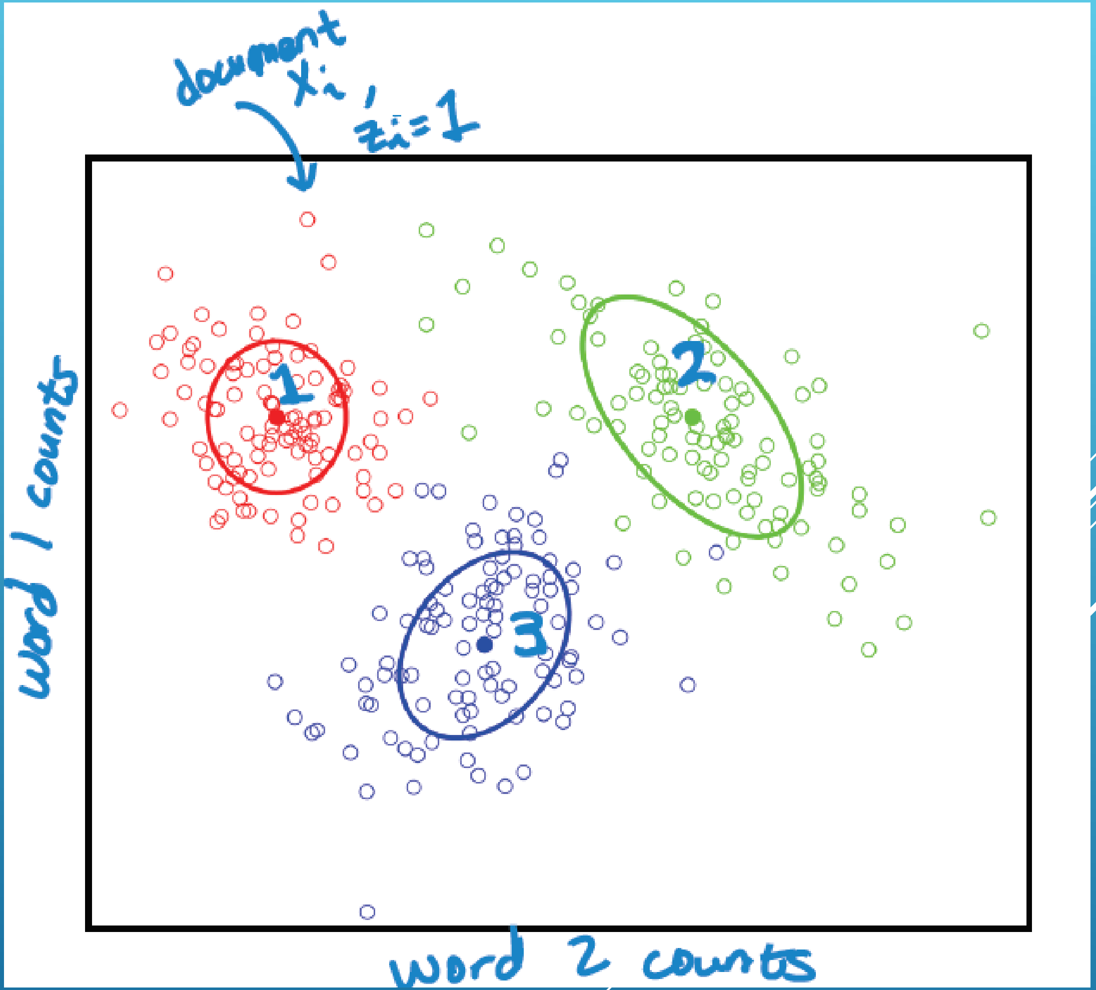
But what if you don't have labels?

Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

AN UNSUPERVISED LEARNING TASK

No labels provided
...uncover cluster structure from input alone

Input: docs as vectors $\mathbf{x}_i$
Output: cluster labels z_i



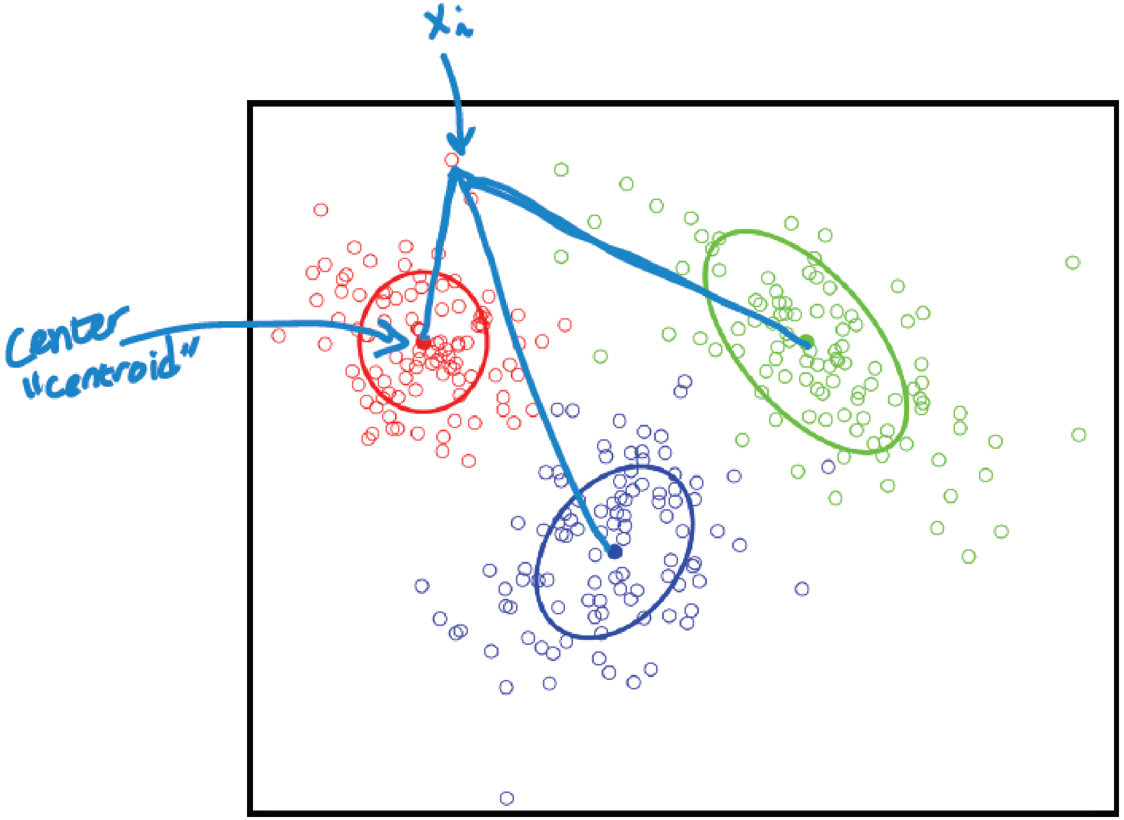
Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

What defines a cluster?

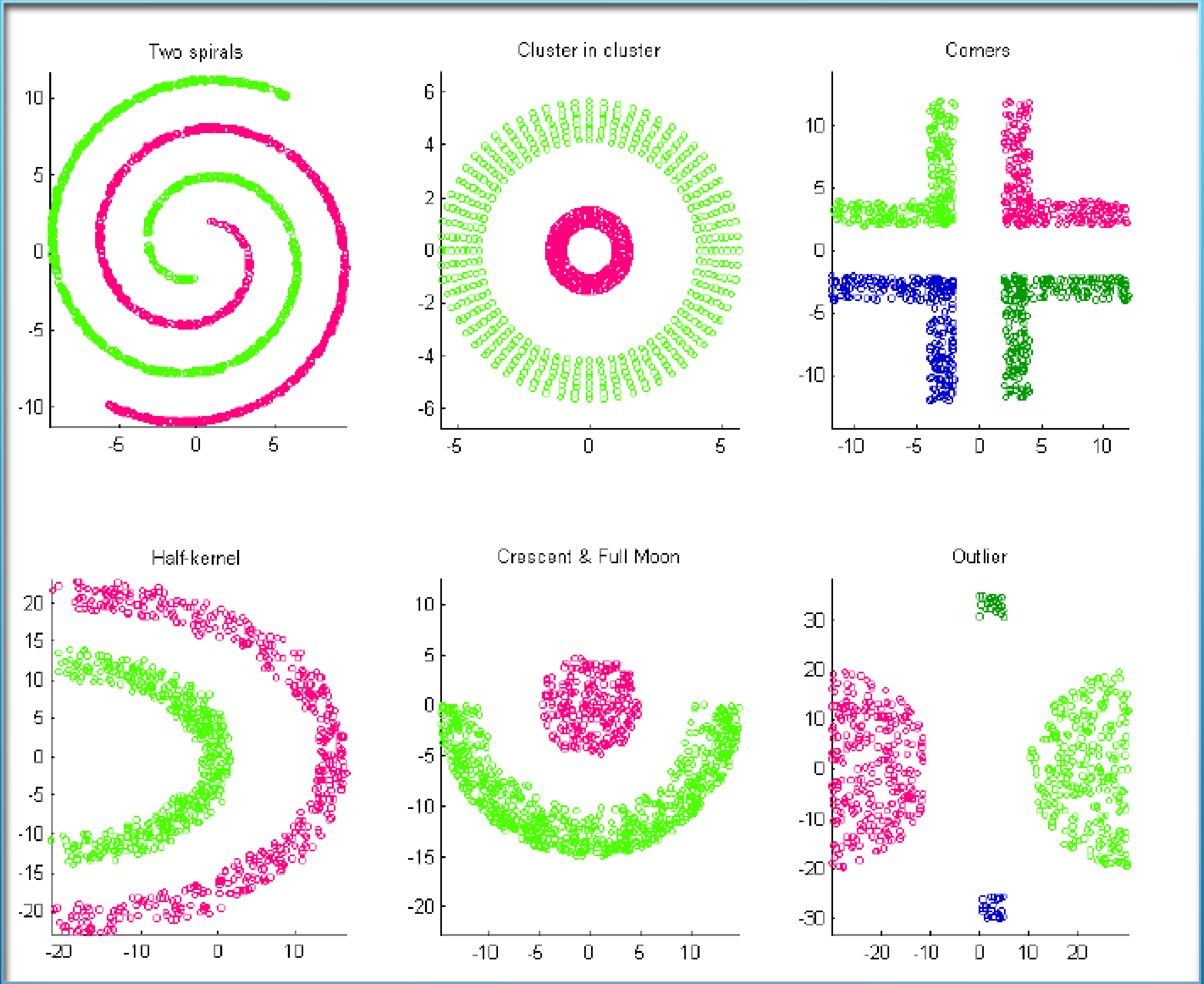
Cluster defined by
center & shape/spread

Assign observation $\mathbf{x}_i$ (doc) to cluster k (topic label) if

- Score under cluster k is higher than under others
- For simplicity, often define score as distance to cluster center (ignoring shape)

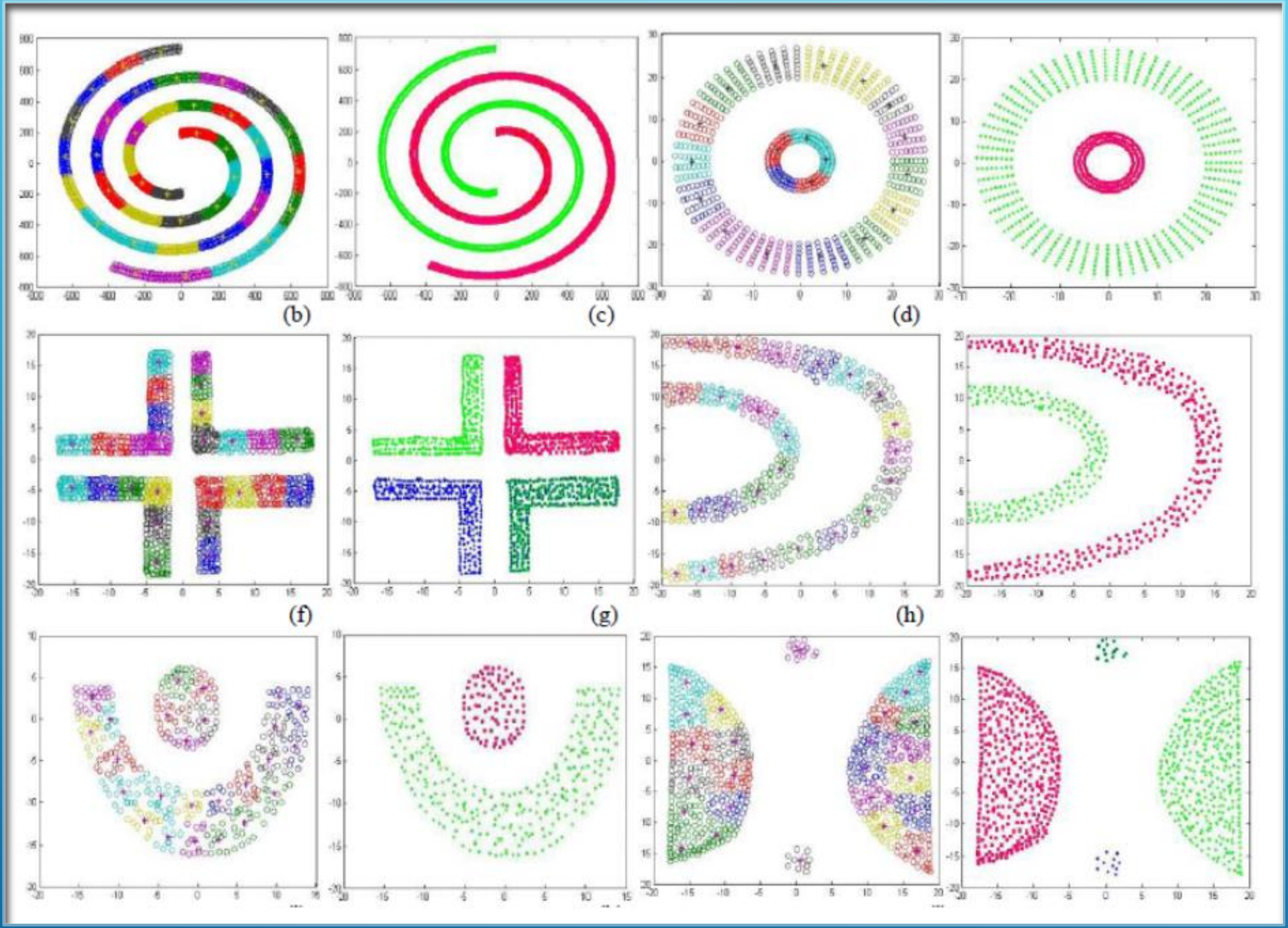


Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin



CHALLENGING CLUSTERS

Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin



MORE CHALLENGING CLUSTERS

Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

K-means Clustering

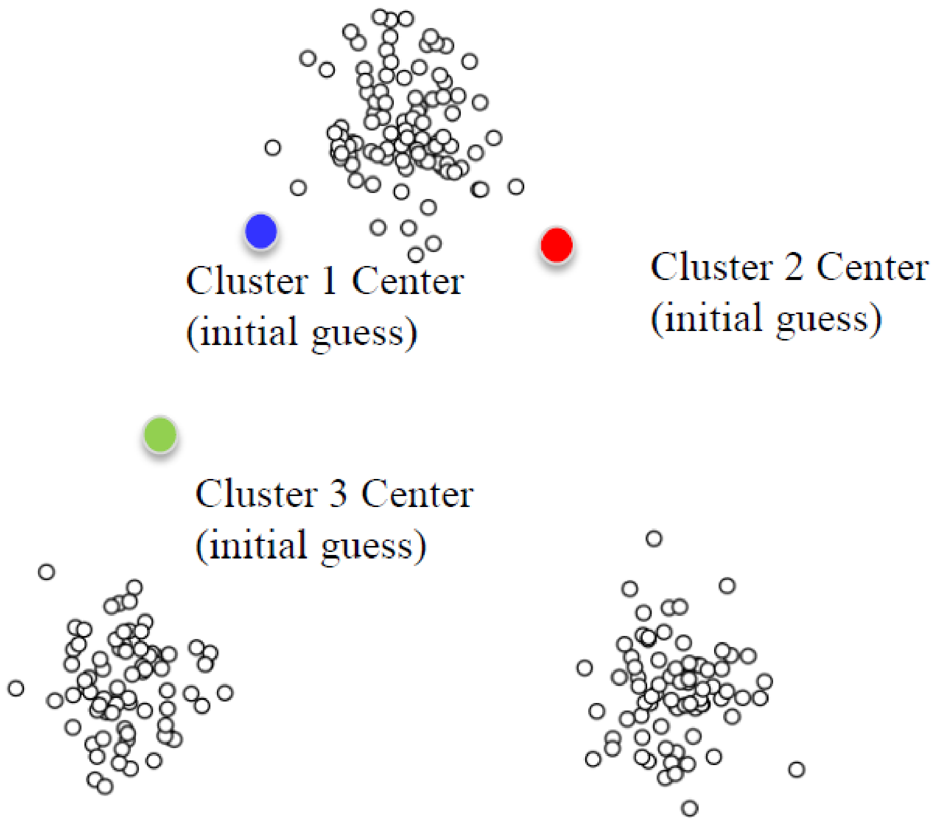
The k-means algorithm

Initialization Pick the number of clusters *k* you want to find. Then pick *k random* points to serve as an initial guess for the cluster centers.

Step A Assign each data point to the nearest cluster center.

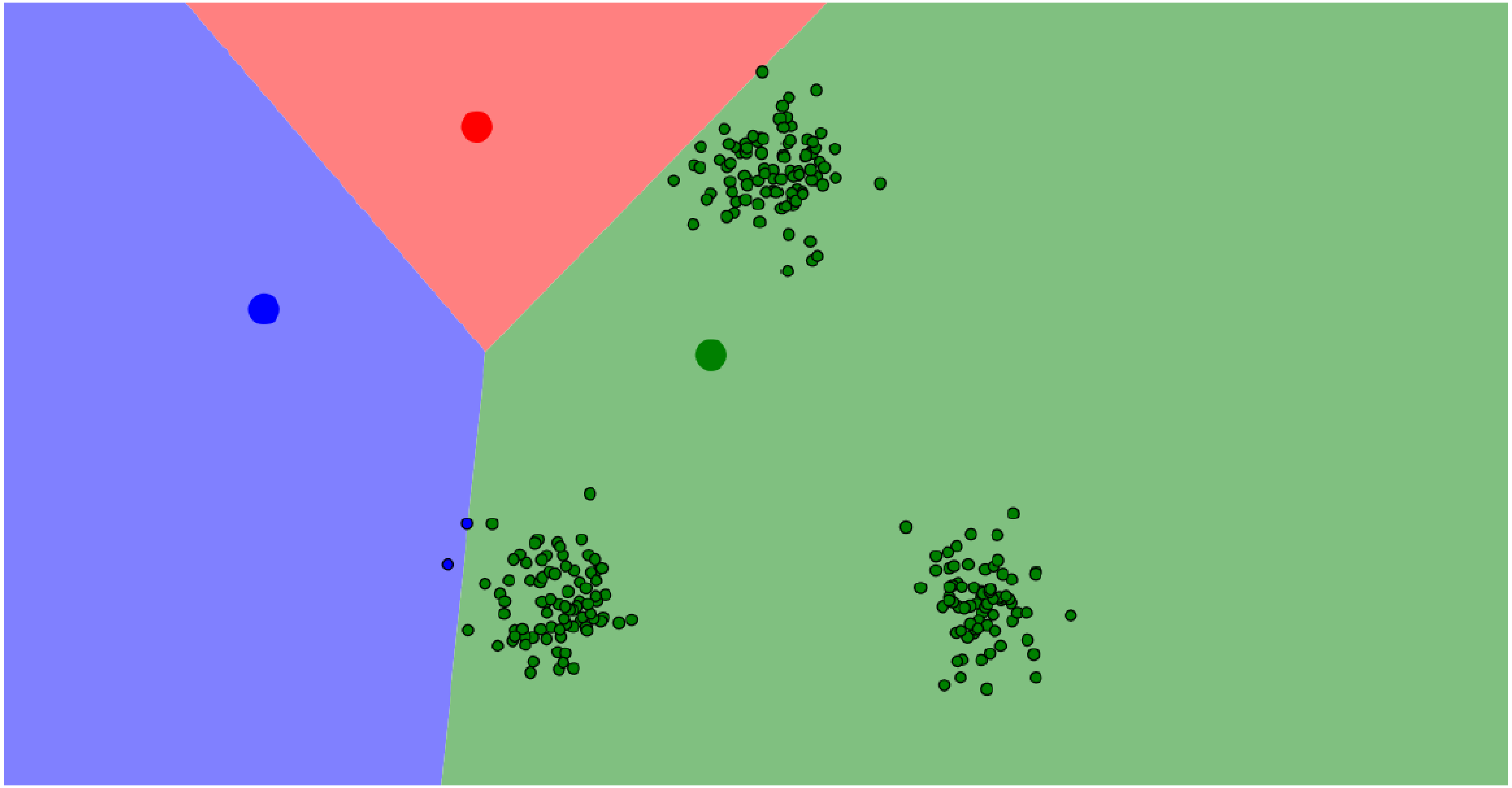
Step B Update each cluster center by replacing it with the mean of all points assigned to that cluster (in step A).

Repeat steps A and B until the centers converge to a stable solution.



Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

K-means Example: Step 1A

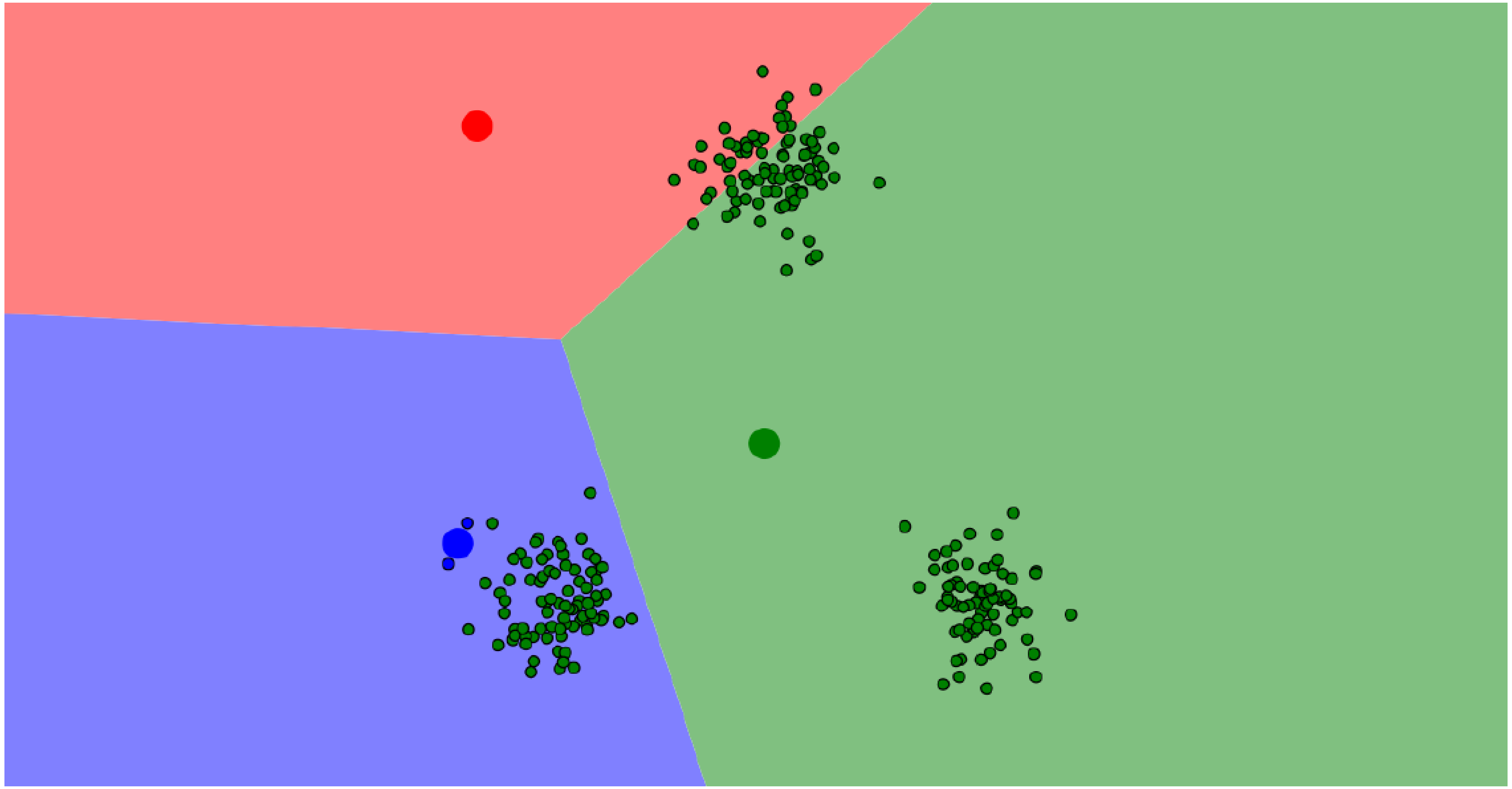


We want three clusters, so three centers are chosen randomly.

Data points are colored according to the closest center.

Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

K-means Example: Step 1B

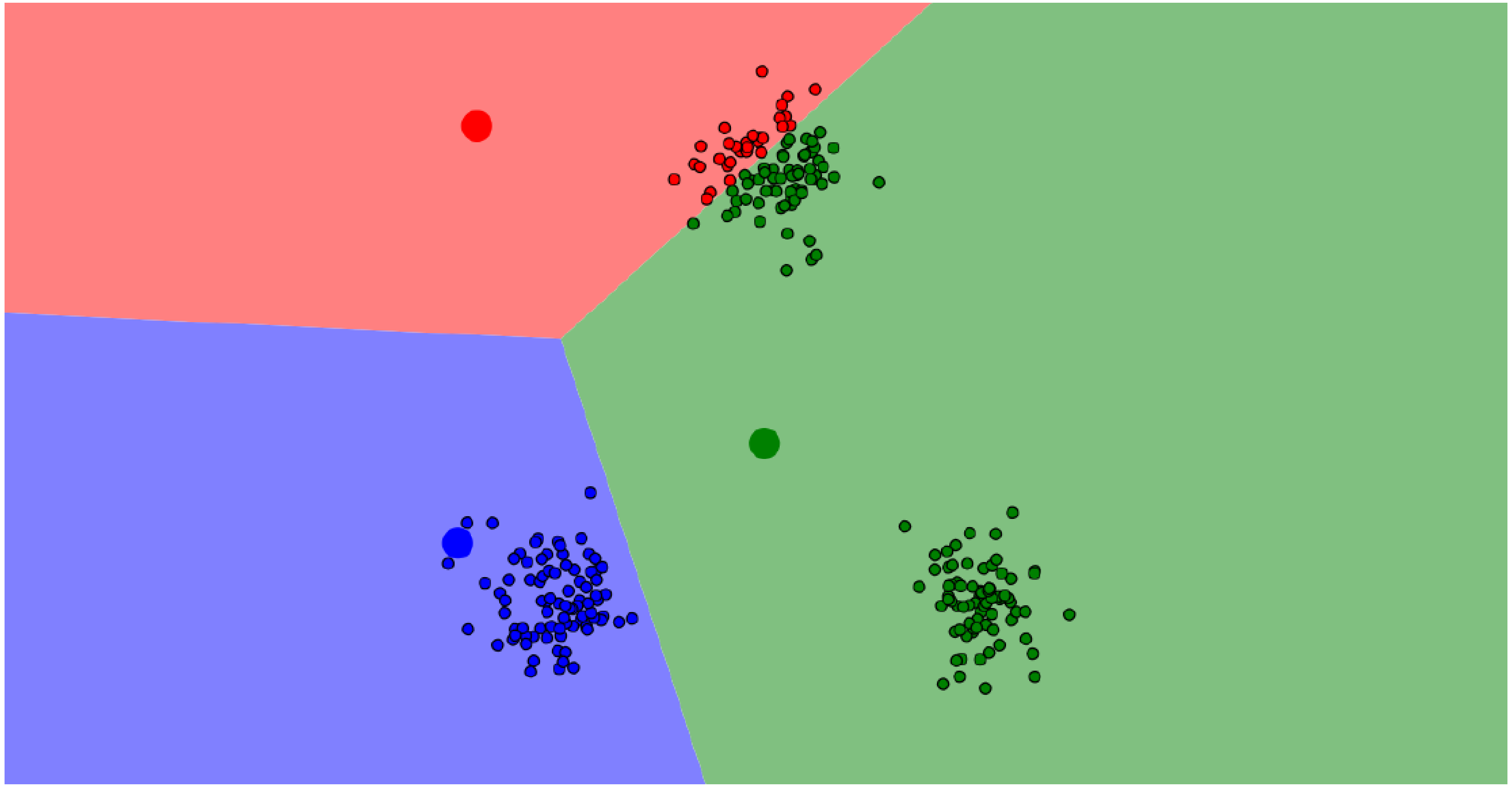


Each center is then updated…

… using the mean of all points assigned to that cluster.

Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

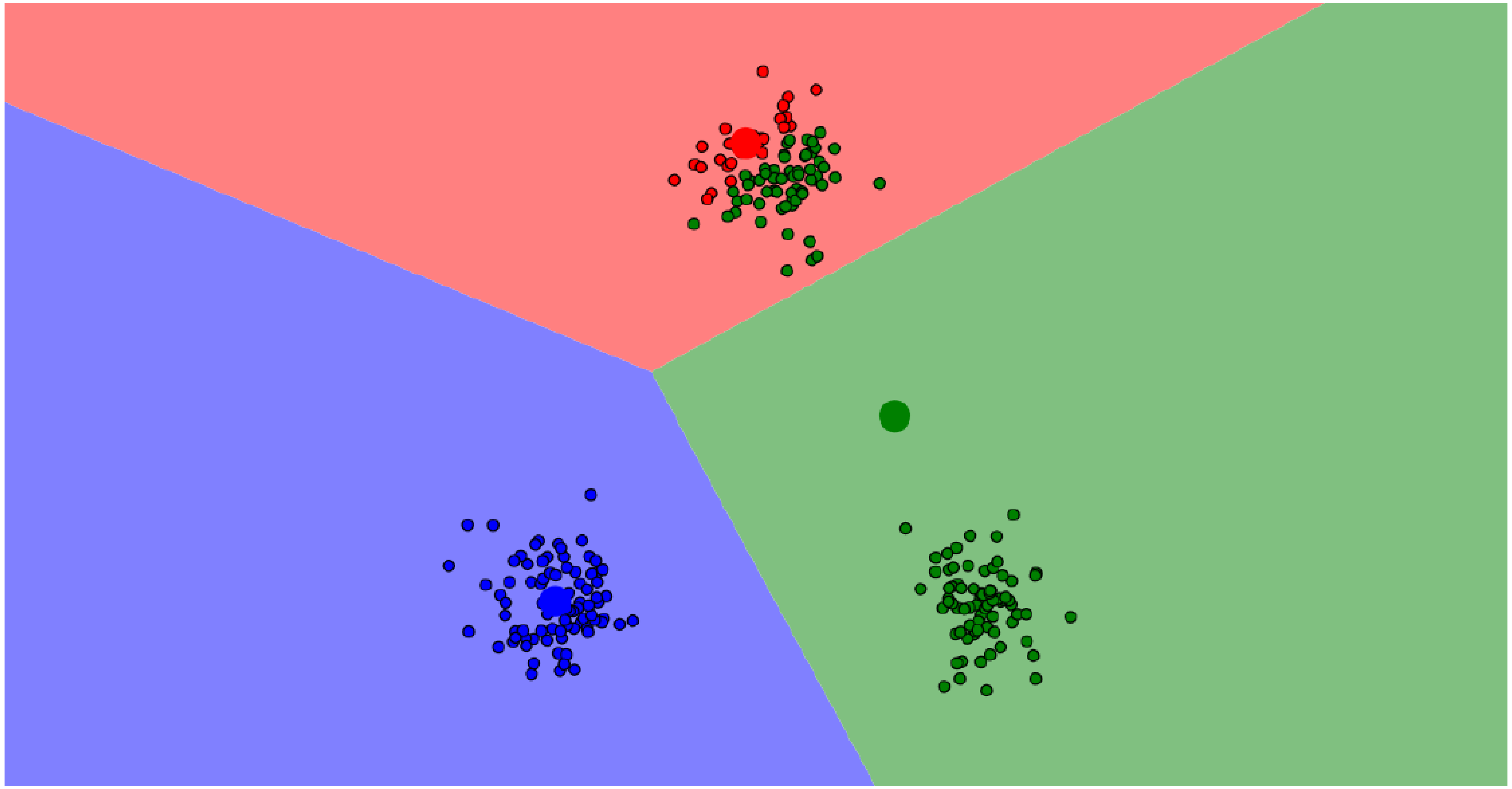
K-means Example: Step 2A



Data points are colored (again) according to the closest center.

Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

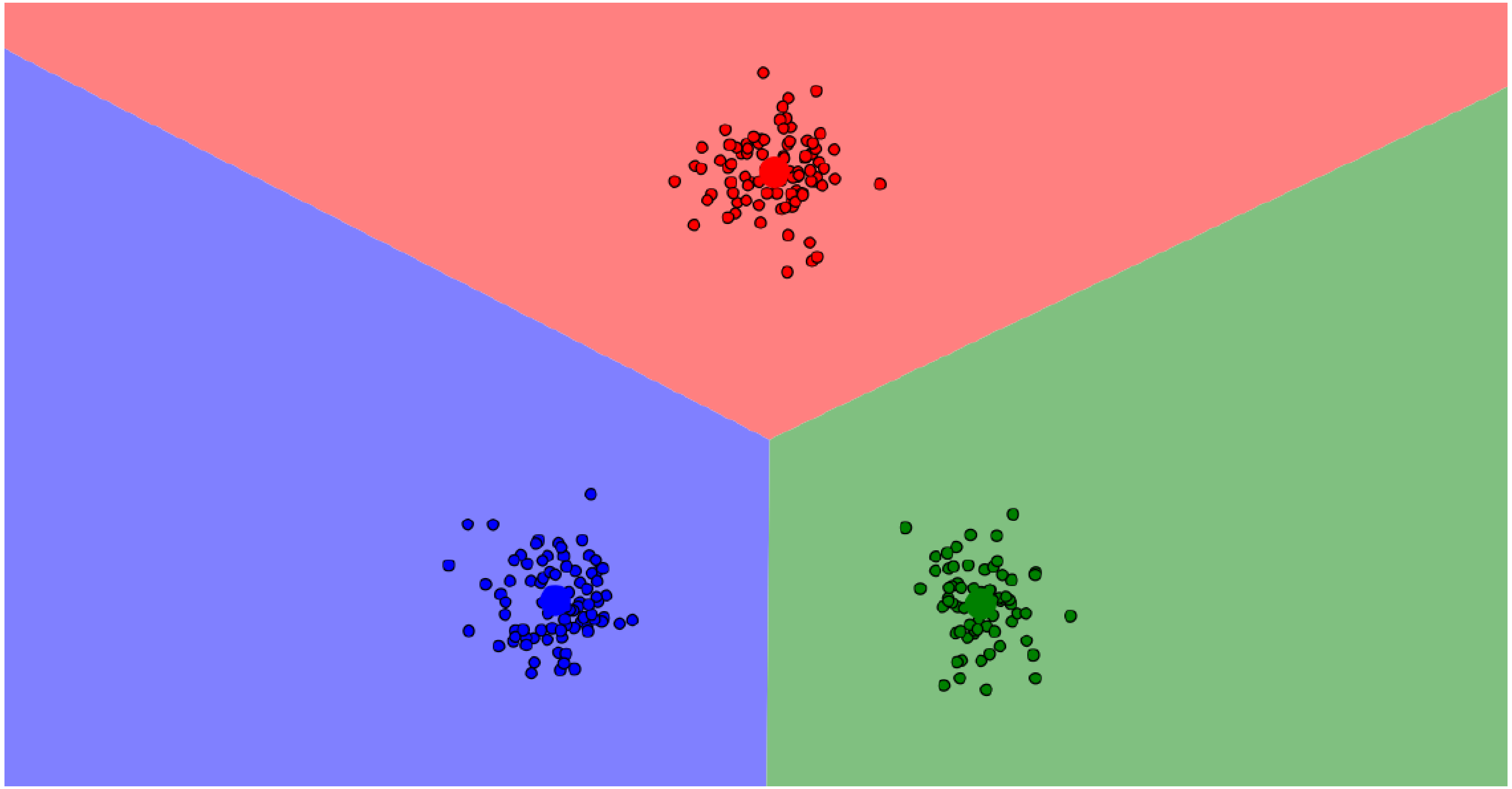
K-means Example: Step 2B



Re-calculate all cluster centers.

Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

K-means Example: Converged



After repeating these steps for several more iterations…

The centers converge to a stable solution!

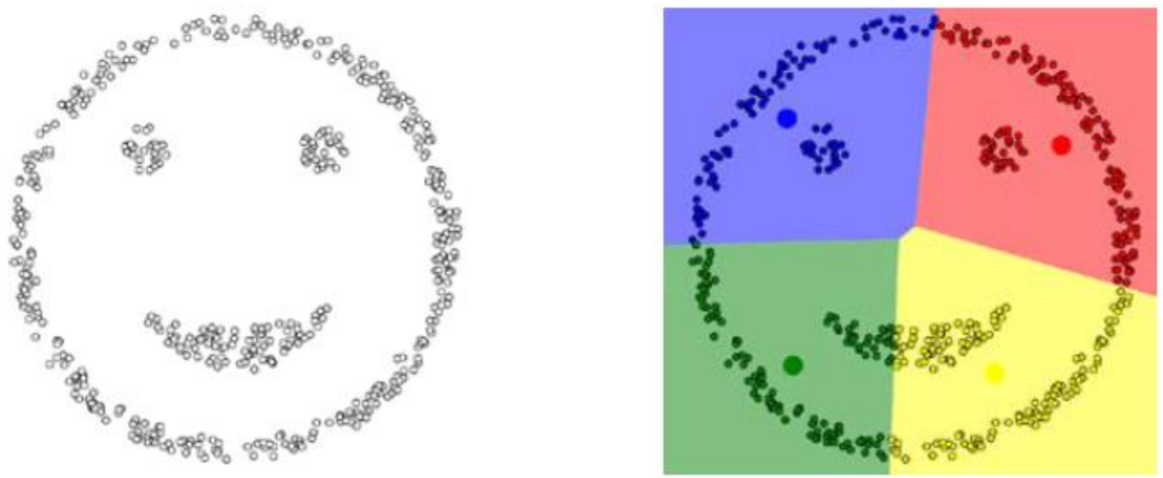
These centers define the final clusters.

Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

LIMITATIONS OF K-MEANS

- Works well for simple clusters that are same size, well-separated, globular shapes.

- Does not do well with irregular, complex clusters.



K-means typically performs poorly with data having complex, irregular clusters.

Source: *Applied Machine Learning in Python*. Coursera. University of Michigan, Prof. Kevin Collins Thompson

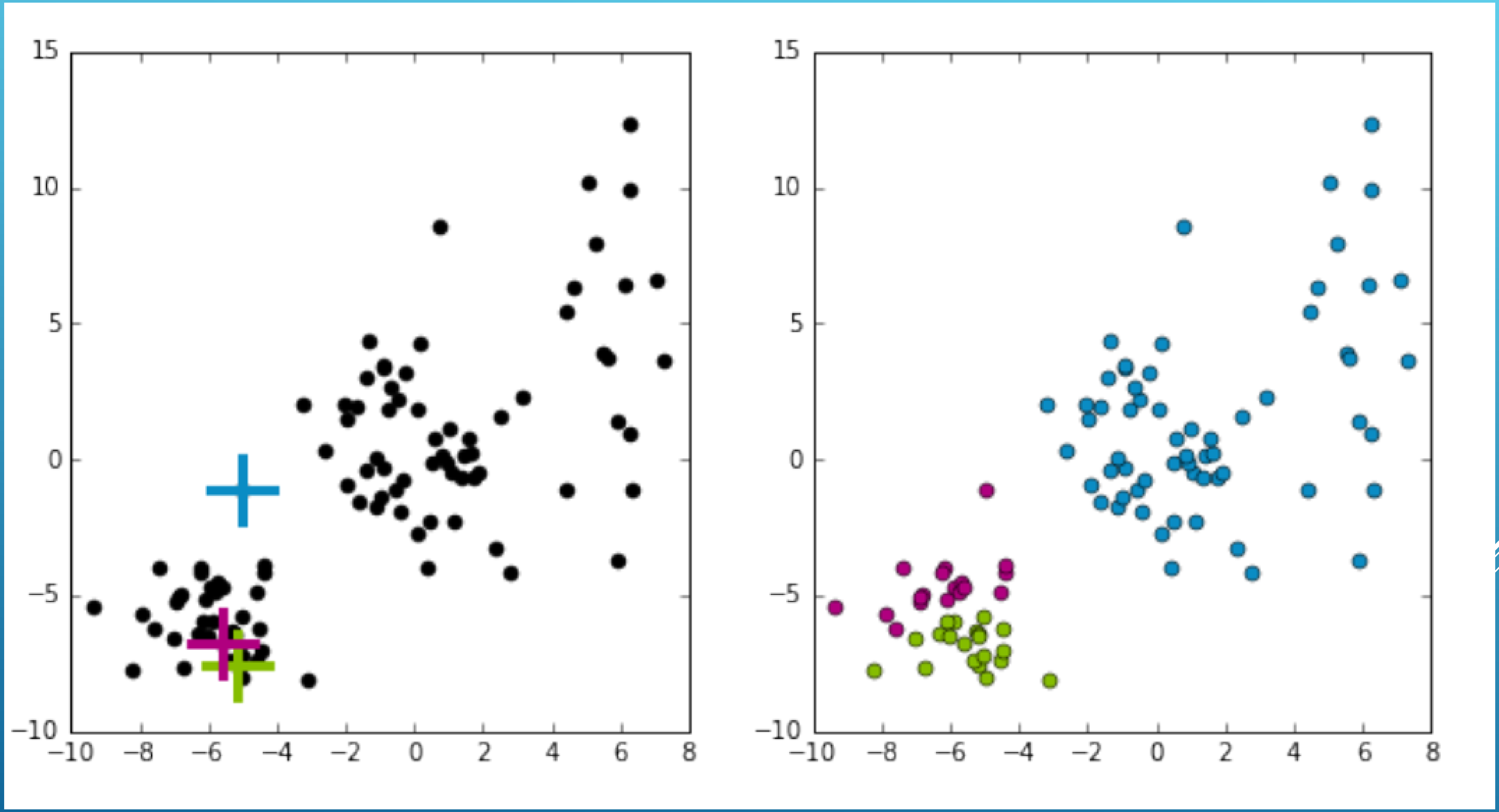
Converges to:

- ~~Global optimum~~
- Local optimum
- ~~neither~~

CONVERGENCE OF K-MEANS

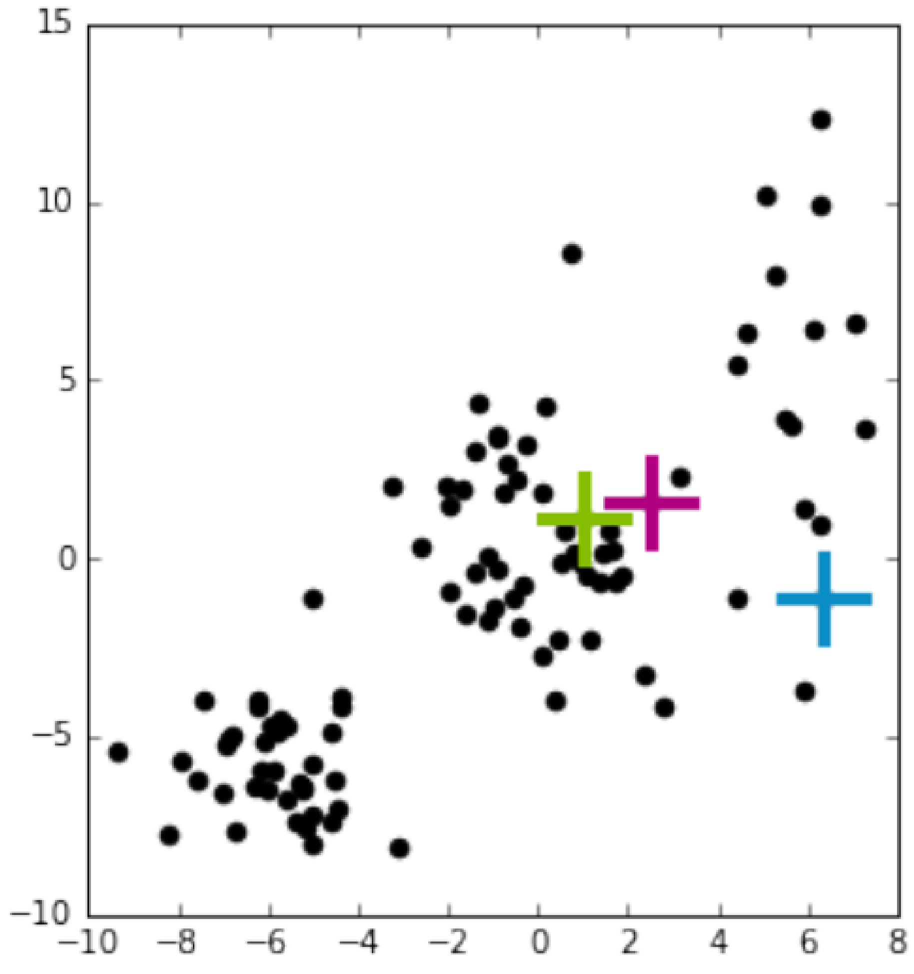
Reference: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

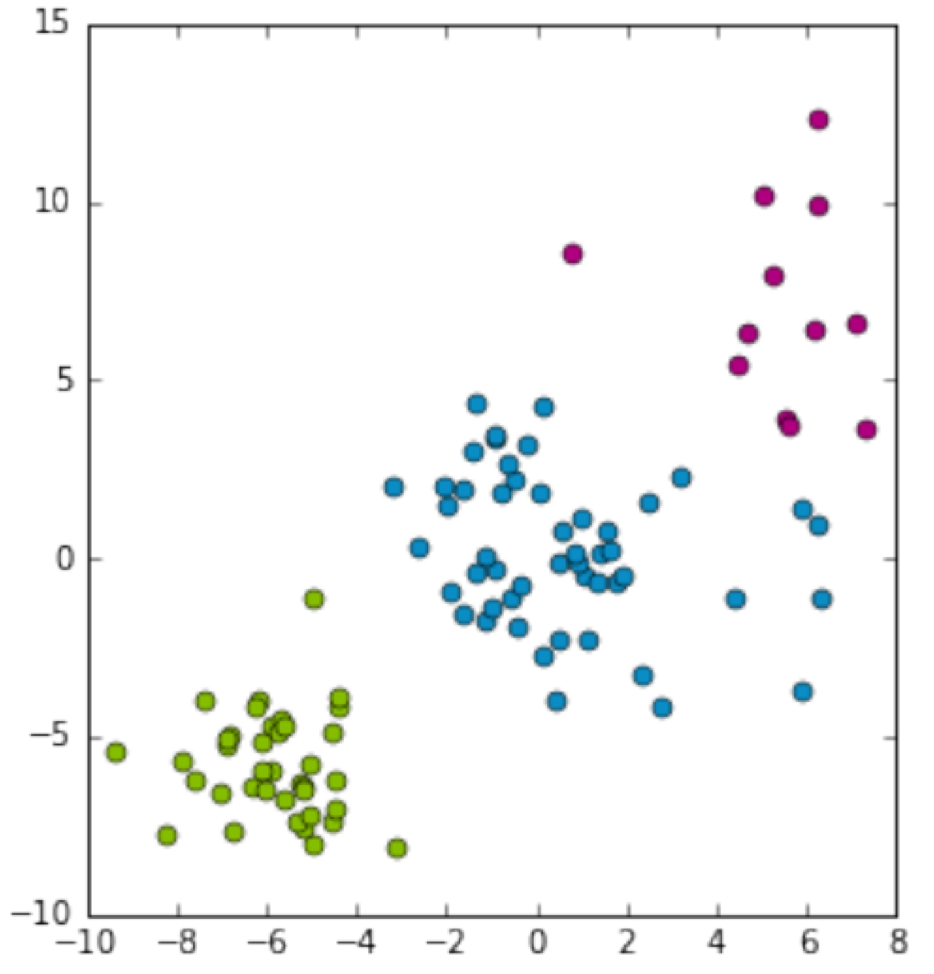
CONVERGENCE OF K-MEANS TO LOCAL OPTIMUM



Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

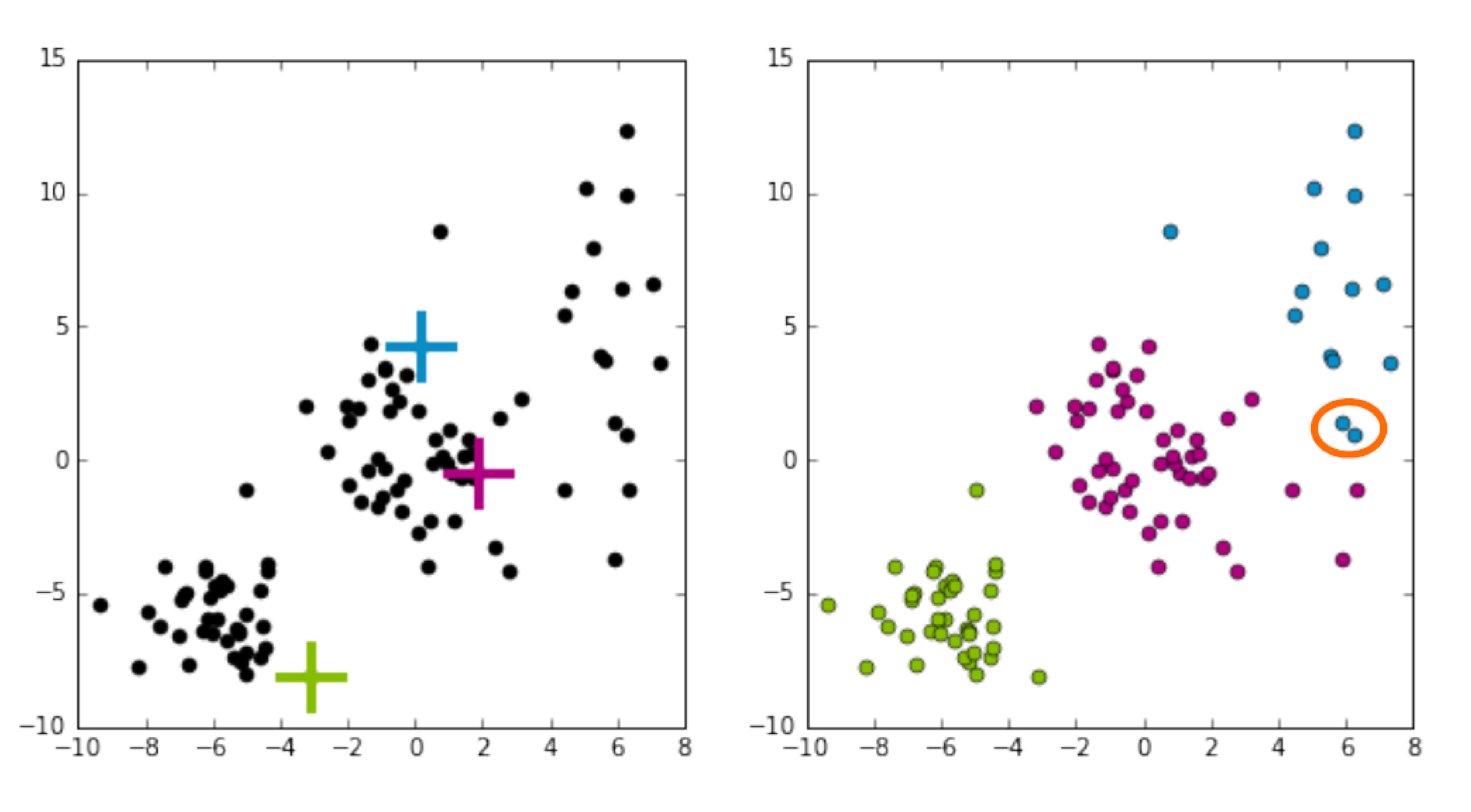
CONVERGENCE OF K-MEANS TO LOCAL OPTIMUM





Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

CONVERGENCE OF K-MEANS TO LOCAL OPTIMUM



Source: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

Smart Initialization with K-Means++

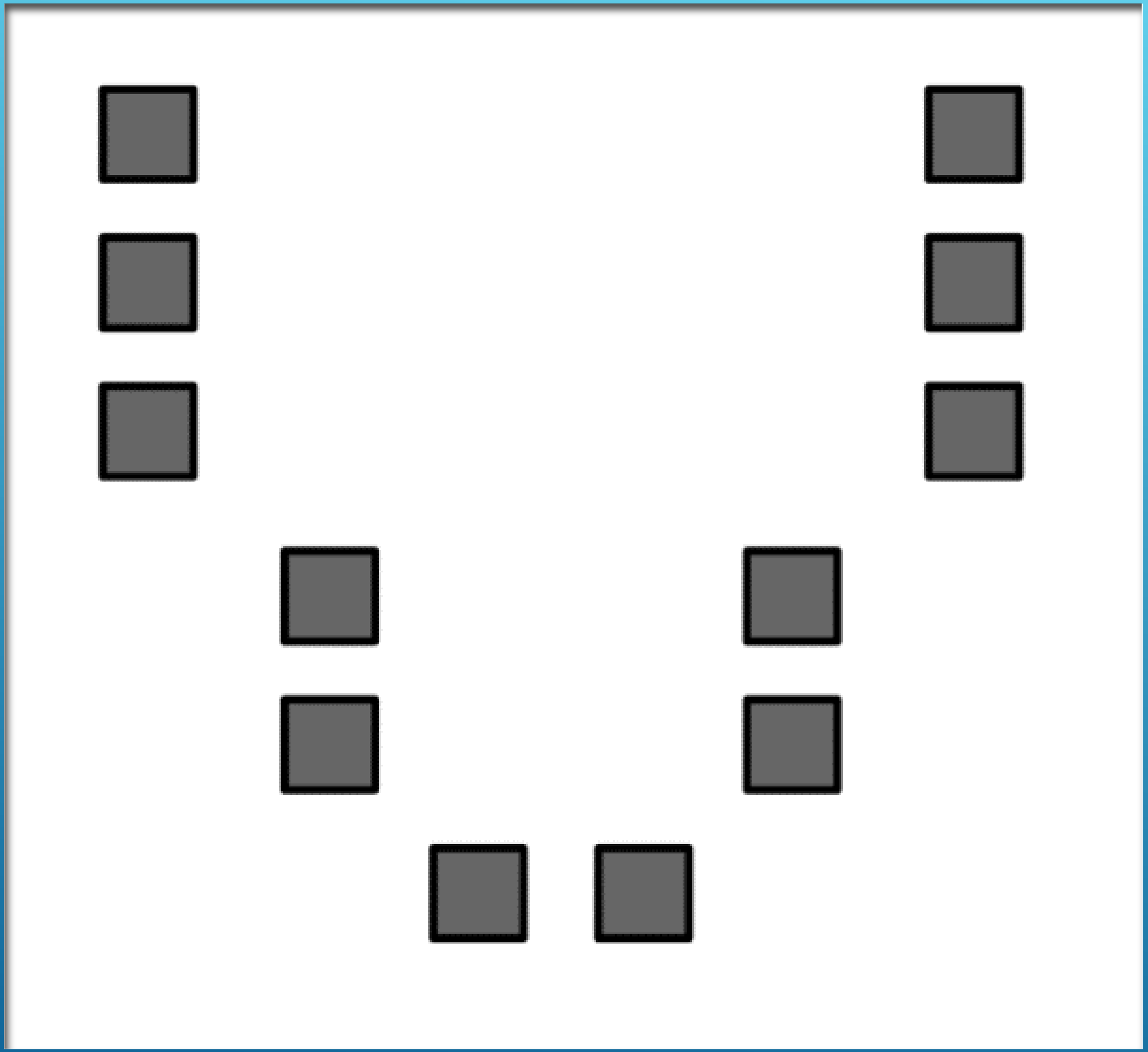
k-means++ overview

Initialization of k-means algorithm is critical to quality of local optima found

Smart initialization:

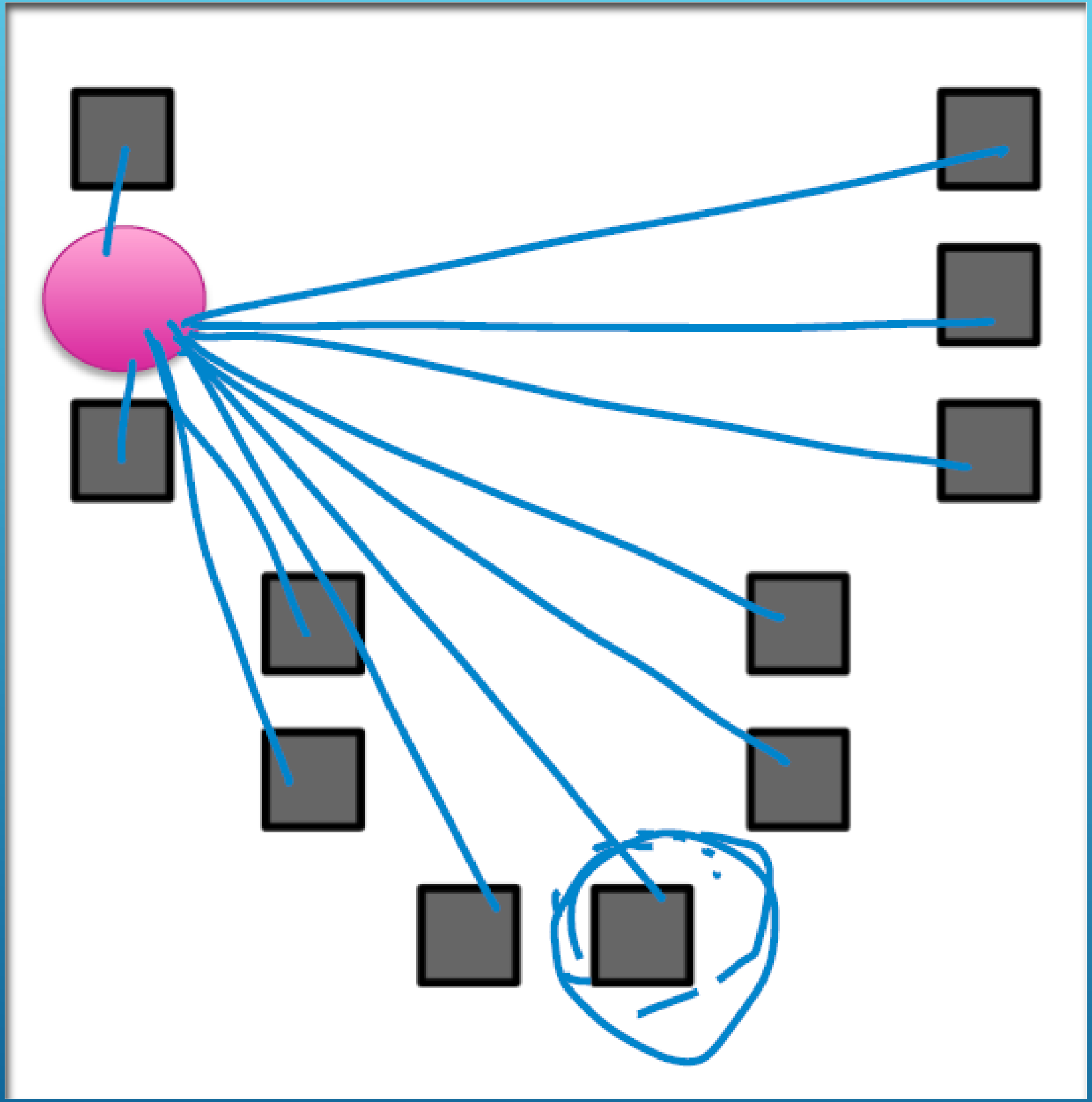
1. Choose first cluster center uniformly at random from data points
2. For each obs **x**, compute distance d(**x**) to nearest cluster center
3. Choose new cluster center from amongst data points, with probability of **x** being chosen proportional to $d(\mathbf{x})^2$
4. Repeat Steps 2 and 3 until k centers have been chosen

Reference: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin



K-MEANS++ VISUALIZED

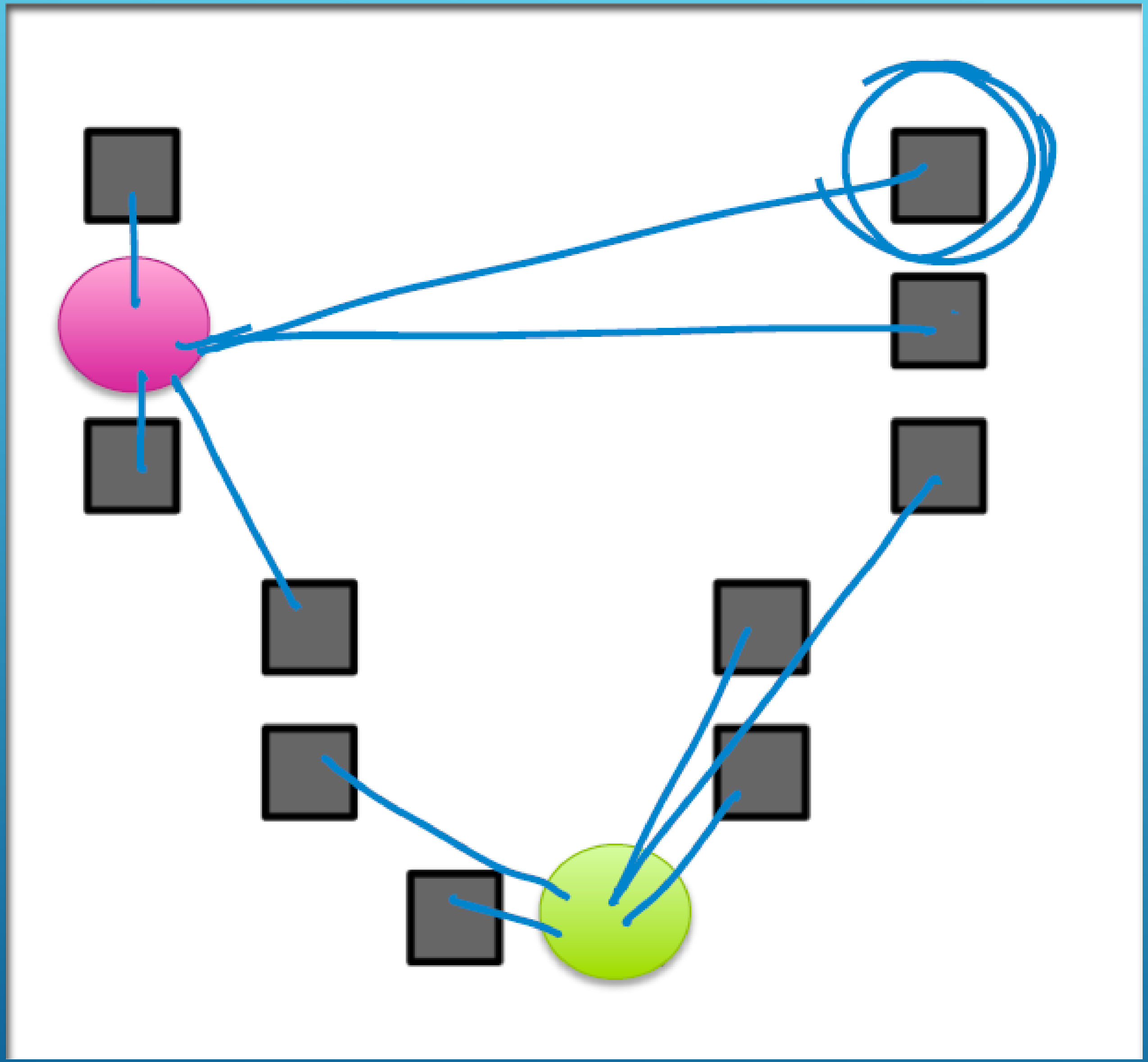
Reference: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin



K-MEANS++ VISUALIZED

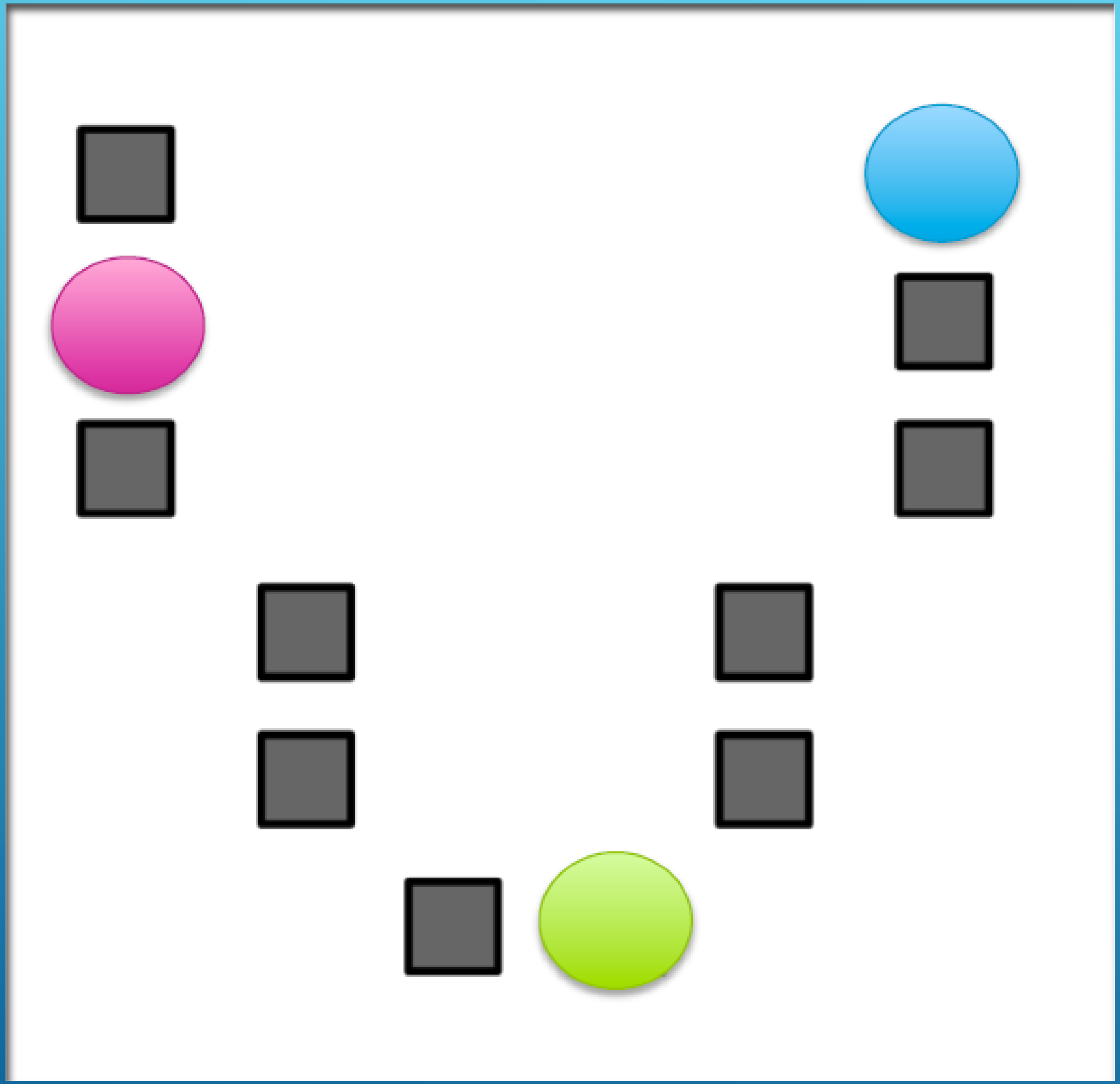
more likely to select a datapoint as a cluster center if that datapoint is far away ($dist^2$ increases this effect)

Reference: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin



K-MEANS++ VISUALIZED

Reference: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin



K-MEANS++ VISUALIZED

Reference: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

FUTURE TOPICS

1. **Density Estimation Topics**
   - Histograms
   - Kernel Density Estimation
2. **Dimensionality Reduction**
   - Principal Component Analysis (PCA)
   - t-SNE
   - UMAP
   - Autoencoders
3. **Outlier Detection Topics**
   - One-Class Classifier Learning
   - Autoencoders
4. **Clustering Topics**
   - Hierarchical Agglomerative Clustering
   - Gaussian Mixture Model
   - DBSCAN
   - HDBSCAN*
   - Cross-validation
5. **Deep Neural Network Approaches**

References:

The Hundred-Page Machine Learning Book. Andriy Burkov.

Applied Machine Learning in Python. Coursera. University of Michigan, Prof. Kevin Collins Thompson

Cluster analysis, https://en.wikipedia.org/w/index.php?title=Cluster_analysis&oldid=1002271612 (last visited Jan. 27, 2021).

Dimensionality reduction, https://en.wikipedia.org/w/index.php?title=Dimensionality_reduction&oldid=1002754996 (last visited Jan. 27, 2021).

INTRODUCING UNSUPERVISED LEARNING

Scott O'Hara
Metrowest Developers Machine Learning Group
02/17/2021

NEW TALK

Reference: Machine Learning: Clustering & Retrieval
University of Washington, Profs. Emily Fox & Carlos Guestrin

Reference: The Hundred-Page Machine Learning Book. Andriy Burkov.

Reference: *Applied Machine Learning in Python*. Coursera.
University of Michigan, Prof. Kevin Collins Thompson