SciTinyML

Scientific use of machine learning on low power devices
Regional Workshop - Africa

Unsupervised Learning and Anomaly Detection

Marcelo Rovai
Professor, UNIFEI - Brazil

Shawn Himel
Senior DevRel Engineer, Edge Impulse

Machine Learning

Supervised learning

Task-driven

- Regression
- Classification
- Object detection

Unsupervised learning

Data-driven

- Clustering
- Segmentation
- Anomaly detection

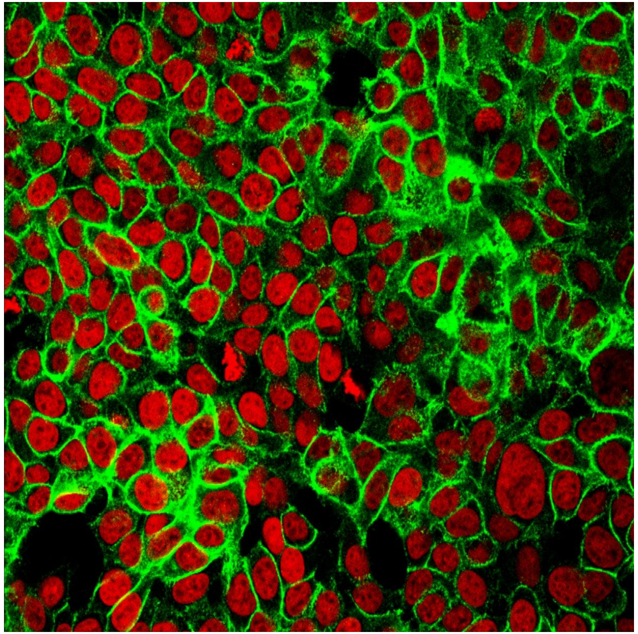
Reinforcement learning

Learn from experience

- Robotics
- Games
- Recommender systems

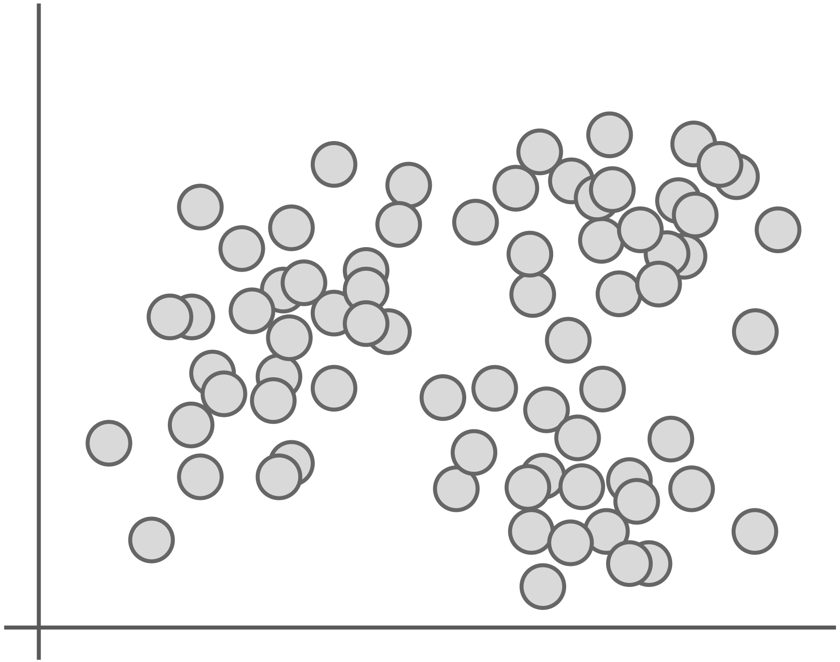
Unsupervised Learning

- No labels!
- Model automatically discovers patterns in the data
- Uses
  - Segmentation
  - Clustering
  - Dimensionality reduction
  - Anomaly detection



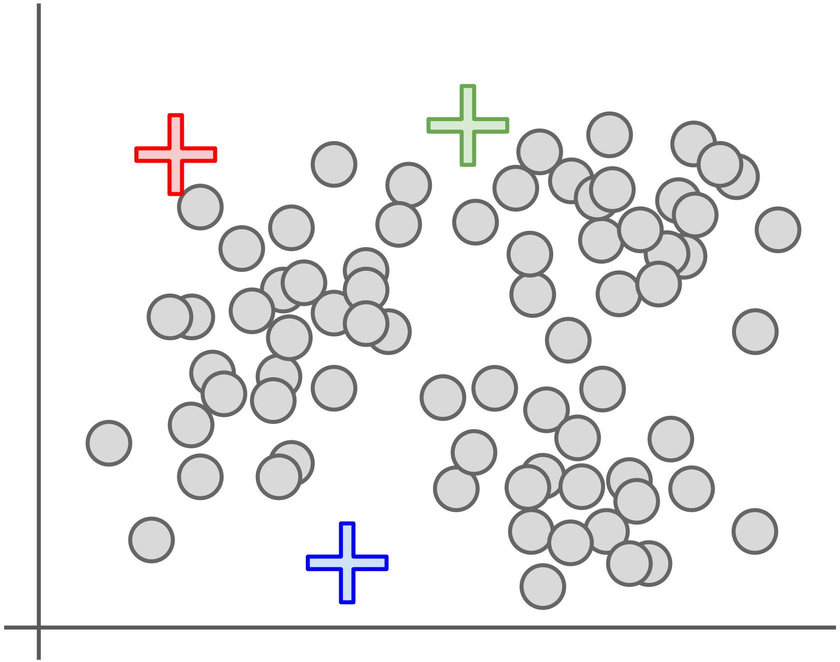
K-means Clustering

1. Define k (e.g. k=3)



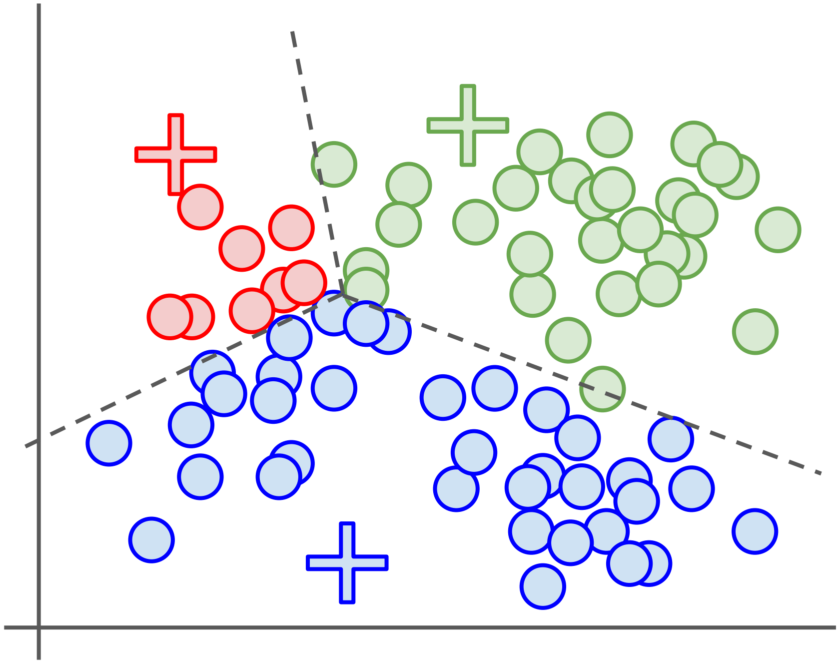
K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster



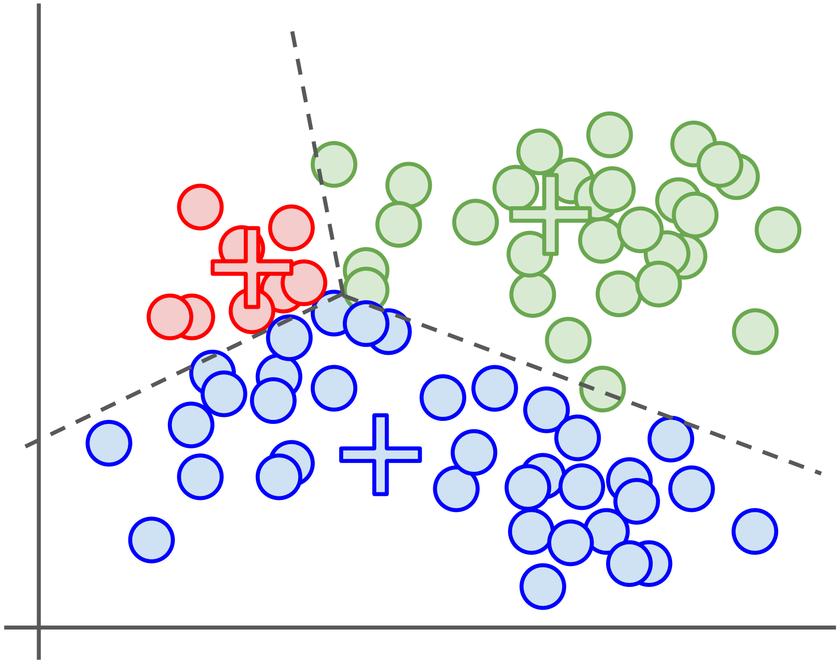
K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance



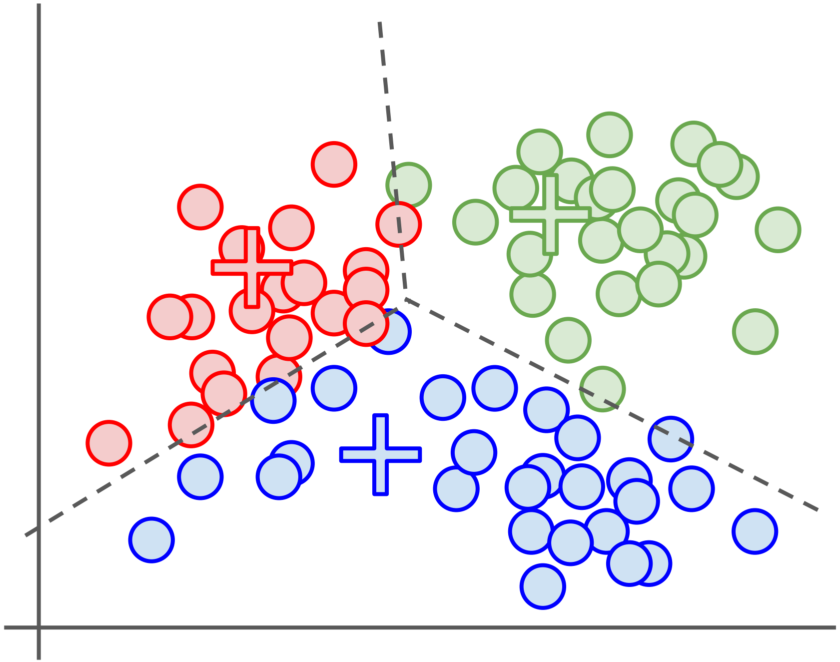
K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance
4. Re-compute the centroid of the cluster



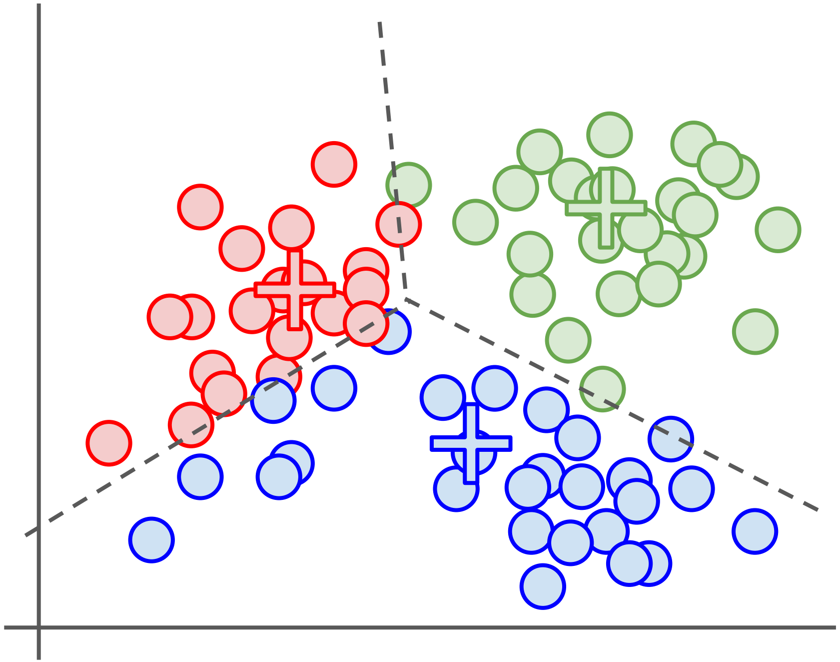
K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance
4. Re-compute the centroid of the cluster
5. Repeat steps 3-4



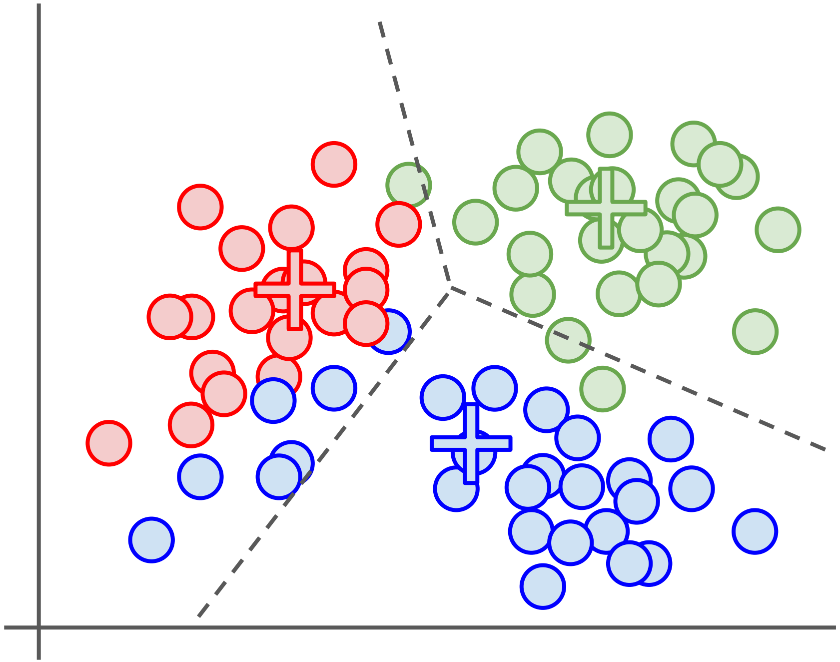
K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance
4. Re-compute the centroid of the cluster
5. Repeat steps 3-4



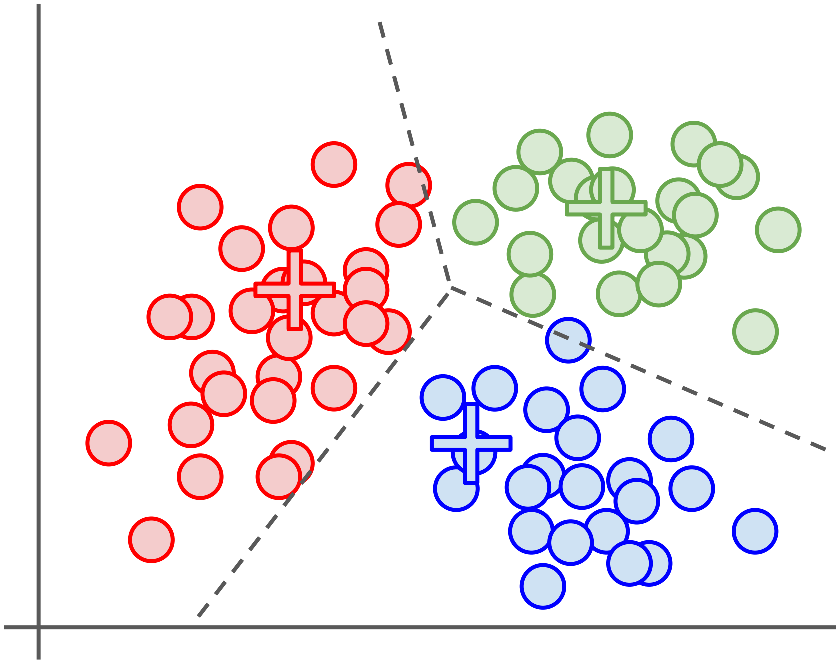
K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance
4. Re-compute the centroid of the cluster
5. Repeat steps 3-4



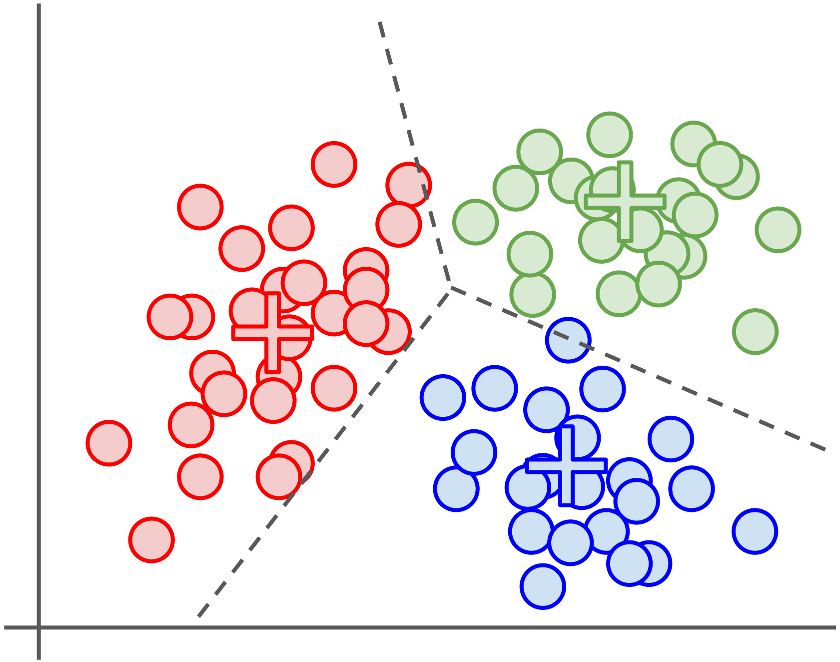
K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance
4. Re-compute the centroid of the cluster
5. Repeat steps 3-4



K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance
4. Re-compute the centroid of the cluster
5. Repeat steps 3-4



K-means Clustering

1. Define k (e.g. k=3)
2. Randomly choose centroid for each cluster
3. Assign every sample to nearest centroid based on Euclidean distance
4. Re-compute the centroid of the cluster
5. Repeat steps 3-4
6. ...until one of:
    a. Sum of distances between data points and corresponding centroid is minimized
    b. No change in centroids
    c. Maximum iterations reached

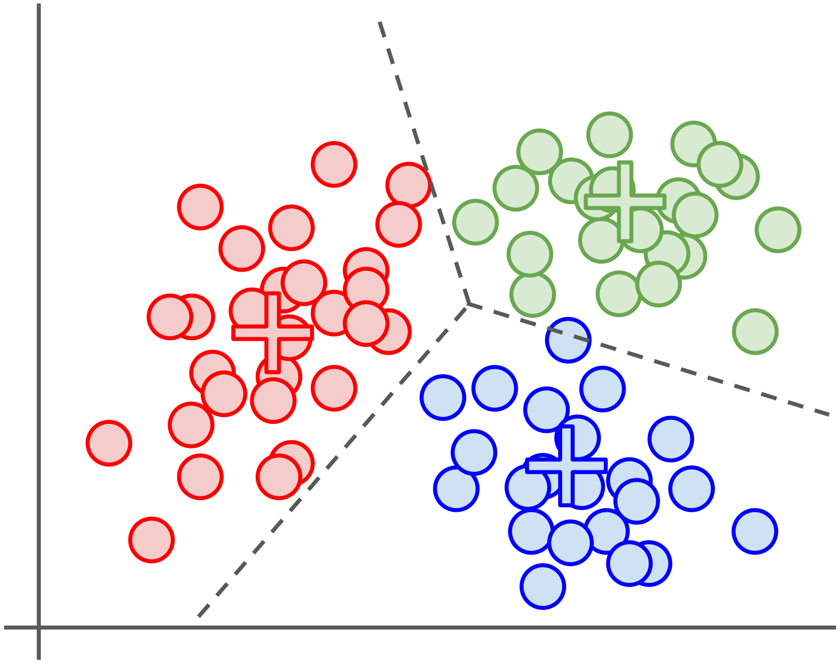
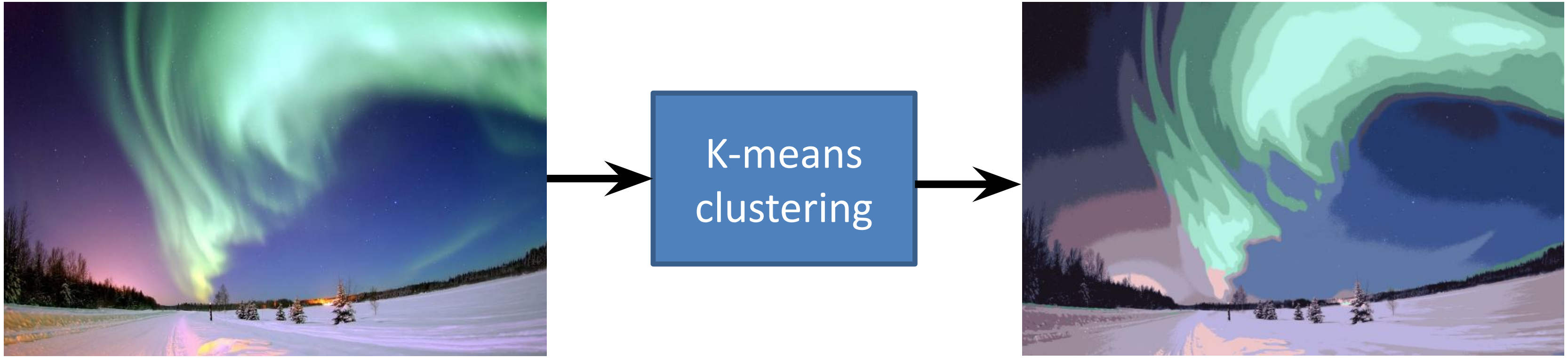


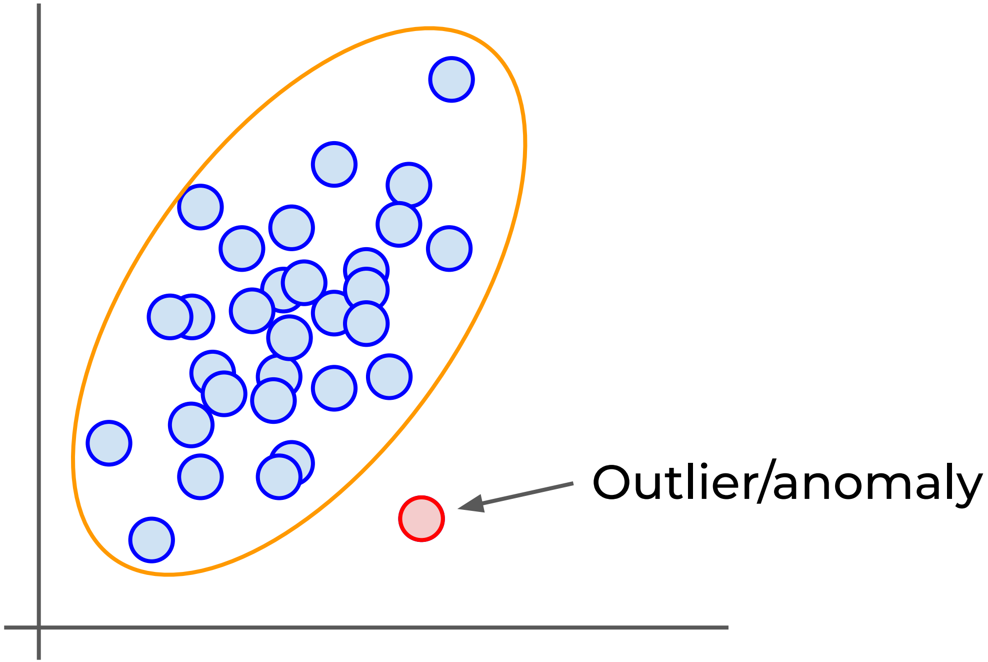
Image Segmentation



Anomaly Detection

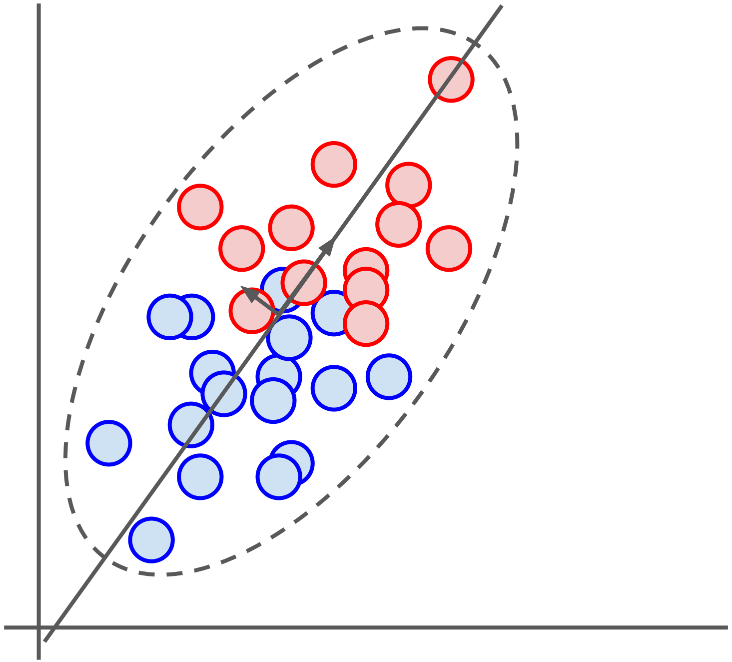
Examples:

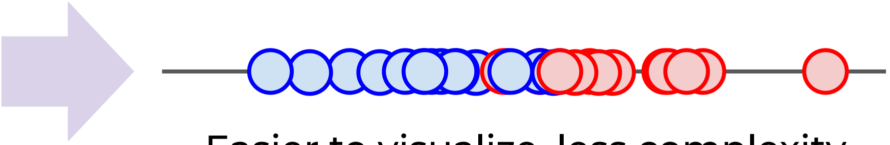
- Email spam
- Credit card fraud
- Motion alarm
- Fault detection



Dimensionality Reduction

Example: principal component analysis (PCA)





Easier to visualize, less complexity

EDGE IMPULSE

hello@edgeimpulse.com

3031 Tisch Way
110 Plaza West
San Jose, CA 95128
USA