BoW Classification

Computer Vision
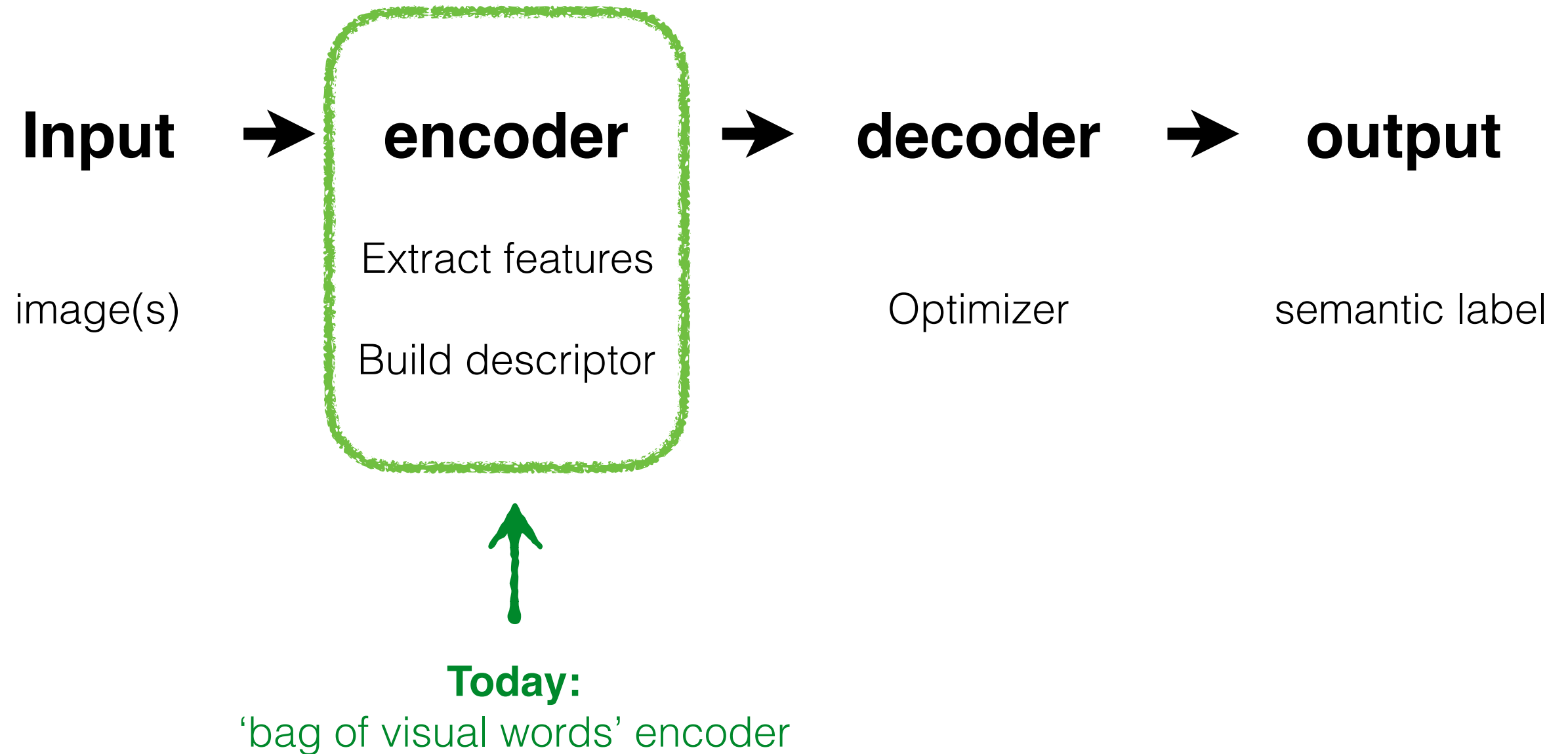**Carnegie Mellon University (Kris Kitani)**

'Classical'
Image Classification Pipeline
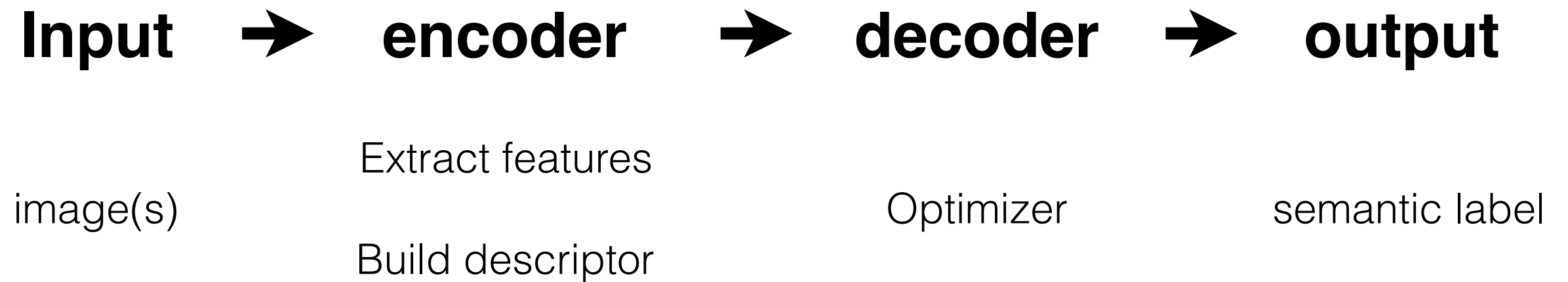
Input ➤ **encoder** ➤ **decoder** ➤ **output**

Extract features

image(s)   Optimizer   semantic label

Build descriptor

'Classical'
Image Classification Pipeline

Input ➔ **encoder** ➔ **decoder** ➔ **output**

image(s)

encoder: Extract features, Build descriptor

Optimizer

semantic label

Today:
'bag of visual words' encoder

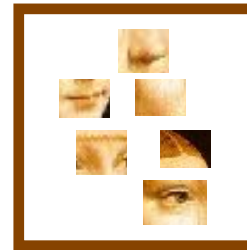# BoW encoder

Encode:

1. Dictionary Learning

2. Build Bags-of-Words
(BOW) descriptor

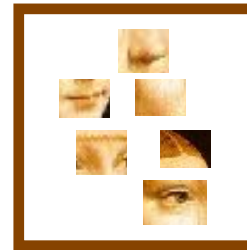**Decode:**

1. Train classifier

2. Classify using BOW
    descriptor

BoW encoder

a. Extract features from image

Encode:

1. Dictionary Learning

BoW encoder

Encode:

1. Dictionary Learning
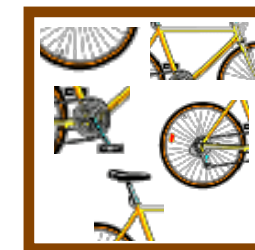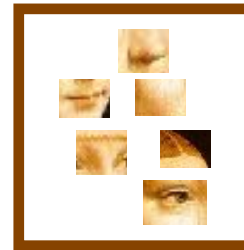
2. Build Bags-of-Words (BOW) descriptor

Decode:

1. Train classifier

2. Classify using BOW descriptor

a. Extract features from image

What kinds of features can we extract?

BoW encoder

Encode:

1. Dictionary Learning

2. Build Bags-of-Words (BOW) descriptor

Decode:

1. Train classifier

2. Classify using BOW descriptor

a. Extract features from image

What kinds of features can we extract?

Regular grid
- Vogel & Schiele, 2003
- Fei-Fei & Perona, 2005

Interest point detector
- Csurka et al. 2004
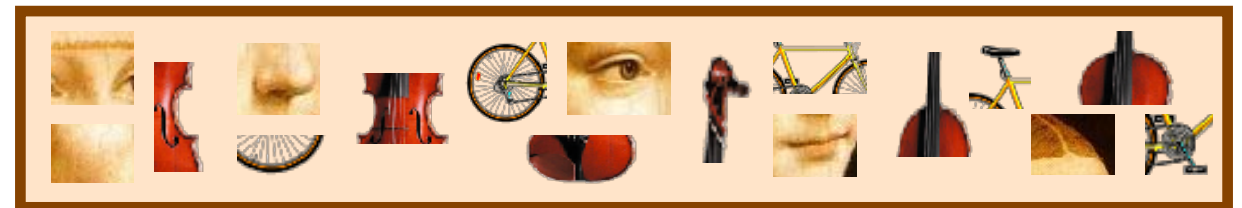- Fei-Fei & Perona, 2005
- Sivic et al. 2005

Other methods
- Random sampling (Vidal-Naquet & Ullman, 2002)
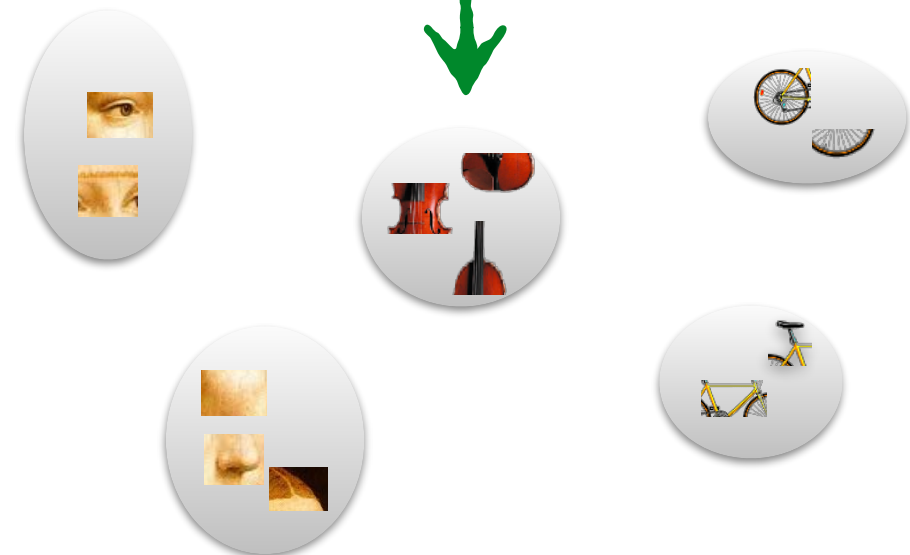- Segmentation-based patches (Barnard et al. 2003)
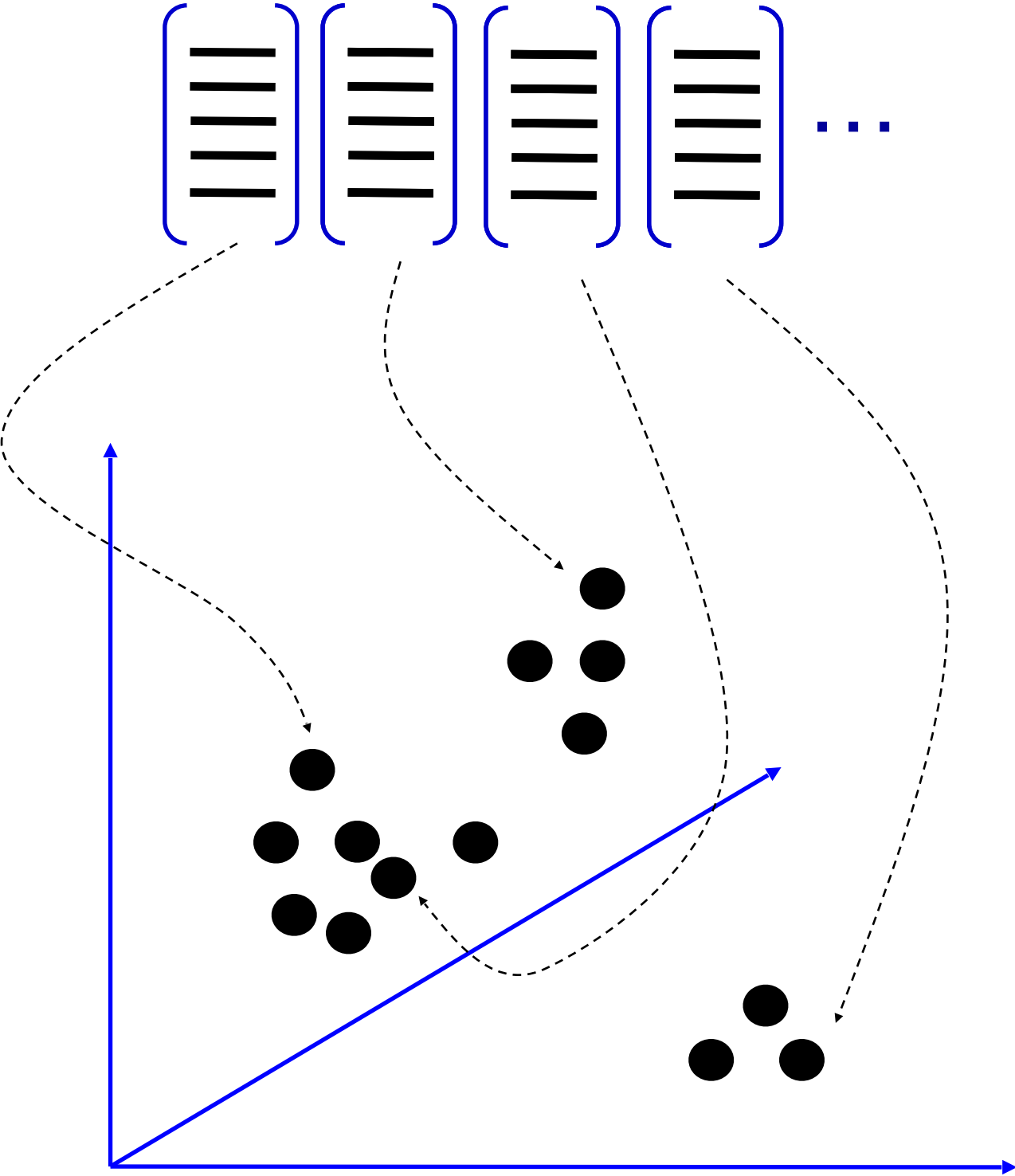
BoW encoder

Encode:

1. Dictionary Learning

2. Build Bags-of-Words (BOW) descriptor

Decode:

1. Train classifier

2. Classify using BOW descriptor

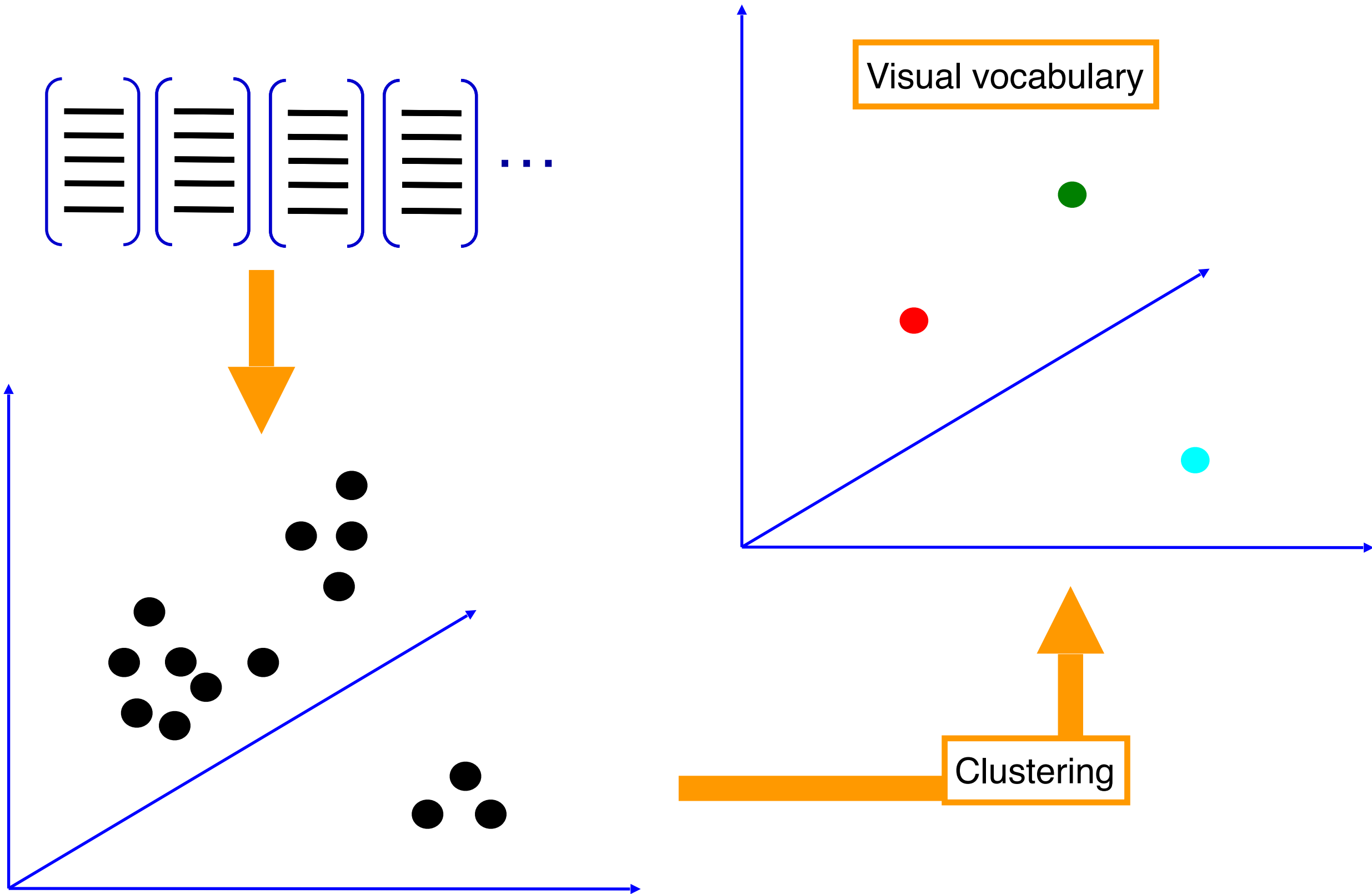a. Extract features from image

b. Unsupervised Clustering

K-means clustering

b. Unsupervised Clustering
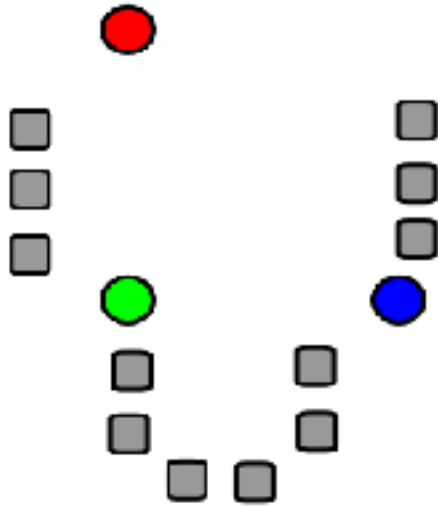
b. Unsupervised Clustering

b. Unsupervised Clustering
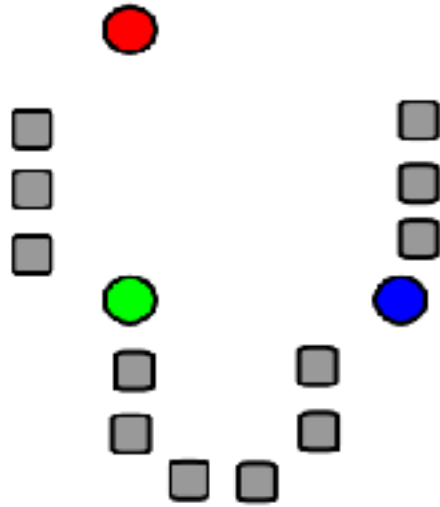

Visual vocabulary
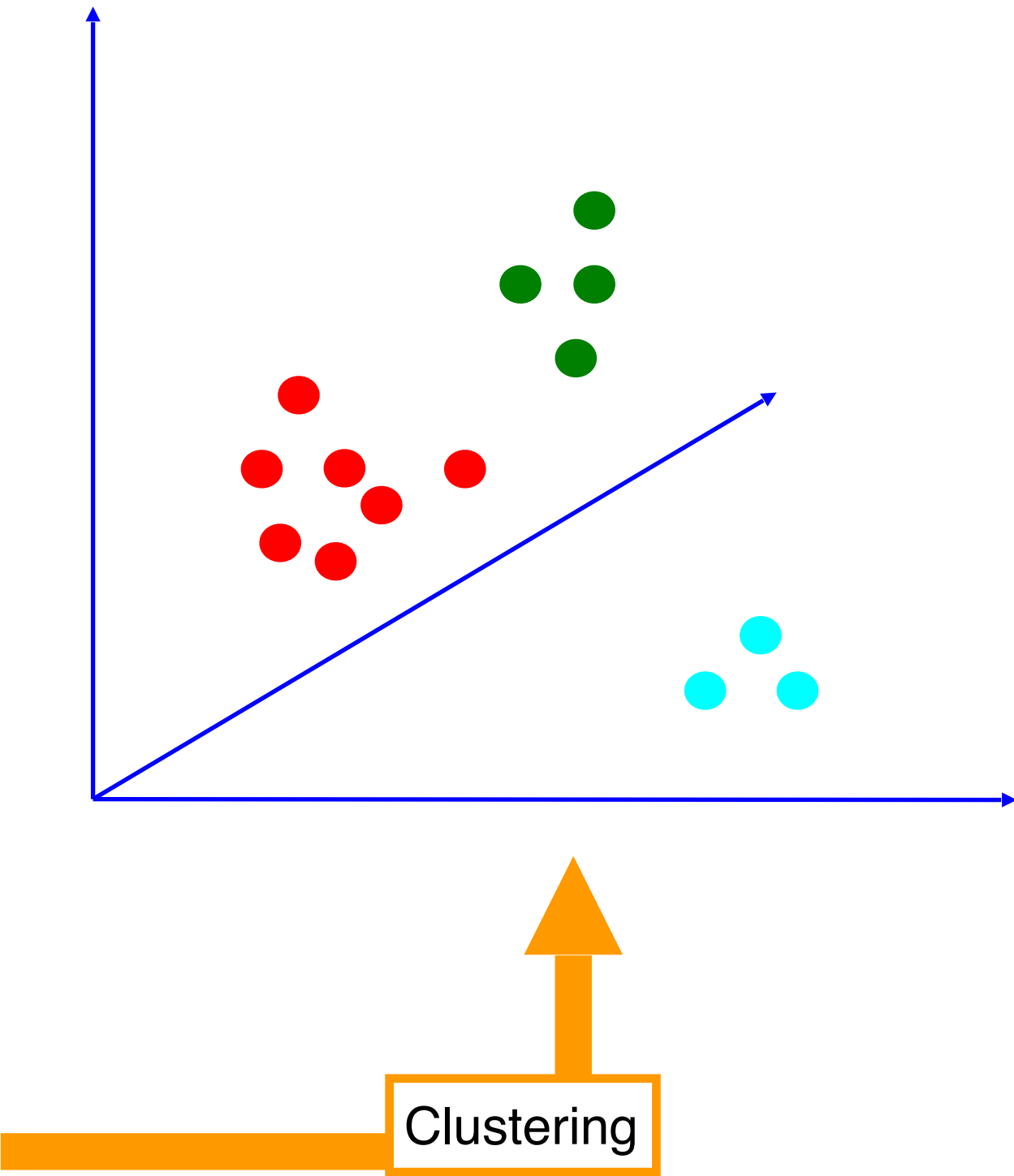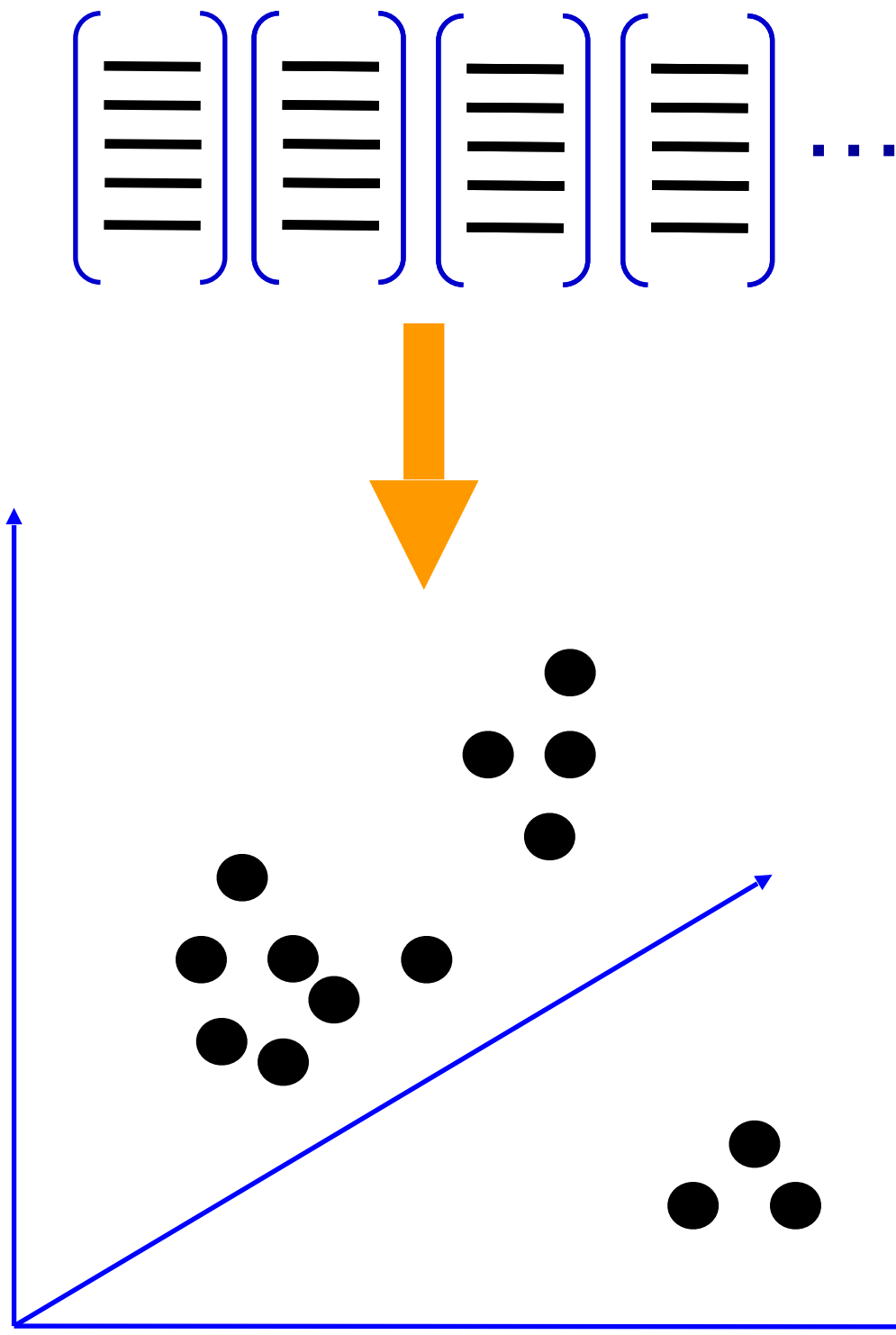
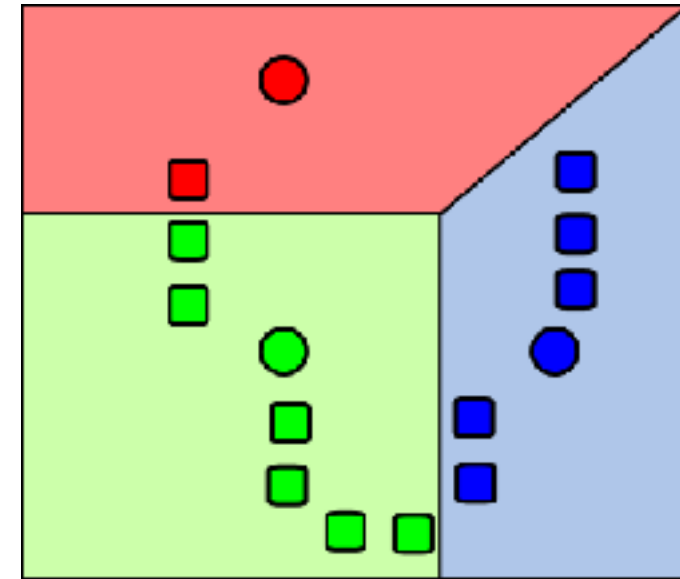Clustering

b. Unsupervised Clustering

1. Select initial
centroids at random
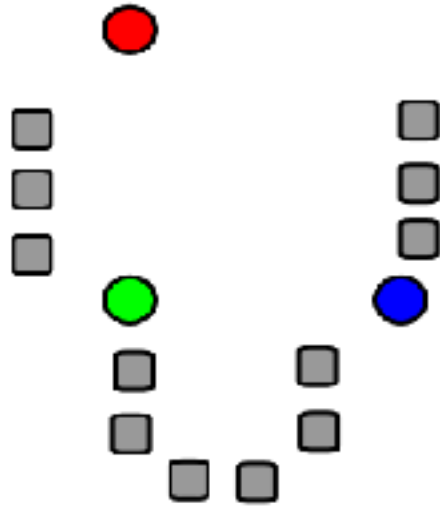
b. Unsupervised Clustering

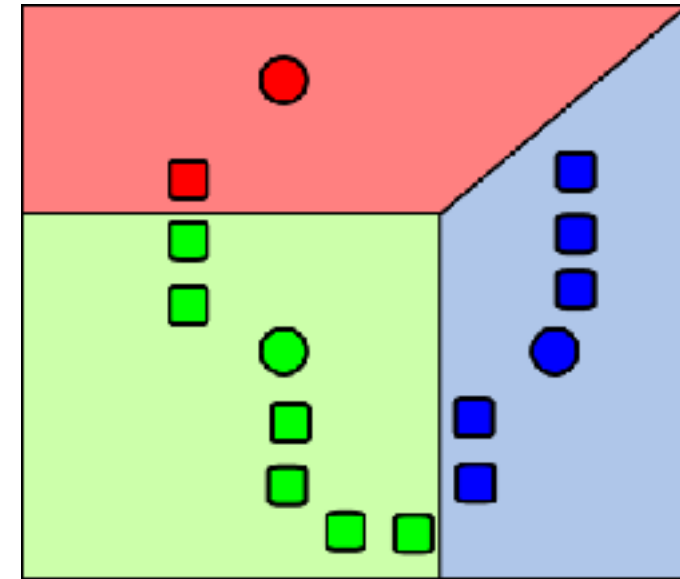1. Select initial
centroids at random

2. Assign each object to
the cluster with the
nearest centroid.

b. Unsupervised Clustering

1. Select initial centroids at random

2. Assign each object to the cluster with the nearest centroid.

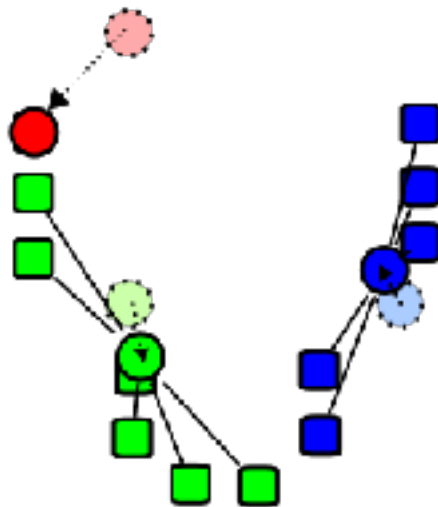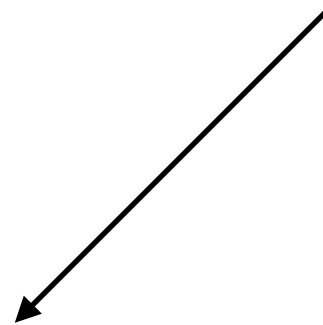3. Compute each centroid as the mean of the objects assigned to it (go to 2)

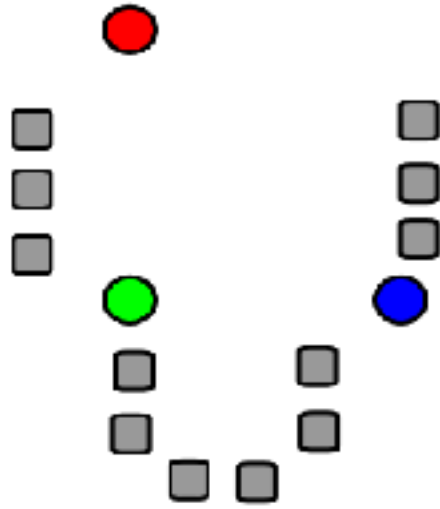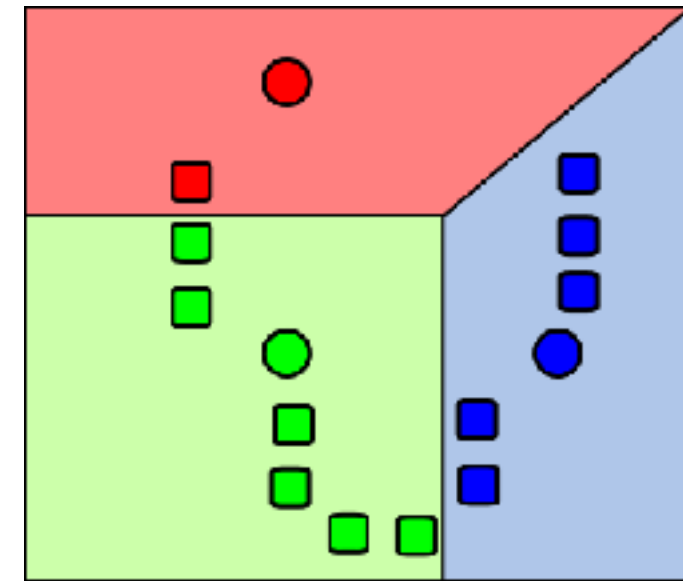# b. Unsupervised Clustering

1. Select initial centroids at random

2. Assign each object to the cluster with the nearest centroid.

3. Compute each centroid as the mean of the objects assigned to it (go to 2)

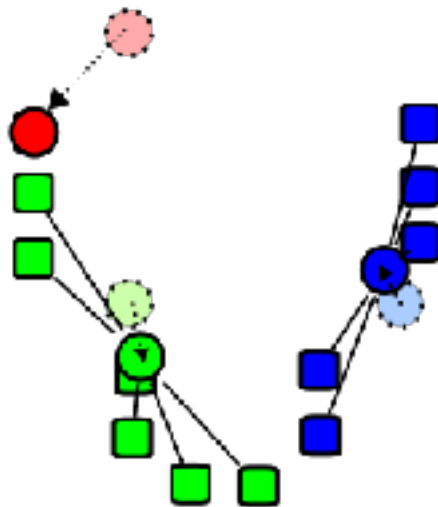2. Assign each object to the cluster with the nearest centroid.
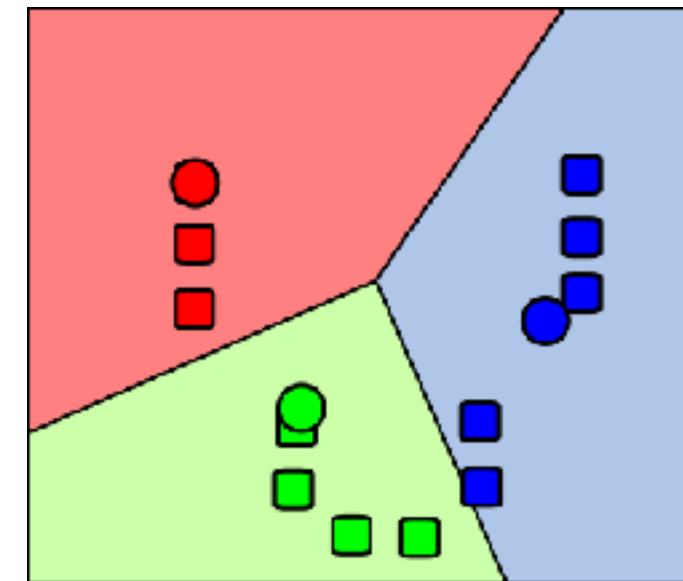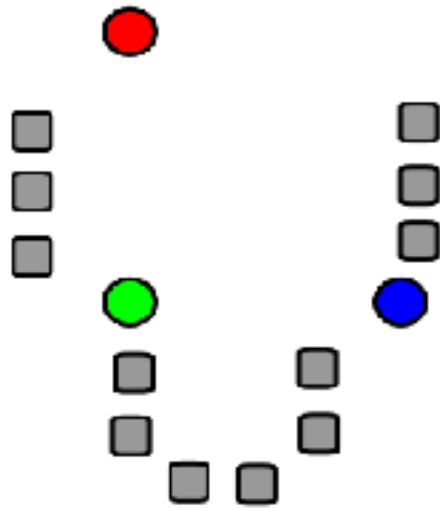
b. Unsupervised Clustering

1. Select initial centroids at random

2. Assign each object to the cluster with the nearest centroid.

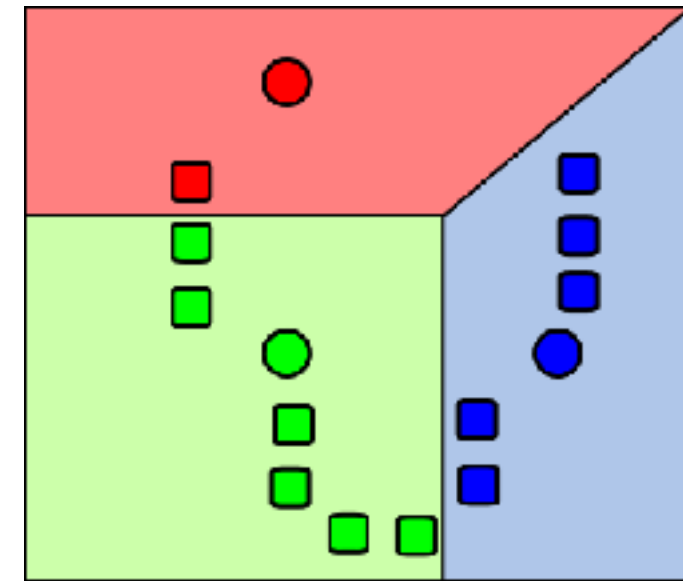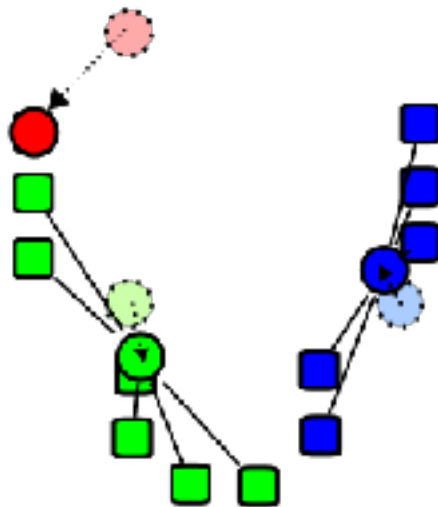3. Compute each centroid as the mean of the objects assigned to it (go to 2)

2. Assign each object to the cluster with the nearest centroid.

Repeat previous 2 steps until no change

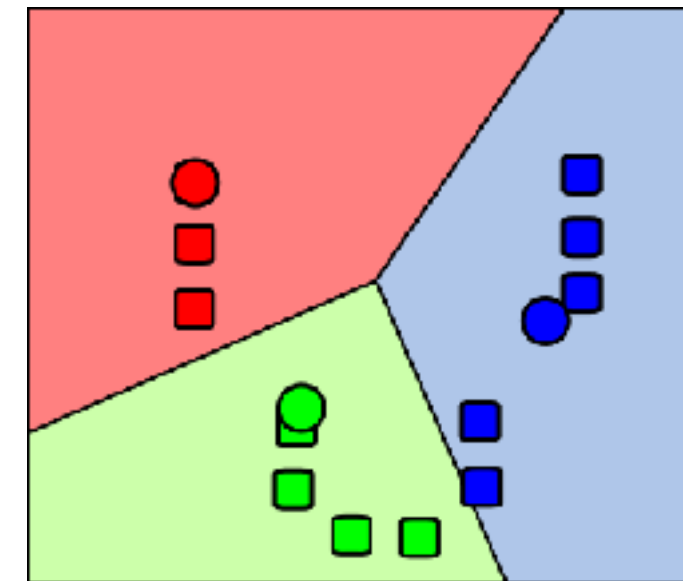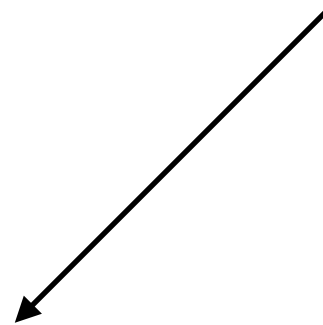# K-means Clustering

Given k:

1. Select initial centroids at random.

2. Assign each object to the cluster with the nearest centroid.

3. Compute each centroid as the mean of the objects assigned to it.

4. Repeat previous 2 steps until no change.

BoW encoder

Encode:

1. Dictionary Learning

2. Build Bags-of-Words (BOW) descriptor

After learning a visual dictionary, we need an image representation.

Let's use a bag-of-words descriptor!

Decode:

1. Train classifier

2. Classify using BOW descriptor

BoW encoder

a. Quantize feature

Encode:

1. Dictionary Learning

2. Build Bags-of-Words
(BOW) descriptor

Decode:
1.   Train classifier

2.   Classify using BOW
     descriptor

BoW encoder

Encode:

1. Dictionary Learning

2. Build Bags-of-Words (BOW) descriptor

Decode:

1. Train classifier

2. Classify using BOW descriptor

a. Quantize feature

Associate each feature to nearest 'word' (cluster)

BoW encoder

Encode:

1. Dictionary Learning

2. Build Bags-of-Words
(BOW) descriptor

Decode:
1. Train classifier

2. Classify using BOW
descriptor

a. Quantize feature

Associate each feature to
nearest 'word' (cluster)

b. Build histogram (for each image)

BoW encoder

Encode:
1. Dictionary Learning

2. Build Bags-of-Words (BOW) descriptor

Decode:
1. Train classifier

2. Classify using BOW descriptor

BoW encoder

Encode:
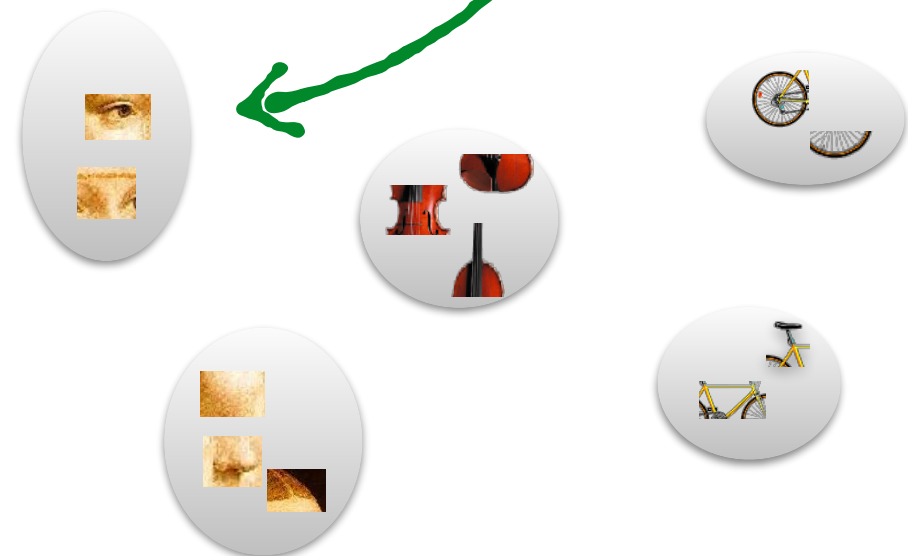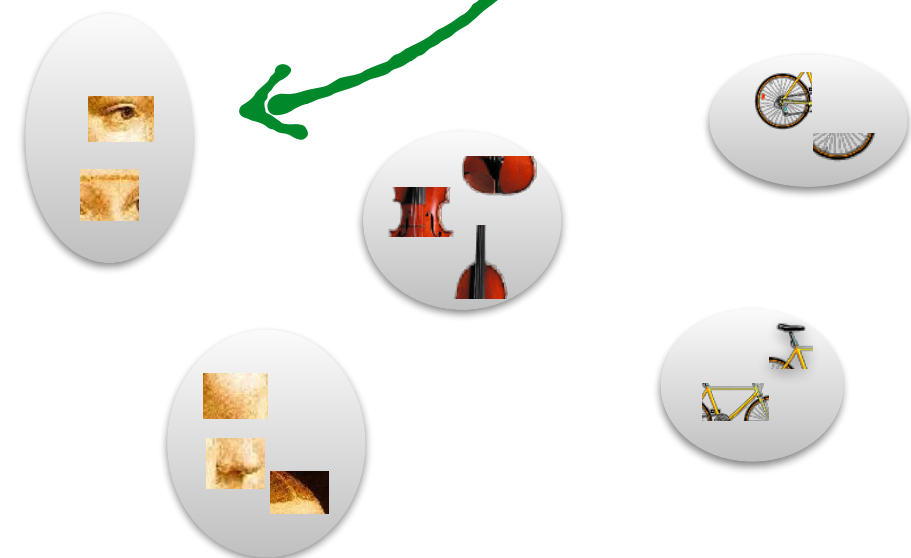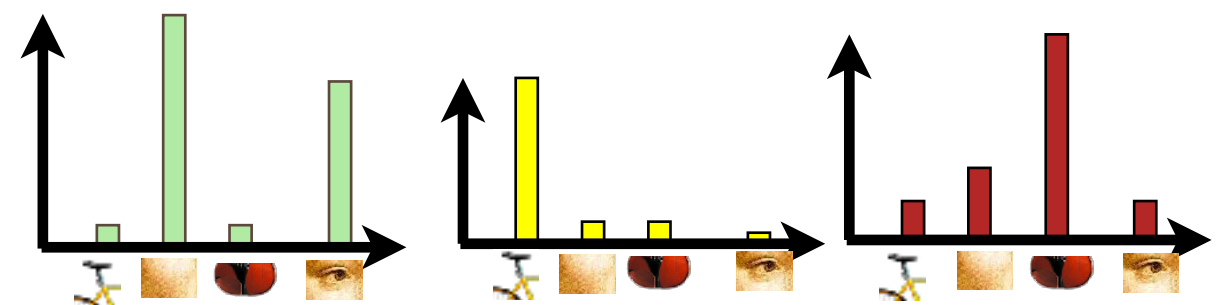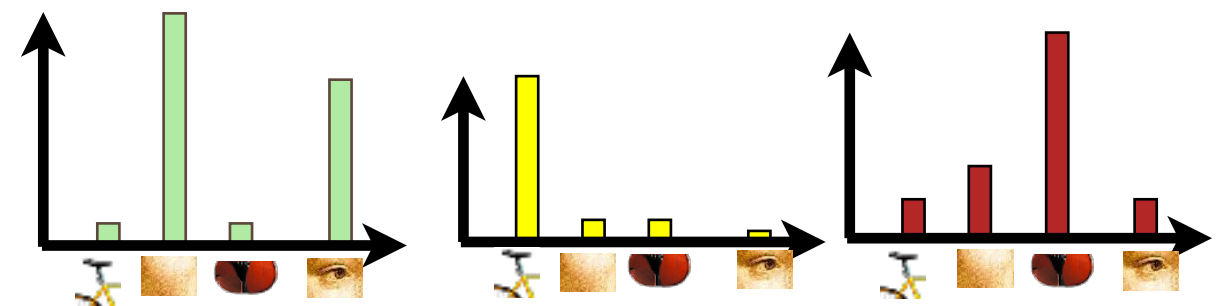1. Dictionary Learning

2. Build Bags-of-Words (BOW) descriptor

Decode:
1. Train classifier

2. Classify using BOW descriptor

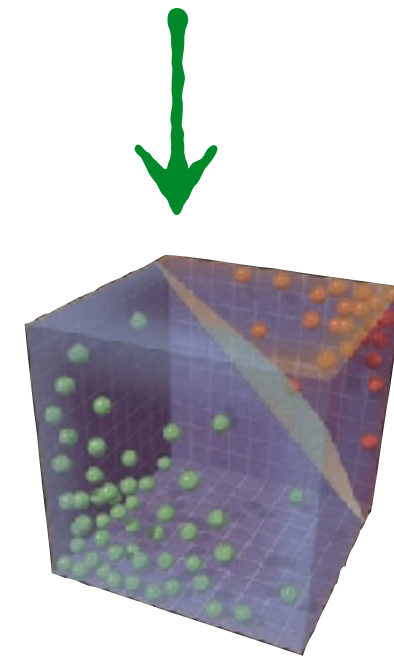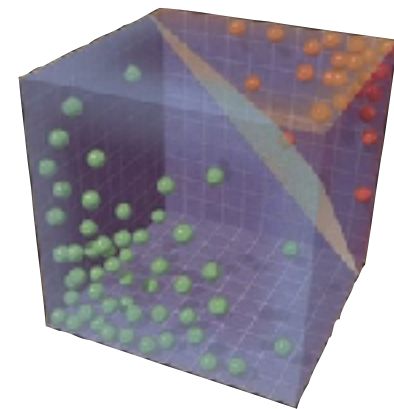Given the bag-of-features representations of images from different classes…

Train some fancy machine learning model

BoW encoder

Decode:

1.   Train classifier

2.   Classify BOW descriptor

Get output from machine learning model

'Classical'
Image Classification Pipeline

Input ➔ **encoder** ➔ **decoder** ➔ **output**

image(s)

Extract features

Build descriptor

Optimizer

semantic label

Next session