Unsupervised Learning

- Find patterns in the data.
- Group the data into clusters.
- Many clustering algorithms.
  - K means clustering
  - EM clustering
  - Graph-Theoretic Clustering
  - Clustering by Graph Cuts
  - etc

Clustering by K-means Algorithm

Form K-means clusters from a set of n-dimensional feature vectors

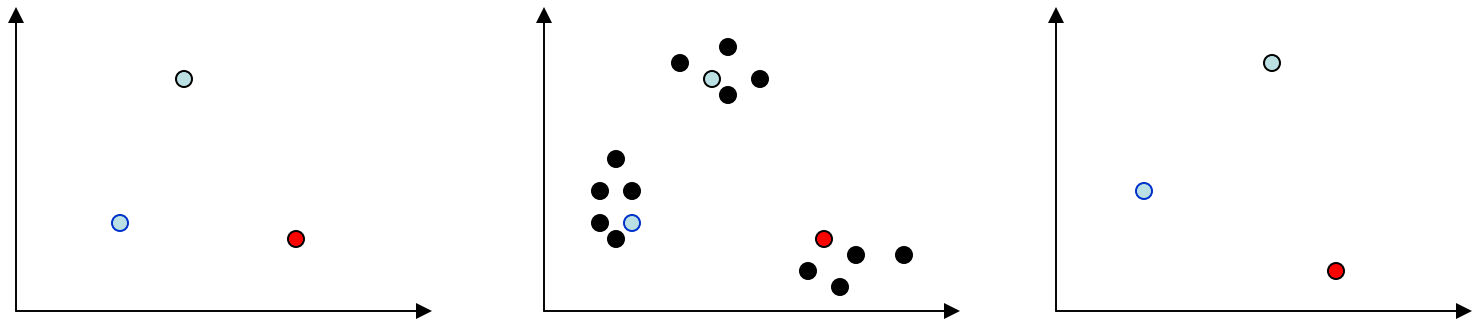
1. Set ic (iteration count) to 1

2. Choose randomly a set of K means $m_1(1), \ldots, m_K(1)$.

3. For each vector x_i, compute $D(x_i, m_k(ic))$, $k=1,\ldots K$
   and assign x_i to the cluster C_j with nearest mean.

4. Increment ic by 1, update the means to get $m_1(ic),\ldots,m_K(ic)$.

5. Repeat steps 3 and 4 until $C_k(ic) = C_k(ic+1)$ for all k.



K-Means Classifier (shown on RGB color data)

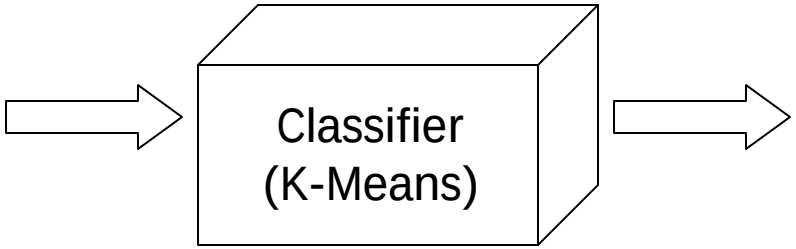
$x_1=\{r_1, g_1, b_1\}$

$x_2=\{r_2, g_2, b_2\}$

…

$x_i=\{r_i, g_i, b_i\}$

…



Classification Results

$x_1 \rightarrow C(x_1)$

$x_2 \rightarrow C(x_2)$

…

$x_i \rightarrow C(x_i)$

…

Cluster Parameters

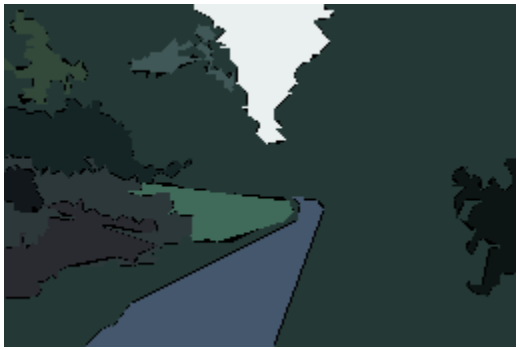
m_1 for C_1

m_2 for C_2

…

m_k for C_k

original data
one RGB per pixel



color clusters

K-Means Classifier (Cont.)

Input (Known)

$x_1=\{r_1, g_1, b_1\}$

$x_2=\{r_2, g_2, b_2\}$

…

$x_i=\{r_i, g_i, b_i\}$

…

Output (Unknown)

Cluster Parameters

m_1 for C_1

m_2 for C_2

…

m_k for C_k

Classification Results

$x_1 \rightarrow C(x_1)$

$x_2 \rightarrow C(x_2)$

…

$x_i \rightarrow C(x_i)$

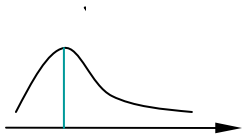
…

K-Means → EM

The clusters are usually Gaussian distributions.

- Boot Step:
  - Initialize K clusters: $C_1, \ldots, C_K$

    (μ_j, Σ_j) and $P(C_j)$ for each cluster j.



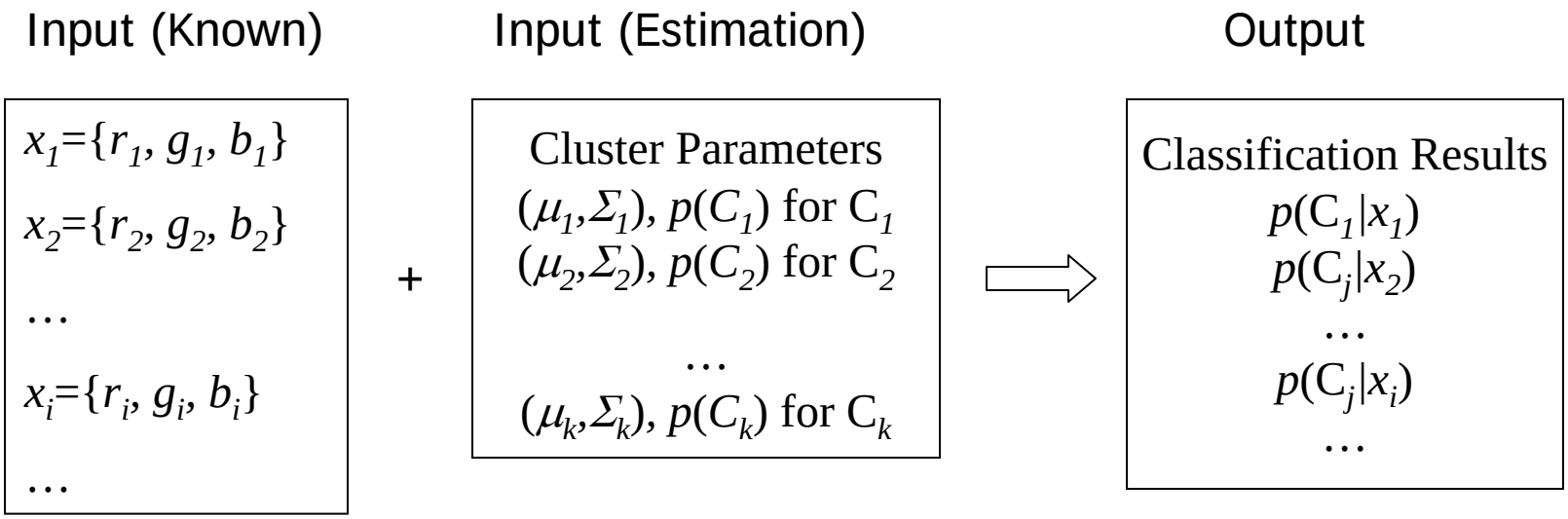
- Iteration Step:
  - Estimate the cluster of each datum

    $p(C_j \mid x_i)$ ➡ Expectation
  - Re-estimate the cluster parameters

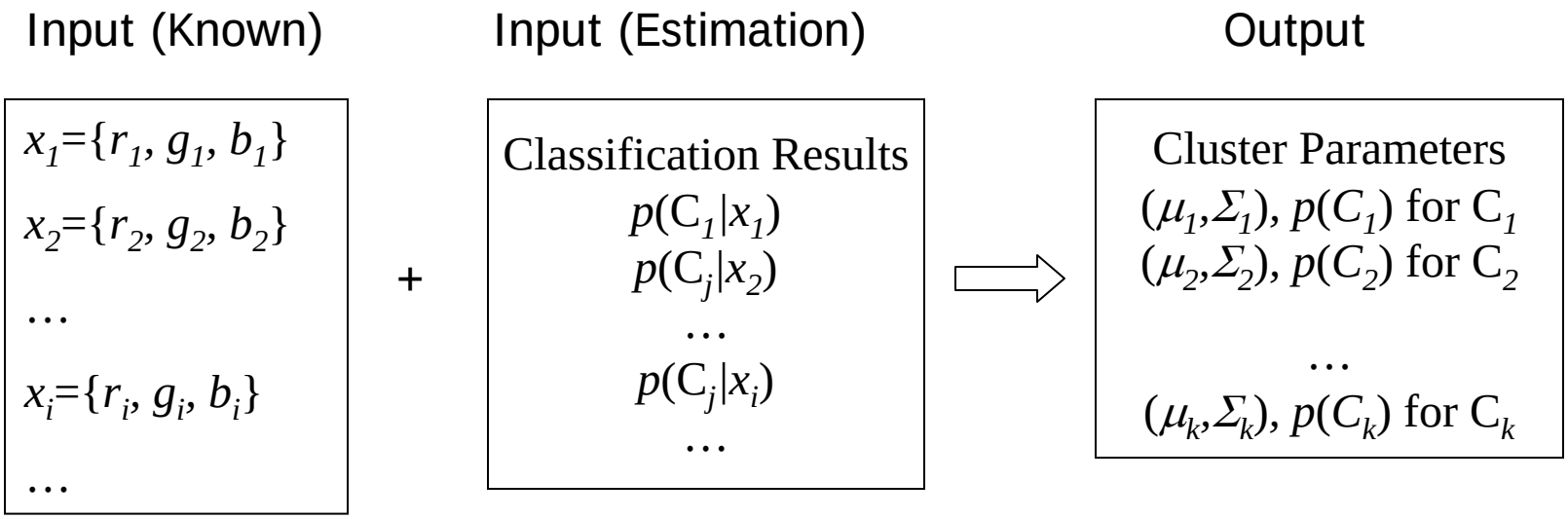
    $(\mu_j, \Sigma_j), p(C_j)$ For each cluster j ➡ Maximization

The resultant set of clusters is called a **mixture model**; if the distributions are Gaussian, it's a Gaussian mixture.

Expectation Step

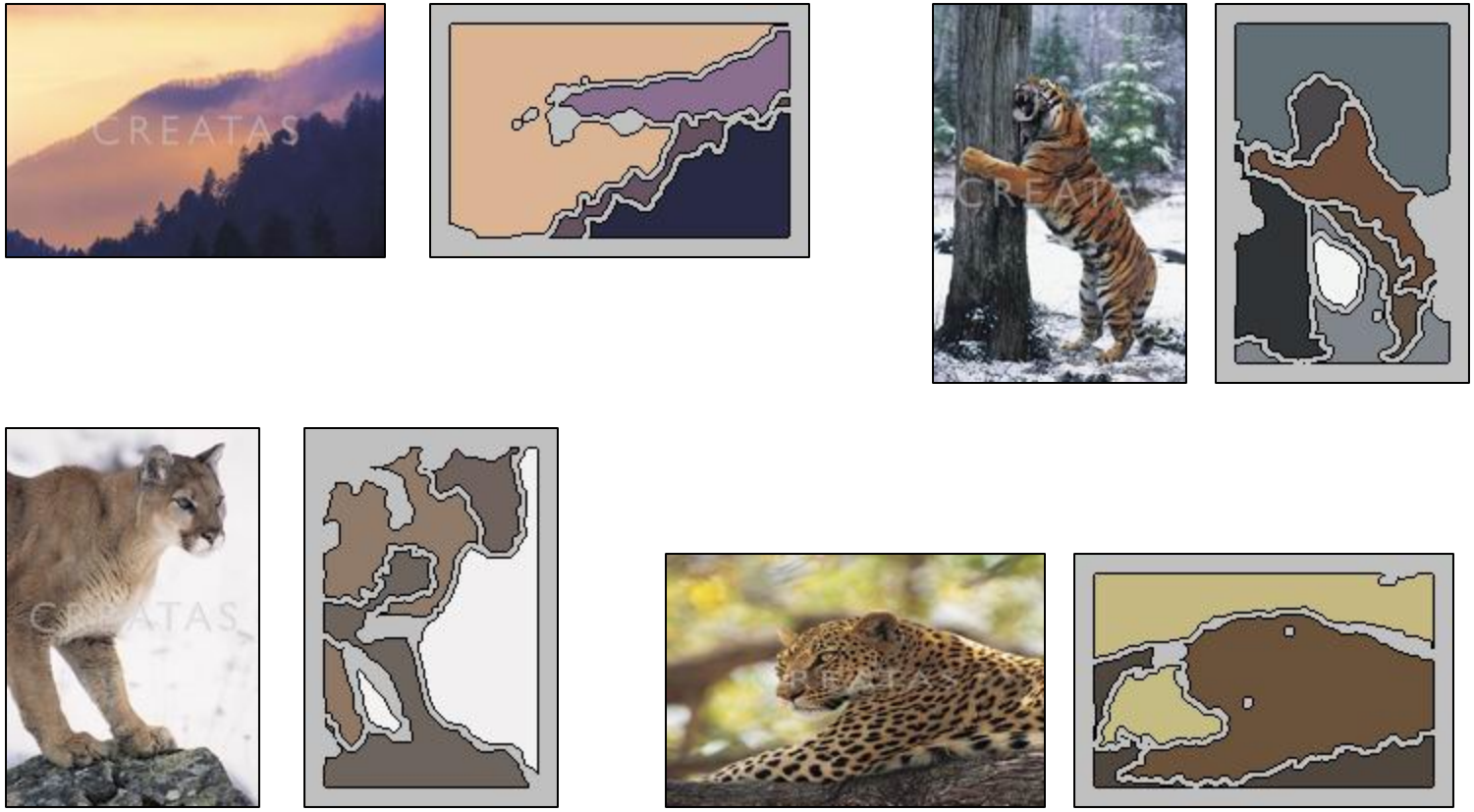


Maximization Step



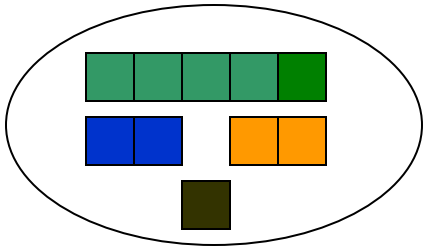
EM Clustering using color and texture information at each pixel

(from Blobworld)

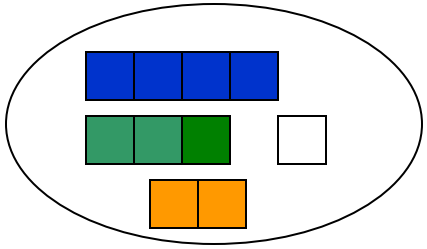


EM for Classification of Images in Terms of their Color Regions

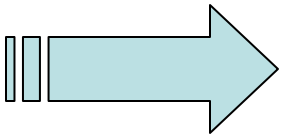
Initial Model for “trees”



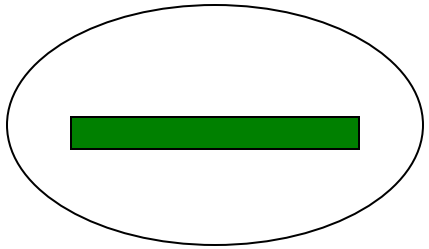
Initial Model for “sky”



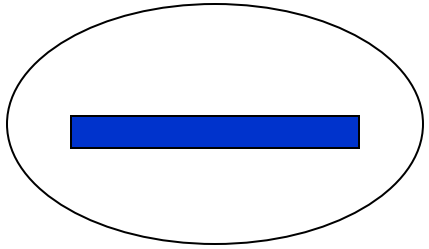
EM



Final Model for “trees”

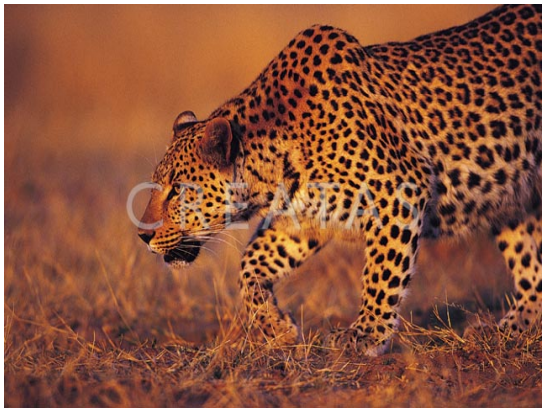


Final Model for “sky”



Sample Results

cheetah





Sample Results (Cont.)

grass

Sample Results (Cont.)

lion