UiO : Department of Technology Systems
University of Oslo

Image Feature Extraction

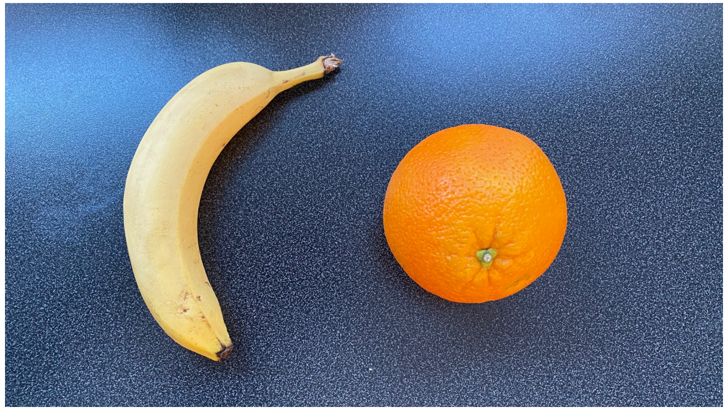
Idar Dyrdal

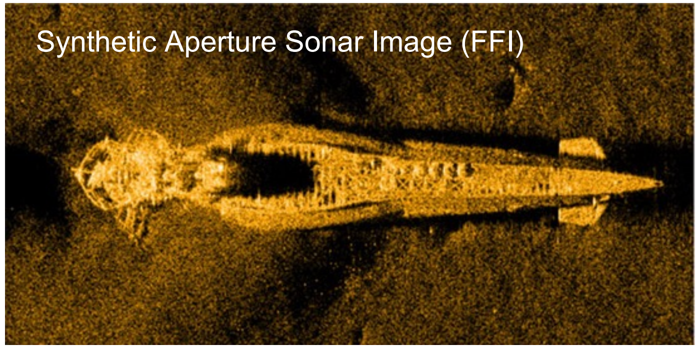
Image Analysis

Typical image analysis pipeline:

- Pre-processing
- Segmentation (e. g. object detection)
- *Feature extraction*
- Feature selection
- Classifier training
- Evaluation of classifier performance.

Features for image analysis





Applications:

- Remote sensing
- Medical imaging
- Character recognition
- Robot Vision
- …

Major goal of image feature extraction:

Given an image, or a region within an image, generate the features that will subsequently be fed to a classifier in order to classify the image in one of the possible classes.

(Theodoridis & Koutroumbas: «Pattern Recognition», Elsevier 2006).

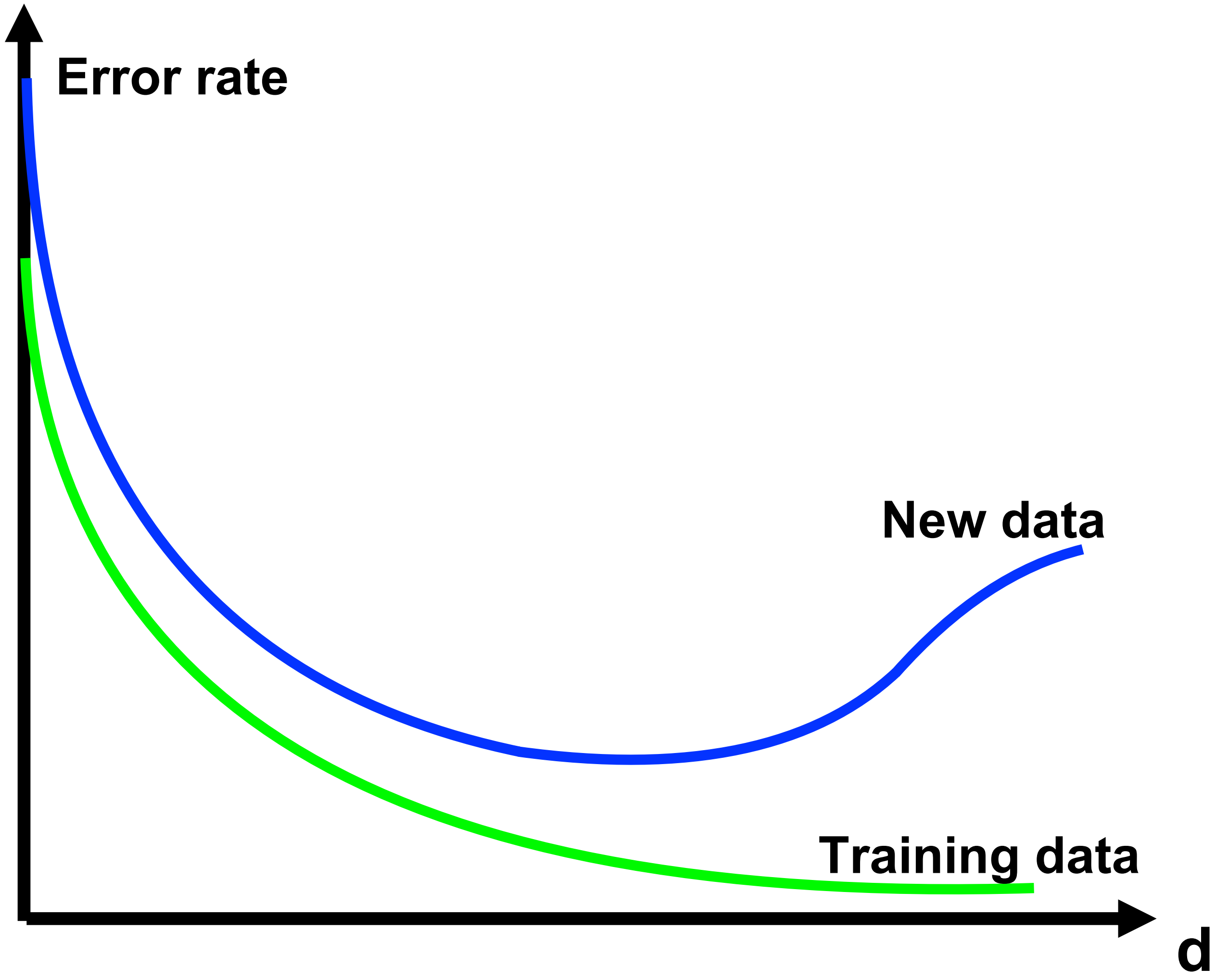
Feature extraction

The goal is to generate features that exhibit high information-packing properties:

- Extract the information from the raw data that is most relevant for discrimination between the classes
- Extract features with low within-class variability and high between class variability
- Discard redundant information

- The information in an image f[i,j] must be reduced to enable reliable classification (generalization)
- A 64x64 image → 4096-dimensional feature space!

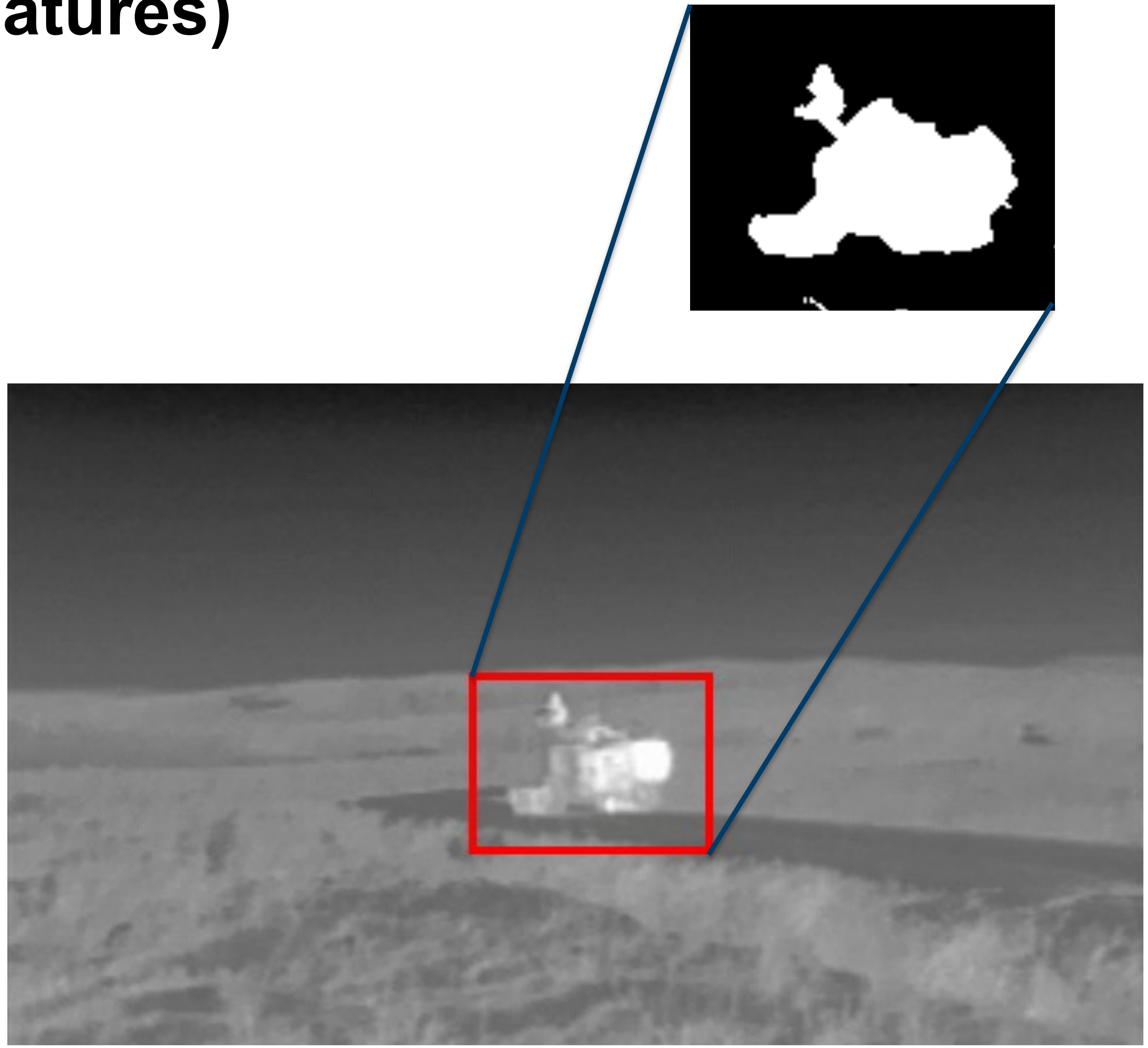
"Curse of dimensionality"



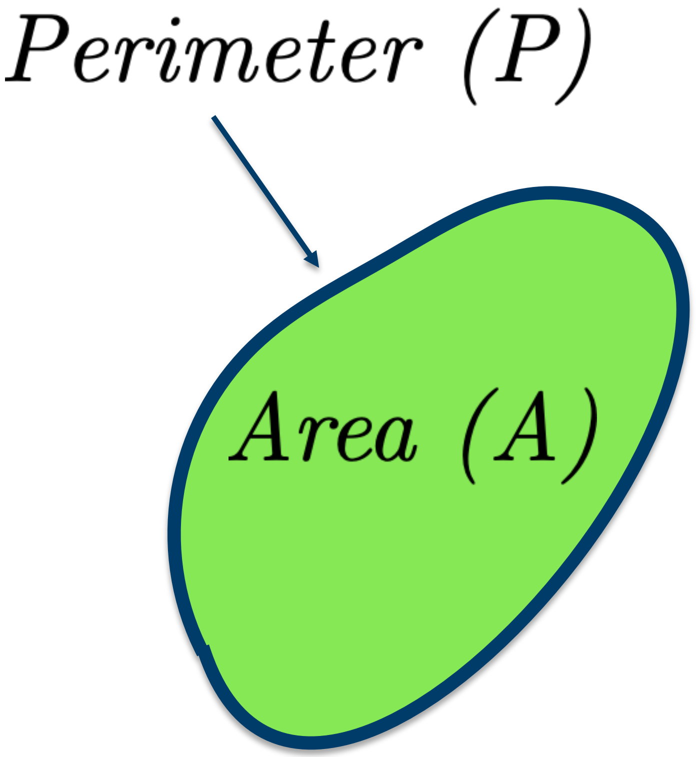
$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix}$$

Feature types (regional features)

- Colour features
- Gray level features
- Shape features
- Histogram (texture) features



Shape features - example



Possible shape feature: $\dfrac{P^2}{A}$

Moments

Geometric moments (order p, q):

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) \mathrm{d}x \mathrm{d}y \approx \sum_i \sum_j i^p j^q f[i,j]$$

Central moments:

$$\mu_{pq} = \sum_i \sum_j (i - \tilde{i})^p (j - \tilde{j})^q f[i,j] \text{ where } \begin{cases} \tilde{i} = \dfrac{m_{10}}{m_{00}} \\ \tilde{j} = \dfrac{m_{01}}{m_{00}} \end{cases}$$

Binary images

$$f[i,j] = \begin{cases} 1 \Rightarrow \text{Object pixel} \\ 0 \Rightarrow \text{Background pixel} \end{cases}$$

$$\text{Area:} \quad m_{00} = \sum_i \sum_j f[i,j]$$

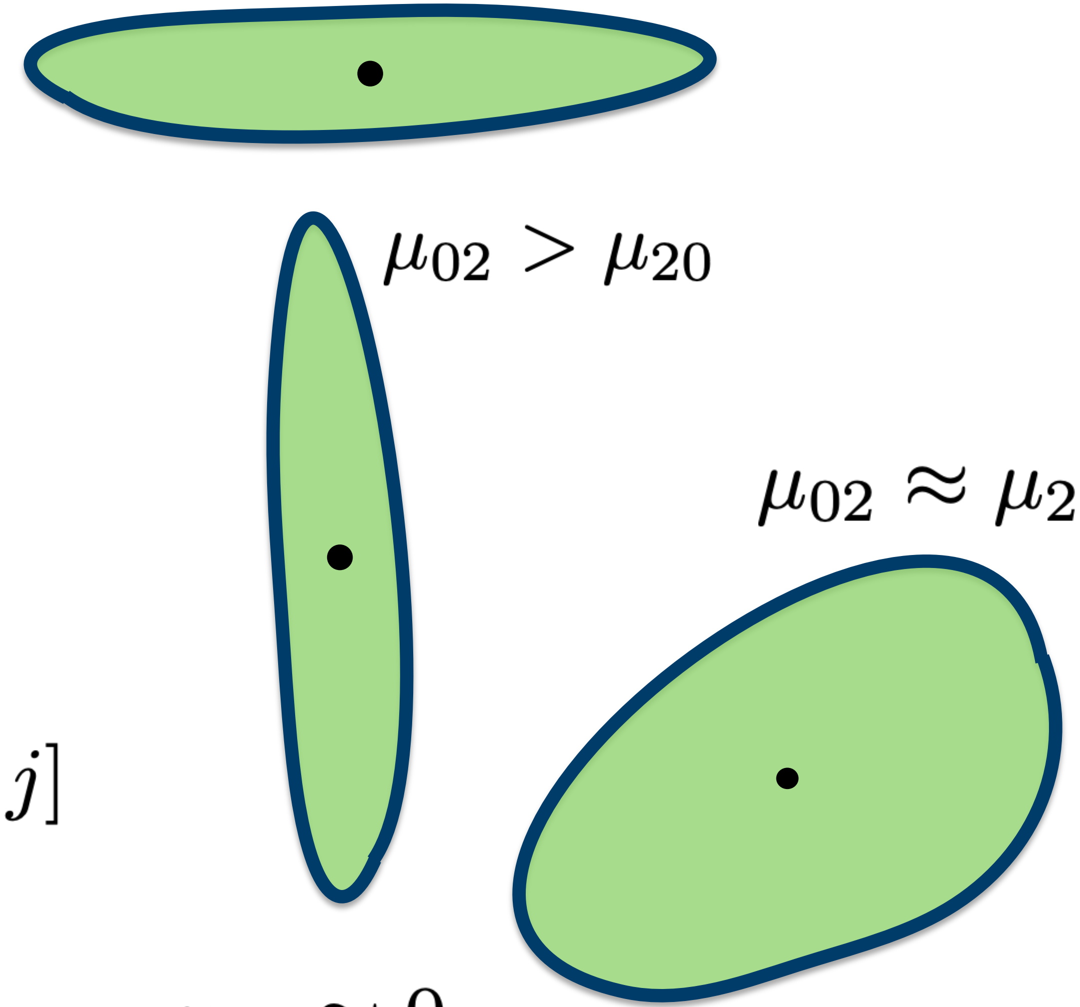
$$\text{Center of mass:} \begin{cases} m_{10} = \sum_i \sum_j i\, f[i,j] \quad \Rightarrow \quad \tilde{i} = \frac{m_{10}}{m_{00}} \\ m_{01} = \sum_i \sum_j j\, f[i,j] \quad \Rightarrow \quad \tilde{j} = \frac{m_{01}}{m_{00}} \end{cases}$$

Moments of inertia

$$\mu_{20} = \sum_i \sum_j (i - \tilde{i})^2 f[i, j]$$

$$\mu_{02} = \sum_i \sum_j (j - \tilde{j})^2 f[i, j]$$

$$\mu_{11} = \sum_i \sum_j (i - \tilde{i})(j - \tilde{j}) f[i, j]$$



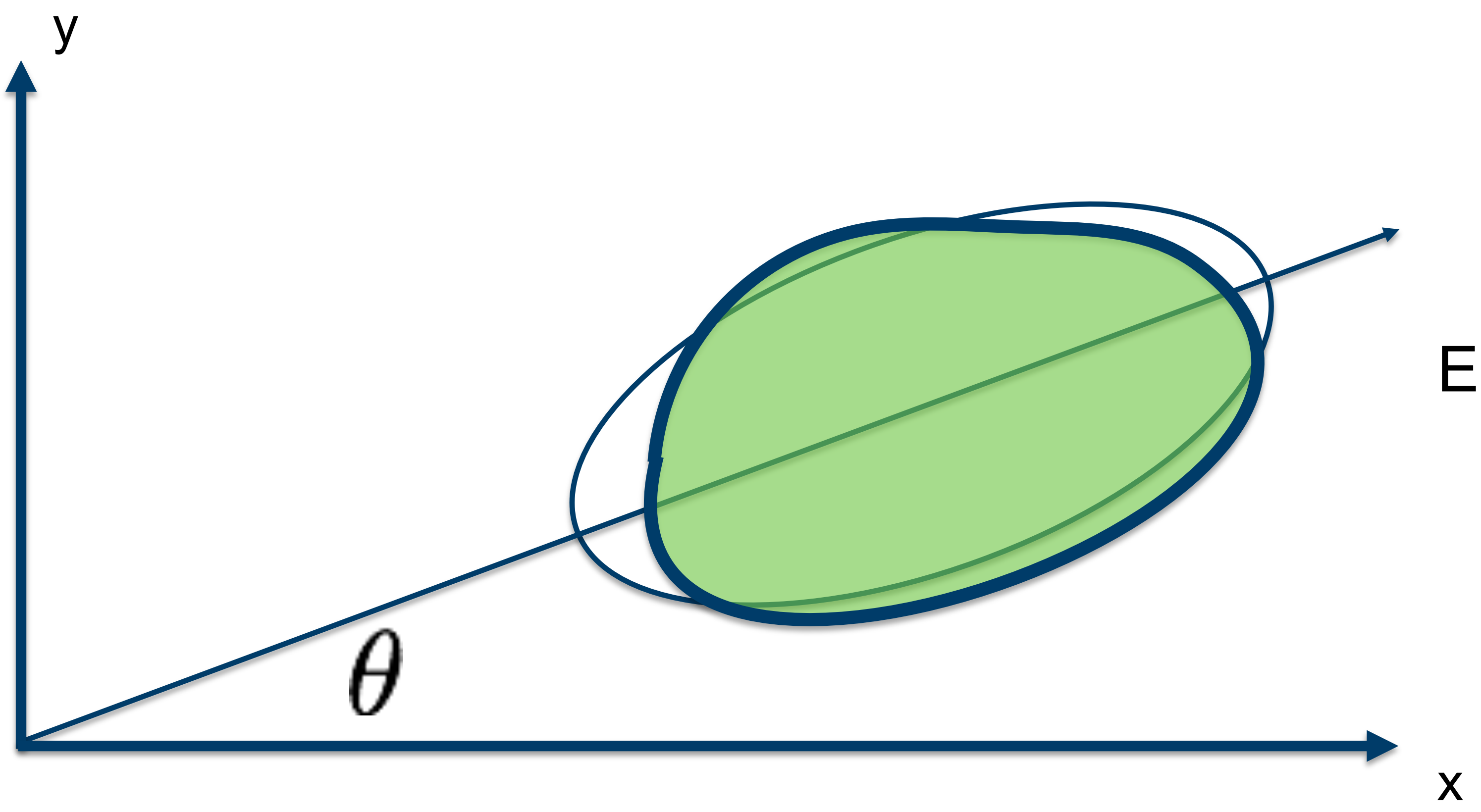
$\mu_{20} > \mu_{02}$

$\mu_{02} > \mu_{20}$

$\mu_{02} \approx \mu_{20}$

$\mu_{11} \approx 0$

Closest fitting ellipse



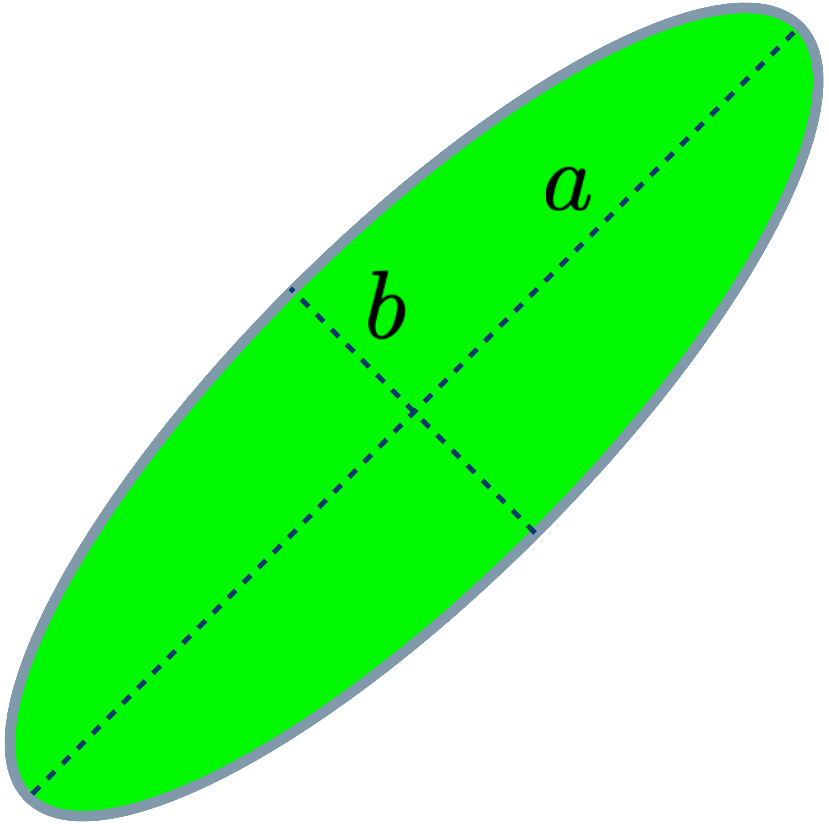
Orientation:

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right]$$

Eccentricity:

$$\epsilon = \frac{(\mu_{20}-\mu_{02})^2 + 4\mu_{11}}{A}$$

Major and minor axes



$$a^2 = \frac{2\left(\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2}\right)}{\mu_{00}}$$

$$b^2 = \frac{2\left(\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2}\right)}{\mu_{00}}$$

Histogram (texture) features

- First order statistics (information related to the gray level distribution)
- Second order statistics (information related to spatial/relative distribution of gray level), i.e. second order histogram, co-occurrence matrix

Histogram:

$$P(I) = \frac{Number\ of\ pixels\ with\ gray\ level\ I}{Total\ number\ of\ pixels\ in\ the\ region}$$

Moments from gray level histogram:

$$m_p = E\{I^p\} \approx \sum_{I=0}^{L-1} I^p P(I), \quad p = 1, 2, \ldots$$

$$m_1 = E(I) = Mean\ value\ of\ I$$

Entropy:

$$H = -E\{\ln P(I)\} \approx -\sum_{I=0}^{L-1} P(I) \ln P(I)$$

Energy:

$$W = E\{P(I)^2\} \approx \sum_{l=0}^{L-1} P(I)^2$$

Histogram (texture) features

Central moments:

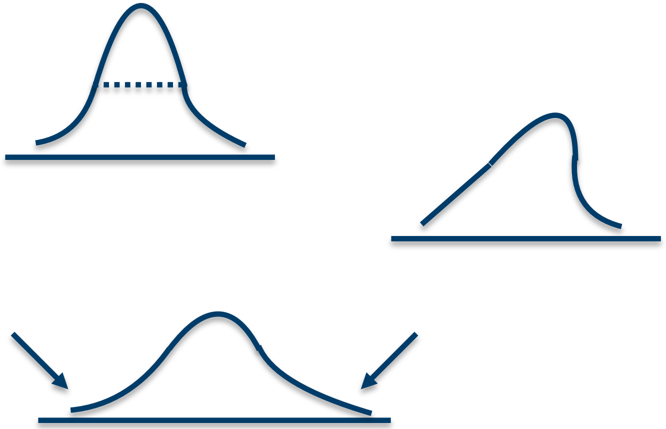
$$\mu_p = E\{(I - E(I))^p\} \approx \sum_{I=0}^{L-1} (I - m_1)^p P(I), \quad p = 1, 2, \ldots$$

Features:

$$\mu_2 = \sigma^2 = variance$$

$$\mu_3/\sigma^3 = skewness$$

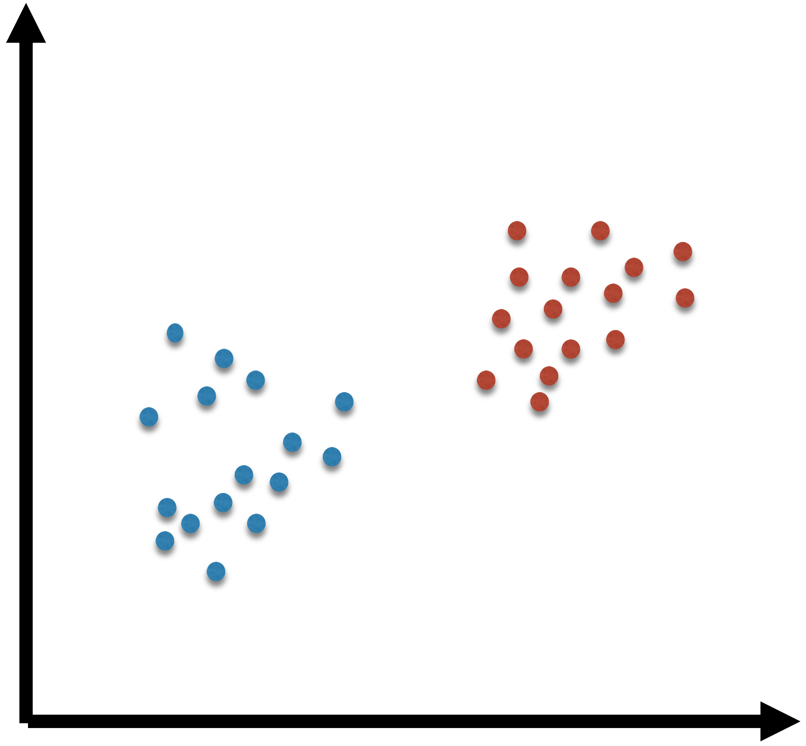
$$\mu_4/\sigma^4 = kurtosis$$



Feature selection

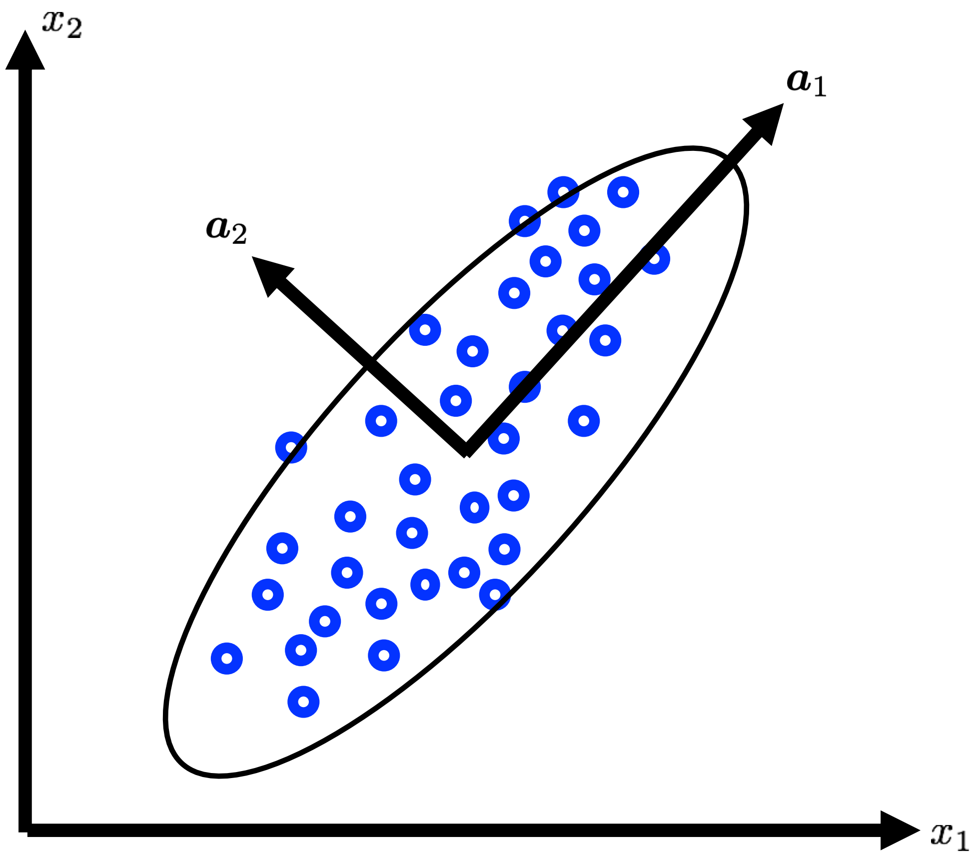
- A number of feature candidates may have been generated.
- Using all candidates will easily lead to over trainig (unreliable classification of new data).
- Dimensionality reduction is required, i.e. feature selection!
- Exhaustive search impossible!
- Trial and error (select feature combination, train classifier, estimate error rate).
- Suboptimal search.
- «Branch and Bound» search.
- Linear or non-linear mappings to lower dimensional feature space.

Scatter plot of features

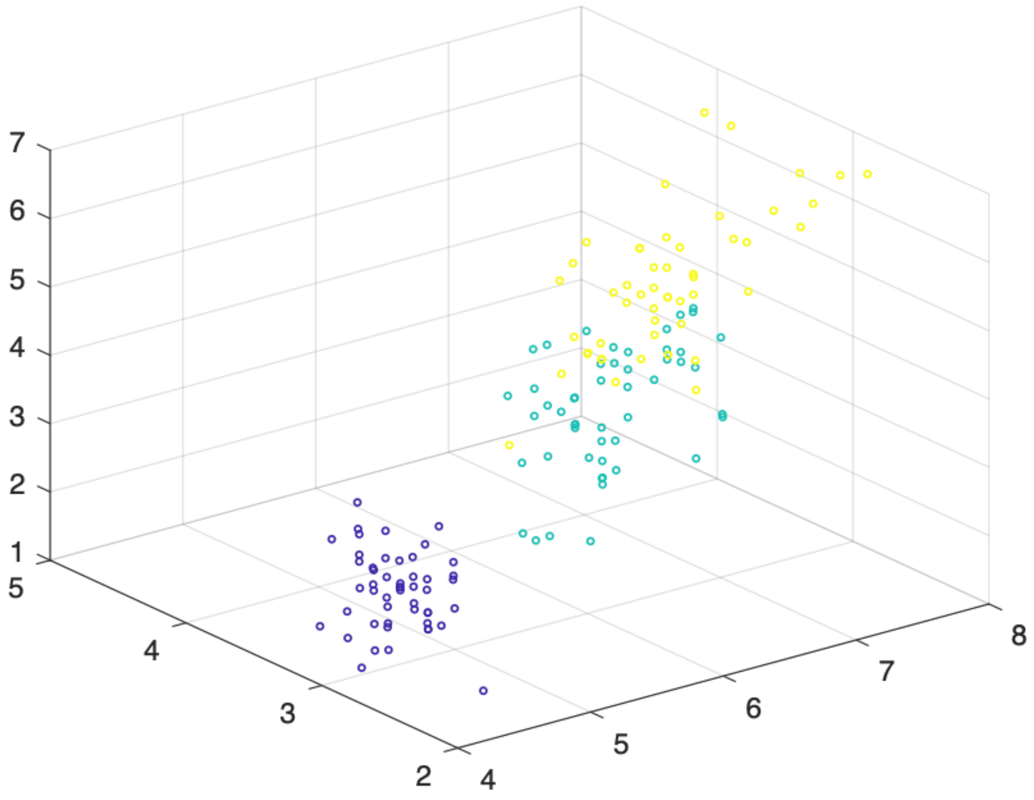


Dimensionality reduction - linear transformations

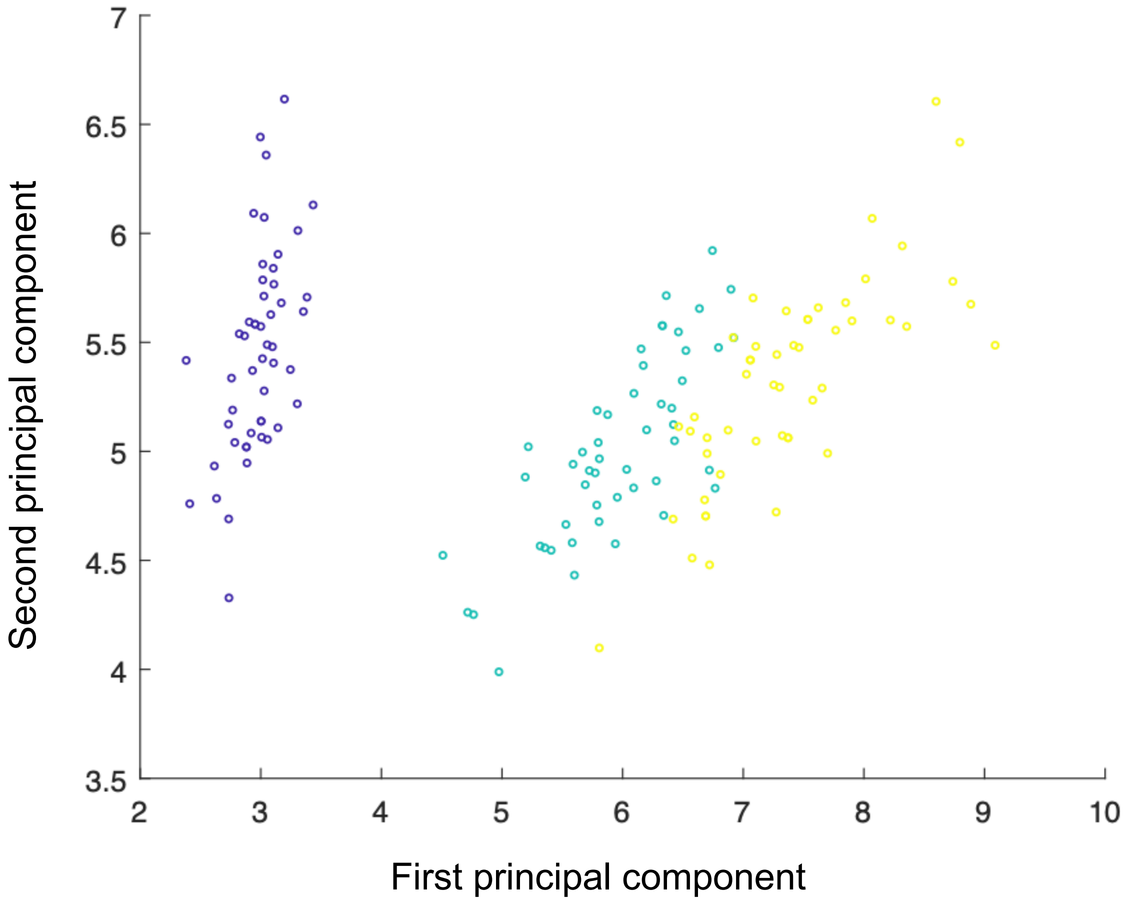
- **PCA**, ICA, LLE, Isomap
- PCA (Principal Components Analysis) is one of the most important techniques for dimensionality reduction
- It takes advantage of correlations between the features to produce the best possible lower dimensional representation of the data with respect to reconstruction error
- The eigenvectors of the lumped covariance matrix defines the new features in the transformed feature space.



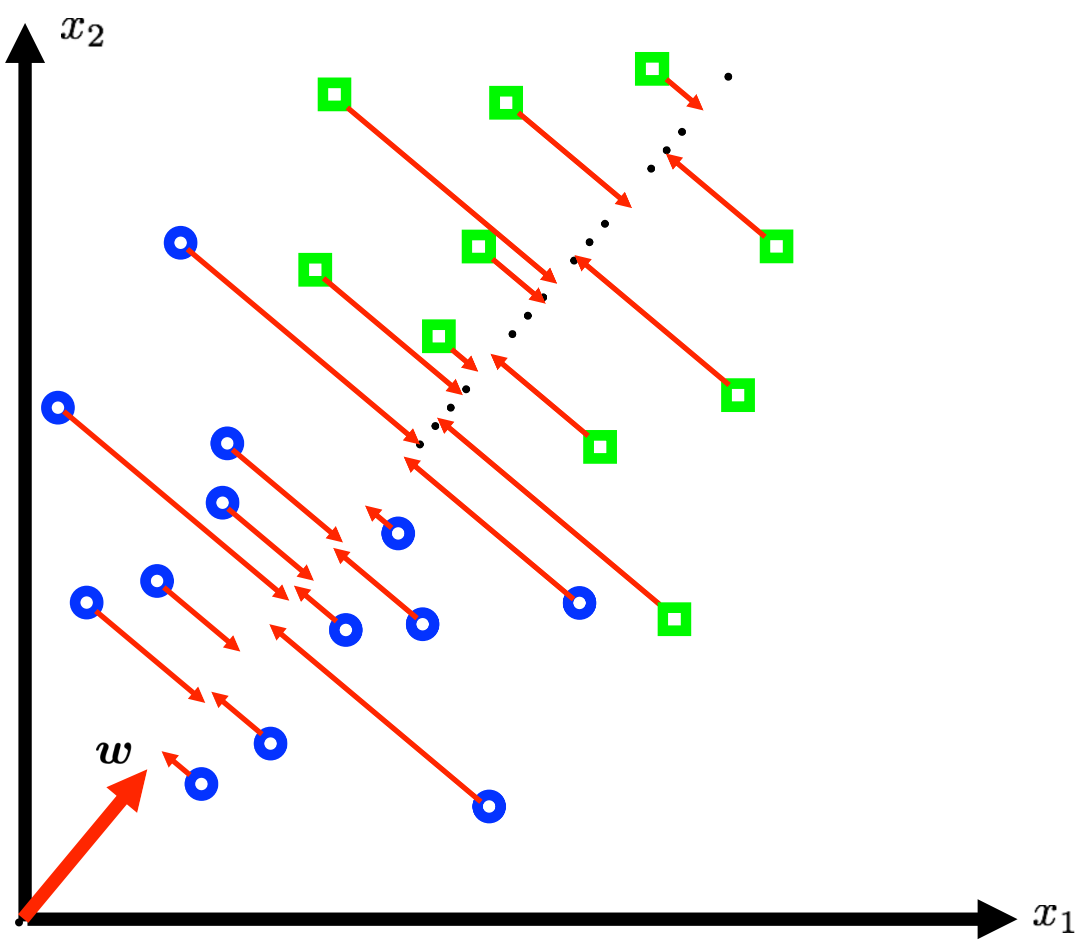
Principle Components Analysis - example



Fisher's Iris dataset (features 1, 2 and 3)



Dimensionality reduction – Fishers linear discriminant



- Projection of multidimensional feature vectors to a lower-dimensional feature space.
- *Fishers linear discriminant* provides a projection from a d-dimensional space (d>1) to a one-dimensional space in such a way that the separation between classes are maximized.

Summary

Image feature extraction:

- Feature extraction
- Feature selection

Additional reading:

- Szeliski 6.2 - 6.3