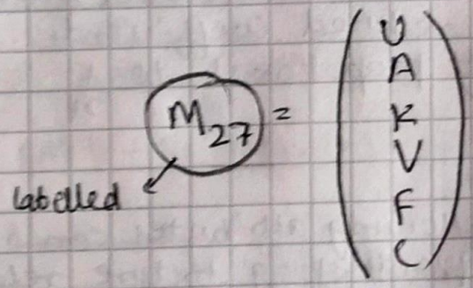
classification

① A process of assigning a given object to a class of similar objects. The metric used for similarity is based on features of the objects.

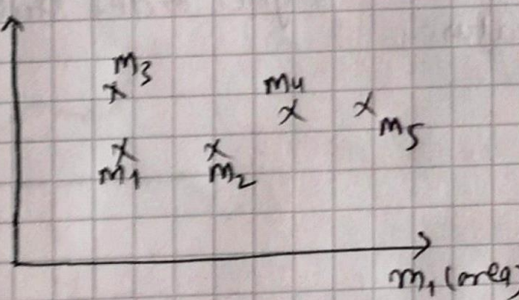
② method → ① First, feature parameters are collected using methods of feature selection and extraction and represented as feature vector.
② Feature vectors are then taken to characterize the objects.

Eg. We have 2 objects with N no. of features each

③ Feature Vectors can have no. of parameters.



④ feature space → the 'N' features in a feature vector span an N-dimensional 'feature space'.



dimension of feature space → no. of feature that i decided to use for my classification

every object ($m_1, m_2 \cdots$) has an area and variance.

problems with feature space → often the N-features are not linearly independent or are not sharply separated.

→ In this case, the dimension of feature space is reduced to the no. of distinctly diff. features.

now → by Karhunen-Loève-Transformation.

⑤ Task of classification → analysis of feature clusters and separation from other clusters.

methods : { distance based classification ("geometrical classifiers")
stochastic classification (by exploiting stochastic properties)
knowledge based classification (eg. fuzzy methods)
Artificial neural networks

feature based
pixel based

⑥ * for training the classifier → sufficiently large and rich dataset.
* the selection can be performed based on goodness criteria.

⑦ mean and spread: ① Two classes are considered. Is a specific feature suitable for separating the classes?

$$G_v = \frac{(\mu_{1v} - \mu_{2v})^2}{\sigma_{1v} - \sigma_{2v}}$$

centre of gravity ↑ with mean and variance:

$$\mu = \frac{1}{n}\sum_i x_i \quad \text{(mean)} \qquad \sigma^2 = \frac{\Sigma(x_i - \mu)^2}{n-1} \quad \text{(variance)}$$

* features with distinctly different means and small variances are good for separating the classes.

⑧ steps for performing classification:
① feature selection
② calculation of features using available training data
③ determination of stochastic properties of the clusters.
④ separation and classification, i.e.
① calculation of n-dimensional feature vectors for an object.
② determination of degree of membership to a certain cluster.

⑨ classifiers: ① Parallelopiped classifier ✓
② minimum distance ✓
③ KNN ✓
④ maximum-likelihood
⑤ NN
⑥ SVMs

Parallelopiped → ① building the clusters
② Drawing a bounding box and calculating the min/max of class limits (boundaries)
③ classification via inclusion / exclusion

pros → ① class boundaries are parallel to coordinate axes
② simple training.

cons → ① useless if there is overlap
② outliers in data can lead to overlap

minimum distance → ① characterizing the classes by their means (centre of mass)
② object is assigned to the class whose mean is closest.

pros → simple calculation, efficient

cons → mean does not suffice to characterize the distribution.

* The application of the minimum distance algorithm is reliable only if all the classes are compact and variances of the feature vectors of the individual classes are about the same.

K-NN Pros → ① non-parametric, thus no training necessary.
② In principle, arbitrarily complex class boundaries possible.

Cons → ① class representants should cover the region well
→ large sample set is necessary.
② search effort of KNN rises with the no. of elements in the clusters (→ using hyper spheres)

Optimal statistical classification → using of Bayes theorem.
→ when we have classes with significant different spreads.