

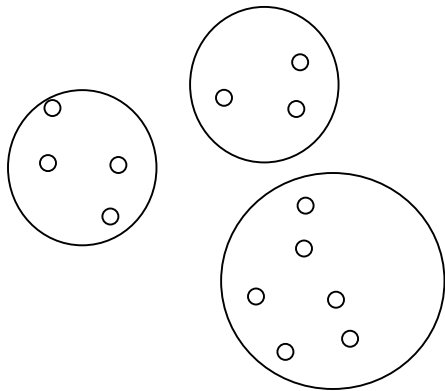
APPLIED MACHINE LEARNING

Classification

Clustering, semi-supervised clustering and classification

Clustering

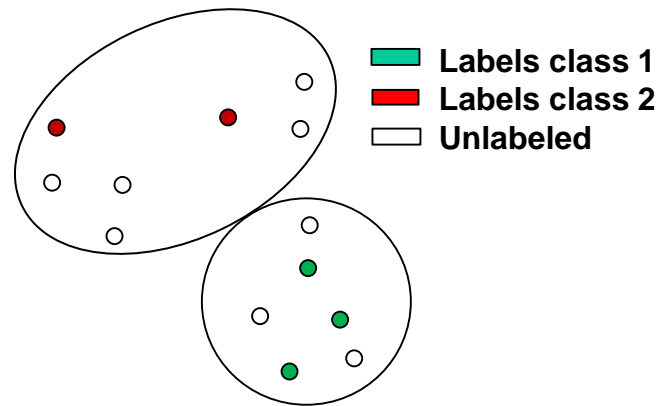
No labels for the points!



Group points according to the geometrical distribution of points

Semi-supervised clustering

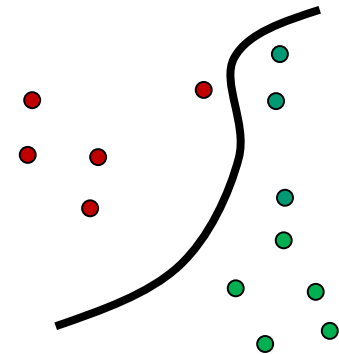
Labels a fraction of the points



Use the labels to choose hyperparameters of clustering using F1-measure.

Classification

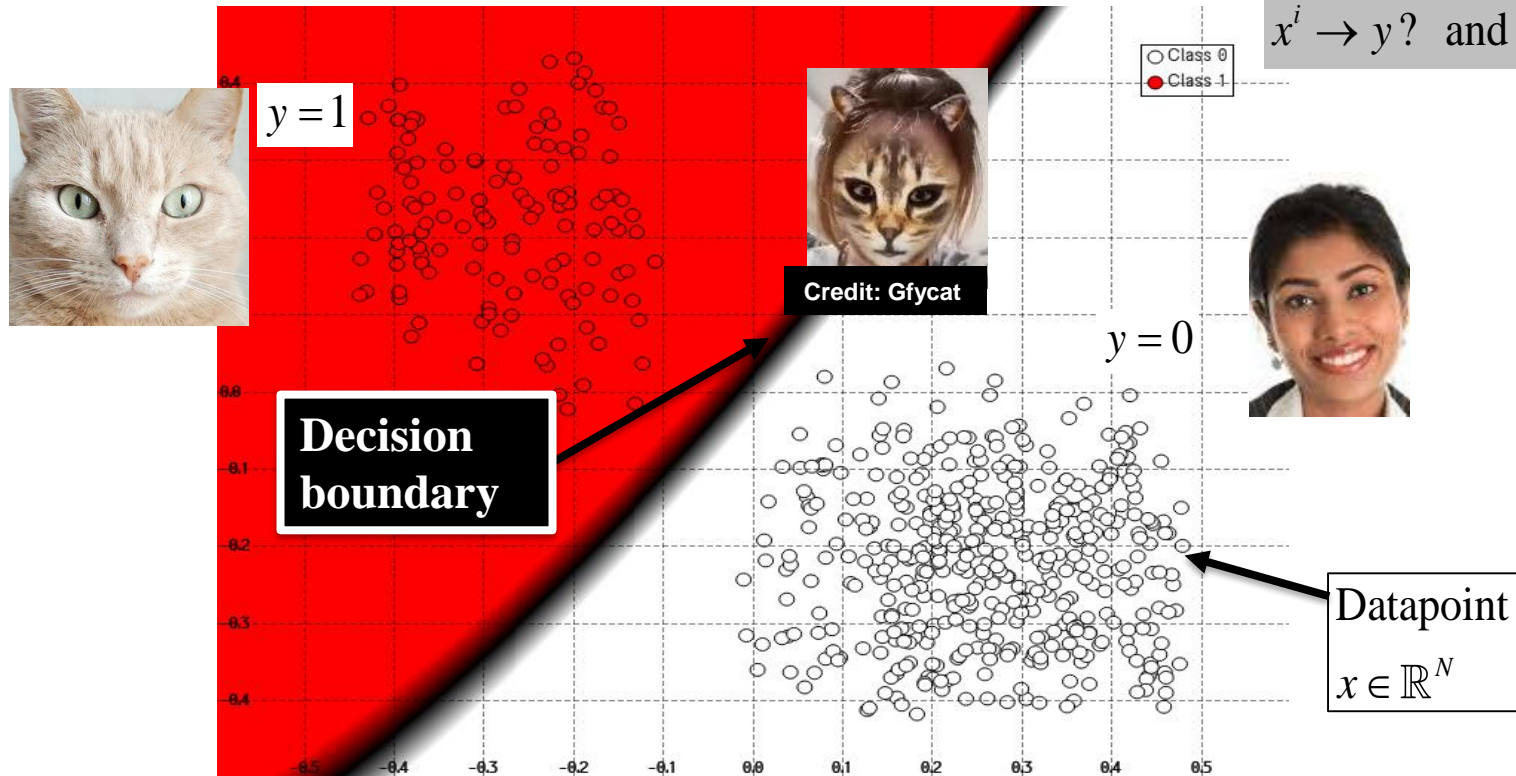
All points are labelled



Use the labels to determine the boundary between the two classes

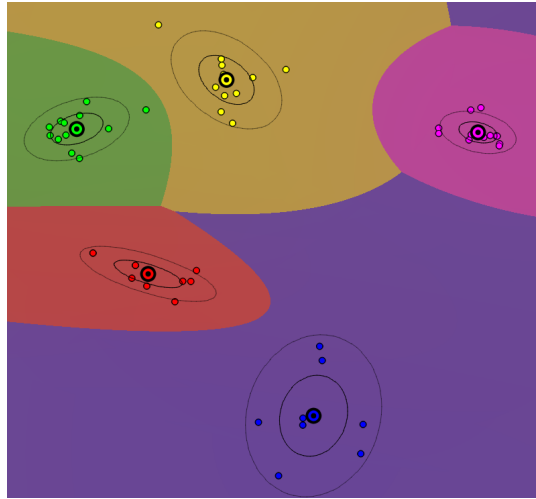
Binary Classification: Formalization

Class label $y \in \mathbb{N}$
 $x^i \rightarrow y?$ and $y = \{0,1\}$

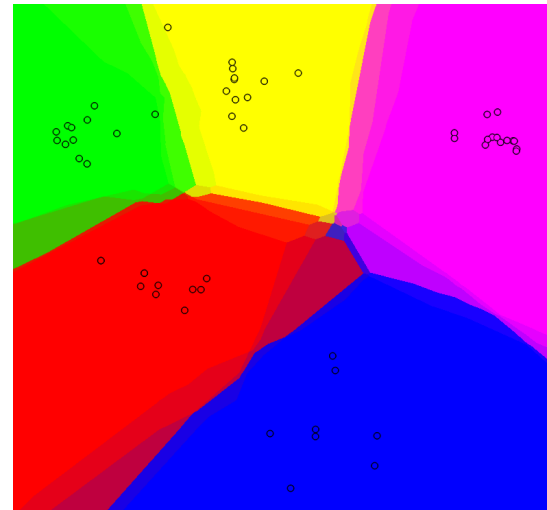


- ❑ Classification is a supervised learning technique
 - Number and class types known, class label given for all training points.
- ❑ Class label is discrete.
- ❑ Goal: find a decision boundary which separates at best the classes.

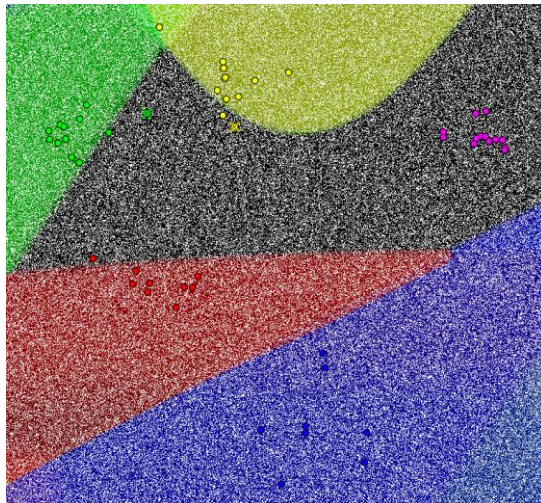
Classification techniques covered in applied machine learning course



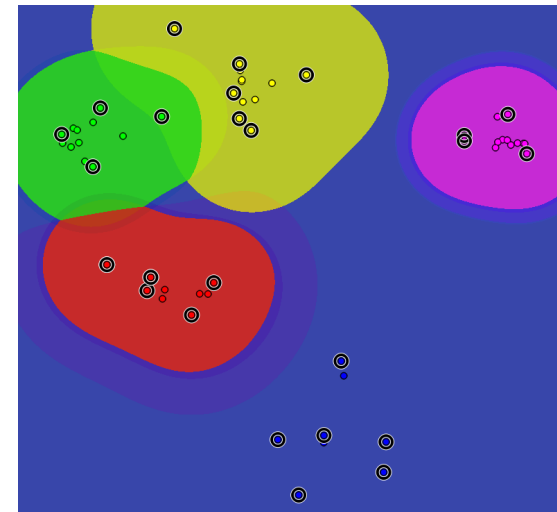
Gaussian Mixture Models (GMM)



K-nearest neighbors (KNN)



Neural Networks (NN)



Support Vector Machine (SVM)

People Detection : /Users/basilio/Desktop/Videos/Balloons1.mp4

Dataset Creation Training/Testing

Gaussian Mixture Models (SLOW TRAINING)

Mixture Count: 1

Cov.Mat. Complexity: 3

Train/Test ratio: 100

Train Validate

Detection Threshold: 0.600

Gaussian Mixture Models:
 Mixture Count: determines the number of gaussians per mixture per class (determines how "complex" the model can be to represent one class)
 Covariance Matrix Complexity: determines the complexity of the covariance matrix should have (1: spherical, 2: diagonal, 3: full)

The screenshot shows a video player interface with a 'Dataset Creation' sidebar on the left. The sidebar is set to 'Gaussian Mixture Models (SLOW TRAINING)' with 'Mixture Count' at 1 and 'Cov.Mat. Complexity' at 3. The 'Train/Test ratio' is 100. The main video area shows a scene with a woman in a floral dress and a red and white van. Numerous red bounding boxes are overlaid on the scene, each containing a small image of a balloon and a percentage value (e.g., 87%, 71%, 22%, 84%, 44%, 25%, 48%, 81%, 20%). A large green arrow points from the top of the van towards the bottom of the frame.

/Users/basilio/Desktop/Videos/Balloons1.mp4

Open Video 00:00:00:11 (1/1)

Make Image Brighter Make Image Darker

Play

Time 1 to 2 (3 sec)

People Detection : /Users/basilio/Desktop/Videos/Balloons1.mp4

Dataset Creation Training/Testing

Support Vector Machines

Kernel Width: 0.250

C-penalty: 100

Train/Test ratio: 100

Train Validate

SVM with 33 Support Vectors

Detection Threshold: 0.600

Support Vector Machines:
 Kernel Width: determines the size of the kernel (how smoothly the objective function climbs around the support vectors)
 C-Penalty: penalty incurred for wrongly classifying a training sample (trade-off between overfitting and overgeneralising)

This screenshot is similar to the first one but uses 'Support Vector Machines' for detection. The sidebar settings are 'Support Vector Machines', 'Kernel Width' at 0.250, and 'C-penalty' at 100. The 'Train/Test ratio' remains 100. The video content and the red bounding boxes around balloons are identical to the first screenshot. A large green arrow points from the top of the van towards the bottom of the frame. At the bottom right of the interface, there are buttons for 'Screenshot', 'Test on current image', 'Continue Training (SVM)', 'Batch Process', and 'Export Video'.

Time 1 to 2 (3 sec)