Mean Shift Clustering

The mean shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters.

Given n data points $\mathbf{x}_i$, $i = 1, ..., n$ on a d-dimensional space R^d, the multivariate kernel density estimate obtained with kernel $K(\mathbf{x})$ and window radius h is

$$f(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right). \tag{1}$$

For radially symmetric kernels, it suffices to define the profile of the kernel $k(\mathbf{x})$ satisfying

$$K(\mathbf{x}) = c_{k,d}k(\|\mathbf{x}\|^2) \tag{2}$$

where $c_{k,d}$ is a normalization constant which assures $K(\mathbf{x})$ integrates to 1. The modes of the density function are located at the zeros of the gradient function $\nabla f(\mathbf{x}) = 0$.

The gradient of the density estimator (1) is

$$\begin{aligned} \nabla f(\mathbf{x}) &= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{x}) g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right) \\ &= \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)\right] \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}\right]. \end{aligned} \tag{3}$$

where $g(s) = -k'(s)$. The first term is proportional to the density estimate at $\mathbf{x}$ computed with kernel $G(\mathbf{x}) = c_{g,d}g(\|\mathbf{x}\|^2)$ and the second term

$$\mathbf{m}_h(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x} \tag{4}$$

is the *mean shift*. The mean shift vector always points toward the direction of the maximum increase in the density. The mean shift procedure, obtained by successive

- computation of the mean shift vector $\mathbf{m}_h(\mathbf{x}^t)$,
- translation of the window $\mathbf{x}^{t+1} = \mathbf{x}^t + \mathbf{m}_h(\mathbf{x}^t)$

is guaranteed to converge to a point where the gradient of density function is zero. Mean shift mode finding process is illustrated in Figure 1.

The mean shift clustering algorithm is a practical application of the mode finding procedure:

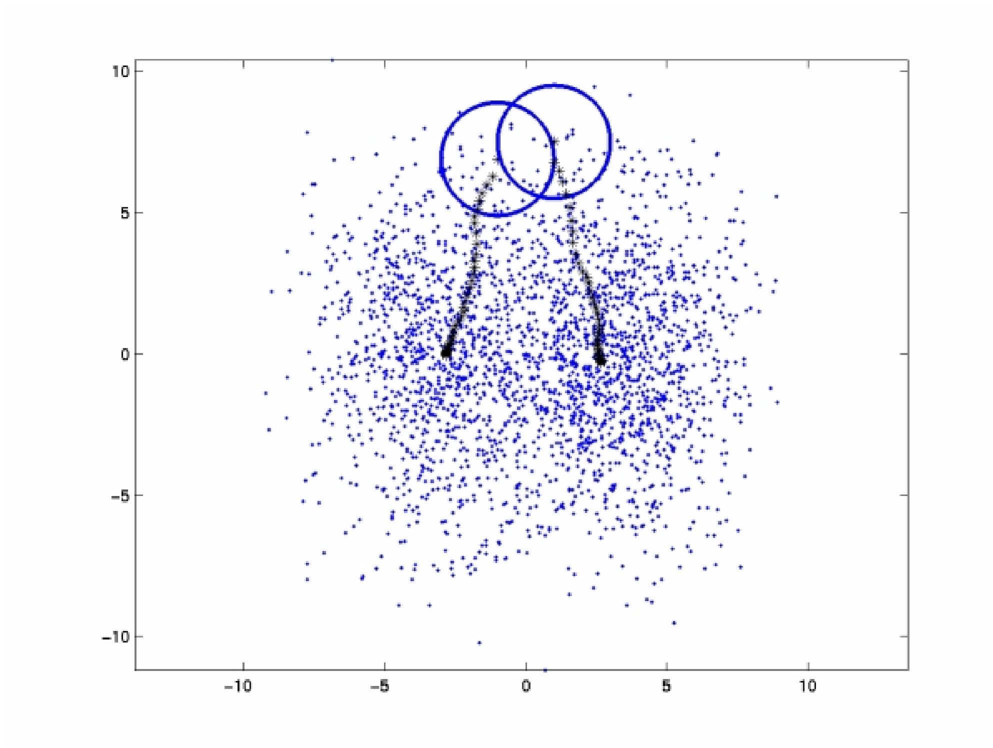


Figure 1: **Mean Shift Mode Finding**

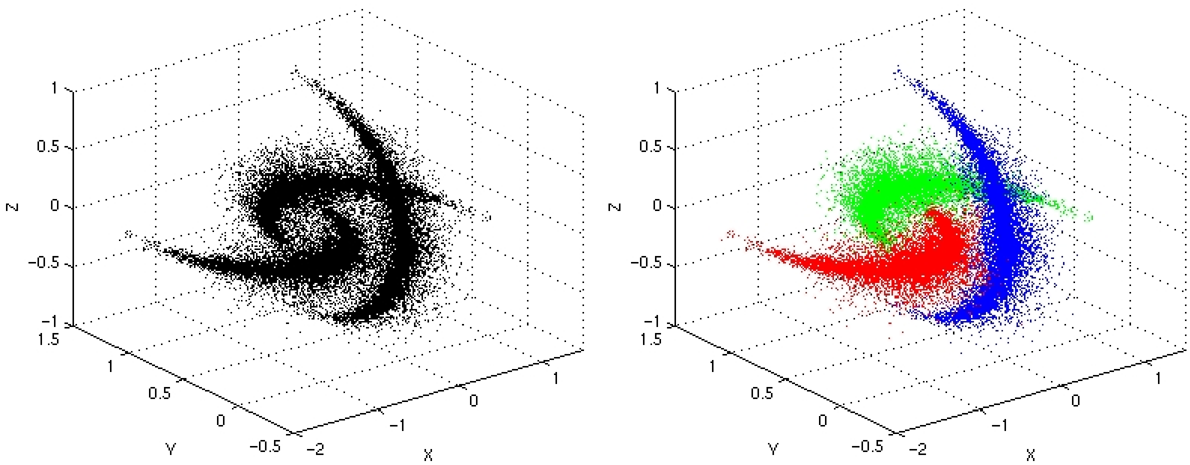
- starting on the data points, run mean shift procedure to find the stationary points of the density function,
- prune these points by retaining only the local maxima.

The set of all locations that converge to the same mode defines the *basin of attraction* of that mode. The points which are in the same basin of attraction is associated with the same cluster. Figure 2 shows two examples of mean shift clustering on three dimensional data.

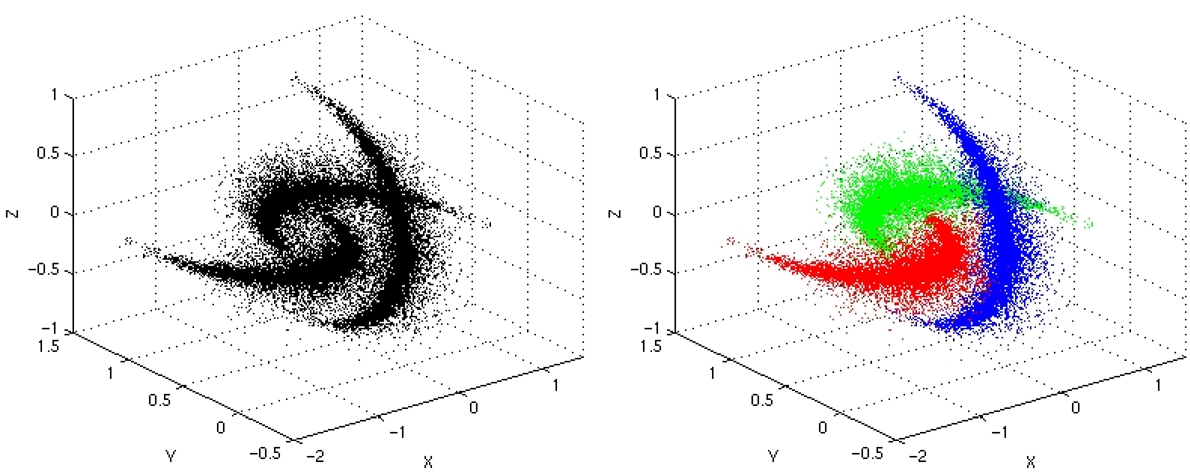
More details on mean shift clustering on Lie Groups can be found in [1].

References

[1] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24:603–619, 2002. Available at `http://www.caip.rutgers.edu/riul/research/papers/pdf/mnshft.pdf`.



(a) Synthetic example of three non-linearly separable clusters (32640 points).



(b) Real example of 14826 points in the LUV color space.

Figure 2: **Mean Shift Clustering**