Mean Shift Algorithm

1. Motives
2. Mean Shift (high level)
3. Kernel Density
4. Mean Shift (details)

Mean Shift Algorithm

Motives

- Clustering (unsupervised)
- Arbitrarily shaped clusters
  - KMeans: Spheres
  - Gaussian Mixtures: Ovals

KMeans

Assumes sphere-shaped clusters:

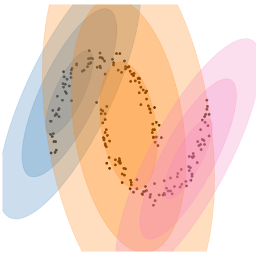
$$\text{cluster}(x) = \underset{c \in C}{\text{argmin}} \, ||x - c||$$

Gaussian Mixture Models

Assumes oval-shaped clusters:

$$\text{cluster}(x) = \underset{c \in C}{\text{argmax}}\; f_c(x)$$

$$c \sim \mathcal{N}(\mu_c, \mathbf{\Sigma}_c)$$



Parametric vs. Nonparametric Models

Parametric Models:

- Described with a fixed number of parameters
- After training: Make predictions from the parameters
- e.g. KMeans, GMMs (unsupervised)
- e.g. Linear Regression, Support Vector Machines (supervised)

Parametric vs. Nonparametric Models

Nonparametric Models:

- Cannot be described with a fixed number of parameters
- No training step: Make predictions directly from training data
- + flexibility, - interpretability
- e.g. MeanShift (unsupervised)
- e.g. LOESS (supervised)

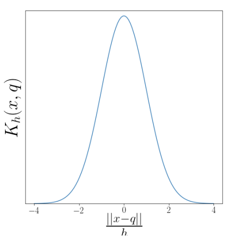
Mean Shift Algorithm

Introduction

Prerequisite: Weighted Mean

$$\mu = \frac{\sum w_i \mathbf{x}_i}{\sum w_i}$$

Prerequisite: Kernels



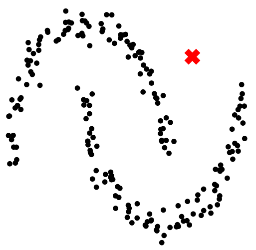
Gaussian (RBF) Kernel

$$K_h(\mathbf{q}, \mathbf{x}) \propto e^{-\frac{1}{2}\left(\frac{||\mathbf{x}-\mathbf{q}||}{h}\right)^2}$$

h is the "bandwidth"

iteration -1
Close Controls

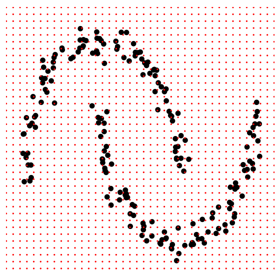
MeanShift



$$\mathbf{q} \leftarrow \frac{\sum_{\mathbf{x}\in X} \mathbf{x} K_h(\mathbf{q},\mathbf{x})}{\sum_{\mathbf{x}\in X} K_h(\mathbf{q},\mathbf{x})}$$

iteration 0
Close Controls

MeanShift



$$\mathbf{q} \leftarrow \frac{\sum_{\mathbf{x}\in X} \mathbf{x} K_h(\mathbf{q},\mathbf{x})}{\sum_{\mathbf{x}\in X} K_h(\mathbf{q},\mathbf{x})}$$

bandwidth 0
Close Controls

Kernel Density

$$\mathbb{P}(\mathbf{q}) = \sum_{\mathbf{x}\in X} K_h(\mathbf{q},\mathbf{x})$$

Kernel Density (Gaussian Kernel)

$$\mathbb{P}(\mathbf{q}) = \sum_{\mathbf{x} \in X} K_h(\mathbf{q}, \mathbf{x})$$

$$\mathbb{P}(\mathbf{q}) = \sum_{\mathbf{x} \in X} e^{-\frac{1}{2}\left(\frac{||\mathbf{x}-\mathbf{q}||}{h}\right)^2}$$

$$\frac{\partial \mathbb{P}}{\partial \mathbf{q}} = \sum_{\mathbf{x} \in X} K_h(\mathbf{q}, \mathbf{x}) \frac{\partial}{\partial \mathbf{q}} \frac{-1}{2} \left(\frac{||\mathbf{x}-\mathbf{q}||}{h}\right)^2$$

$$||\mathbf{x}-\mathbf{q}||^2 = \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{q} + \mathbf{q}^T\mathbf{q}$$

$$\frac{\partial \mathbb{P}}{\partial \mathbf{q}} = \frac{1}{h^2} \sum_{\mathbf{x} \in X} K_h(\mathbf{q}, \mathbf{x})(\mathbf{x}-\mathbf{q})$$

Kernel Density (Gaussian Kernel)

$$\frac{\partial \mathbb{P}}{\partial \mathbf{q}} = \frac{1}{h^2} \sum_{\mathbf{x} \in X} K_h(\mathbf{q}, \mathbf{x})(\mathbf{x}-\mathbf{q})$$

$$\frac{\sum_{\mathbf{x} \in X} \mathbf{x} K_h(\mathbf{q}, \mathbf{x})}{\sum_{\mathbf{x} \in X} K_h(\mathbf{q}, \mathbf{x})} = \mathbf{q} + \frac{\partial \mathbb{P}}{\partial \mathbf{q}} \frac{h^2}{\sum_{\mathbf{x} \in X} K_h(\mathbf{q}, \mathbf{x})}$$

Mean Shift == Gradient Ascent on the Kernel Density Function!

Convergence of Mean Shift

¯_(ツ)_/¯

Recap

Mean Shift:

- Clustering (unsupervised)
- Arbitrarily shaped clusters
- Probably definitely converges
- Gives the Modes of the Kernel Density Function