PA 2018 04

Small addendum to density estimation

Converting the unsupervised Density estimation task to a supervised task(from HTF sec 14.2.4)

⇒ Perform a regression on the joint sample space to model the density (relative to the reference density) (14.2.4)

Mean Shift Algorithm

Method for finding the modes of a PDF.
Applications are essentially applications of clustering:
- data exploration
- image segmentation,denoising,...
- dimensionality/complexity reduction

Let $S = x_1, x_2, \ldots, x_N$ denote N d-dimensional observations x_i. The multivariate kernel estimate is $p(x) = \frac{1}{N} * \sum_{i=1}^{N} K_\lambda(x_i, x)$
The maxima of p(x) are characterized by the locations x where $\nabla p(x)$ vanishes.

$\nabla p(x) = \nabla \frac{1}{N} * \sum_{i=1}^{N} K_\lambda(x_i, x) = \frac{1}{N} * \sum_{i=1}^{N} \nabla K_\lambda(x_i, x)$

Let us assume that the kernel is radial symmetric
$K_\lambda(x_i, x) = C_d * k_lambda(||x_i - x||_2^2)$
Derive the kernel:
$\frac{\partial k_\lambda(s)}{\partial s} = k'_\lambda(s)$ derivative of the kernel

$\frac{\partial s}{\partial x} = \frac{(x_i - x)^T (x_i - x)}{x} = 2(x_i - x)$ application of the chain rule

$\nabla p(x) = \frac{1}{N} * \sum_{i=1}^{N} C_d * k'_\lambda(||x_i - x||_2^2) * (-2(x_i - x)) = 0 \Leftrightarrow \sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2) * x_i - \sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2) * x = 0$

$\Rightarrow \frac{\sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2) * x_i}{\sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2)} - x = 0$ Mean shift equation

To perform gradient ascent, we have to compute the gradient and to "walk upwards " along the gradient direction until the gradient vanishes

Algorithm:
1) compute the mean shift rector
$m(x^{(t)}) = \frac{\sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2) * x_i}{\sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2)} - x = 0$

2) update $x^{(t+1)} = x^{(t)} + m(x^{(t)}) = \frac{\sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2) * x_i}{\sum_{i=1}^{N} k'_\lambda(||x_i - x||_2^2)}$

Why "MEAN SHIFT"?

If the kernel is chosen as the Epanechnikov kernel, then the update simplifies to computing the weighted "mean" of all feature vectors a sphere around x

Epanechnikov Kernel: $K_E(x) = c * (1 - x^T x) \; if \; x^T x \;\leq\; 1 \; otherwise \; 0$

Paper: Comaniciu, Meer:"Mean Shift: A robust Approach Toward Feature Space Analysis"