Machine Learning 2
Exercise Sheet 7

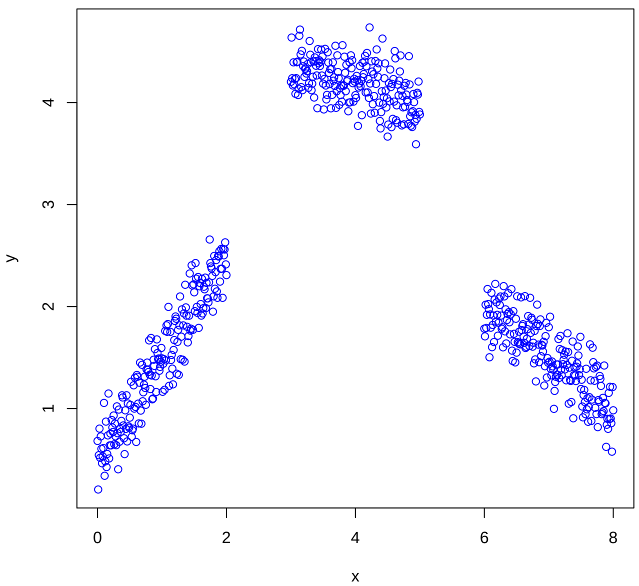
Prof. Dr. Dr. Lars Schmidt-Thieme, Nicolas Schilling
Information Systems and Machine Learning Lab
University of Hildesheim

June 1st, 2016

Submission until June 8th, 18.00 to schilling@ismll.de

Exercise 12: Mixture of Linear Regressions (6 Points)

In R, create a univariate data set that looks approximately like this:



a) Think of three linear functions $\gamma_i^\top x$ for $i = 1, 2, 3$ such that their softmax is as gating function for the respective data clouds.

b) Learn three (unregularized) linear regressions (one per data cloud) and use the gating function of a) to build a mixture of experts. Use the mixture of experts on all data points and plot the result.

Exercise 13: Product of Gaussian Process Experts (6 Points)

Product of Experts work similar to mixtures of experts. They split the data into n subsets as:

$$X = (X^{(1)}, ..., X^{(n)}) \qquad y = (y^{(1)}, ..., y^{(n)}) \ ,$$

and then model the joint likelihood of y given univariate data X and model parameters θ as a product of individual experts:

$$p(y \,|\, X, \theta) = \prod_{i=1}^{n} p_i \left(y^{(i)} \,|\, X^{(i)}, \theta^{(i)} \right) \ .$$

Show that for p_i being the predictive density of a Gaussian process, the product of two GP experts for a new instance $x_\star$ is **proportional to** a Gaussian Distribution with mean $\mu_\star$ and variance $\sigma_\star^2$:

$$\mu_\star = \sigma_\star^2(\sigma_1^{-2}\mu_1 + \sigma_2^{-2}\mu_2)$$
$$\sigma_\star^2 = (\sigma_1^{-2} + \sigma_2^{-2})^{-1}$$

Note: Of course μ_i and σ_i are also dependent on $x^\star$! It is simply not written because then formulas are too cumbersome.