Machine Learning Group
Faculty IV – Electrical Engineering and Computer Science
Technische Universität Berlin

Exercise Sheet 11

Exercise 1: Mixture Density Networks (30 + 30 + 40 P)

In this exercise, we prove some of the results from the paper Mixture Density Networks by Bishop (1994). The mixture density network is given by

$$p(\boldsymbol{t}|\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i(\boldsymbol{x})\phi_i(\boldsymbol{t}|\boldsymbol{x})$$

with the mixture elements

$$\phi_i(\boldsymbol{t}|\boldsymbol{x}) = \frac{1}{(2\pi)^{c/2}\sigma_i(\boldsymbol{x})^c} \exp\Big(-\frac{\|\boldsymbol{t} - \boldsymbol{\mu}_i(\boldsymbol{x})\|^2}{2\sigma_i(\boldsymbol{x})^2}\Big).$$

The contribution to the error function of one data point q is given by

$$E^q = -\log\Big\{\sum_{i=1}^{m} \alpha_i(\boldsymbol{x}^q)\phi_i(\boldsymbol{t}^q|\boldsymbol{x}^q)\Big\}$$

We also define the posterior distribution

$$\pi_i(\boldsymbol{x}, \boldsymbol{t}) = \frac{\alpha_i\phi_i}{\sum_{j=1}^{m} \alpha_j\phi_j}$$

which is obtained using the Bayes theorem. We would like to compute the gradient of the error E^q w.r.t. the mixture parameters

(a) *Show* that $\dfrac{\partial E^q}{\partial \alpha_i} = -\dfrac{\pi_i}{\alpha_i}$

(b) Show that $\dfrac{\partial E^q}{\partial \mu_{ik}} = \pi_i\Big(\dfrac{\mu_{ik} - t_k}{\sigma_i^2}\Big)$

(c) We now assume that the neural network produces the mixture coefficients as:

$$\alpha_i = \frac{\exp(z_i^\alpha)}{\sum_{j=1}^{M} \exp(z_j^\alpha)}$$

where z^α denotes the outputs of the neural network (after the last linear layer) associated to these mixture coefficients. *Compute* using the chain rule for derivatives (i.e. by reusing some of the results in the first part of this exercise) the derivative $\partial E^q/\partial z_i^\alpha$.