Machine Learning—Tutorial Notes

Feed Forward Neural Networks

1 Multiple targets

Problem 1: Consider a regression problem involving multiple target variables in which it is assumed that the distribution of the targets, conditioned on the input vector $\boldsymbol{x}$, is a Gaussian of the form

$$p(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{w}) = \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{y}(\boldsymbol{x}, \boldsymbol{w}), \boldsymbol{\Sigma})$$

where $\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{w})$ is the output of a neural network with input vector $\boldsymbol{x}$ and a weight vector $\boldsymbol{w}$, and $\boldsymbol{\Sigma}$ is the covariance of the assumed Gaussian noise on the targets. Given a set of independent observations of $\boldsymbol{x}$ and $\boldsymbol{z}$, write down the error function that must be minimised in order to find the maximum likelihood solution for $\boldsymbol{w}$, if we assume that $\boldsymbol{\Sigma}$ is fixed and known. Now assume that $\boldsymbol{\Sigma}$ is also to be determined from the data and write down an expression for the maximum likelihood solution for $\boldsymbol{\Sigma}$. Note that the optimisations of $\boldsymbol{w}$ and $\boldsymbol{\Sigma}$ are now coupled, in contrast to the case of independent target variables discussed in the exercise above.

2 Error functions

Problem 2: Show that maximising likelihood for a multi-class neural network model in which the network outputs have the interpretation $y_k(\boldsymbol{x}, \boldsymbol{w}) = p(z_k = 1 \mid \boldsymbol{x})$ is equivalent to the minimisation of the cross-entropy error function.

Problem 3: Show that the derivative of the error function

$$E(\boldsymbol{w}) = -\sum_{n=1}^{N} \left\{ z_n \ln y_n + (1 - z_n) \ln(1 - y_n) \right\}$$

(y_n denotes $y(\boldsymbol{x}_n, \boldsymbol{w})$) with respect to the activation a_k for the output unit having a logistic sigmoid activation function satisfies

$$\frac{\partial E}{\partial a_k} = y_k - z_k$$

Problem 4: Show that the derivative of the standard multi-class error function

$$E(\boldsymbol{w}) = -\sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \ln y_k(\boldsymbol{x}_n, \boldsymbol{w})$$

with respect to the activation a_k for output units having a softmax activation function satisfies

$$\frac{\partial E}{\partial a_k} = \sum_{n=1}^{N} y_k(\boldsymbol{x}_n, \boldsymbol{w}) - z_{nk}$$

3 Robust classification

Problem 5: Consider a binary classification problem in which the target values are $z \in \{0, 1\}$, with a network output $y(\boldsymbol{x}, \boldsymbol{w})$ that represents $p(z = 1 \mid \boldsymbol{x})$, and suppose that there is a probability ε that the class label on a training data point has been incorrectly set. Assuming independent and identically distributed data, write down the error function corresponding to the negative log likelihood. Verify that the well known error function for binary classification is obtained when $\varepsilon = 0$. Note that this error function makes the model robust to incorrectly labelled data, in contrast to the usual error function.