Machine Learning Worksheet 12

Neural Networks

Problem 1: Consider a two-layer network function of the form in which the hidden-unit nonlinear activation functions $\phi(\cdot)$ are given by logistic sigmoid functions of the form

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

Show that there exists an equivalent network, which computes exactly the same function, but with hidden unit activation functions given by $\tanh(x)$.

Problem 2: Show that the derivative of the logistic sigmoid activation function can be expressed in terms of the function value itself. Also derive the corresponding result for the tanh activation function.

Problem 3: If we have multiple target variables $\boldsymbol{z}$, and we assume that they are independent conditional on $\boldsymbol{x}$ and $\boldsymbol{w}$ with shared noise precision β then the conditional distribution of the target values is given by

$$p(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{w}) = \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{y}(\boldsymbol{x}, \boldsymbol{w}), \beta^{-1}\boldsymbol{I})$$

Show that minimising the resulting likelihood function under the above conditional distribution for a multi-output neural network is equivalent to minimising a sum-of-squares error function.

Problem 4: In the uploaded Jupyter notebook neuralnetworks1.ipynb, implement the momentum. See that learning with the momentum can be considerably faster than without.

Don't submit the implementation or learning curves. Rather, if you did the above, write "Done." as answer to this problem.