# COMP9418 Quiz 2

Advanced Topics in Statistical Machine Learning, 17s2, UNSW Sydney

Last Update: Tuesday 12$^{\text{th}}$ September, 2017 at 16:25

**Submission deadline**: Monday September 18th, 2017 at 23:59:59

**Late Submission Policy**: One mark will be deducted from the total for each day late, up to a total of four days. If five or more days late, a zero mark will be given.

**Form of Submission**: You should submit your solution in one single file in pdf format with the name `solution.pdf`. No other formats will be accepted (scanned versions of legible handwritten answers are accepted). There is a maximum file size cap of 2MB so make sure your submission does not exceed this size.

Submit your files using give. On a CSE Linux machine, type the following on the command-line:

```
$ give cs9418 quiz2 solution.pdf
```

Alternative, you can submit your solution via the course website
`https://webcms3.cse.unsw.edu.au/COMP9418/17s2/resources/12517`

*Recall the guidance regarding plagiarism in the course introduction: this applies to this homework and if evidence of plagiarism is detected it may result in penalties ranging from loss of marks to suspension.*

1. Consider the hyperbolic secant distribution with parameters $\mu$ and $\sigma$:

$$f(z) = \frac{1}{2\sigma} \operatorname{sech}\left(\frac{\pi}{2} \frac{z-\mu}{\sigma}\right), \tag{1}$$

where

$$\operatorname{sech}(z) \stackrel{\text{def}}{=} \frac{1}{\cosh(z)} = \frac{2}{\exp(z) + \exp(-z)}. \tag{2}$$

This distribution has mean $\mu$ and variance $\sigma^2$. In this exercise, we will consider an instance of the distribution in Equation (1) as the source distribution for an ICA model:

$$p_k(z) = \frac{2}{\pi\left(\exp(z) + \exp(-z)\right)}. \tag{3}$$

In particular, we will assume iid observations, $\mathcal{D} = \{\mathbf{x}^{(n)}\}_{n=1}^{N}$ and the ICA model:

$$p_z(\mathbf{z}^{(n)}) = \prod_{k=1}^{K} p_k(z_k^{(n)}), \tag{4}$$

$$\mathbf{x}^{(n)} = \mathbf{V}^{-1}\mathbf{z}^{(n)}, \tag{5}$$

where each $p_k$ is given by Equation (3) and $\mathbf{V}$ is an orthogonal matrix.

(a) [0.5 marks] Generate a plot of $p_k(z)$ along with that of a Gaussian with the same mean and variance.

(b) [0.5 marks] Is $p_k(z)$ a suitable source distribution for ICA? Explain your reasoning.

(c) [1 mark] write down the log likelihood $\mathcal{L} \stackrel{\text{def}}{=} \log p(\mathcal{D}|\mathbf{V})$ .

(d) [3 marks] Show that the gradients of the log likelihood above wrt the model parameters can be written as:

$$\frac{\partial \mathcal{L}}{\partial V}_{ij} = NV_{ij} - \sum_{n=1}^{N} \tanh(\mathbf{v}_i^T \mathbf{x}^{(n)}) x_j^{(n)}, \tag{6}$$

where $V_{ij}$ is the $(i, j)$th entry of the matrix $\mathbf{V}$, $\mathbf{v}_i$ is the $i$th row of matrix $\mathbf{V}$ and

$$\tanh(z) \stackrel{\text{def}}{=} \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)}. \tag{7}$$