September 6, 2000

ERRATA: Metrics that learn relevance

Samuel Kaski and Janne Sinkkonen

Neural Networks Research Centre, Helsinki University of Technology
P. O. Box 5400, FIN-02015 HUT, Finland
samuel.kaski@hut.fi, janne.sinkkonen@hut.fi

The paper needs clarification in few places. We decided therefore to write a brief errata. Page numbers start from 1.

- p.2, after eqn (4): $p(c_i|v_j)$ is actually defined by $p(c_i|v_j) = p(c_i, v_j)/p(v_j)$, where $p(c_i, v_j) = \int p(c_i|\mathbf{x})p(v_j|\mathbf{x})p(\mathbf{x})d\mathbf{x}$.
- p.3: Note that the distribution $\hat{p}(c|\mathbf{x})$ is not a proper distribution since it does not sum up to unity. We were going to add a footnote about this. In our next papers we will show that the distribution

$$\hat{p}(c_i|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp \sum_j y_j(\mathbf{x}) \log p(c_i|v_j) \ ,$$

  where $Z(\mathbf{x})$ normalizes the sum over i to unity, has useful properties.
- p.4: It is possible to obtain much better results by computing the gradient in the new metric, by using the "natural gradient." The adaptation rule then becomes the familiar SOM adaptation rule,

$$\mathbf{m}_k(t+1) = \mathbf{m}_k(t) + h_{ci}(\mathbf{x}(t) - \mathbf{m}_k(t)) \ .$$