Compactness of variational approximations

David J. C. MacKay Richard Turner Maneesh Sahani

June 19, 2008

Let's approximate a complicated distribution $P(\mathbf{x})$ by a simpler distribution $Q(\mathbf{x})$, possibly a separable distribution. It's often the case that variational free energy minimization (also known as mean field) leads to an approximating distribution Q that is 'more compact' than the true distribution.

Is there in fact a **theorem** that we could prove along the lines of 'optimized Q is **always** more compact'? Or can we find a counterexample?

Let's put the conjecture down in a precise form. Let Q be a separable approximation. Let Q^* be the separable approximation that minimizes the variational free energy

$$F(Q) = \langle E \rangle_Q - S_Q.$$

The conjecture states that the entropy S_{Q^*} of the approximation is smaller than the true distribution's entropy S_P.

Counterexample to this conjecture

Let x_1 and x_2 take values in the set $\{1, \ldots, N\}$, and let

$$E(x_1, x_2) = \begin{cases} -\epsilon & x_1 = x_2 \\ 0 & \text{otherwise} \end{cases}$$

We assume that Q^* will take the form of either Q_A: a spike on one state such as $(x_1, x_2) = (1, 1)$; or Q_U: the uniform distribution over all states. These have free energies $F_A = -\epsilon$ and $F_U = -\epsilon/N - \log N$ respectively. [To do: prove that these are indeed the only forms taken by Q^*.]

The true distribution P has entropy

$$S_P = N(N-1)p_{\text{on}} \log \frac{1}{p_{\text{on}}} + N p_{\text{off}} \log \frac{1}{p_{\text{off}}}$$

where

$$p_{\text{on}} = e^{\epsilon}/Z,$$

$$p_{\text{off}} = 1/Z,$$

and

$$Z = N(N-1) + Ne^{\epsilon}.$$

To get a counterexample, we set $N = 16$ and $\epsilon = 4$. (Any value of ϵ less than 5.8 would do the trick.)

Under these conditions, the uniform Q_U is the better of the approximations, with entropy 8 bits, whereas the true distribution P has an entropy of 5.6 bits.

Conclusion: the folk theorem about variational approximations being 'more compact' is not always true.

Acknowledgments

We thank Manfred Opper for helpful discussions and the Newton Institute for hospitality.