TNN is a special case of TBB

C. Kemp, T. L. Griffiths, S. Stromsten & J. B. Tenenbaum
Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139
{ckemp,gruffydd,sean_s,jbt}@mit.edu

TBB (Tree-Based Bayes) and TNN (Tree Nearest Neighbor) are two algorithms for semi-supervised learning described in [1]. Both take a tree with a small number of labeled leaves and classify all remaining leaves. TBB includes a single free parameter λ, and TBB becomes equivalent to TNN once λ is set sufficiently high:

Theorem 1 *For each ultrametric tree $\mathcal{T}$, there is a λ such that TNN and TBB produce identical classifications for all examples with a unique nearest neighbor.*

We prove the theorem here.

The proof depends critically on the mutation model used to define TBB. Let L be a variable corresponding to the class label. The probability that L changes value along a branch b of length $|b|$ is

$$p(L \text{ changes along } b) = \frac{1 - e^{-2\lambda|b|}}{2}. \tag{1}$$

Note that the mutation process is conservative: it is always more likely for L to stay the same than to switch along any branch.

Let the 'skeleton' of $\mathcal{T}$ be the subtree consisting of all paths from the labeled leaves to the root. Since the mutation process is conservative, the classification of any node N_i according to TBB is the most likely value at the node where the path from N_i meets the skeleton. Let N_j be any labeled node, N_L be the set of all labeled nodes, and N_{L-j} be the set of all labeled nodes except N_j.

Let N_a be the most recent ancestor of N_j with two labeled descendants (if there is no such ancestor, then N_j is the only labeled node, and both algorithms will label all nodes with n_j, the value at N_j). Without loss of generality, suppose that $n_j = 1$, and that the distance between N_j and N_a is 1. We establish the theorem by showing that every node in the skeleton between N_j and N_a has a posterior distribution that favors $n_j = 1$ once λ grows large.

Suppose N_m is a node in the skeleton between N_l and N_a. The posterior probability at N_m is:

$$\begin{aligned} p(n_m = 1|n_L) &= p(n_m = 1|n_j, n_{L-j}) \\ &= \frac{p(n_j = 1|n_m = 1, n_{L-j})p(n_m = 1|n_{L-j})}{p(n_j = 1|n_{L-j})}. \end{aligned}$$

The denominator does not depend on n_m. Thus:

$$\begin{aligned} p(n_m = 1|n_L) &\propto p(n_j = 1|n_m = 1, n_{L-j})p(n_m = 1|n_{L-j}) \\ &\propto p(n_j = 1|n_m = 1) \sum_{n_a \in \{0,1\}} p(n_m = 1|n_a)p(n_a|n_{L-j}). \end{aligned}$$

Let $q = p(n_a = 0|n_{L-j})$. Then:

$$p(n_m = 1|n_L) \propto p(n_j = 1|n_m = 1)(p(n_m = 1|n_a = 0)q + p(n_m = 1|n_a = 1)(1 - q)).$$

Assume that the distance between N_j and N_m is d (and thus that the distance between N_m and N_a is $1 - d$). Using Equation 1, $p(n_j = 1|n_m = 1) = \frac{1+e^{-2\lambda d}}{2}$, $p(n_m = 1|n_a = 0) = \frac{1-e^{-2\lambda(1-d)}}{2}$, and $p(n_m = 1|n_a = 1) = \frac{1+e^{-2\lambda(1-d)}}{2}$. Thus:

$$p(n_m = 1|n_L) \propto \frac{1 + e^{-2\lambda d}}{2} \left(\frac{1 - e^{-2\lambda(1-d)}}{2} q + \frac{1 + e^{-2\lambda(1-d)}}{2} (1 - q) \right).$$

Similarly,

$$p(n_m = 0|n_L) \propto \frac{1 - e^{-2\lambda d}}{2} \left(\frac{1 + e^{-2\lambda(1-d)}}{2} q + \frac{1 - e^{-2\lambda(1-d)}}{2} (1 - q) \right).$$

Assume that $q > 0.5$, otherwise every skeleton node between N_j and N_a has a posterior that favors 1. It is now straightforward to show that $p(n_m = 1|n_L) > p(n_m = 0|n_L)$ if and only if $d < d_{max} = \frac{1}{2} + \frac{1}{4\lambda} \log(\frac{1}{2q-1})$.

We give a worst-case analysis to show that $\lim_{\lambda \to \infty} d_{max} = 1$. For a given λ, d_{max} will be smallest when q is largest: in other words, when there is the best evidence possible that $n_a = 0$. Assume that the tree has $k + 1$ external nodes. Then q will be largest when all external nodes except N_j are set to 0, and have N_a as their parent. Since the tree is ultrametric, the distance between any leaf and N_a is at least 1. It follows that $q < q_{max} = p(n'_a = 0|n_L)$, where N'_a is the root of the tree in Figure 1.

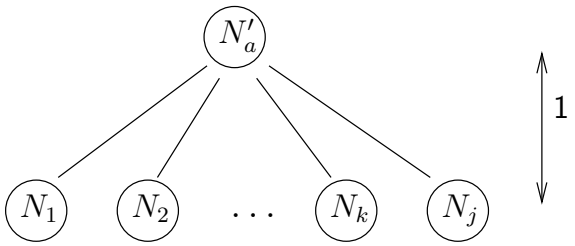

Figure 1: The $k + 1$-leaf tree which produces the largest value of q for fixed λ

Now

$$\begin{aligned} p(n'_a = 0|n_L) &= \frac{p(n_L|n'_a = 0)p(n'_a = 0)}{p(n_L)} \\ &= \frac{p(n_L|n'_a = 0)}{2p(n_L)} \\ &\propto p(n_L|n'_a = 0). \end{aligned}$$

since the prior distribution at the root of the tree is uniform.

Each leaf value is independent of all the others given n'_a, so

$$p(n'_a = 0|n_L) \propto p(n_1 = 0|n'_a = 0) \dots p(n_k = 0|n'_a = 0).$$

Using the mutation model again, $p(n_a = 0|n_L) \propto \left(1 + e^{-2\lambda}\right)^k$, and $p(n_a = 1|n_L) \propto \left(1 - e^{-2\lambda}\right)^k$. Thus

$$q_{max} = \frac{1}{1 + \left(\frac{1-e^{-2\lambda}}{1+e^{-2\lambda}}\right)^k}.$$

Now

$$\begin{aligned}
\lim_{\lambda\to\infty} \frac{1}{\lambda} \log\left(\frac{1}{2q_{max} - 1}\right) &= \lim_{\lambda\to\infty} \frac{1}{\lambda} \log\left(\frac{(1 - e^{-2\lambda})^k - (1 + e^{-2\lambda})^k}{(1 + e^{-2\lambda})^k + (1 + e^{-2\lambda})^k}\right) \\
&= \lim_{\lambda\to\infty} \frac{1}{\lambda} \log\left(\frac{(1 - e^{-2\lambda})^k - (1 + e^{-2\lambda})^k}{2}\right) \\
&= \lim_{\lambda\to\infty} \frac{1}{\lambda} \log((1 - e^{-2\lambda})^k - (1 + e^{-2\lambda})^k).
\end{aligned}$$

Using the binomial series

$$(1 + x)^k = 1 + kx + \binom{k}{2}x^2 + \binom{k}{3}x^3 \dots$$

we have

$$\begin{aligned}
\lim_{\lambda\to\infty} \frac{1}{\lambda} \log\left(\frac{1}{2q_{max} - 1}\right) &= \lim_{\lambda\to\infty} \frac{1}{\lambda} \log((1 - e^{-2\lambda})^k - (1 + e^{-2\lambda})^k) \\
&= \lim_{\lambda\to\infty} \frac{1}{\lambda} \log(-2ke^{-2\lambda} - 2\binom{k}{3}e^{-6\lambda} + \dots) \\
&= \lim_{\lambda\to\infty} \frac{1}{\lambda} \log(-2ke^{-2\lambda}) \\
&= \lim_{\lambda\to\infty} \frac{1}{\lambda} (\log(2k) + \log(-e^{-2\lambda})) \\
&= \lim_{\lambda\to\infty} \frac{1}{\lambda} 2\lambda \\
&= 2.
\end{aligned}$$

Thus $\lim_{\lambda\to\infty} d_{max} = 1$ as required. It follows that λ can be set sufficiently high that all nodes in the skeleton between N_j and N_a have a posterior that favors 1.

References

[1] C. Kemp, T. L. Griffiths, S. Stromsten, and J. B. Tenenbaum. Semi-supervised learning with trees. Submitted to NIPS 2003.