Supplementary Document for BNB

Bryan Hooi, Christos Falousos

1 Proof of Theorem 6.1

1.1 Preliminaries Given a tree $\mathcal{T}_k$, the tree partitions the data space, mapping each data point X_t into a leaf of $\mathcal{T}_k$: let leaf(X_t) map X_t to its leaf. Define $\mathcal{X}$ as the set of leaves, and $d(x, y)$ as the shortest path distance (in terms of number of graph hops) between leaves x and y.

LEMMA 1.1. *$(\mathcal{X}, d)$ is a metric space.*

1.2 Separation Depth

THEOREM 1.1. *The separation depth $s_{k,t}$ can be rewritten in terms of distances in this metric space:*

$$s_{k,t} = d_{lim} - \frac{1}{2} \min_{x \in L, y \in R} d(\mathit{leaf}(x), \mathit{leaf}(y)) + 1 \tag{1.1}$$

Proof. Consider the pair of points $x \in L$ and $y \in R$ minimizing Eq. (6.2). Their shortest path distance is $d(\text{leaf}(x), \text{leaf}(y))$, so their deepest common ancestor (call it z) must be $d(\text{leaf}(x), \text{leaf}(y))/2$ steps above them. Since leaf(x) and leaf(y) are at depth d_{lim}, thus z is at depth $d_{lim} - d(\text{leaf}(x), \text{leaf}(y))/2$. Moreover, since x and y were the closest pair of points, z must be the deepest node with descendants in both L and R. Thus, L and R are fully separated exactly at depth $d_{lim} - d(\text{leaf}(x), \text{leaf}(y))/2 + 1$, but not at any lower depth (due to node z). ∎

2 Proof of Theorem 6.2

THEOREM 2.1. *For any $\lambda > 0$, letting $p_\lambda = 2^{\lambda-1}$, we have*[1]*:*

$$P(s_{k,t} \leq \lambda) \leq (1/2)^{w-p_\lambda}. \tag{2.2}$$

Proof. Fix an arbitrary ordering $<$ on the data space (e.g. lexicographic ordering), and let $X^1, \cdots, X^{2w}$ be the original data points $(X_1, \cdots, X_{2w})$ sorted according to the order $<$. Since the data points $X_1, \cdots, X_{2w}$ are i.i.d., if we condition on $X^1, \cdots, X^{2w}$, by symmetry, the conditional probability that the original data is mapped to $X^1, \cdots, X^{2w}$ by any particular permutation is equal, which is $1/(2w)!$.

Define Z_i as a 0-1 random variable, taking value 0 if X^i is one of $X_1, \cdots, X_w$, and 1 otherwise. Then exactly w of the random variables $Z_1, \cdots, Z_{2w}$ are 1, and since we earlier showed that every permutation of the $X^1, \cdots, X^{2w}$ is equally likely, thus now each assignments of 0s and 1s to $Z_1, \cdots, Z_{2w}$ with exactly w 1s is also equally likely. Thus, by symmetry, each such assignment has probability $1/\binom{2w}{w}$.

For any λ, $P(s_{k,t} \leq \lambda)$ is the probability that the sets L and R were fully separated by the time the tree reached level λ. Note that at level λ, our tree partitions the data space into $p_\lambda = 2^{\lambda-1}$ parts. The number of assignments to the Z variables for which L and R are fully separated is then at most 2^{p_λ}, since each of the p_λ parts we have to choose to assign either 0 or 1 to all variables in that part. Moreover, we earlier showed that each of the $\binom{2w}{w}$ assignments has the same probability. Thus, the probability that L and R are fully separated at level λ is at most

$$P(s_{k,t} \leq \lambda) \leq \frac{2^{p_\lambda}}{\binom{2w}{w}} \tag{2.3}$$

$$= \frac{2^{p_\lambda}(1)\dots(w)}{(w+1)\dots(2w)} \tag{2.4}$$

$$\leq 2^{p_\lambda}(1/2)^w \tag{2.5}$$

$$\leq (1/2)^{w-p_\lambda}. \tag{2.6}$$

In fact, this can be tightened significantly by using stronger bounds for central binomial coefficients: it is known that $\binom{2w}{w} \geq 4^w/\sqrt{4w}$ for any positive integer w [1]. Substituting this instead gives

$$P(s_{k,t} \leq \lambda) \leq \frac{2^{p_\lambda}}{\binom{2w}{w}} \tag{2.7}$$

$$\leq \frac{2^{p_\lambda}\sqrt{4w}}{4^w} \tag{2.8}$$

$$= 2^{p_\lambda - 1 - 2w}\sqrt{w} \tag{2.9}$$

$$= (1/2)^{2w-1-p_\lambda}\sqrt{w}. \tag{2.10}$$

∎

[1] Note that some theorem statements have been amended from their previous versions. We will correct this fully in the final version of the paper.

3 Proof of Theorem 6.3

The next theorem concerns our final change score c_t. For any error threshold $\varepsilon > 0$, we have:

THEOREM 3.1.

$$P(c_t \geq d_{lim} + 1 - \log(1 + \log \frac{\varepsilon}{N} + w)) \leq N(1/2)^{w-p_\lambda} \tag{3.11}$$

Proof. Applying union bound on the (weaker) result of the previous theorem gives

$$P(s_{k,t} \leq \lambda) \leq N(1/2)^{w-p_\lambda} \ \forall \ k \in \{1, \cdots, N\} \tag{3.12}$$

Recalling that $c_t = d_{lim} + 1 - \frac{1}{N}\sum_{j=1}^{N} s_{j,t}$, substituting this gives:

$$P(c_t \geq d_{lim} + 1 - \lambda) \leq N(1/2)^{w-p_\lambda} \tag{3.13}$$

Finally, if $p_\lambda = \log \frac{\varepsilon}{N} + w$, then $N(1/2)^{w-p_\lambda} \leq \varepsilon$. Substituting this into (3.13) (and using the fact that $\lambda = \log p_\lambda + 1$) gives the result.

■

References

[1] N. D. Kazarinoff. *Analytic inequalities.* Courier Corporation, 2014.