Patient Risk Prediction Model via Top-k Stability Selection

Jiayu Zhou, Jimeng Sun, Yashu Liu, Jianying Hu, Jieping Ye
Jiayu.Zhou@asu.edu

Appendix A: Proof of Theorem 3.1

Before proceeding to the proof of Theorem 3.1, we firstly seek some insights by analyzing simultaneous selection probability from random splittings. Let D_1 and D_2 be two disjoint subsets of T generated by random splitting with sample size $\lfloor n/2 \rfloor$. The simultaneously selected set is then given by:

$$\hat{S}^{\text{smlt},\lambda} = \hat{S}^{\lambda}(D_1) \bigcap \hat{S}^{\lambda}(D_2).$$

The corresponding simultaneous selection probability for any set $F \subseteq \{1, \ldots, p\}$ is defined as $\hat{\Pi}_F^{\text{smlt},\lambda} = P^* \left(F \subseteq \hat{S}^{\text{smlt},\lambda} \right)$.

LEMMA 5.1. *For any subset $F \subseteq \{1, \ldots, p\}$, a lower bound for the average of top k maximum simultaneous selection probabilities is given by*

$$\mathcal{M}_k^{\lambda\in\Lambda} \left(\hat{\Pi}_F^{smlt,\lambda} \right) \geq 2 \cdot \mathcal{M}_k^{\lambda\in\Lambda} \left(\hat{\Pi}_F^{\lambda} \right) - 1.$$

Proof: According to the Lemma 1 in [20], it follows that $\hat{\Pi}_F^{\text{smlt},\lambda} \geq 2 \cdot \hat{\Pi}_F^{\lambda} - 1$ for any set $F \subseteq \{1, \ldots, p\}$. We therefore have:

$$\begin{aligned}
\mathcal{M}_k^{\lambda\in\Lambda} \left(\hat{\Pi}_F^{\text{smlt},\lambda} \right) &= \frac{1}{k} \sum_{i=1}^{k} \max_{\lambda \in \Lambda_F^{-i}} \left(\hat{\Pi}_F^{\text{smlt},\lambda} \right) \\
&\geq \frac{1}{k} \sum_{i=1}^{k} \left(2 \cdot \max_{\lambda \in \Lambda_F^{-i}} \left(\hat{\Pi}_F^{\lambda} \right) - 1 \right) \\
&= \frac{2}{k} \sum_{i=1}^{k} \left(\max_{\lambda \in \Lambda_F^{-i}} \left(\hat{\Pi}_F^{\lambda} \right) \right) - 1 \\
&= 2\mathcal{M}_k^{\lambda\in\Lambda} \left(\hat{\Pi}_F^{\lambda} \right) - 1.
\end{aligned}$$

This completes the proof of Lemma 5.1. □

LEMMA 5.2. *For a subset of features $F \subseteq \{1, \ldots, p\}$, if $P\left(F \in \hat{S}^{\Lambda_F^{-i}} \right) \leq \epsilon_i$, for $i = 1, 2, \ldots, k$, where Λ_F^{-i} is defined the same as in the definition of top-k stable features, $\hat{S}^{\lambda}$ is estimated from $\lfloor n/2 \rfloor$ samples, then:*

$$P\left(\mathcal{M}_k^{\lambda\in\Lambda} \left(\hat{\Pi}_F^{smlt,\lambda} \right) \geq \xi \right) \leq \frac{1}{k \cdot \xi} \sum_{i=1}^{k} \epsilon_i^2.$$

Proof: Let $D_1, D_2 \subseteq \{1, \ldots, n\}$ be two subsamples of T with size $\lfloor n/2 \rfloor$ generated from random splitting. Define $B_F^{\lambda} = \mathbf{1}\Big\{ F \subseteq \Big\{ \hat{S}^{\lambda}(D_1) \bigcap \hat{S}^{\lambda}(D_2) \Big\} \Big\}$, and the simultaneous selection probability is given by $\hat{\Pi}_F^{\text{smlt},\lambda} = \mathbb{E}^* \left(B_F^{\lambda} \right) = \mathbb{E}\left(B_F^{\lambda} | T \right)$, where the expectation $\mathbb{E}^*$ is with respect to the random splitting. Hence for $i = 1, 2, 3, \ldots, k$ we have:

$$\max_{\lambda \in \Lambda_F^{-i}} \left(\hat{\Pi}_F^{\text{smlt},\lambda} \right) = \max_{\lambda \in \Lambda_F^{-i}} \mathbb{E}^* \left(B_F^{\lambda} \right) = \max_{\lambda \in \Lambda_F^{-i}} \mathbb{E}\left(B_F^{\lambda} | T \right).$$

It follows immediately that:

$$\mathcal{M}_k^{\lambda\in\Lambda} \left(\hat{\Pi}_F^{\text{smlt},\lambda} \right) = \mathcal{M}_k^{\lambda\in\Lambda} \left(\mathbb{E}^* \left(B_F^{\lambda} \right) \right) = \mathcal{M}_k^{\lambda\in\Lambda} \left(\mathbb{E}\left(B_F^{\lambda} | T \right) \right).$$

The inequality $P\left(F \in \hat{S}^{\Lambda_F^{-i}}\right) \leq \epsilon_i$ (for sample size $\lfloor n/2 \rfloor$) implies that:

$$\max_{\lambda \in \Lambda_F^{-i}} P\left(B_F^\lambda = 1\right) \leq P\left(F \in \hat{S}^{\Lambda_F^{-i}}(D_1)\right)^2 \leq \epsilon_i^2.$$

That is, for $i = 1, 2, \dots, k$, $\max_{\lambda \in \Lambda_F^{-i}} \mathbb{E}\left(B_F^\lambda\right) \leq \epsilon_i^2$. Thus,

$$\begin{aligned}
\mathbb{E}\left(\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_F^{\text{smlt},\lambda}\right)\right) =& \mathbb{E}\left[\mathcal{M}_k^{\lambda\in\Lambda}\left(\mathbb{E}\left(B_F^\lambda | T\right)\right)\right] \\
=& \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k} \max_{\lambda \in \Lambda_F^{-i}}\left(\mathbb{E}\left(B_F^\lambda | T\right)\right)\right] \\
=& \frac{1}{k}\sum_{i=1}^{k} \max_{\lambda \in \Lambda_F^{-i}}\left(\mathbb{E}\left(B_F^\lambda\right)\right) \\
=& \mathcal{M}_k^{\lambda\in\Lambda}\left(\mathbb{E}\left(B_F^\lambda\right)\right) \\
\leq& \frac{1}{k}\sum_{i=1}^{k}\epsilon_i^2.
\end{aligned}$$

Using the Markov-type inequality [2], we have:

$$\begin{aligned}
\xi P\left(\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_F^{\text{smlt},\lambda}\right) \geq \xi\right) \leq& \mathbb{E}\left[\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_F^{\text{smlt},\lambda}\right)\right] \\
\leq& \frac{1}{k}\sum_{i=1}^{k}\epsilon_i^2,
\end{aligned}$$

thus $P\left(\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_F^{\text{smlt},\lambda}\right) \geq \xi\right) \leq \frac{1}{k\cdot\xi}\sum_{i=1}^{k}\epsilon_i^2$. This completes the proof of the lemma. □

Proof of Theorem 3.1 (Top-k Error Control):
From Theorem 1 in [20] we have that: $P\left(f \in \hat{S}^{\Lambda}\right) \leq u_\Lambda/p$ for $f \in N$, it follows immediately that for all $f \in N$, $i = 1, 2, \dots, k$ we have $P\left(f \in \hat{S}^{\Lambda_f^{-i}}\right) \leq u_{\Lambda_f^{-i}}/p$.
Using Lemma 5.2, we have:

$$P\left(\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_f^{\text{smlt},\lambda}\right) \geq \xi\right) \leq \frac{1}{k\cdot\xi}\sum_{i=1}^{k}\left(u_{\Lambda_f^{-i}}/p\right)^2.$$

By Lemma 5.1, it follows that:

$$\begin{aligned}
&P\left\{\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_f^{\lambda}\right) \geq \pi_{\text{thr}}\right\} \\
\leq& P\left\{\left(\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_f^{\text{smlt},\lambda}\right) + 1\right)/2 \geq \pi_{\text{thr}}\right\} \\
\leq& \frac{\sum_{i=1}^{k} u_{\Lambda_f^{-i}}^2}{k\cdot p^2\cdot(2\pi_{\text{thr}} - 1)}.
\end{aligned}$$

Therefore we have:

$$\mathbb{E}(V_k) = \sum_{f\in N} P\left\{\mathcal{M}_k^{\lambda\in\Lambda}\left(\hat{\Pi}_f^{\lambda}\right) \geq \pi_{\text{thr}}\right\} \leq \frac{\sum_{i=1}^{k} u_{\Lambda,i}^2}{k\cdot p\cdot(2\pi_{\text{thr}} - 1)},$$

where $u_{\Lambda,i}^2 = \mathbb{E}_f[u_{\Lambda_f^{-i}}]^2$. This completes the proof. □