Supplement to "A model-free feature screening approach based on kernel density estimation"

Xiangjie Li †, Lei Wang†,Jingxiao Zhang†*

†Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China.

May 31, 2017

The supplementary gives the proof of Proposition 2.1:

Proof. (1) If X_j is independent of Y, then X_j will be independent for any H, and so $f(X|H=l)=f(X),$for any $l \in \{1,\cdots,S\}$, thus $\omega_j^{\mathcal{S}}=0$. Now suppose $\omega_j^{\mathcal{S}}=0$ for all choices of $\mathcal{S}$, say $|g_l^o(x_j)-g_m^o(x_j)|=0$ a.s.for any $l,m \in \{1,\cdots,S\}$. We consider $H=1$ if $Y \le y$ and $H=2$ otherwise for any y. Because $\omega_j^{\mathcal{S}}=0$ and X is independent of H, $f(X_j|Y \le y)=f(X_j|Y>y)$ for all y, that is, if $\Pr(Y \le y)=0$ or $\Pr(Y>y)=0$,the result holds naturally and if $\Pr(Y \le y)>0$ and $\Pr(Y>y)>0$,then

$$
\begin{aligned}
&\frac{f(X_j, Y \le y)}{\Pr(Y \le y)} = \frac{f(X_j, Y>y)}{1-\Pr(Y \le y)}\\
\Leftrightarrow&\big(f(X_j, Y \le y)+f(X_j, Y>y)\big)\Pr(Y \le y) = f(X_j, Y \le y)\\
\Leftrightarrow&\Pr(Y \le y)f(X_j) = f(X_j, Y \le y)\\
\Leftrightarrow&\Pr(Y \le y)f(X_j) = f(X_j|Y \le y)\Pr(Y \le y)\\
\Leftrightarrow&f(X_j) = f(X_j|Y \le y) \text{ for any } y.
\end{aligned}
$$

*Corresponding author. Email: zhjxiaoruc@163.com

Consequently, X_j is independent of Y.

(2) Suppose $\omega_j^{\mathcal{S}} = 0$, then X_j is independent of H according to the previous proposition. Therefore, $\Pr(Y \leq a_1 | X_j) = \Pr(H = 1 | X_j) = \mathrm{P}(H = 1)$ is a constant irrelated to x_j, which contradicts with our assumption. As a result, we must have $\omega_j^{\mathcal{S}}$.

(3) For the last conclusion, we define

$$\omega_j^* = \sup_{y_1, y_2} \int \left| g_{y_1}^o(x_j) - g_{y_2}^o(x_j) \right| \mathrm{d}x_j, \tag{1}$$

where

$$\begin{aligned} g_{y_1}^o(x_j) &= \begin{cases} f(x_j | Y = y_1), & x_j \in \Omega_{y_1} \\ 0, & x_j \in \Omega_{y_2} \setminus \Omega_{y_1}. \end{cases} \\ g_{y_2}^o(x_j) &= \begin{cases} f(x_j | Y = y_2), & x_j \in \Omega_{y_2} \\ 0, & x_j \in \Omega_{y_1} \setminus \Omega_{y_2}. \end{cases} \end{aligned}$$

and

$$\omega_j^{\mathcal{S}} = \max_{l,m} \int_{\Omega_l \cup \Omega_m} |g_l^o(x_j) - g_m^o(x_j)| \mathrm{d}x_j, \tag{2}$$

where

$$\begin{aligned} g_l^o(x_j) &= \begin{cases} f(x_j | H = l), & x_j \in \Omega_l \\ 0, & x_j \in \Omega_m \setminus \Omega_l \end{cases} \\ g_m^o(x_j) &= \begin{cases} f(x_j | H = m), & x_j \in \Omega_m \\ 0, & x_j \in \Omega_l \setminus \Omega_m. \end{cases} \end{aligned}$$

It is noteworthy that $g_{y_1}^o$ and $g_{y_2}^o$ corresponds to the random variable Y, but g_l^o and g_m^o corresponds to the random variable H, however we abuse the notation a little by writing the $g_\cdot^o$ for both two cases. We can obtain $\omega_j^* > 0$ in that X_j is not independent of $Y, \omega_j^* > 0$. So we just need to prove $\omega_j^{\mathcal{S}} \to \omega_j^*$ as $S \to \infty$. By the definition ω_j^*, for $\forall\ \varepsilon > 0$, there exists y_1^*, y_2^* such that

$$\left| \omega_j^* - \int |g_{y_1^*}^o(x_j) - g_{y_2^*}^o(x_j)| \mathrm{d}x_j \right| < \varepsilon.$$

Because $f(x_j|y)$ is continuous in y and is bounded by a fixed constant M and

$$\int |g^o_{y_1}(x_j) - g^o_{y_2}(x_j)|\mathrm{d}x_j \le \int \left[|g^o_{y_1}(x_j)| + |g^o_{y_2}(x_j)|\right]\mathrm{d}x_j = 2.$$

there exists $\delta > 0$ such that $\int |g^o_y(x_j) - g^o_{y_1^*}(x_j)|\mathrm{d}x_j$ for any $|y - y_1^*| < \delta$. Take $\Delta = \Pr(|y - y_1^*| < \delta)$, due to $\max_{l=1,\cdots,S}\Pr(H = l) \to 0$, there exists S_0 such that $\Pr(H = l) \le \frac{\Delta}{2}$ for $S > S_0$. In such cases, there exists $[a_{l_1}, a_{l_1+1}) \subset (y_1^* - \delta, y_1^* + \delta)$. We can easily prove that

$$\int |g^o_{l_1}(x_j) - g^o_{y_1^*}(x_j)|\mathrm{d}x_j \le \varepsilon \tag{3}$$

Similarly,there exists l_2 such that

$$\int |g^o_{l_2}(x_j) - g^o_{y_2^*}(x_j)|\mathrm{d}x_j \le \varepsilon \tag{4}$$

In addition,

$$\int |g^o_{l_1}(x_j) - g^o_{l_2}(x_j)|\mathrm{d}x_j \le \omega_j^{\mathcal{S}} \le \omega_j^*$$

Hence,

$$\begin{aligned}
|\omega_j^* - \omega_j^{\mathcal{S}}| =& \omega_j^* - \omega_j^{\mathcal{S}} \\
\le& \int |g^o_{y_1^*}(x_j) - g^o_{y_2^*}(x_j)|\mathrm{d}x_j + \varepsilon - \int |g^o_{l_1}(x_j) - g^o_{l_2}(x_j)|\mathrm{d}x_j \\
\le& \int |g^o_{y_1^*}(x_j) - g^o_{y_2^*}(x_j) - g^o_{l_1}(x_j) + g^o_{l_2}(x_j)|\mathrm{d}x_j + \varepsilon \\
\le& \int \left[|g^o_{l_1}(x_j) - g^o_{y_1^*}(x_j)| + |g^o_{l_2}(x_j) - g^o_{y_2^*}(x_j)|\right]\mathrm{d}x_j + \varepsilon \\
=& \sum_{i=1,2} \int |g^o_{l_i}(x_j) - g^o_{y_i^*}(x_j)|\mathrm{d}x_j + \varepsilon \\
\le& 3\varepsilon.
\end{aligned}$$

Due to ε is arbitrary, the conclusion follows. $\square$