Modification of the MDR method related to generalized penalty function

Nickolay A. Slepow

3rd International Workshop
"Analysis, Geometry and Probability"
Ulm, 28.09.15 - 2.10.15
Problem formulation

Set $\mathbb{Y} = \{-m, -m + 1, \ldots, m\}$ and $\mathbb{X} = \{0, 1, \ldots, s\}^n$, where $n, s, m \in \mathbb{N}$. Let

$$X = (X_1, \ldots, X_n)$$

be a random vector with components

$$X_j : \Omega \rightarrow \{0, 1, \ldots, s\}, j = 1, \ldots, n,$$

$$Y : \Omega \rightarrow \mathbb{Y}$$

be a response variable,

$$f : \mathbb{X} \rightarrow \mathbb{Y}$$

be a non-random function.
Auxiliary results

Choose a penalty function
\( \psi : X \times Y \to \mathbb{R}_+ \) and
introduce the error function
\[
 Err(f) := E|Y - f(X)|\psi(X, Y)
\]

In previous works it was considered:
\( \psi : Y \to \mathbb{R}_+ \) and
error function \( Err(f) := E|Y - f(X)|\psi(Y) \)

\[ \operatorname{Err}(f) \rightarrow \min_{f} \implies f_{\text{opt}} \]

Introduce the
\[
\operatorname{Err}(x, k) := \sum_{y=-m}^{m} |k-y| \psi(x, y) P(X = x, Y = y)
\]

One of optimal functions \( f(x) \) is

\[
f_{\text{opt}}(x) = \begin{cases} 
  m, & \text{Err}(m, x) < \text{Err}(m - 1, x), \\
  k \neq \pm m, & \text{Err}(k, x) \leq \text{Err}(k + 1, x) \text{ and } \\
  \text{Err}(k, x) < \text{Err}(k - 1, x), \\
  -m, & \text{Err}(-m, x) \leq \text{Err}(-m + 1, x). 
\end{cases}
\]

(1)
Lemma

For each \( x \in X \) the function \( \text{Err}(\cdot, x) \) has the non-decreasing increments

\[
\text{Err}(k + 1, x) - \text{Err}(k, x),
\]

\( k = -m, \ldots, m - 1 \).

Thus function \( f_{opt}(x) \) is defined correctly.

Nickolay A. Slepow

Modification of the MDR method
The distribution of \((X, Y)\) is unknown:

\[
f_{\text{opt}}(\cdot) \implies \hat{f}_{PA}(\cdot),
\]

\[
\psi(\cdot) \implies \hat{\psi}(\cdot),
\]

\[
\text{Err}(\cdot) \implies \hat{\text{Err}}(\cdot).
\]

**K-fold cross-validation.** \(K \in \mathbb{N}, (K > 1)\)

\(k = 1, \ldots, K\)

\[
S_k(N) := \{(k - 1)[N/K] + 1, \ldots, k[N/K]I\{k < K\} + NI\{k = K\}\}.
\]
Non-regularized estimate of $f_{opt}(\cdot)$:

$$
\tilde{f}_{PA}(x, S_k(N)) = \begin{cases} 
  m, & \tilde{\text{Err}}(m, x, S_k(N)) < \tilde{\text{Err}}(m-1, x, S_k(N)), \\
  k \neq \pm m, & \tilde{\text{Err}}(k, x, S_k(N)) \leq \tilde{\text{Err}}(k+1, x, S_k(N)) \text{ and } \\
                  & \tilde{\text{Err}}(k-1, x, S_k(N)) < \tilde{\text{Err}}(k, x, S_k(N)), \\
  -m, & \tilde{\text{Err}}(-m, x, S_k(N)) \leq \tilde{\text{Err}}(-m+1, x, S_k(N)),
\end{cases}
$$

(2)

where

$$
\tilde{\text{Err}}(k, x, S_k(N)) := \sum_{y \in Y} \sum_{j \in S_k(N)} |k - y| \hat{\psi}(x, y, S_k(N)) \frac{\mathbb{I}\{Y^j = y, X^j = x\}}{\#S_k(N)}.
$$

(3)
For
\[ Err(f) = E|Y - f(X)|\psi(X, Y) = \]
\[ = \sum_{x\in X, y\in Y} |f(x) - y|\psi(x, y) P(X = x, Y = y) \]
we put
\[ \hat{Err}_K(f_{PA}, \xi_N) := \]
\[ = \frac{1}{K} \sum_{k=1}^{K} \sum_{x\in X, y\in Y} \sum_{j\in S_k(N)} |f_{PA}(x, \xi_N(S_k(N))) - y| \times \]
\[ \hat{\psi}(x, y, S_k(N)) \frac{\mathbb{1}\{Y^j = y, X^j = x\}}{\#S_k(N)}. \]
Let \( \hat{\psi}(x, y, S_k(N)) \) be a strongly consistent estimator of \( \psi(x, y) \) as \( N \to \infty \) for all \( x \in \mathbb{X}, y \in \mathbb{Y} \).

It was proved that Theorem 1 and Corollary 2 from [1] (A.Bulinski, A.Rakitko: MDR method for nonbinary response variable. 2014) hold for \( \psi(x, y) \) (not only for \( \psi(y) \)).

Thus the choice of \( f_{opt}(x) \) and \( \tilde{f}_{PA}(x, S_k(N)) \) from (1) and (2) respectively implies the following relation

\[
\hat{\text{Err}}_K(f_{PA}, \xi_N) \to \text{Err}(f) \text{ a.s., } N \to \infty. \quad (4)
\]
Theorem (Bulinski, Rakitko 2014)

Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d. random vectors with the same law as $(X, Y)$, $\psi$ be a penalty function, $f : \mathbb{X} \rightarrow \mathbb{Y}$ and $f_{PA}$ define the prediction algorithm. Assume that there exists nonempty set $U \subset \mathbb{X}$ such that for each $x \in U$ and every $k = 1, \ldots, K$ one has

$$f_{PA}(x, \xi_N(S_k(N))) \rightarrow f(x) \text{ a.s., } N \rightarrow \infty. \quad (5)$$

Then relation (4) holds if and only if

$$\sum_{k=1}^{K} \sum_{x \in \mathbb{X} \setminus U} w^{T}(x) Q_{\delta}(N, x, k) \rightarrow 0 \text{ a.s., } N \rightarrow \infty. \quad (6)$$
Now we apply the theorem with

\[ w(x) = (\psi(x, -m) P(Y = -m, X = x), \ldots, \psi(x, m) P(Y = m, X = x))^\top \]

instead of

\[ w(x) = (\psi(-m) P(Y = -m, X = x), \ldots, \psi(m) P(Y = m, X = x))^\top. \]
Corollary (analogue of Bulinski, Rakitko 2014)

Let $\psi$ be a penalty function, $f : \mathbb{X} \to \mathbb{Y}$ and the prediction algorithm be defined by a function (family) $f_{PA}$. Suppose that for some set $U \subset \mathbb{X}$ condition (5) is satisfied. Assume that for each $x \in M \setminus U$ exist $i=i(x)$ and $j=j(x)$ belonging to $\mathbb{Y}$, $i < j$, such that

$$i \leq f_{PA}(x, \xi_N(S_k(N))) \leq j \text{ a.s. for } k = 1, \ldots, K$$

when $N$ is large enough. Then the condition

$$\text{Err}(\min\{f(x), i\}, x) = \text{Err}(\min\{f(x), i\} + 1, x) = \ldots$$

$$= \text{Err}(\max\{f(x), j\}, x)$$

implies that (6) holds.
Our results are applicable to every submodel in which one considers a subset of factors $X^\alpha_l = (X_{i_1}, X_{i_2}, \ldots, X_{i_l})$.

Let fix a value of $l$.

Thus the choice of (1),(2) and condition (4) for submodels implies the condition

$$\hat{\text{Err}}_K(\hat{f}^\alpha_{PA}, \xi_N) \leq \hat{\text{Err}}_K(\hat{f}^\beta_{PA}, \xi_N) + \varepsilon \text{ a.s.}$$

for the most significant subsequence of predictors $X^\alpha_l$ for any $\varepsilon > 0$ for all $N$ large enough.
Above-mentioned results are hold for regularized estimate of $f_{opt}(\cdot)$:

$$\tilde{f}_{PA}(x, S_k(N)) = \begin{cases} 
  m, & \tilde{\text{Err}}(m, x, S_k(N)) + \varepsilon_N < \tilde{\text{Err}}(m - 1, x, S_k(N)), \\
  k, & \tilde{\text{Err}}(k, x, S_k(N)) \leq \tilde{\text{Err}}(k + 1, x, S_k(N)) + \varepsilon_N \text{ and } \tilde{\text{Err}}(k - 1, x, S_k(N)) < \tilde{\text{Err}}(k, x, S_k(N)) + \varepsilon_N, \\
  -m, & \tilde{\text{Err}}(-m, x, S_k(N)) \leq \tilde{\text{Err}}(-m + 1, x, S_k(N)) + \varepsilon_N, 
\end{cases}$$

\[ (7) \]

where $k \neq \pm m$, $\varepsilon_N \to 0$ and $N^{1/2} \varepsilon_N \to \infty$ as $N \to \infty$. 

Nickolay A. Slepov

Modification of the MDR method
To consider the model with different personal influence of patients we introduce a random vector $Z$ (for j-th patient - $Z^j$).

Let $\{Z^j\}_{j=1}^N$ be i.i.d., but $Z^j$ depends on $X^j$, $Y^j$.

Let $F(z)$ be a non-random function, specifying the patient’s characteristics (e.g. age, weight and etc.).

$0 \leq F(Z^j) \leq C = \text{const}$ a.s.
Put

$$\psi(x, y) = \frac{EF(Z) \mathbb{I}\{X = x, Y = y\}}{P(X = x, Y = y)EF(Z)\mathbb{I}\{Y = y\}}.$$  \hspace{1cm} (8)

It corresponds to the choice of $\psi(x, y) = \frac{1}{P(Y = y)}$ on changed by $F(Z)$ (characteristics of patients) values of probability:

$$P(X = x, Y = y) \mapsto EF(Z)\mathbb{I}\{X = x, Y = y\}.$$
\begin{equation}
\hat{\psi}(x, y, S_k(N)) = \frac{1}{\#S_k(N)} \sum_{i \in S_k(N)} F(Z_i) \mathbb{1}\{Y_i = y\} \times \sum_{j \in S_k(N)} F(Z_j) \mathbb{1}\{Y_j = y, X_j = x\} \times \frac{\sum_{p \in S_k(N)} \mathbb{1}\{X_p = x, Y_p = y\}}{\sum_{p \in S_k(N)} \mathbb{1}\{X_p = x, Y_p = y\}},
\end{equation}

here \(0/0 := 0\).
Theorem

Let $\psi(x, y)$, $\hat{\psi}(x, y, S_k(N))$, $f_{\text{opt}}(x)$ and $\tilde{f}_{\text{PA}}(x, S_k(N))$ be from (8), (9), (1) and (7) respectively. Then the following relation holds

$$\sqrt{N}(\hat{\text{Err}}_K(\tilde{f}_{\text{PA}}, \xi_N) - \text{Err}(f)) \xrightarrow{\text{law}} \zeta \sim N(0, \sigma^2), \quad N \to \infty.$$ 

One can apply this theorem for every submodel.


