Supplementary Material for Causal Inference via Nonlinear Variable Decorrelation for Healthcare Applications

No Author Given

No Institute Given

1 Algorithm

We combine algorithm 1 with object function (2) (in our main paper), to select the robust rules and prune the redundant items. In the *RulesSelection* function, we delete one rule each time with lowest $\|w\|_2^2$ and save the rule sets with the highest accuracy. In the *ItemReduce* function, we apply cross-validation to train SVM model and save the item sets with best accuracy.

Algorithm 1 Rules Selection and Item Reduction

Input: $Rules\{X_i\}$ are the association rules obtained by Apriori algorithm with training datasets. $data$ is EHR datasets.
Output: $Bestrules$

function RULESSELECTION($Rules, data$)
    $Bestrules \leftarrow Rules$
    $Objfunction$ is objective function
    $Select \leftarrow Bestrules$
    $Bestaccuracy \leftarrow Select$
    $Lastrules \leftarrow \varnothing$
    while $Select \neq Lastrules$ **do**
        $Lastrules \leftarrow Select$
        $w \leftarrow \text{argmin}\, Objectfunction(Select, data)$
        $Selected \leftarrow \text{argmin}\, w_i^2$
        $Temprules \leftarrow \{Bestrules\}/\{Selected\}$
        $Tempaccuracy \leftarrow Temprules$
        if $Tempaccuracy > Bestaccuracy$ **then**
            $Bestaccuracy \leftarrow Tempaccuracy$
            $Select \leftarrow Temprules$
        end if
    end while
    return $Bestrules$
end function
function ITEMREDUCE($Bestrules, data$)
    $Bestauc \leftarrow SVM(Bestrules, data)$
    $Lastrules \leftarrow \varnothing$
    while $Bestrules \neq Lastrules$ **do**
        $Item \leftarrow \text{argmax}\, SVM(\{Bestrules\}/\{Item\})$
        $Accuracy \leftarrow SVM(\{Bestrules\}/\{Item\})$
        if $Accuracy \geq Bestauc$ **then**
            $Bestauc \leftarrow Accuracy$
            $Bestrules \leftarrow \{Bestrules\}/\{Item\}$
        end if
    end while
    return $Bestrules$
end function

Table 1: Introduction of individual features on different datasets.

Features	Explanation
Heart Disease	
age middle	Patients between the ages of 40 and 60
#major vessels0	The number of major vessels (0-3) colored by flourosopy is 0
fixed defect	Thalium stress test result is fixed defect
pressure normal	Blood pressure within the normal range
ST-T wave abnormality	Resting electrocardiography result is ST-T wave abnormality
cholesterol edge	Serum cholesterol is in range $(200, 220]$ mg/dl
lower than 120mg/ml	Fasting blood sugar is lower than 120mg/ml
non-anginal pain	Chest pain type is non-angina
cholesterol high	Serum cholesterol is higher than 220 mg/dl
no exercise induced angina	not Exercise induced angina
downsloping	Slope of peak exercise ST segment is downsloping
heart disease	It refers to the presence of heart disease in the patient
Esophageal Cancer	
Modified Ryan Score 2.0	(near complete response): single cells or rare small groups of cancer cells
Esophagectomy Procedure 4	Complete MIS/Robotic McKeown (Three-Hole) esophagectomy
tobacco use	Use tobacco
Alcohol Use	Use Alcohol
Neoadjuvant Radiation	Patient underwent neoadjuvant radiation
Histological Grade 2	How differentiated the tumor is: Moderately Differentiated
Final Histology 1	History: Adenocarcinoma
Histological Grade 3	How differentiated the tumor is: Poorly Differentiated
clinical m Stage 1	Details any spread (metastasis) to other sites of the body: M0
esoph tumor location 4	Lower Thoracic, including GE junction
Esophagectomy Procedure 5	Hybrid (Laparoscopy + Thoracotomy) McKeown (Three-Hole) esophagectomy
recurrence	Details whether the patient experience recurrence of their cancer
Cauda Equina Syndrome	
elixsum	Elixhauser comorbidity sum for that patient is high
beds	Number of beds in the hospital is small
procedure 03 09	ICD-9-CM Procedure Codes: 03.09
Emergency	The patient requires immediate medical intervention as a result of severe
diagnosis 344 60	ICD9 indicators
complication 240days	Indicators for complication within 240 days of discharge
life threatening	The patient's condition is very dangerous
if aa	The racial of the patient is `African American`
die360	Patient died within 360 days

2 Proof

Lemma 1. *If the number of features in the datasets and the terms in the Taylor expansion are fixed, when $n \to \infty$ there exists $W \succeq 0$ such that*

$$\lim_{n \to \infty} \|\mathscr{F}^{(i)}_{p_2 \to p_1, i>0}\|_2^2$$

Proof. Based on our regularizer, we know that

$$\begin{pmatrix} n & \sum_i w_i x_{ip_2} & \cdots & \sum_i w_i^k x_{ip_2}^k \\ \sum_i w_i x_{ip_2} & \sum_i w_i^2 x_{ip_2}^2 & \cdots & \sum_i w_i^{k+1} x_{ip_2}^{k+1} \\ \vdots & \vdots & & \sum_i \\ \sum_i w_i^k x_{ip_2}^k & \sum_i w_i^{k+1} x_{ip_2}^{k+1} & \ddots & \sum_i w_i^{2k} x_{ip_2}^{2k} \end{pmatrix} \begin{pmatrix} f_{p_1 p_2}\left(x_{p_2}(0)\right) \\ f_{p_1 p_2}\left(x_{p_2}(0)\right) \\ \vdots \\ f^{(p)}_{p_1 p_2}\left(x_{p_2}(0)\right) \end{pmatrix} = \begin{pmatrix} \sum_i y_i \\ \sum_i w_i x_{ip_2} y_i \\ \vdots \\ \sum_i w_i^k x_{ip_2}^k y_i \end{pmatrix}$$

We assume that the covariance is 0:

$$\operatorname{cov}\left(\hat{x}_{ip_2}, y_i\right) = \operatorname{cov}\left(\hat{x}_{ip_2}^2, y_i\right) = \operatorname{cov}\left(\hat{x}_{ip_2}^3, y_i\right) = \cdots = \operatorname{cov}\left(\hat{x}_{ip_2}^k, y_i\right) = 0$$

Combine with the following equation, we can get

$$\begin{aligned} n \to \infty : \frac{1}{n} \sum_n \hat{x}_{ip_2} y_i - \frac{1}{n^2} \sum_n \hat{x}_{ip_2} \sum_n y_i &= \frac{1}{n} \sum_n \hat{x}_{ip_2}^2 y_i - \frac{1}{n^2} \sum_n \hat{x}_{ip_2}^2 \sum_n y_i \\ &= \frac{1}{n} \sum_n \hat{x}_{ip_2}^k y_i - \frac{1}{n^2} \sum_n \hat{x}_{ip_2}^k \sum_n y_i = 0 \end{aligned}$$

$$\begin{pmatrix} n & \sum_i \hat{x}_{ip_2} & \cdots & \sum_i \hat{x}_{ip_2}^k \\ \sum_i \hat{x}_{ip_2} & \sum_i \hat{x}_{ip_2}^2 & \cdots & \sum_i \hat{x}_{ip_2}^{k+1} \\ \vdots & \vdots & & \vdots \\ \sum_i \hat{x}_{ip_2}^k & \sum_i \hat{x}_{ip_2}^{k+1} & \ddots & \sum_i \hat{x}_{ip_2}^{2k} \end{pmatrix} \begin{pmatrix} f_{p_1 p_2}\left(x_{p_2}(0)\right) \\ f'_{p_1 p_2}\left(x_{p_2}(0)\right) \\ \vdots \\ {f^{(p)}}_{p_1 p_2}\left(x_{p_2}(0)\right) \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \sum_i y_i \hat{x}_{ip_2} \sum_i y_i \\ \vdots \\ \sum_i \hat{x}_{ip_2}^k \sum_i y_i \end{pmatrix}$$

$$\begin{pmatrix} \frac{\sum_i \hat{x}_{ip_2}^2}{\sum_i \hat{x}_{ip_2}} - \sum_i \hat{x}_{ip_2} & \frac{\sum_i \hat{x}_{p_2}^3}{\sum_i \hat{x}_{ip_2}} - \sum_i \hat{x}_{ip_2}^2 & \cdots & \frac{\sum_i \hat{x}_{ip_2}^{k+1}}{\sum_i \hat{x}_{ip_2}} - \sum_i \hat{x}_{ip_2}^k \\ \sum_i \hat{x}_{ip_2}^3 & & & \\ \frac{\sum_i \hat{x}_{ip_2}^2}{2} - \sum_i \hat{x}_{ip_2} & \frac{\sum_i \hat{x}_{ip_2}^4}{\sum_i \hat{x}_{ip_2}^2} - \sum_i \hat{x}_{ip_2}^2 & \cdots & \frac{\sum_i \hat{x}_{ip_2}^{k+2}}{\sum_i \hat{x}_{ip_2}^2} - \sum_i \hat{x}_{ip_2}^k \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_i \hat{x}_{ip_2}^{k+1}}{\sum_i \hat{x}_{ip_2}^k} - \sum_i \hat{x}_{ip_2}^k & \frac{\sum_i \hat{x}_{ip_2}^{k+2}}{\sum_i \hat{x}_{ip_2}^k} - \sum_i \hat{x}_{ip_2}^k & \cdots & \frac{\sum_i \hat{x}_{ip_2}^{2k}}{\sum_i \hat{x}_{ip_2}^k} - \sum_i \hat{x}_{ip_2}^k \end{pmatrix} \begin{pmatrix} f_{p_1 p_2}\left(x_{p_2}(0)\right) \\ f'' p_1 p_2 \left(x_{p_2}(0)\right) \\ \vdots \\ f^{(p)}_{p_1 p_2}\left(x_{p_2}(0)\right. \end{pmatrix} = 0$$

$$\begin{vmatrix} \frac{\sum_i \hat{x}_{ip_2}^2}{\sum_i \hat{x}_{ip_2}} - \sum_i \hat{x}_{ip_2} & \frac{\sum_i \hat{x}_{ip_2}^3}{\sum_i \hat{x}_{ip_2}} - \sum_i \hat{x}_{ip_2}^2 & \cdots & \frac{\sum_i \hat{x}_{ip_2}^{k+1}}{\sum_i \hat{x}_{ip_2}} - \sum_i \hat{x}_{ip_2}^k \\ \frac{\sum_i \hat{x}_{ip_2}^3}{\sum_i \hat{x}_{ip_2}^2} - \sum_i \hat{x}_{ip_2} & \frac{\sum_i \hat{x}_{ip_2}^4}{\sum_i \hat{x}_{ip_2}^2} - \sum_i \hat{x}_{ip_2}^2 & \cdots & \frac{\sum_i \hat{x}_{ip_2}^{k+2}}{\sum_i \hat{x}_{ip_2}^2} - \sum_i \hat{x}_{ip_2}^k \\ \vdots & \vdots & & \vdots \\ \frac{\sum_i \hat{x}_{ip_2}^{k+1}}{\sum_i \hat{x}_{ip_2}^k} - \sum_i \hat{x}_{ip_2}^k & \frac{\sum_i \hat{x}_{ip_2}^{k+2}}{\sum_i \hat{x}_{ip_2}^k} - \sum_i \hat{x}_{ip_2}^k & \cdots & \frac{\sum_i \hat{x}_{ip_2}^{2k}}{\sum_i \hat{x}_{ip_2}^k} - \sum_i \hat{x}_{ip_2}^k \end{vmatrix} \neq 0$$

$\hat{x}_{ip_2}^2$ is influenced by the w_i which can be adjusted, and the determinant of matrix is not equal to 0, hence the equation has only the trivial solution. We can get

$$f'_{p_1p_2}\left(x_{p_2}(0)\right) = f''_{p_1p_2}\left(x_{p_2}(0)\right) = \cdots = f^{(p)}_{p_1p_2}\left(x_{p_2}(0)\right) = 0$$

If we can prove under our regularizer, we can prove our method can work:

$$n \rightarrow \infty : \left(\hat{x}_{ip_2}, y_i\right) = \operatorname{cov}\left(\hat{x}_{ip_2}^2, y_i\right) = \operatorname{cov}\left(\hat{x}_{ip_2}^3, y_i\right) = \cdots = \operatorname{cov}\left(\hat{x}_{ip_2}^k, y_i\right) = 0$$

We set $\left(\hat{x}_{ip_2}, \hat{x}_{ip_2}^2, \ldots, \hat{x}_{ip_2}^k\right)$ is kernel density estimators: $g(x_{ip_2}$. We set the weight w_i is:

$$w_i = \frac{\prod_{iq} g\left(x_{ij}^q\right)}{\hat{G}\left(g\left(x_{i1}\right), g\left(x_{i2}\right), \ldots, g\left(x_{ip}\right)\right)}$$

$$\begin{aligned} n \rightarrow \infty : E\left[\hat{x}_{p_1}^q\right] &= \frac{1}{n}\sum_i x_{ip_1}^q \frac{\prod_{iq} g\left(x_{ij}^q\right)}{\hat{G}\left(g\left(x_{i1}\right), g\left(x_{i2}\right), \ldots, g\left(x_{ip}\right)\right)} \\ &= \int \ldots \int x_{ij}^q \prod_l g\left(x_{il}^q\right) dx_{i1} dx_{i1}^1 \ldots dx_{ip}^q + o(1) = \int x_{il}^{q_1} g\left(x_{il}^{q_1}\right) dx_{il}^{q_1} + o(1) \end{aligned}$$

$$\begin{aligned} n \rightarrow \infty : E\left[\hat{x}_{p_1}^q, \hat{x}_{p_2}\right] &= \frac{1}{n}\sum_i x_{ip_1}^q x_{ip_1} \left(\frac{\prod_{iq} g\left(x_{ij}^q\right)}{\hat{G}\left(g\left(x_{i1}\right), g\left(x_{i2}\right), \ldots, g\left(x_{ip}\right)\right)}\right)^2 \\ &= \iint x_{il}^{q_1} x_{im} g\left(x_{il}^{q_1}\right) g\left(x_{im}\right) dx_{il}^{q_1} dx_{im} + o(1) \\ &= \int x_{il}^{q_1} g\left(x_{il}^{q_1}\right) dx_{il}^{q_1} \int x_{im} g\left(x_{im}\right) dx_{im} + o(1) \end{aligned}$$

$$n \rightarrow \infty : \operatorname{cov}\left(\hat{x}_{ip_1}^q, \hat{x}_{p_1}\right) = E\left[\hat{x}_{p_1}^q\right] E\left[\hat{x}_{p_1}\right] - E\left[\hat{x}_{p_1}^q, \hat{x}_{p_2}\right] = 0$$

We can get:

$$f'_{p_1p_2}\left(x_{p_2}(0)\right) = f''_{p_1p_2}\left(x_{p_2}(0)\right) = \cdots = f^{(p)}_{p_1p_2}\left(x_{p_2}(0)\right) = 0$$