Dimension Reduction for Regression with Reproducing Kernels

Kenji Fukumizu
Institute of Statistical Mathematics, Japan
Visiting UC Berkeley

Statistical Colloquium. March 18, 2003

Joint work with Michael Jordan and Francis Bach in Berkeley
Outline

- Introduction
  - Dimension reduction for regression

- Conditional Independence and RKHS
  - Dimension reduction and conditional independence
  - Reproducing kernel Hilbert space
  - Conditional covariance operator

- Kernel Dimension Reduction for Regression
  - Algorithm and experimental results

- Extension to Variable Selection

- Summary
Introduction

Dimension reduction for regression

- Regression

\[ Y \sim f(X, Z) \quad \text{or} \quad p(Y \mid X) \]

\( Y \): response variable, \( X \): \( m \)-dim. explanatory variable, \( Z \): noise

- Goal: Find effective subspace defined by \( B \).

\[ p(Y \mid B^T X) = p(Y \mid X) \quad B: m \times d \quad \text{matrix} \quad d \text{ is fixed.} \]

- Effective subspace to explain \( Y \).
- Compact representation of the statistical relation.
  - data analysis: what determines \( Y \)?.
  - preprocessing of regression: accuracy of regression, computational efficiency.
– Example

\[ Y = \frac{2}{1 + \exp(-2X_1)} + N(0; 0.1^2) \]
**Semi-parametric problem**

Assume

\[ p_{Y|X}(Y \mid X) = \tilde{p}(Y \mid B_0^T X) \]

\( B_0: m \times d \) matrix

i.i.d. sample \((X^{(1)}, Y^{(1)}), \ldots, (X^{(n)}, Y^{(n)})\) given.

Find the subspace \( B_0 \) without knowing anything about \( p_{Y|X} \) (or \( \tilde{p} \)).

There is the infinite degree of freedom on unestimated \( p \).

\rightarrow Semiparametric problem.

**Approach**

- Formulate the problem by conditional independence.
- Use reproducing kernel Hilbert spaces as functional spaces for the infinite degree of freedom.
Existing Methods

- **Sliced Inverse Regression (SIR, Li 1991)**
  - PCA of $E[X|Y] \rightarrow$ use slice of $Y$.
  - Semiparametric method: no assumption on $p(Y|X)$.
  - Elliptic assumption on the distribution of $X$ is necessary.

- **Principle Hessian Direction (pHd, Li 1992)**
  - Average Hessian $\Sigma_{jxx} \equiv E[(Y - \bar{Y})(X - \bar{X})(X - \bar{X})^T]$ is used.
  - If $X$ is Gaussian, eigenvectors gives the effective directions.
  - Gaussian assumption on $X$. $Y$ must be one-dimensional.

- **Projection pursuit approach (e.g. Friedman et al. 1981)**
  - Additive model is used for regressor.

- **Canonical Correlation Analysis (CCA) / Partial Least Square (PLS)**
  - Linear assumption on the regression.
Conditional Independence

- **Dimension reduction and conditional independence**

  \[(U, V) = (B^T X, C^T X) \quad \text{for} \quad (B, C) \in O(m)\]

  \(B\) gives the effective subspace \(\iff p_{Y|X}(y | x) = p_{Y|U}(y | B^T x)\)

  \(\iff p_{Y|U,V}(y | u, v) = p_{Y|U}(y | u) \quad \text{for all} \quad y, u, v\)

  \(\iff \text{Conditional independence} \quad Y \perp V | U\)

- **Characterization of conditional independence**

  Reproducing kernel Hilbert space (RKHS)
Reproducing Kernel Hilbert Space

Definition

\[ \Omega: \text{set.} \quad H: \text{Hilbert space } \subset \{ f : \Omega \to \mathbb{R} \} \]

\( H: \) reproducing kernel Hilbert space (RKHS)

\[ \iff \ \exists \ k : \Omega \times \Omega \to \mathbb{R} \quad \text{symmetric function (reproducing kernel) s.t.} \]

1) \( k(\cdot, x) \in H \) for all \( x \in \Omega. \)

2) \( \langle k(\cdot, x), f \rangle_H = f(x) \) for \( \forall f \in H, x \in \Omega. \) reproducing property

Reproducing property makes computation easy and feasible.

\[
\text{e.g.) For } f = \sum_{i=1}^{n} a_i k(\cdot, X_i), \ g = \sum_{j=1}^{m} b_j k(\cdot, X_j) \\
\langle f, g \rangle_H = \sum_{i,j} a_i b_j k(X_i, X_j)
\]

- Example: Gaussian kernel

\[ k : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}, \quad k(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right) \]

There is a RKHS on \( \mathbb{R}^m \) with reproducing kernel \( k. \)
RKHS and Independence

**Independence and characteristic functions**

Random variables $X$ and $Y$ are independent

\[ \iff E_{XY}[e^{-i\omega^T X} e^{-i\eta^T Y}] = E_X[e^{-i\omega^T X}] E_Y[e^{-i\eta^T Y}] \quad \text{for all } \omega \text{ and } \eta. \]

$e^{-i\omega^T X}$ and $e^{-i\eta^T Y}$ work as test functions which account for the infinite degree of freedom ($L^2$).

**RKHS characterization**

$H_X$ and $H_Y$ are RKHS on $\Omega_X$ and $\Omega_Y$, respectively.

Random variables $X \in \Omega_X$ and $Y \in \Omega_Y$ are independent

\[ \iff E_{XY}[f(X)g(Y)] = E_X[f(X)] E_Y[g(Y)] \quad \text{for all } f \in H_X, \ g \in H_Y \]

This is true if $H_X$ and $H_Y$ are RKHS for Gaussian kernels.

(Bach & Jordan 2002)
Cross-covariance Operator

■ Definition

$X$ and $Y$: random variable on $\Omega_X$ and $\Omega_Y$, respectively.
$H_X$ and $H_Y$: RKHS on $\Omega_X$ and $\Omega_Y$, respectively, with bounded kernels.
We can define a bounded operator $\Sigma_{YX} : H_X \to H_Y$ by

$$\langle g, \Sigma_{YX} f \rangle_{H_Y} = E_{XY}[f(X)g(Y)] - E_X[f(X)]E_Y[g(Y)] \quad (= \text{Cov}[f(X), g(Y)])$$
for all $f \in H_X, \ g \in H_Y$

$\Sigma_{YX}$ is called cross-covariance operator.

■ Cross-covariance operator and Independence

Theorem

$H_X$ and $H_Y$: RKHS with Gaussian kernel.

$X$ and $Y$ are independent $\iff \Sigma_{YX} = 0$
RKHS and Conditional Independence

Conditional covariance

$X$ and $Y$ are random vectors. $H_X, H_Y :$ RKHS with kernel $k_X, k_Y,$ resp.

Assumption: $\exists \Sigma_{XX}^{-1}, \ E_{Y|X}[g(Y) \mid X] \in H_X$ for all $g \in H_Y.$

$$\left\langle f, \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} g \right\rangle = E_X \left[ \text{Cov}_{Y|X} \left[ f(Y), g(Y) \mid X \right] \right]$$

Def. $\Sigma_{YY|X} \equiv \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$ : conditional covariance operator

\textit{c.f.} For Gaussian $\text{Cov}_{Y|X} \left[ a^T Y, b^T Y \mid X = x \right] = a^T \left( V_{YY} - V_{YX} V_{XX}^{-1} V_{XY} \right) b$

- Monotonicity of conditional covariance operators

$Y, X = (U, V) :$ random vectors

$$\Sigma_{YY|U} \geq \Sigma_{YY|X}$$

$\geq :$ in the sense of self-adjoint operators
RKHS and Conditional Independence

### Conditional independence

Theorem

\[ X = (U, V) \] and \( Y \) are random vectors.

\( H_X, H_U, H_Y : \text{RKHS with Gaussian kernel} \ k_X, k_U, k_Y, \text{resp.} \)

\[ E_{Y|X}[g(Y) | X] \in H_X \text{ and } E_{Y|U}[g(Y) | U] \in H_U \text{ for all } g \in H_Y. \]

\[ \iff \]

\[ Y \perp V | U \iff \Sigma_{YY|U} = \Sigma_{YY|X} \]

### Minimization of conditional covariance operator

\[ \min_{B: U = B^T X} \Sigma_{YY|U} \iff B \text{ gives the effective subspace} \]

- Evaluation
  - Operator norm -- maximum eigenvalue.
  - Trace norm -- sum of eigenvalues
  - Determinant -- product of eigenvalues
Kernel Dimension Reduction

Estimation of conditional covariance operator

\((X^{(1)}, Y^{(1)}), \ldots, (X^{(n)}, Y^{(n)}) : \text{i.i.d. sample from the true joint probability.}\)

The space is restricted in the linear hull of \( \{k(\cdot, X^{(i)}) | 1 \leq i \leq n\} \)
and \( \{k(\cdot, Y^{(i)}) | 1 \leq i \leq n\} \)

Replace \( \Sigma_{YY|U} \) by \( n \times n \) matrix

\[
\hat{\Sigma}_{YY|U} = \hat{\Sigma}_{YY} - \hat{\Sigma}_{UY} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{UY}
\]

where

\[
\hat{\Sigma}_{UU} = (G_U + \epsilon I_n)^2, \quad \hat{\Sigma}_{YY} = (G_{YY} + \epsilon I_n)^2, \quad \hat{\Sigma}_{UY} = G_U G_Y
\]

\( \epsilon : \) regularization coefficient

\[
G_U = (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \left( k_U(U^{(i)}, U^{(j)}) \right) (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T)
\]

\[
G_Y = (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \left( k_Y(Y^{(i)}, Y^{(j)}) \right) (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T)
\]

reproducing property and empirical average
Kernel Dimension Reduction

Kernel dimension reduction (KDR)

\[ \min_B \hat{\Sigma}_{YY|U} = \hat{\Sigma}_{YY} - \hat{\Sigma}_{YU} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{UY} \quad \text{where} \quad U = B^TX \]

\[ \Leftrightarrow \min_B \det[I_n - \hat{\Sigma}_{YY}^{-1/2} \hat{\Sigma}_{YU} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{UY} \hat{\Sigma}_{YY}^{-1/2}] \]

\[ \Leftrightarrow \min_B \frac{\det \hat{\Sigma}_{[YU][YU]}}{\det \hat{\Sigma}_{YY} \det \hat{\Sigma}_{UU}} \]

Kernel generalized variance (KGV, Bach & Jordan 2002)

Kernel Dimension Reduction (KDR) = minimization of KGV

Kernel Dimension Reduction

- **Extension of Kernel ICA**
  - Kernel ICA (Bach & Jordan 02): kernel method for independence.
    - KDR: kernel method for conditional independence.

- **Wide applicability of KDR**
  - Semiparametric method: no assumptions on $p(Y|X)$.
  - KDR needs no strong assumption on the distribution of $X$, $Y$ and dimensionality of $Y$.
    - c.f. other method; SIR, pHd, CCA, PLS, etc.

- **Computational cost**
  - Multiplication of $n \times n$ matrices is computationally hard.
    - Incomplete Cholesky decomposition
  - Local minimum $\rightarrow$ annealing is used in gradient method.
Experiments

Synthesized data

- Data

\[ X: \text{2 dim}, \quad Y: \text{1 dim} \]

100 data

\[ Y \sim 2 \exp(-X_1^2) + N(0; 0.1^2) \]

- Results

<table>
<thead>
<tr>
<th>Angle (deg.)</th>
<th>SIR</th>
<th>pHd</th>
<th>CCA</th>
<th>PLS</th>
<th>KDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>-86.522</td>
<td>57.015</td>
<td>-10.416</td>
<td>-26.093</td>
<td>0.298</td>
<td></td>
</tr>
</tbody>
</table>
Wine data

- Data
  - 13 dim. 178 data.
  - 3 classes
  - 2 dim. projection

Experiments
Classification accuracy

- Purpose:
  to see how much information on $Y$ is maintained in the
  low-dimensional subspace of $X$.

- Test classification accuracy of Support Vector Machine after
  reducing dimensionality.

- Data sets for binary classification from UCI repository.

- Comparison with pHd.
  Many methods are NOT applicable for binary classification tasks.
Breast-cancer-Wisconsin

X: 30 dim.
# training data = 200
# test data = 369
Heart-disease

X: 13 dim.
# training data=149,
# test data=148
Experiments

Ionosphere

X: 34 dim.
# training data=151
# test data=200
Extension to Variable Selection

Variable selection by KGV

- Select subset \((X_{i_1}, ..., X_{i_d})\) from \(\{X_1, ..., X_m\}\).
- Principle
  \[
  Y \perp V \mid U \iff \Sigma_{YY|U} = \Sigma_{YY|X}
  \]
- KGV gives an objective function for variable selection.
  \[
  \min_U \frac{\det \hat{\Sigma}_{[YU][YU]}}{\det \hat{\Sigma}_{YY} \det \hat{\Sigma}_{UU}}
  \]
  \(\min\) is taken over subsets \(U = (X_{i_1}, ..., X_{i_d})\) where \(1 \leq i_1 < \cdots < i_d \leq m\)

- Problem: combinatorial explosion
  - \(mC_d\) evaluations are needed.
  - Calculation of all the combinations is possible only for small \(m\) and \(d\).
Experiments of Variable Selection

- **Small data set**
  - *Boston Housing:*
    - X :13 dim.,
    - Y = house price,
    - 506 data.
  - 4 variables are selected. 
    \[ _{13}C_4 = 715. \]
  - 4 variables are selected. 

ACE: Breiman & Friedman (1985)

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>ACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRIM</td>
<td></td>
<td>O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDUS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RM</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>AGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIS</td>
<td></td>
<td>O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAX</td>
<td>O</td>
<td></td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>PTRATIO</td>
<td>O</td>
<td>O</td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSTAT</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Variable Selection for Large Data Sets

- **Computational issue**
  - Combinatorial explosion
    If \( m \) and \( d \) are large, e.g. \( m=1000, d=20 \), evaluation of all the subsets is intractable.

- **Efficient optimization**
  - Greedy algorithm
    1. Start from one variable.
    2. For already chosen \( t \) variables \( S_t = \{X_{i_1}, ..., X_{i_t}\} \), evaluate KGV of \( S_t \cup \{X_j\} \) for all \( j \), and select the best one.
    3. Repeat this to \( d \) variables.

  - Random optimization
    Genetic algorithm
Application: Gene Selection

AML/ALL classification (Golub et al. 1999)
- Microarray data: 6817 dim. 38 data.
- Class label:
  AML (acute myeloid leukemia) / ALL (acute lymphoblastic leukemia).

Results
- 50 genes are selected by the kernel method and compared with previous works.
## Application: Gene Selection

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Golub99</th>
<th>Lee03</th>
<th>Szabo02</th>
<th>Li02</th>
<th>Fuj</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukotriene C4 synthase (LTC4S)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Zyxin</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>FAH Fumarylacetoacetate</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>LYN V-yes-1 Yamaguchi sarcoma</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>LEPR Leptin receptor</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CD33 CD33 antigen (differentiation)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Interleukin-8 mRNA for interferon-gamma</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>PRG1 Proteoglycan 1, secretory</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>DF D component of complement (ad)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>INTERLEUKIN-8 PRECURSOR</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>INDUCED MYELOID LEUKEMIA</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>PEPTIDYL-PROLYL CIS-TRANS</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Phosphotyrosine independent ligase</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>ATP6C Vacuolar H+ ATPase proton</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CST3 Cystatin C (Amyloid angiopathy)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>interleukin 8 (IL8) gene</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CTSD Cathepsin D (lysosomal aspa)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>TGAX Integrin, alpha X (antig)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>LALG3 Lectin, galactoside-bind</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Epb72 gene exon 1</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>MAJOR HISTOCOMPATIBILITY</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>LYZ Lysosome</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Azurocidin gene</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>PFC Properdin P factor, complement</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Lysocepholipase homolog (HU-K5)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>PGB Protective protein for beta</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Catalase (EC 1.1.1.6) 5'flank</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>FTH1 Ferritin heavy chain</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CD36 CD36 antigen (collagen type)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>EUKARYOTIC PEPTIDE CHAIN</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>GB DEF CD36 gene exon 15</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Colony-stimulating factor 1</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CA2 Carbonic anhydrase II</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Hepatocyte growth factor-like protein</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Mpo Myeloperoxidase</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CHRNA7 Cholinergic receptor, nI</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>AFX-HUMTFRRM1507_M_at</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>AFX-HUMTFRRM1507_M_at</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CINH Complement component 1 Inh</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>GB DEF Glycoprotein Sta (type)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>GYPE Glycoprotein E</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>AFX-HUMTFRRM1507_3_at</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Metabotropic glutamate receptor</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>GB DEF = Neutrophil elastase ge</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>ELA2 Elastase 2, neutrophil</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>GB DEF Kazal-type serine prote</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>LCAT Lechthin-cholesterol acyltr</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>ALDH2 Alcohol dehydrogenase 2</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>ANX8 Annexin VIII</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>PRSS3 Protease, serine, 3 (tryp</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

#agree/#selected 25/50 10/28 4/9 8/10 29/50
Summary

- **Kernel method for dimension reduction in regression**
  - Dimension reduction for regression = conditional independence.
  - Conditional covariance operators gives the criterion for the conditional independence.

- **Kernel dimension reduction / variable selection**
  - Have wide applicability to dimension reduction / variable selection. *c.f.* other methods have some restrictions.
  - Find effective subspaces / variables in practical problems.

- **Future/ongoing studies**
  - Theoretical analysis of the estimator: consistency etc.
  - How to choose the number of dimensions.
  - More efficient optimization techniques for variable selection.
  - Mixture of effective subspaces.