Machine Learning for Personalized Medicine

Jean-Philippe Vert

MascotNum 2014, ETH Zurich, April 24, 2014
1 body = $10^{14}$ cells
1 cell = $6 \times 10^9$ ACGT coding for 20,000 genes
Sequencing revolution

Cost per Genome

NIH National Human Genome Research Institute

genome.gov/sequencingcosts
A flood of *omics* data

- Interactome
- Genome
- Phenome
- Transcriptome
- Epigenome
- Publications
- Mutations
- Structural variations
Cancer

NORMAL CELL

FIRST MUTATION
Cell seems normal but is predisposed to proliferate excessively

SECOND MUTATION
Cell begins to proliferate too much but is otherwise normal

THIRD MUTATION
Cell proliferates more rapidly; it also undergoes structural changes

MALIGNANT CELL

FOURTH OR LATER MUTATION
Cell grows uncontrollably and looks obviously deranged
A cancer cell
A cancer cell
A cancer cell
Opportunities

- What is your risk of developing a cancer? *(prevention)*
- After diagnosis and treatment, what is the risk of relapse? *(prognosis)*
- What specific treatment will cure your cancer? *(personalized medicine)*
1. Learning molecular classifiers with network information

2. Kernel bilinear regression for toxicogenomics
Outline

1. Learning molecular classifiers with network information

2. Kernel bilinear regression for toxicogenomics
Breast cancer prognosis

Gene-Expression Profiling

St. Gallen Criteria

No. at Risk

Good signature 60 57 54 45 31 22 12
Poor signature 91 72 55 41 26 17 9
Low risk 120 107 88 68 48 34 19
High risk 2 2 1 17 9 5 2
Learning with regularization

Given a training set \((x_i, y_i)_{i=1,...,n}\) where \(x_i \in \mathbb{R}^p\) (typically, \(n = 200, p = 20,000\)), we estimate a linear predictor

\[
f_\beta(x) = \beta^\top x
\]

by solving

\[
\min_{\beta \in \mathbb{R}^p} R(\beta) + \lambda \Omega(\beta)
\]

where:

- \(R(\beta)\) is a convex empirical risk, typically

\[
R(\beta) = \frac{1}{n} \sum_{i=1}^{n} \ell(\beta^\top x_i, y_i)
\]

for some loss function \(\ell\) (squared error, logistic loss, hinge loss...)

- \(\Omega(\beta)\) is a regularization term, typically \(\| \beta \|_2\) (ridge regression, SVM...) or \(\| \beta \|_1\) (lasso...)
The idea

- We look for a **limited set** of genes that are sufficient for prediction.
- Selected genes should inform us about the underlying biology.
Lack of stability of signatures

Haury et al. (2011)
Basic biological functions usually involve the **coordinated action of several proteins:**
- Formation of **protein complexes**
- Activation of metabolic, signalling or regulatory **pathways**

Many pathways and protein-protein interactions are **already known**

**Hypothesis:** the weights of the classifier should be “coherent” with respect to this **prior knowledge**
Graph based penalty

\[ f_\beta(x) = \beta^T x \quad \min_\beta R(f_\beta) + \lambda \Omega(\beta) \]

Prior hypothesis

Genes near each other on the graph should have similar weights.

An idea (Rapaport et al., 2007)

\[
\Omega(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2, \\
\min_{\beta \in \mathbb{R}^p} R(f_\beta) + \lambda \sum_{i \sim j} (\beta_i - \beta_j)^2.
\]
Graph based penalty

\[ f_\beta(x) = \beta^T x \]

\[ \min_{\beta} R(f_\beta) + \lambda \Omega(\beta) \]

Prior hypothesis

Genes near each other on the graph should have similar weights.

An idea (Rapaport et al., 2007)

\[ \Omega(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2, \]

\[ \min_{\beta \in \mathbb{R}^p} R(f_\beta) + \lambda \sum_{i \sim j} (\beta_i - \beta_j)^2. \]
The Laplacian of the graph is the matrix $L = D - A$.

$$L = D - A = \begin{pmatrix}
1 & 0 & -1 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 \\
-1 & -1 & 3 & -1 & 0 \\
0 & 0 & -1 & 2 & -1 \\
0 & 0 & 0 & 1 & 1
\end{pmatrix}$$
Theorem

The function \( f(x) = \beta^T x \) where \( \beta \) is solution of

\[
\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \ell \left( \beta^T x_i, y_i \right) + \lambda \sum_{i \sim j} (\beta_i - \beta_j)^2
\]

is equal to \( g(x) = \gamma^T \Phi(x) \) where \( \gamma \) is solution of

\[
\min_{\gamma \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \ell \left( \gamma^T \Phi(x_i), y_i \right) + \lambda \gamma^T \gamma,
\]

and where

\[
\Phi(x)^T \Phi(x') = x^T K_G x'
\]

for \( K_G = L^* \), the pseudo-inverse of the graph Laplacian.
Example

\[ L^* = \begin{pmatrix}
0.88 & -0.12 & 0.08 & -0.32 & -0.52 \\
-0.12 & 0.88 & 0.08 & -0.32 & -0.52 \\
0.08 & 0.08 & 0.28 & -0.12 & -0.32 \\
-0.32 & -0.32 & -0.12 & 0.48 & 0.28 \\
-0.52 & -0.52 & -0.32 & 0.28 & 1.08 \\
\end{pmatrix} \]
Rapaport et al

Classifiers

N Glycan biosynthesis

Protein kinases

DNA and RNA polymerase subunits

Nitrogen, asparagine metabolism

Oxidative phosphorylation, TCA cycle

Purine metabolism

Phenylalanine, tyrosine and tryptophan biosynthesis

Lysine biosynthesis

Biosynthesis of steroids, ergosterol metabolism

Folate biosynthesis

Riboflavin metabolism

Sulfur metabolism

Porphyrin and chlorophyll metabolism

Glycolysis / Gluconeogenesis

N Glycan biosynthesis
Spectral analysis of gene expression profiles using gene networks

Fig. 5. [Diagram of gene networks]

This work was supported by the grant ACI-IMPBIO-2004-47 of the French Ministry for Research and New Technologies.
Other penalties with kernels

\[ \Phi(x)^\top \Phi(x') = x^\top K_G x' \]

with:

- \( K_G = (c + L)^{-1} \) leads to

\[ \Omega(\beta) = c \sum_{i=1}^{p} \beta_i^2 + \sum_{i \sim j} (\beta_i - \beta_j)^2. \]

- The diffusion kernel:

\[ K_G = \exp_M(-2tL). \]

penalizes high frequencies of \( \beta \) in the Fourier domain.
Other penalties without kernels

- Gene selection + Piecewise constant on the graph

\[ \Omega(\beta) = \sum_{i \sim j} |\beta_i - \beta_j| + \sum_{i=1}^{p} |\beta_i| \]

- Gene selection + smooth on the graph

\[ \Omega(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2 + \sum_{i=1}^{p} |\beta_i| \]
Graph lasso

Two solutions

\[ \Omega_{\text{intersection}}(\beta) = \sum_{i \sim j} \sqrt{\beta_i^2 + \beta_j^2}, \]

\[ \Omega_{\text{union}}(\beta) = \sup_{\alpha \in \mathbb{R}^p : \forall i \sim j, \|\alpha_i^2 + \alpha_j^2\| \leq 1} \alpha^\top \beta. \]
Generalization: Group lasso with overlapping groups

\[ \Omega_{\text{latent}}^G (w) \triangleq \begin{cases} \min_v \sum_{g \in G} \| v_g \|_2^2 \\ w = \sum_{g \in G} v_g \\ \text{supp} (v_g) \subseteq g. \end{cases} \]

Properties

- Resulting support is a union of groups in \( G \).
- Possible to select one variable without selecting all the groups containing it.
- Equivalent to group lasso when there is no overlap.
Consistency in group support (Jacob et al., 2009)

- Let \( \tilde{w} \) be the true parameter vector.
- Assume that there exists a unique decomposition \( \tilde{v}_g \) such that \( \tilde{w} = \sum_g \tilde{v}_g \) and \( \Omega_{\text{latent}}^G (\tilde{w}) = \sum \| \tilde{v}_g \|_2 \).
- Consider the regularized empirical risk minimization problem \( L(w) + \lambda \Omega_{\text{latent}}^G (w) \).

Then

- under appropriate mutual incoherence conditions on \( X \),
- as \( n \to \infty \),
- with very high probability,

the optimal solution \( \hat{w} \) admits a unique decomposition \( (\hat{v}_g)_{g \in G} \) such that

\[
\{ g \in G | \hat{v}_g \neq 0 \} = \{ g \in G | \tilde{v}_g \neq 0 \}.
\]
### Consistency in group support (Jacob et al., 2009)

- Let $\bar{w}$ be the true parameter vector.
- Assume that there exists a unique decomposition $\bar{v}_g$ such that $\bar{w} = \sum_g \bar{v}_g$ and $\Omega^g_{\text{latent}}(\bar{w}) = \sum \|\bar{v}_g\|_2$.
- Consider the regularized empirical risk minimization problem $L(\bar{w}) + \lambda \Omega^g_{\text{latent}}(\bar{w})$.

Then

- under appropriate mutual incoherence conditions on $X$,
- as $n \to \infty$,
- with very high probability,

the optimal solution $\hat{w}$ admits a unique decomposition $(\hat{v}_g)_{g \in G}$ such that

$$\{g \in G | \hat{v}_g \neq 0\} = \{g \in G | \bar{v}_g \neq 0\}.$$
Experiments

Synthetic data: overlapping groups

- 10 groups of 10 variables with 2 variables of overlap between two successive groups: \{1, \ldots, 10\}, \{9, \ldots, 18\}, \ldots, \{73, \ldots, 82\}.
- Support: union of 4\textsuperscript{th} and 5\textsuperscript{th} groups.
- Learn from 100 training points.

Frequency of selection of each variable with the lasso (left) and \(\Omega^G_{\text{latent}}\) (middle), comparison of the RMSE of both methods (right).
Lasso signature (accuracy 0.61)
Graph Lasso signature (accuracy 0.64)
Outline

1. Learning molecular classifiers with network information
2. Kernel bilinear regression for toxicogenomics
Pharmacogenomics / Toxicogenomics

Patients with same condition

DNA Profiling

Good responders

Bad side effects

No Responders
Genotypes from the 1000 genome project
RNASEq from the Geuvadis project
Cell line $X$, chemical $Y$, toxicity $Z$.

Bilinear regression model:

$$ Z = f(X, Y) + b(Y) + \epsilon, $$

Estimation by kernel ridge regression:

$$ \min_{f \in \mathcal{H}, b \in \mathbb{R}^p} \sum_{i=1}^n \sum_{j=1}^p (f(x_i, y_j) + b_j - z_{ij})^2 + \lambda \| f \|^2, $$
Solving in $O(\text{max}(n, p)^3)$

**Theorem 1.** Let $Z \in \mathbb{R}^{n \times p}$ be the response matrix, and $K_X \in \mathbb{R}^{n \times n}$ and $K_Y \in \mathbb{R}^{p \times p}$ be the kernel Gram matrices of the $n$ cell lines and $p$ chemicals, with respective eigenvalue decompositions $K_X = U_X D_X U_X^\top$ and $K_Y = U_Y D_Y U_Y^\top$. Let $\gamma = U_X^\top 1_n$ and $S \in \mathbb{R}^{n \times p}$ be defined by $S_{ij} = 1/(\lambda + D_X^i D_Y^j)$, where $D_X^i$ (resp. $D_Y^i$) denotes the $i$-th diagonal term of $D_X$ (resp. $D_Y$). Then the solution $(f^*, b^*)$ of (2) is given by

$$b^* = U_Y \text{Diag} \left( S^\top \gamma^2 \right)^{-1} \left( S^\top \circ \left( U_Y^\top Z^\top U_X \right) \right) \gamma \quad (3)$$

and

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \quad f^*(x, y) = \sum_{i=1}^n \sum_{j=1}^p \alpha^*_{i,j} K_X(x_i, x) K_Y(y_i, y), \quad (4)$$

where

$$\alpha^* = U_X \left( S \circ \left( U_X^\top \left( Z - 1_n b^* \right) U_Y \right) \right) U_Y^\top \quad (5)$$
Kernel Trick

cell line descriptors

drug descriptors
Kernel Trick

Cell line descriptors → Kernelized → Kcell

Drug descriptors → Kernelized → Kdrug
Kernel Trick

- Cell line descriptors
- Drug descriptors
- Kernelized
- Kernel bilinear regression
- \( \hat{f} \)
- Kernel choice?
- Descriptors
- Data integration
- Missing data
Kernel choice

1. **$K_{\text{cell}}$** :
   - 29 cell line kernels tested
   - 1 kernel that *integrate all information*
   - deal with missing data

2. **$K_{\text{drug}}$** :
   - 48 drug kernels tested
   - *multi-task* kernels
Kernel choice

1. $K_{\text{cell}}$:
   - 29 cell line kernels tested
   - 1 kernel that integrate all information
   - deal with missing data

2. $K_{\text{drug}}$:
   - 48 drug kernels tested
   - multi-task kernels
Cell line data integration

**Covariates**
- linear kernel

**SNPs**
- 10 gaussian kernels

**RNA-seq**
- 10 gaussian kernels
Cell line data integration

Covariates
- linear kernel

SNPs
- 10 gaussian kernels

RNA-seq
- 10 gaussian kernels

Integrated kernel
Multi-task drug kernels

1. Dirac
2. Multi-Task
3. Feature-based
4. Empirical
5. Integrated

Independent regression for each drug
Multi-task drug kernels

- **Dirac**
- **Multi-Task**
- Feature-based
- Empirical
- Integrated

sharing information across drugs
Multi-task drug kernels

1. Dirac
2. Multi-Task
3. Feature-based
4. Empirical
5. Integrated

Linear kernel and 10 gaussian kernels based on features:

- CDK (160 descriptors) and SIRMS (9272 descriptors)
- Graph kernel for molecules (2D walk kernel)
- Fingerprint of 2D substructures (881 descriptors)
- Ability to bind human proteins (1554 descriptors)
Multi-task drug kernels

Dirac
Multi-Task
Feature-based
Empirical
Integrated
Multi-task drug kernels

1. Dirac
2. Multi-Task
3. Feature-based
4. Empirical
5. Integrated

Integrated kernel:
- Combine all information on drugs
29x48 kernel combinations: CV results

Color Key and Histogram

Value

0 0.5 0.51 0.52 0.53

Count

0 150

CI

KsnpRbf6.txt
KsnpRbf7.txt
KsnpRbf8.txt
KrnaseqRbf1.txt
KsnpRbf5.txt
KsnpRbf4.txt
KcovariatesSex.txt
KsnpRbf2.txt
KsnpRbf3.txt
KsnpRbf1.txt
KrnaseqRbf3.txt
KrnaseqRbf2.txt
KsnpMean
uniform
dirac
KrnaseqRbf6.txt
KrnaseqRbf7.txt
KrnaseqRbf5.txt
KrnaseqRbf8.txt
KrnaseqRbf10.txt
KrnaseqRbf9.txt
KrnaseqRbf4.txt
KrnaseqMean
KcovariatesBatch.txt
KcovariatesPopulation
Kint
Kcovariates.txt
29x48 kernel combinations: CV results

Color Key and Histogram

Value

CI

integrated and covariates kernels

KsnpRbf6.txt
KsnpRbf7.txt
KsnpRbf8.txt
KrnaseqRbf1.txt
KsnpRbf5.txt
KsnpRbf4.txt
KcovariatesSex.txt
KsnpRbf2.txt
KsnpRbf3.txt
KsnpRbf1.txt
KrnaseqRbf3.txt
KrnaseqRbf2.txt
KsnpMean
uniform
dirac
KrnaseqRbf6.txt
KrnaseqRbf7.txt
KrnaseqRbf5.txt
KrnaseqRbf8.txt
KrnaseqRbf10.txt
KrnaseqRbf9.txt
KrnaseqRbf4.txt
KrnaseqMean
KcovariatesBatch.txt
KcovariatesSex.txt
KcovariatesOperatio
Kint
Kcovariates.txt

Ksubstructure.txt
Kchemcpp.txt
Kmultitask1
KsirmsRbf1.txt
KcdkRbf1.txt
KpredtargetRbf1.txt
KsirmsRbf2.txt
KcdkRbf2.txt
KpredtargetRbf2.txt
KcdkRbf3.txt
KpredtargetRbf3.txt
KcdkRbf4.txt
KpredtargetRbf4.txt
KcdkRbf5.txt
KpredtargetRbf5.txt
KcdkRbf6.txt
KpredtargetRbf6.txt
KcdkRbf7.txt
KpredtargetRbf7.txt
KcdkRbf8.txt
KpredtargetRbf8.txt
KcdkRbf9.txt
KpredtargetRbf9.txt
KcdkRbf10.txt
KpredtargetRbf10.txt
KsnpMean
KsnpRbf1.txt
KsnpRbf2.txt
KsnpRbf3.txt
KsnpRbf4.txt
KsnpRbf5.txt
KsnpRbf6.txt
KsnpRbf7.txt
KsnpRbf8.txt
KsnpRbf9.txt
KsnpRbf10.txt
29x48 kernel combinations: CV results

Color Key and Histogram

Value

CI

Count

0

0.5

0.51

0.52

0.53

sightly multi-task on drugs

covariates kernel on cell lines

Kcovariates.txt

KcovariatesMean

KcovariatesSex.txt

KcovariatesPopulation.txt

KcovariatesBatch.txt

KrnaseqMean

KrnaseqRbf4.txt

KrnaseqRbf9.txt

KrnaseqRbf10.txt

KrnaseqRbf8.txt

KrnaseqRbf7.txt

KrnaseqRbf6.txt

KpredtargetRbf6.txt

KpredtargetRbf5.txt

KpredtargetRbf4.txt

KpredtargetMean

KsnpMean

KsnpRbf6.txt

KsnpRbf7.txt

KsnpRbf8.txt

KsnpRbf5.txt

KsnpRbf4.txt

Kemploi.txt

dirac

uniform

count

0

150

Color Key

and Histogram

Value
Kernel on cell lines: CV results

- integrated kernel
- batch effect

Mean CI for cell line kernels

0.50 0.51 0.52 0.53 0.54

integrated kernel
batch effect

jeudi 7 novembre 13
Mean CI for chemicals kernels

KpredtargetRbf10.txt
Kmultitask10
KpredtargetRbf7.txt
KsirmsRbf8.txt
KcdkRbf8.txt
KpredtargetRbf6.txt
KpredtargetRbf5.txt
KsirmsRbf7.txt
KpredtargetRbf4.txt
Ksubstructure.txt
Kempirical
KpredtargetMean
KcdkRbf7.txt
KpredtargetRbf3.txt
KpredtargetMean
KcdkRbf5.txt
KcdkRbf6.txt
Kmultitask7
KcdkRbf4.txt
Kmultitask9
Kmultitask8
KsirmsRbf5.txt
KsirmsMean
Kmultitask6
KcdkMean
Kmultitask2
Kmultitask4
Kint
KsirmsRbf4.txt
Kmultitask5
Kmultitask3
KcdkRbf2.txt
KsirmsRbf3.txt
KcdkRbf3.txt
KsirmsRbf1.txt
KsirmsRbf8.txt
KpredtargetRbf2.txt
Kmultitask1
KsirmsRbf2.txt
KcdkRbf1.txt
KpredtargetRbf11.txt
Ksubstructure.txt
Kmultitask11
Kchemcpp.txt

0.50 0.51 0.52 0.53 0.54
Thanks