1 Introduction

2 Support Vector Machine Classifiers

3 Heat Induced Cellular Death

4 Toxicological Evaluation of Titania Nanoparticles

5 Current & Future Research
Cancer causes about 13% of all human deaths
7.6 million deaths worldwide due to cancer in 2007

**CURRENT TREATMENTS**

- Radiation
  - Toxic for other tissues
- Chemotherapy
  - High toxicity
- Surgery
  - Complication and high cost
- Gene/bone marrow transplantation
  - High cost

The cost of the therapy is estimated to $18,000 - $37,000 per patient in a total of $72B annually.
Toxicity

- A shop in Beijing, China: Polyacrylic coating application on polystyrene boards
  - Coatings used by the workers contain nanoparticles
- Seven women workers (aged 18-47) were identified with shortness of breath
  - Clinical examinations reveal hypoxemia (low oxygen saturation in blood), pleural effusion (fluid in thoracic cavity), and pericardial effusion (fluid around heart)
  - Severe skin rash from intense itching of their faces, hands, and forearms

**Motivation**

**Toxicity**

- A shop in Beijing, China: Polyacrylic coating application on polystyrene boards
  - Coatings used by the workers contain nanoparticles
- Seven women workers (aged 18-47) were identified with shortness of breath
  - Clinical examinations reveal hypoxemia (low oxygen saturation in blood), pleural effusion (fluid in thoracic cavity), and pericardial effusion (fluid around heart)
  - Severe skin rash from intense itching of their faces, hands, and forearms
- Two other women from the same shop died (ages 19 and 29)
  - Pathology samples from the workers lungs identified 30 nm (nanometer) in diameter particles

Criticisms

“I think the paper should never have been published without characterizing the exposure and the toxicological reactivity of the nanoparticles before blaming the effects on them.”

“The real tragedy here is that these workers could have been protected if a conventional chemical hygiene plan had been implemented that included a working ventilation system and personal protective equipment. Preventing inhalation of 30-nm nanoparticles can be as simple as the proper use of an inexpensive mask sold by your neighborhood home improvement store.”

These researchers are funded by DuPont, Intel, L’oreal, Lockheed Martin, Proctor&Gamble, Unidym.
Cell Death Discrimination

Investigate *in-situ* the mechanism that the particles are inducing by Titania nanoparticles and heat

- Lung epithelia cancer cells (A549 from ATCC) are grown for 24 hours
- 785nm high power near infrared diode laser for spectroscopy analysis (this light does not damage cells even after 60min at 115mW power)
  - Cells have been grown on MgF$_2$ crystal for minimizing background radiation
  - After reaching 80% confluency, cells are treated with the testing material
- Drug or heat exposure
- Recording for the next 24 hours
Cell Death Discrimination

- Helps understanding pathological processes
  - induced by a disease
  - induced by pharmaceutical treatments such as anti-cancer drugs
- Raman spectroscopy is employed to assess the potential toxicity of chemical substances
  - noninvasive and does not require chemicals or markers
  - measurements can be taken rapidly and in real time
  - it is possible to analyze the health of either a single cell or the entire population
  - has no interference with water and CO₂
Raman Spectroscopy

Observes vibrational, rotational, and other low-frequency modes
Relies on Raman scattering of monochromatic light from a laser
Laser light interacts with molecular vibrations, phonons or other excitations
The energy levels of the laser photons are shifted up or down
Drawback: complexity of the obtained spectra

Traditionally, peak fitting has been used to analyze Raman spectra
- time consuming due to large number of overlapping peaks
- series of mathematical procedures to remove the baseline, the fluorescence, to normalize the spectra, and calculate the average and the standard deviation
- prior knowledge needed on which peaks are discriminant with limited interference from background noise
Outline

1. Introduction

2. Support Vector Machine Classifiers

3. Heat Induced Cellular Death

4. Toxicological Evaluation of Titania Nanoparticles

5. Current & Future Research
## An Example: Flu Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Blood type</th>
<th>Body Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>F</td>
<td>A(-)</td>
<td>99</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>M</td>
<td>AB(+)</td>
<td>104</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>M</td>
<td>B(+)</td>
<td>103</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>M</td>
<td>0(+)</td>
<td>98</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>F</td>
<td>B(-)</td>
<td>105</td>
<td>?</td>
</tr>
</tbody>
</table>

Patients 1-4: Training Set  
Patient 5: Test Set  
Goal: Minimize the generalization error
## An Example: Flu Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Blood type</th>
<th>Body Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>A(-)</td>
<td>99</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>M</td>
<td>AB(+)</td>
<td>104</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>B(+)</td>
<td>103</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>M</td>
<td>0(+)</td>
<td>98</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>F</td>
<td>B(-)</td>
<td>105</td>
<td>?</td>
</tr>
</tbody>
</table>

Patients 1-4: Training Set

Patient 5: Test Set

Goal: Minimize the generalization error
### An Example: Flu Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Blood type</th>
<th>Body Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>A(-)</td>
<td>99</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>M</td>
<td>AB(+)</td>
<td>104</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>B(+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>M</td>
<td>0(+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>F</td>
<td>B(-)</td>
<td>105</td>
<td></td>
</tr>
</tbody>
</table>
### An Example: Flu Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Blood type</th>
<th>Body Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>A(-)</td>
<td>99</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>M</td>
<td>AB(+)</td>
<td>104</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>B(+)</td>
<td>103</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>F</td>
<td>B(-)</td>
<td>105</td>
<td></td>
</tr>
</tbody>
</table>

Patients 1-4: Training Set  
Patient 5: Test Set  
Goal: Minimize the generalization error
### An Example: Flu Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Blood type</th>
<th>Body Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>A(-)</td>
<td>99</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>M</td>
<td>AB(+)</td>
<td>104</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>B(+)</td>
<td>103</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>M</td>
<td>0(+)</td>
<td>98</td>
<td>-</td>
</tr>
</tbody>
</table>

Patients 1-4: Training Set
Patient 5: Test Set

Goal: Minimize the generalization error
### An Example: Flu Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Blood type</th>
<th>Body Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>A(-)</td>
<td>99</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>M</td>
<td>AB(+)</td>
<td>104</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>B(+)</td>
<td>103</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>M</td>
<td>0(+)</td>
<td>98</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>F</td>
<td>B(-)</td>
<td>105</td>
<td>?</td>
</tr>
</tbody>
</table>
## An Example: Flu Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Blood type</th>
<th>Body Temperature</th>
<th>FLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>A(-)</td>
<td>99</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>M</td>
<td>AB(+)</td>
<td>104</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>B(+)</td>
<td>103</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>M</td>
<td>0(+)</td>
<td>98</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>F</td>
<td>B(-)</td>
<td>105</td>
<td>?</td>
</tr>
</tbody>
</table>

- Patients 1-4: *Training Set*
- Patient 5: *Test Set*
- Goal: Minimize the generalization error
Support Vector Machines

Feature 1

Feature 2

Control Cell
Death Cell

Erhun Kundakcioglu (University of Houston) SVMs for Toxic Evaluation of Nanoparticles March 30, 2011 11 / 45
Support Vector Machines

Feature 1

Feature 2

Control Cell

Death Cell
Support Vector Machines

Feature 1

Feature 2

Control Cell
Death Cell

Erhun Kundakcioglu (University of Houston) SVMs for Toxic Evaluation of Nanoparticles March 30, 2011 11 / 45
Support Vector Machines

Feature 1

Feature 2

Control Cell
Death Cell
Support Vector Machines

Feature 1

Feature 2

Control Cell
Death Cell
Support Vector Machines

Feature 1

margin

Feature 2

Control Cell
Death Cell
Support Vector Machines

Support Vector Machine Classifiers

Erhun Kundakcioglu (University of Houston)

SVMs for Toxic Evaluation of Nanoparticles

March 30, 2011 11 / 45
Support Vector Machines

Feature 1

Feature 2

margin

Control Cell
Death Cell

Erhun Kundakcioglu (University of Houston) SVMs for Toxic Evaluation of Nanoparticles March 30, 2011 11 / 45
Support Vector Machines

Feature 1

Feature 2

Control Cell
Death Cell

Erhun Kundakcioglu (University of Houston) SVMs for Toxic Evaluation of Nanoparticles

March 30, 2011 11 / 45
Developed by Vladimir Vapnik in 1995

State-of-the-art supervised learning methods

Classify two linearly/nonlinearly separable sets of pattern vectors
  - Kernel trick

Quadratic convex optimization problem
  - Efficiently solved
  - Can also be formulated as a SOCP problem

Minimizes the generalization error

Widely used; e.g., pattern recognition, text categorization, biomedicine, brain-computer interface, financial applications.
Hard Margin Classifiers

- Given the data $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, +1\}$ for $i = 1, \ldots, n$
Support Vector Machine Classifiers

Hard Margin Classifiers

- Given the data $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, +1\}$ for $i = 1, \ldots, n$
- Separating hyperplane is defined as $\langle w, x \rangle + b = 0$
Hard Margin Classifiers

- Given the data \( x_i \in \mathbb{R}^d \) and labels \( y_i \in \{-1, +1\} \) for \( i = 1, \ldots, n \)
- Separating hyperplane is defined as \( \langle w, x \rangle + b = 0 \)
- The geometric margin between the hyperplane and the nearest data point \( x^* \) is

\[
\frac{|\langle w, x^* \rangle + b|}{\|w\|}
\]
Hard Margin Classifiers

- Given the data $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, +1\}$ for $i = 1, \ldots, n$
- Separating hyperplane is defined as $\langle w, x \rangle + b = 0$
- The geometric margin between the hyperplane and the nearest data point $x^*$ is
  \[
  \frac{|\langle w, x^* \rangle + b|}{\|w\|} \rightarrow \text{functional margin}
  \]
Hard Margin Classifiers

- Given the data \( x_i \in \mathbb{R}^d \) and labels \( y_i \in \{-1, +1\} \) for \( i = 1, \ldots, n \)
- Separating hyperplane is defined as \( \langle w, x \rangle + b = 0 \)
- The geometric margin between the hyperplane and the nearest data point \( x^* \) is

\[
\frac{|\langle w, x^* \rangle + b|}{\|w\|} \rightarrow \text{functional margin}
\]

\[
\downarrow
\]

\[
\begin{align*}
\text{max} & \quad \text{geometric margin} \\
\text{subject to} & \quad \text{functional margin} = 1
\end{align*}
\]
Hard Margin Classifiers

- Given the data $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, +1\}$ for $i = 1, \ldots, n$
- Separating hyperplane is defined as $\langle w, x \rangle + b = 0$
- The geometric margin between the hyperplane and the nearest data point $x^*$ is

$$\frac{|\langle w, x^* \rangle + b|}{\|w\|} \rightarrow \text{functional margin}$$

$$\downarrow$$

$$\max \text{ geometric margin}$$
subject to $\text{functional margin} = 1$

$$\downarrow$$

$$\min_{w,b} \frac{1}{2} \|w\|^2$$
subject to $y_i(\langle w, x_i \rangle + b) \geq 1 \quad i = 1, \ldots, n$
Support Vector Machine Classifiers

Soft Margin Classifiers

Feature 1

Feature 2

Control Cell
Death Cell
Support Vector Machine Classifiers

Soft Margin Classifiers

Feature 1

Feature 2

error ($\xi$)

Control Cell
Death Cell

Erhun Kundakcioglu (University of Houston) SVMs for Toxic Evaluation of Nanoparticles March 30, 2011 14 / 45
Soft Margin Classifiers

In the context of Support Vector Machines (SVMs), soft margin classifiers allow for some data points to be on the wrong side of the decision boundary, introducing a small amount of error.

Feature 1

Feature 2

classification error ($\xi$)

Control Cell

Death Cell
Soft Margin Classifiers

\[ \min_{w, b, \xi} \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{n} \xi_i^2 \]

subject to \[ y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i \quad i = 1, \ldots, n \]
Outline

1. Introduction
2. Support Vector Machine Classifiers
3. Heat Induced Cellular Death
4. Toxicological Evaluation of Titania Nanoparticles
5. Current & Future Research
Each row represents a lung epithelia cell
Each column represents the Raman outputs at different energy levels
There are \( \sim 1300 \) columns
Some cells are subject to either Etoposide or Triton X-100
- Etoposide is a strong chemotherapeutic drug, used for Ewing’s sarcoma, lung cancer, testicular cancer, lymphoma, non-lymphocytic leukemia, and glioblastoma multiforme
- Triton X-100 raptures the cellular membrane and results in the necrotic death of the cells

The effect of abnormal heat (45 °C for 2 hours)
Effect of Abnormal Heat - Heat Induced Cellular Death

- Training Set:
  - Etoposide vs. control cells
  - Triton X-100 vs. control cells
  - Triton X-100 vs. Etoposide

- Test Set: Cells that are subject to 45 °C for 2 hours
Computational Results for Death Cell Discrimination

Distance from the Hyperplane

Sample ID's

- Control A549 (24 Hrs)
- Etoposide (24Hrs)
- Heating at 45°C
Computational Results for Death Cell Discrimination

- Control (24 hrs)
- Triton-X (24 hrs)
- Heating at 45°C

Distance from the Hyperplane

Sample ID's
Computational Results for Death Cell Discrimination

Distance from the Hyperplane

Sample ID's

- Triton-X (24 hrs)
- Etoposide (24 hrs)
- Heating 45°C
Heat treatment can be used as a form of *programmed cell death*

Foundation for developing diagnostic tools for cancer or other genetic diseases

Powerful monitoring of the cellular response and disease treatment with chemotherapy and radiation
Outline

1. Introduction
2. Support Vector Machine Classifiers
3. Heat Induced Cellular Death
4. Toxicological Evaluation of Titania Nanoparticles
5. Current & Future Research
Experiment / Data Set

- Each row represents a lung epithelia cell
- Each column represents the Raman outputs at different energy levels
- There are \( \sim 1300 \) columns
- Some cells are subject to either Etoposide or Triton X-100
  - Etoposide is a strong chemotherapeutic drug, used for Ewing’s sarcoma, lung cancer, testicular cancer, lymphoma, non-lymphocytic leukemia, and glioblastoma multiforme
  - Triton X-100 raptures the cellular membrane and results in the necrotic death of the cells
- The effect of TiO\(_2\)
Titania Nanoparticles

- Claim: Titania Nanoparticles / Anatase (i.e., TiO$_2$) is potentially toxic
- Training Set:
  - Etoposide vs. control cells
  - Triton X-100 vs. control cells
- Test set: Cells that are subject to TiO$_2$
Results: Cross Validation

Figure: Cross-validation results with leave-one-out method: apoptotic vs. healthy.
Results: Cross Validation

Figure: Cross-validation results with leave-one-out method: necrotic vs. healthy.
Results: Apoptotic-Healthy

Figure: Progression from healthy to apoptotic along the course of the 36 hrs.
Results: Necrotic-Healthy

Figure: Inconclusive since the distances from the separation plane are scattered along both regions.
Remarks

- Classification between apoptotic and healthy cells is in agreement with previous literature.
- There is a progression from healthy to apoptotic for Titania exposed cells in 36 hrs.
- Classification between necrotic and healthy cells did not show any particular trend.
- The MTT assay shows the cytotoxicity of Titania nanoparticles, which is in agreement with other reports claiming that titania nanoparticles are potentially toxic.
  - MTT assay: Laboratory test for measuring the activity of enzymes that reduce MTT to formazan giving a purple color.
Outline

1. Introduction
2. Support Vector Machine Classifiers
3. Heat Induced Cellular Death
4. Toxicological Evaluation of Titania Nanoparticles
5. Current & Future Research

Erhun Kundakcioglu (University of Houston) SVMs for Toxic Evaluation of Nanoparticles March 30, 2011 31 / 45
Observation: SVM classifier is extremely sensitive to outliers

- Robust techniques: SVMs with hard margin loss
  - Minimize the number of misclassified points

\[
\min_{w, b, z} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} z_i \\
\text{subject to} \quad y_i (\langle w, x_i \rangle + b) \geq 1 \quad \text{if } z_i = 0, \ i = 1,\ldots,n \\
z_i \in \{0, 1\} \quad \quad \quad \quad i = 1,\ldots,n
\]
SVMs with Hard Margin Loss

- The formulation can accommodate nonlinear kernel functions
- Convex quadratic IP formulation
- Proven to be $NP$-hard
- More robust (i.e., less sensitive to outliers compared to Hinge loss function)
- Universally consistent
  - A classifier is *consistent* if the probability of misclassification converges in expectation to a Bayes optimal rule as sample size is increased
  - A classifier is *universally consistent* if it is consistent for all distributions of data
SVMs with Hinge Loss

\[
\begin{align*}
\min_{w, b, \xi} & \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{n} \xi_i^2 \\
\text{subject to} & \quad y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \\
& \quad i = 1, \ldots, n
\end{align*}
\]
SVMs with Hinge Loss

\[
\begin{align*}
\text{min}_{w, b, \xi} & \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{n} \xi_i^2 \\
\text{subject to} & \quad y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i \\
& \quad i = 1, \ldots, n
\end{align*}
\]
SVMs with Hinge Loss

\[
\begin{align*}
\min_{w,b,\xi} & \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{n} \xi_i^2 \\
\text{subject to} & \quad y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \quad i = 1, \ldots, n
\end{align*}
\]
SVMs with Hinge Loss

\[
\begin{align*}
\min_{w, b, \xi} & \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{n} \xi_i^2 \\
\text{subject to} & \quad y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i \\
& \quad \xi_i \geq 0, \quad i = 1, \ldots, n
\end{align*}
\]
SVMs with Hard Margin Loss

\[
\min_{w, b, z} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} z_i \\
\text{subject to} \quad y_i(\langle w, x_i \rangle + b) \geq 1 \quad \text{if } z_i = 0, \ i = 1, \ldots, n \\
\quad z_i \in \{0, 1\} \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \q
SVMs with Hard Margin Loss

$$\min_{w,b,z} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} z_i$$

subject to

$$y_i(\langle w, x_i \rangle + b) \geq 1$$

if $$z_i = 0, \quad i = 1, \ldots, n$$

$$z_i \in \{0, 1\}$$

$$i = 1, \ldots, n$$
## Multiple Instance Learning - Drug Activity Prediction

<table>
<thead>
<tr>
<th>Drug</th>
<th>Conformation</th>
<th>Naproxen</th>
<th>Sodium</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c1</td>
<td>50</td>
<td>50</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c2</td>
<td>20</td>
<td>30</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c3</td>
<td>30</td>
<td>20</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c4</td>
<td>20</td>
<td>40</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c5</td>
<td>30</td>
<td>30</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c6</td>
<td>60</td>
<td>45</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c7</td>
<td>5</td>
<td>5</td>
<td>Ineffective</td>
</tr>
<tr>
<td></td>
<td>c8</td>
<td>10</td>
<td>40</td>
<td>Ineffective</td>
</tr>
<tr>
<td></td>
<td>c9</td>
<td>40</td>
<td>10</td>
<td>Ineffective</td>
</tr>
<tr>
<td></td>
<td>c10</td>
<td>60</td>
<td>50</td>
<td>Ineffective</td>
</tr>
<tr>
<td></td>
<td>c11</td>
<td>10</td>
<td>10</td>
<td>Ineffective</td>
</tr>
<tr>
<td></td>
<td>c12</td>
<td>10</td>
<td>50</td>
<td>Ineffective</td>
</tr>
<tr>
<td></td>
<td>c13</td>
<td>50</td>
<td>10</td>
<td>Ineffective</td>
</tr>
</tbody>
</table>

Erhun Kundakcioglu (University of Houston)  SVMs for Toxic Evaluation of Nanoparticles  March 30, 2011  36 / 45
### Multiple Instance Learning - Drug Activity Prediction

<table>
<thead>
<tr>
<th>Drug</th>
<th>Conformation</th>
<th>Naproxen</th>
<th>Sodium</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c1</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c2</td>
<td>20</td>
<td>30</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c3</td>
<td>30</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c4</td>
<td>20</td>
<td>40</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c5</td>
<td>30</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>c6</strong></td>
<td><strong>60</strong></td>
<td><strong>45</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c7</td>
<td>5</td>
<td>5</td>
<td>Effective</td>
</tr>
<tr>
<td></td>
<td>c8</td>
<td>10</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c9</td>
<td>40</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>c10</strong></td>
<td><strong>60</strong></td>
<td><strong>50</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c11</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c12</td>
<td>10</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c13</td>
<td>50</td>
<td>10</td>
<td>Ineffective</td>
</tr>
</tbody>
</table>
Multiple Instance Learning - Drug Activity Prediction

Effective (+)

Ineffective (−)
Multiple Instance Learning - Drug Activity Prediction

Naproxen

Sodium

Effective (+)

Ineffective (−)

Erhun Kundakcioglu (University of Houston) SVMs for Toxic Evaluation of Nanoparticles March 30, 2011 37 / 45
Multiple Instance Learning - Drug Activity Prediction

Erhun Kundakcioglu (University of Houston)  SVMs for Toxic Evaluation of Nanoparticles  March 30, 2011  37 / 45
Multiple Instance Learning - Drug Activity Prediction

Select at least one from each positive bag such that the margin is maximized.
Multiple Instance Learning - Drug Activity Prediction

Keep all negative instances

Select at least one from each positive bag such that the margin is maximized.
Multiple Instance Learning - Drug Activity Prediction

- Effective (+)
- Ineffective (−)

Naproxen
Sodium
Current & Future Research

Other Applications of Multiple Instance Learning

- Drug/molecular activity prediction
- Medical image annotation
- Analysis of noisy time series data (e.g., local field potential, EEG)
- Nanotoxicity
Hard Margin Multiple Instance Classification: Formulation 1

\[ \eta_i = \begin{cases} 
1 & \text{if instance } i \text{ is selected} \\
0 & \text{otherwise} 
\end{cases} \]

\[
\min \frac{1}{2} \|w\|^2 + C \sum_i v_i \\
\text{s.t.} \quad -\langle w, x_i \rangle - b \geq 1 - Mv_i \quad \forall i \in I^-
\]

\[
\langle w, x_i \rangle + b \geq 1 - Mv_i - M(1 - \eta_i) \quad \forall i \in I^+
\]

\[
\sum_{i \in I} \eta_i = 1 \\
\quad \forall j \in J^+
\]

\[ v_i \in \{0, 1\} \quad \forall i \in I^+ \]

\[ \eta_i \in \{0, 1\} \quad \forall i \in I^+ \]
Hard Margin Multiple Instance Classification: Formulation 2

- $\hat{v}_j$: Minimum misclassification for each positive bag.

$$\begin{align*}
\min & \quad \frac{1}{2} \|w\|^2 + C \sum_{i \in I^-} v_i + C \sum_{j \in J^+} \hat{v}_j \\
\text{s.t.} & \quad - (\langle w, x_i \rangle + b) \geq 1 - Mv_i; \quad \forall i \in I^- \\
& \quad \langle w, x_i \rangle + b \geq 1 - Mv_i; \quad \forall i \in I^+ \\
& \quad \hat{v}_j = \sum_{i \in I_j} \eta_i v_i; \quad \forall j \in J^+ \\
& \quad \sum_{i \in I_j} \eta_i = 1; \quad \forall j \in J^+ \\
& \quad 0 \leq \eta_i \leq 1 \quad \forall i \in I^+ \\
& \quad \hat{v}_j \in \{0, 1\} \quad \forall j \in J^+ \\
& \quad v_i \in \{0, 1\} \quad \forall i
\end{align*}$$
Hard Margin Multiple Instance Classification: Linearized Formulation 2

\[
\begin{align*}
\min & \quad \frac{1}{2} \|w\|^2 + C \sum_{i \in I^-} v_i + C \sum_{j \in J^+} \hat{v}_j \\
\text{s.t.} & \quad -(\langle w, x_i \rangle + b) \geq 1 - Mv_i; \quad \forall i \in I^- \\
& \quad \langle w, x_i \rangle + b \geq 1 - Mv_i; \quad \forall i \in I^+ \\
& \quad \hat{v}_j = \sum_{i \in I_j} \hat{z}_i; \quad \forall j \in J^+ \\
& \quad \hat{z}_i \geq -1 + \eta_i + v_i; \quad \forall i \in I^+ \\
& \quad \hat{z}_i \leq \eta_i; \quad \forall i \in I^+ \\
& \quad \hat{z}_i \leq v_i; \quad \forall i \in I^+ \\
& \quad \sum_{i \in I_j} \eta_i = 1; \quad \forall j \in J^+ \\
& \quad 0 \leq \hat{z}_i \leq 1 \quad \forall i \in I^+ \\
& \quad \hat{v}_j \in \{0, 1\} \quad \forall j \in J^+ \\
& \quad v_i \in \{0, 1\} \quad \forall i \in I^+
\end{align*}
\]
Hard Margin Multiple Instance Classification: Formulation 3

\[
\begin{align*}
\min & \quad \frac{1}{2} \|w\|^2 + C \sum_{i \in I^-} v_i + C \sum_{j \in J^+} \hat{v}_j \\
\text{s.t.} & \quad - (\langle w, x_i \rangle + b) \geq 1 - M v_i; \quad \forall i \in I^- \\
& \quad \langle w, x_i \rangle + b \geq 1 - M v_i; \quad \forall i \in I^+ \\
& \quad \hat{v}_j \geq \sum_{i \in I_j} v_i - |I_j| + 1; \quad \forall j \in J^+ \\
& \quad v_i \in \{0, 1\} \quad \forall i
\end{align*}
\]
Molecular activity prediction


<table>
<thead>
<tr>
<th># of feat.</th>
<th># of ins.</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>CP1</th>
<th>CP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>40</td>
<td>0.052</td>
<td>0.092</td>
<td>0.056</td>
<td>0.368</td>
<td>0.074</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>0.064</td>
<td>0.091</td>
<td>0.068</td>
<td>0.218</td>
<td>0.057</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>0.128</td>
<td>0.157</td>
<td>0.135</td>
<td>0.28</td>
<td>0.107</td>
</tr>
<tr>
<td>80</td>
<td>40</td>
<td>0.371</td>
<td>0.328</td>
<td>0.302</td>
<td>0.625</td>
<td>0.283</td>
</tr>
<tr>
<td>166</td>
<td>40</td>
<td>0.777</td>
<td>0.681</td>
<td>0.661</td>
<td>1.075</td>
<td>0.786</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>0.414</td>
<td>0.722</td>
<td>0.506</td>
<td>2.676</td>
<td>0.788</td>
</tr>
<tr>
<td>40</td>
<td>80</td>
<td>1.164</td>
<td>1.618</td>
<td>1.476</td>
<td>4.916</td>
<td>0.454</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
<td>3.606</td>
<td>4.648</td>
<td>3.028</td>
<td>8.23</td>
<td>1.586</td>
</tr>
<tr>
<td>10</td>
<td>160</td>
<td>N/A</td>
<td>127.44</td>
<td>91.31</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>20</td>
<td>160</td>
<td>62.455</td>
<td>5.51</td>
<td>45.555</td>
<td>128.82</td>
<td>10.555</td>
</tr>
<tr>
<td>40</td>
<td>160</td>
<td>42.65</td>
<td>18.76</td>
<td>53.975</td>
<td>71.695</td>
<td>9.825</td>
</tr>
<tr>
<td>80</td>
<td>160</td>
<td>59.655</td>
<td>66.14</td>
<td>45.035</td>
<td>1767.635</td>
<td>8.025</td>
</tr>
<tr>
<td>166</td>
<td>160</td>
<td>152.17</td>
<td>114.505</td>
<td>242.31</td>
<td>583.665</td>
<td>20.395</td>
</tr>
</tbody>
</table>
## Results

### Molecular activity prediction


<table>
<thead>
<tr>
<th># of feat.</th>
<th># of ins.</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>CP1</th>
<th>CP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>200</td>
<td>N/A</td>
<td>371.2</td>
<td>70.35</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>65.14</td>
<td>13.625</td>
<td>6.265</td>
<td>N/A</td>
<td>251.465</td>
</tr>
<tr>
<td>40</td>
<td>200</td>
<td>30.035</td>
<td>805.17</td>
<td>58.815</td>
<td>87.33</td>
<td>12.72</td>
</tr>
<tr>
<td>80</td>
<td>200</td>
<td>191.87</td>
<td>105.38</td>
<td>N/A</td>
<td>N/A</td>
<td>21.805</td>
</tr>
<tr>
<td>166</td>
<td>200</td>
<td>677.84</td>
<td>584.21</td>
<td>572.89</td>
<td>770.43</td>
<td>35.055</td>
</tr>
<tr>
<td>10</td>
<td>300</td>
<td>N/A</td>
<td>124.45</td>
<td>87.48</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>20</td>
<td>300</td>
<td>95.54</td>
<td>357.99</td>
<td>N/A</td>
<td>N/A</td>
<td>331.77</td>
</tr>
<tr>
<td>40</td>
<td>300</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>375.79</td>
</tr>
<tr>
<td>80</td>
<td>300</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>152</td>
</tr>
<tr>
<td>166</td>
<td>300</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>400</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>40</td>
<td>400</td>
<td>435.8633</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>644.69</td>
</tr>
<tr>
<td>80</td>
<td>400</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>166</td>
<td>400</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>469.43</td>
</tr>
</tbody>
</table>

Erhun Kundakcioglu (University of Houston)  SVMs for Toxic Evaluation of Nanoparticles  March 30, 2011  44 / 45
Acknowledgments

- **Georgios Pyrgiotakis**
  Particle Engineering Research Center, University of Florida

- **Mohammad H. Poursaeidi**
  Department of Industrial Engineering, University of Houston

Q&A