An Introduction to Support Vector Machines

Ming-Chang Lee

Department of Information Management
Yu Da College of Business
Outline

1. Support Vector Machines
2. Binary classification
3. R demo
4. References
1. Support Vector Machines

- History:
  Statistical Learning Theory (Vapnik 1998)

- Development:
  - Binary classification SVM
  - Multi-class SVM
Application:
- Text categorization
- Image recognition
- Hand-written Digit Recognition
- Bioinformatics
2. Binary classification

- Consider a two-class, linearly separable classification problem:
  - training data \((x_i, y_i), i = 1, \ldots, l\). \(\{x_i, y_i\}, x_i \in \mathbb{R}^n\)

\[
y_i = \begin{cases} 
+1, & x_i \in \text{class 1} \\
-1, & x_i \in \text{class 2}
\end{cases}
\]
Making decision boundary

Class 1: $y_i = +1$
Class 2: $y_i = -1$
Large-margin Decision Boundary

\[ \text{margin} = (d+) \, \text{与} \, (d-) \rightarrow \text{Goal: Max \{margin\}} \]

\[ y_i = +1 \]

\[ y_i = -1 \]

\[ w^T x + b = 0 \]
Find maximum margin

\[ w^T x_i + b \geq 1 \quad \text{if} \quad y_i = +1 \]
\[ w^T x_i + b \leq 1 \quad \text{if} \quad y_i = +1 \]

\[ m_{\text{argin}} = (d_+) + (d_-) = \frac{1}{\|w\|} + \frac{1}{\|w\|} = \frac{2}{\|w\|} = \frac{2}{\sqrt{w^T w}} \]

Minimize cost function \( \Phi(w) = \frac{1}{2} w^T w \)
Subject to \( y_i(w^T x_i + b) \geq 1 \quad \forall i = 1, \ldots, l. \)
Lagrange multipliers function

Minimize  \[ J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{l} \alpha_i \left[ y_i (w^T x_i + b) - 1 \right] \]

Subject to  \[ \alpha_i \geq 0 \quad \forall i = 1, \ldots, l. \]

\[
\frac{dJ(w, b, \alpha)}{dw} = 0, \quad w = \sum_{i=1}^{l} \alpha_i y_i x_i
\]

\[
\frac{dJ(w, b, \alpha)}{db} = 0, \quad \sum_{i=1}^{l} \alpha_i y_i = 0
\]
Dual problem

\[ J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{l} \alpha_i y_i w^T x_i - b \sum_{i=1}^{l} \alpha_i y_i + \sum_{i=1}^{l} \alpha_i \]

\[ w^T w = \sum_{i=1}^{l} \alpha_i y_i w^T x_i = \sum_{i=1}^{l} \sum_{i=1}^{l} \alpha_i \alpha_j y_i y_j x_i^T x_j \]

Maximize \[ Q(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} \alpha_i \alpha_j y_i y_j x_i^T x_j \]

Subject to

1. \[ \sum_{i=1}^{l} \alpha_i y_i = 0 \]
2. \[ \alpha_i \geq 0 \quad \forall i = 1, 2, \ldots, l \]
Solution - Support vectors

- Many of the $\alpha_i$ are zero
- Support vectors (SV) $x_i$ :
  - Non-zero $\alpha_i$
  - The decision boundary is determined only by the SV
  - $w = \sum_{i=1}^{l} \alpha_i y_i x_i$
- Testing with a new data $x_{new}$ :
  - Compute $w^T x_{new} + b$
  - If the result is positive, then classify $x_{new}$ as class 1, class 2 otherwise
3. R demo

- Data set: Glass
- Numbers 214, input variables: 9.
- Training data: 143(2/3)
- Testing data: 71(1/3)
- Use SVM to classify
How to get the R?

The Comprehensive R Archive Network

Precompiled Binary Distributions
Base system and contributed packages. Windows and Mac users most likely want these versions of R.

- Linux
- MacOS X (10.2 x and above) This version of R for the Mac is actively maintained.
- MacOS (System 8.6 to 9.1 and MacOS X up to 10.1.x) Last supported version of R is 1.7.1, there will be no more updates.
- Windows (95 and later)

Source Code for all Platforms
Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

What are R and CRAN?

R is ‘GNU S’, a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN mirror nearest to you to minimize network load.
### Index of /bin

<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td>14-Oct-2005 19:19</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>linux/</td>
<td>21-Feb-2005 11:27</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>macos/</td>
<td>19-Apr-2005 09:48</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>package/</td>
<td>19-Oct-2005 15:31</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>windows/</td>
<td>17-Dec-2004 16:38</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

*Apache/1.3.33 Server at cran.r-project.org Port 80*
R for Windows

This directory contains binaries for a base distribution and packages to run on Windows (NT, 95 and later) on Intel and clones (but not NT on Alpha and other platforms).

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions with downloaded executables.

Subdirectories:
- base
  - Binaries for base distribution (managed by Duncan Murdoch)
- contrib
  - Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions/suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Last modified: April 4, 2004, by Friedrich Leisch
R-2.2.0 for Windows

This directory contains a binary distribution of R-2.2.0 to run on Windows 95, 98, ME, NT4.0, 2000 and XP on Intel/clone chips.

Patches to this release are incorporated in the r-patched snapshot build.

A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.

In this directory:

- README.R-2.2.0: Installation and other instructions.
- CHANGES: New features of this Windows version.
- NEWS: New features of all versions.
- R-2.2.0-win32.exe: Setup program (about 25 megabytes). Please download this from a mirror near you.
  This corresponds to the file named SetupR.exe or rwXXXX.exe in pre-R-2.2.0 releases.
- old: The previous release.

Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information, including upgrade advice.

Note to webmasters: A stable link which will redirect to the current Windows binary release is <CRAN MIRROR>-bin/windows-base/release.htm.

Last change: 2005-10-06, by Duncan Murdoch
4. References


THANKS

Q & A

Web:  http://web.ydu.edu.tw/~alan9956/
Email: alan9956@webmail.ydu.edu.tw