Introduction to Machine Learning
General Concepts

Alina Sîrbu

Institute for Scientific Interchange, Turin, Italy
alina.sirbu@isi.it

Big Dive
10th October 2012
Introduction: Basic concepts (10 Oct)

Machine Learning Methods (10, 11 Oct)
  ▶ Doodle poll: http://doodle.com/8quzeweqfk6r4bza3

Machine Learning Practical (15, 17 Oct)
Definition

- “Machine learning is a scientific discipline concerned with the design and development of algorithms that take as input empirical data, such as that from sensors or databases, and yield patterns or predictions thought to be features of the underlying mechanism that generated the data” (Wikipedia)
- Automatically find patterns in the data in order to understand it (knowledge) and make predictions.
Why?

Everyday life: our brain uses learned patterns to help us recognise objects, make decisions, etc.

Mainly two purposes:
- Build algorithms that mimic human behaviour in some way (face recognition, ....)
- Understand data that our brain in not able to process (high dimensionality, ....)

So it all boils down to ....
Why?

Example:

- Target Supermarket chain - Revenue increased by about 50% (Charles Duhigg, "The Power of Habit: Why We Do What We Do in Life and Business")
Learning

- Process of analysing the data to find patterns
  - Unsupervised: no a priori knowledge about data.
  - Supervised: with data for which we have some knowledge.
Supervised machine learning

- Classification: separate group of objects into (discrete) classes.
  - Classes are already known.
  - For some example objects, we know the class they belong to.
  - Classes can be 2 or more.
  - E.g. Target classify clients into pregnant/not pregnant.
Supervised machine learning

- Regression: predict (continuous) value of interest based on other attribute values
  - Statistics tool - express the value of a dependent variable as a function of some independent variables
  - When used to make predictions - machine learning (focus here)
  - Also used to infer relationships between variables, causality, etc.
  - Machine Learning e.g. Predicting the price of housing based on surface, location, previous trends.
Unsupervised machine learning

- Clustering: find groups in the data based on similarities between objects.
  - Clusters (groups) are not previously known (not even their number).
  - There are no objects for which cluster membership is known.
  - E.g. cultural clusters, network community detection, recommendation systems.

- Note: supervised clustering methods exist, but here, for simplicity, we will consider only unsupervised clustering.

- Clustering vs. classification?
Learning: General Method

Existing Data → Train → Model → New Data

Predictions / Knowledge
Data

- “objects”, “data points”, “instances”
- Measurements of a process/situation..
- Features: attributes of the objects/instances
- Toy e.g. A bank is trying to classify clients into “elligible for loan (E)” or “not elligible (NE)”.
  - Two types of features
  - Qualitative (a few possible discrete values)
  - Quantitative (numerical value)
## Bank example

D(Degree), N(Nationality), S(Sex), PC(Previous Client), CR(Criminal Record), RP(Rent Payed), No dep(Number of dependents), MP(Missed Payments), Res(Result of classification)

<table>
<thead>
<tr>
<th>Id</th>
<th>D</th>
<th>N</th>
<th>S</th>
<th>PC</th>
<th>CR</th>
<th>Age</th>
<th>Salary</th>
<th>RP</th>
<th>No dep</th>
<th>MP</th>
<th>Res</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PhD</td>
<td>1</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>50</td>
<td>3000</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>BSc</td>
<td>2</td>
<td>M</td>
<td>1</td>
<td>0</td>
<td>28</td>
<td>2700</td>
<td>600</td>
<td>0</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>HS</td>
<td>1</td>
<td>F</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>2000</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>MSc</td>
<td>1</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>55</td>
<td>3500</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>5</td>
<td>HS</td>
<td>3</td>
<td>F</td>
<td>0</td>
<td>0</td>
<td>49</td>
<td>2200</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>HS</td>
<td>1</td>
<td>M</td>
<td>1</td>
<td>0</td>
<td>75</td>
<td>2000</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>NE</td>
</tr>
<tr>
<td>7</td>
<td>BSc</td>
<td>1</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>38</td>
<td>2000</td>
<td>700</td>
<td>3</td>
<td>10</td>
<td>NE</td>
</tr>
<tr>
<td>8</td>
<td>HS</td>
<td>2</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>83</td>
<td>1800</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NE</td>
</tr>
<tr>
<td>9</td>
<td>MSc</td>
<td>1</td>
<td>F</td>
<td>1</td>
<td>1</td>
<td>44</td>
<td>1500</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>NE</td>
</tr>
<tr>
<td>10</td>
<td>PhD</td>
<td>3</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>28</td>
<td>1200</td>
<td>600</td>
<td>1</td>
<td>0</td>
<td>NE</td>
</tr>
</tbody>
</table>

**New entry to classify:**

<table>
<thead>
<tr>
<th>Id</th>
<th>D</th>
<th>N</th>
<th>S</th>
<th>PC</th>
<th>CR</th>
<th>Age</th>
<th>Salary</th>
<th>RP</th>
<th>No dep</th>
<th>MP</th>
<th>Res</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>BSc</td>
<td>1</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>2500</td>
<td>600</td>
<td>1</td>
<td>0</td>
<td>?</td>
</tr>
</tbody>
</table>
Learning: General Method

Existing Data → Train → Model

New Data → Apply → Prediction / Knowledge
Training

- Training/Learning/Inference/Fitting/reverse engineering
- Supervised methods
  - Separate data into *train* and *test* datasets
  - Apply some algorithm to the train dataset to obtain model
  - Verify the results on the test data (new objects)
  - Two errors: training and testing errors
- Unsupervised methods
  - Apply algorithm to all data.
Cross-validation

Helps with assessing

- Overfitting - training error very small, test error very high
- Generalisation - training and testing errors comparable

Methods:

- Leave-one-out cross validation
- K-fold cross validation - use $\frac{1}{K}$ of data points for testing
### Feature selection

- Back to our bank example

<table>
<thead>
<tr>
<th>Id</th>
<th>D</th>
<th>N</th>
<th>S</th>
<th>PC</th>
<th>CR</th>
<th>Age</th>
<th>Salary</th>
<th>RP</th>
<th>No dep</th>
<th>MP</th>
<th>Res</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PhD</td>
<td>1</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>50</td>
<td>3000</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>BSc</td>
<td>2</td>
<td>M</td>
<td>1</td>
<td>0</td>
<td>28</td>
<td>2700</td>
<td>600</td>
<td>0</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>HS</td>
<td>1</td>
<td>F</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>2000</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>MSc</td>
<td>1</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>55</td>
<td>3500</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>5</td>
<td>HS</td>
<td>3</td>
<td>F</td>
<td>0</td>
<td>0</td>
<td>49</td>
<td>2200</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>HS</td>
<td>1</td>
<td>M</td>
<td>1</td>
<td>0</td>
<td>75</td>
<td>2000</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>NE</td>
</tr>
<tr>
<td>7</td>
<td>BSc</td>
<td>1</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>38</td>
<td>2000</td>
<td>700</td>
<td>3</td>
<td>10</td>
<td>NE</td>
</tr>
<tr>
<td>8</td>
<td>HS</td>
<td>2</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>83</td>
<td>1800</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NE</td>
</tr>
<tr>
<td>9</td>
<td>MSc</td>
<td>1</td>
<td>F</td>
<td>1</td>
<td>1</td>
<td>44</td>
<td>1500</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>NE</td>
</tr>
<tr>
<td>10</td>
<td>PhD</td>
<td>3</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>28</td>
<td>1200</td>
<td>600</td>
<td>1</td>
<td>0</td>
<td>NE</td>
</tr>
</tbody>
</table>
Feature selection

- Sometimes features are redundant or irrelevant
- Removing these → improved performance, faster training, less costs (e.g. bio), !black-box
- How?
  - Filter methods: Rank features based on some criteria, use only top $n$.
  - Wrapper methods: Generate subsets of features and apply Machine Learning technique repeatedly, optimisation algorithm (minimise error).
  - Combination: e.g. Start from top ranked features and increase $n$ until error stops decreasing.
Learning: General Method

Existing Data

Train

Model

New Data

Apply

Prediction/Knowledge
Prediction/Analysis

- Classification, regression: Apply model to new data, to find class or new value.
- Clustering: use obtained clusters to extract other knowledge from data, process objects in clusters in different ways.
  - Send customized leaflets to each cluster
  - Make recommendations based on past purchases
  - Dimensionality reduction (genetic data) - consider cluster as a single entity in further analyses
Next up: Machine Learning Methods