CSCI 590: Machine Learning

Lecture 1: Introduction to Machine Learning
Instructor: Murat Dundar

Acknowledgement:
Some of these slides are taken from course textbook website
http://research.microsoft.com/~cmbishop/prml/
What is Machine Learning?

Machine learning is about teaching a computer to interpret real-world observations.

Learning takes place when the computer can transfer knowledge learned from past observations to interpreting new observations.
Workflow Diagram

Training Data → Domain Knowledge → Expert Knowledge

Future Data → Model → Output
The fundamental problem in ML is how to make sure the computer is learning, not memorizing?

Computers learn quantitatively!

We need to integrate robust models into computers’ memories
- to accurately represent real-world observations
- to accurately match representations with domain knowledge
APPLICATIONS
Can we use machine learning to detect lesions in MR/CT images with a higher sensitivity?
Is This Breast Cancer or Not?

Can we use machine learning to help improve accuracy of diagnosis?
WHICH BACTERIA IS MORE DANGEROUS?

Optical “Fingerprints” of bacteria

A

B

C

D

Can we use machine learning for biosurveillance?
REMOTE SENSING AND SATELLITE IMAGERY

What does the bottom part of the image contain different than the top part? Can we use ML to detect anomalies?
What percentage of crops is corn? Can we use ML for agricultural planning?
What minerals exist on MARS? Can we find water-carrying minerals? Can we use ML to better understand how MARS was formed?
Can we use ML to automatically learn what each image contain?

Source: http://cs.stanford.edu/people/karpathy/cnnembed/cnn_embed_1k.jpg
Can we use ML to automatically learn topics of documents in a corpus?
Can we use ML to automatically tag a new medical abstract using one of the medical subject headers? MeSH has over 27,000 headers.
Can we use ML to identify disease causing genes?
Can we use ML to identify special interest groups such as people with similar interests and hobbies as well as criminals and terrorists.

Source: http://intelligentmeasurement.files.wordpress.com/2011/02/linkedin.gif
Can we use ML to recommend a customer which movie to watch or which product to buy?

Image Source: http://www.bridgewell.com/content_portal.html
Learning

\[ \alpha^T x = 0 \]
Many Disciplines Involved

Programming
Computing
Statistics
Probability
Linear Algebra
Optimization
Application Domains
Algorithms

Machine Learning
Data Mining
The Fundamental Problem

The fundamental problem in ML is how to make sure the computer is learning, not memorizing?

Computers learn quantitatively!

We need to integrate robust models into computers’ memories
  - to accurately represent real-world observations
  - to accurately match representations with domain knowledge
Challenges (1)

How to represent observations
  - quantitative modeling and characterization
How to filter out noise in representations
  - what if computer treats noise as information
How to choose a model
  - trade-off between a simple vs. complex model
  - when to prefer a simple model over a complex one
Challenges (2)

How to evaluate learning
- testing computers’ learning skills before deploying them in real-world environments

How to acquire reliable domain knowledge
- is our own understanding of the observations correct?
- what do we do if we have an ambiguous understanding?

How can we guide learning
- incorporating background information into learning
Application Areas (1)

- Biomarker identification: gene-protein-function-disease-drug relations
- Drug discovery and repurposing
- Safer diagnosis, better prognosis: Computer-aided diagnosis/prognosis
- Predict stock market, commodity prices, future predictions about economy
- Automate tasks: handwritten digit/text/speech recognition, image annotation, web site categorization, topic modeling
Application Areas (2)

- Understand networks and relational data to discover implicit relationships between objects, people etc.
  - Social networks (Facebook, Twitter)
  - Computer networks
  - Electric circuits
  - Protein-protein-gene interaction networks
- Analyze satellite imagery, video streams for real-time surveillance
- Predict major disease outbreaks
- Detect anomalies, intruders, hackers etc.
Learning vs. Overfitting

\[ y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j \]
Sum-of-Squares Error Function

\[ E(w) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 \]
0\textsuperscript{th} Order Polynomial
$1^{st}$ Order Polynomial

\[ M = 1 \]
3rd Order Polynomial
9\textsuperscript{th} Order Polynomial
Over-fitting

Root-Mean-Square (RMS) Error: $E_{RMS} = \sqrt{\frac{2E(w^*)}{N}}$
## Polynomial Coefficients

<table>
<thead>
<tr>
<th></th>
<th>$M = 0$</th>
<th>$M = 1$</th>
<th>$M = 3$</th>
<th>$M = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0^*$</td>
<td>0.19</td>
<td>0.82</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>$w_1^*$</td>
<td></td>
<td>-1.27</td>
<td>7.99</td>
<td>232.37</td>
</tr>
<tr>
<td>$w_2^*$</td>
<td></td>
<td></td>
<td>-25.43</td>
<td>-5321.83</td>
</tr>
<tr>
<td>$w_3^*$</td>
<td></td>
<td></td>
<td>17.37</td>
<td>48568.31</td>
</tr>
<tr>
<td>$w_4^*$</td>
<td></td>
<td></td>
<td></td>
<td>-231639.30</td>
</tr>
<tr>
<td>$w_5^*$</td>
<td></td>
<td></td>
<td></td>
<td>640042.26</td>
</tr>
<tr>
<td>$w_6^*$</td>
<td></td>
<td></td>
<td></td>
<td>-1061800.52</td>
</tr>
<tr>
<td>$w_7^*$</td>
<td></td>
<td></td>
<td></td>
<td>1042400.18</td>
</tr>
<tr>
<td>$w_8^*$</td>
<td></td>
<td></td>
<td></td>
<td>-557682.99</td>
</tr>
<tr>
<td>$w_9^*$</td>
<td></td>
<td></td>
<td></td>
<td>125201.43</td>
</tr>
</tbody>
</table>
Data Set Size: $N = 15$

9th Order Polynomial
Data Set Size: $N = 100$

9th Order Polynomial
Regularization

Penalize large coefficient values

$$\tilde{E}(w) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2} \|w\|^2$$
Regularization: $\ln \lambda = -18$
Regularization: $\ln \lambda = 0$
Regularization: $E_{RMS}$ vs. $\ln \lambda$
# Polynomial Coefficients

<table>
<thead>
<tr>
<th>$w_i$</th>
<th>$\ln \lambda = -\infty$</th>
<th>$\ln \lambda = -18$</th>
<th>$\ln \lambda = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0$</td>
<td>0.35</td>
<td>0.35</td>
<td>0.13</td>
</tr>
<tr>
<td>$w_1$</td>
<td>232.37</td>
<td>4.74</td>
<td>-0.05</td>
</tr>
<tr>
<td>$w_2$</td>
<td>-5321.83</td>
<td>-0.77</td>
<td>-0.06</td>
</tr>
<tr>
<td>$w_3$</td>
<td>48568.31</td>
<td>-31.97</td>
<td>-0.05</td>
</tr>
<tr>
<td>$w_4$</td>
<td>-231639.30</td>
<td>-3.89</td>
<td>-0.03</td>
</tr>
<tr>
<td>$w_5$</td>
<td>640042.26</td>
<td>55.28</td>
<td>-0.02</td>
</tr>
<tr>
<td>$w_6$</td>
<td>-1061800.52</td>
<td>41.32</td>
<td>-0.01</td>
</tr>
<tr>
<td>$w_7$</td>
<td>1042400.18</td>
<td>-45.95</td>
<td>-0.00</td>
</tr>
<tr>
<td>$w_8$</td>
<td>-557682.99</td>
<td>-91.53</td>
<td>0.00</td>
</tr>
<tr>
<td>$w_9$</td>
<td>125201.43</td>
<td>72.68</td>
<td>0.01</td>
</tr>
</tbody>
</table>