Machine Learning
Machine Learning

• Learning: improving with experience at some task
  • Improve over task
  • with respect to some performance measure
  • based on some experience
• Writing computer programs that write computer programs

Learning definition by Tom Mitchell
Outline

• Three machine learning stories/cautionary tales
• Deep learning definition
• Types of machine learning
• Best practices
Machine Learning Story 1
Face Detection & Recognition
Find photos by what’s in them

Looking for that photo of your pup? Just tap “dog” or the place you took it to find it faster.
What Does a Human Face Look Like?
Apple II image from wikipedia.com.
Eyes added digitally.
Apple II image from wikipedia.com.
Eyes added digitally.
if pixel153 > 128 & pixel154 > 128 &
pixel155 > 128 & pixel156 < 64 &
sqrt(pixel157) < 82 &
log(pixel1132 * pixel1133) > 1 ....
then image is a face*

* (not a real face recognition program)

Apple II image from wikipedia.com.
Eyes added digitally.
Machine Learning Story 2
Recommender Systems
Recommended for You

These recommendations are based on items you own and more.

Cybertext: Perspectives on Ergodic Literature
by Espen J. Aarseth (Aug 6, 1997)
Average Customer Review: ★★★★★ (3)
In Stock

List Price: $22.96
Price: $19.55

Add to cart
Add to cart

Recommended because you added Hamlet on the Holodeck to your Shopping Cart and more (Fix this)
Applications of Recommendation

- Movies
- Books
- Music
- Medicine
- Education
- Jobs
Applications of Recommendation

- Movies
- Books
- Music
- Medicine
- Education
- Jobs
Machine Learning Story 3
Housing Markets
Wall Street in the mid-1980s turned to the quants – brainy financial engineers – to invent new ways to boost profits. They and their managers, though laziness and greed, built a huge financial bubble on foundations that they did not understand. It was a recipe for disaster. The journalist Felix Salmon won the American Statistical Association's Excellence in Statistical Reporting Award for 2010. We reprint his article, first published as the cover story of Wired magazine, because it brilliantly conveys complex statistical concepts.
In the years before 2008, it was hardly unthinkable that a math wizard like David X. Li might some day earn a Nobel Prize. After all, financial economists even Wall Street quants – have received the Nobel in economics before, and Li’s work on measuring risk has had more impact, more quickly, than previous Nobel Prize-winning contributions to the field. Though, as dazed bankers, politicians, regulators and investors survey the wreckage of the biggest financial meltdown since the Great Depression, Li is probably thankful he still has a job in finance at all. Not that his achievement should be dismissed. He took a notoriously tough nut – determining correlation, or how seemingly disparate events are related – and cranked
The formula that killed so many pension plans: David X. Li's Gaussian copula, as first published in 2000. Investors exploited it as a quick – and fatally flawed – way to assess risk.

<table>
<thead>
<tr>
<th>Probability</th>
<th>Survival times</th>
<th>Equality</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifically, this is a joint default probability – the likelihood that any two members of the pool (A and B) will both default. It’s what investors are looking for, and the rest of the formula provides the answer.</td>
<td>The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone’s life expectancy when their spouse dies.</td>
<td>A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.</td>
<td></td>
</tr>
<tr>
<td>Copula</td>
<td>Distribution functions</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.</td>
<td>The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.</td>
<td>The all-powerful correlation parameter, which reduces correlation to a single constant – something that should be highly improbable, if not impossible. This is the magic number that made Li’s copula function irresistible.</td>
<td></td>
</tr>
</tbody>
</table>
Machine Learning Stories

- Face recognition
- Recommender systems
- Finance
What is deep learning?

raw image input → image preprocessing → edge detection → object detection → object identification

raw image input → learnable component (neural network) → another neural network → another neural network → object identification
Deep Learning

• Using machine learning to simultaneously train every part of the process from raw input to raw output

• Considered “deep” when compared to “shallow” approach of training/designing each component on its own
Types of Machine Learning

- Types of learning settings
  - Supervised learning
  - Unsupervised learning
- Types of learning algorithms
  - Batch learning
  - Online learning
Example: Digit Classification

http://ufldl.stanford.edu/housenumbers/
Example: Airline Price Prediction
Example: Airline Price Prediction

Advice: BUY  Confidence: 80%
Prices may rise within 7 days

Create a price alert

Sort by: price (low to high)

$367 Honolulu Round Trip
cheapoair.com/Honolulu-Cheap-Flight
Batch Supervised Learning

• Draw data set \( D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \) from distribution \( \mathcal{D} \)

• Algorithm \( A \) learns hypothesis \( h \in H \) from set \( H \) of possible hypotheses \( A(D) = h \)

• We measure the quality of \( h \) as the expected loss:

\[
E_{(x,y) \in \mathcal{D}} [\ell(y, h(x))]
\]

• This quantity is known as the risk

• E.g., loss could be the Hamming loss

\[
\ell_{\text{Hamming}}(a, b) = \begin{cases} 
0 & \text{if } a = b \\
1 & \text{otherwise}
\end{cases}
\]
Online Supervised Learning

- In step $t$, draw data point $x$ from distribution $\mathcal{D}$
- Current hypothesis $h$ guesses the label of $x$
- Get true label from oracle $O$
- Pay penalty if $h(x)$ is wrong (or earn reward if correct)
- Learning algorithm updates to new hypothesis based on this experience
  - Does not store history
Learning Settings

• Supervised or unsupervised (or semi-supervised, weakly supervised, transductive…)

• Online or batch (or reinforcement…)

• Classification, regression

  • (or structured output, clustering, dimensionality reduction…)
Best Practices

- Try range of models with different capacity
- Split data into training, validation, and testing sets
- Measure performance on evaluation set to tune parameters
- Measure performance on testing set as final check
Held-out Validation
# Held-out Validation

<table>
<thead>
<tr>
<th>Type</th>
<th>Accuracy on training data</th>
<th>Accuracy on validation data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>0.91</td>
<td>0.83</td>
</tr>
<tr>
<td>Medium</td>
<td>0.95</td>
<td>0.88</td>
</tr>
<tr>
<td>Complex</td>
<td>0.99</td>
<td>0.79</td>
</tr>
<tr>
<td>Super Complex</td>
<td>1.0</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Summary

• Three machine learning stories
  • One cautionary tale
• Deep learning definition
• Types of machine learning
• Best practices