Overview

• Machine Learning in Practise
  • Probabilities
  • Finite Resource

• Machine Learning @ Amazon
  • Forecasting
  • Machine Translation
  • Visual Systems

• Conclusions and Challenges
Overview

• Machine Learning in Practise
  • Probabilities
  • Finite Resource
• Machine Learning @ Amazon
  • Forecasting
  • Machine Translation
  • Visual Systems
• Conclusions and Challenges
Machine Learning: Formal Definition

- Labelled Data

\[ \{P(y|x, w)\}_{w \in \mathcal{W}} + \{(x_i, y_i)\}_{i=1}^n \mapsto (x \mapsto P(y|x)) \]

- Unlabelled Data

\[ \{P(x|z, w)\}_{w \in \mathcal{W}} + \{x_i\}_{i=1}^n \mapsto (x \mapsto P(z|x)) \]

- Probability is a central concept in Machine Learning!
Why Probability?

1. Mathematics of Uncertainty (Cox’ axioms)
Cox Axioms: Probabilities and Beliefs

• Design: System must assign degree of plausability $p(A)$ to each logical statement $A$.

• Axiom:
  • $p(A)$ is a real number
  • $p(A)$ is independent of Boolean rewrite
  • $p(A|C') > p(A|C)$ \(\land\) \(\Rightarrow\) \(p(B|AC') = p(B|AC')\)
  \(\Rightarrow\) \(p(AB|C') \geq P(AB|C')\)

\[\textbf{P must be a probability measure!}\]
Why Probability?

1. Mathematics of Uncertainty (Cox’ axioms)

2. Variables and Factors map to Memory & CPU
**Factor Graphs**

- **Definition:** Graphical representation of product structure of a function (Wiberg, 1996)
  - Nodes: □ = Factors  ○ = Variables
  - Edges: Dependencies of factors on variables.

- **Semantic:**

  \[ p(x) = \prod_{f} f(x_{V(f)}) \]

  - Local variable dependency of factors

\[ p(a, b, c) = f_1(a) \cdot f_2(b) \cdot f_3(a, b, c) \]
Factor Graphs and Cloud Computing

\[ p(\theta|X, Y) \propto \prod_i p(y_i|\theta, x_i) \cdot \prod_j p(\theta_j) \]

Belief Store
(“Memory”)

Message Passing
(“Communicate”)

Data Messages
(“Compute”)

6/29/17
Factor Graphs and MXNet

In [118]:
   import mxnet as mx

   # A simple network
   x = mx.sym.Variable('x')
   y = mx.sym.Variable('y')
   z = x + y
   net = mx.sym.LinearRegressionOutput(z, name='squaredloss')
   mx.viz.plot_network(net)

Out[118]:

![Diagram of a factor graph showing a simple network with variables x and y, and a linear regression output]

![Graph showing optimization path computed by gradient descent]
Overview

- Machine Learning in Practise
  - Probabilities
  - Finite Resource
- Machine Learning @ Amazon
  - Forecasting
  - Machine Translation
  - Visual Systems
- Conclusions and Challenges
**Finite Resource: Cost**

**Economics 101**

- Profit = Revenue – Cost
- In the long run, a business that generates negative profits is not viable!

<table>
<thead>
<tr>
<th>Facebook</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Revenue</td>
<td>$17,928,000,000.00*</td>
</tr>
<tr>
<td>Daily Revenue</td>
<td>$49,117,808.22</td>
</tr>
<tr>
<td>Number of DAU</td>
<td>1,038,000,000**</td>
</tr>
<tr>
<td>Number of Story Candidates</td>
<td>1,500***</td>
</tr>
<tr>
<td>Number of Daily Stories</td>
<td>1.557E+12</td>
</tr>
<tr>
<td>Maximum Cost per Story Candidate</td>
<td>$0.0000315</td>
</tr>
</tbody>
</table>


---

*It’s power, stupid!*

Some constraints might not be obvious: building new datacenters and **powering** them is non-trivial.

Example: 1 GPU box = 20 CPU boxes
(in terms of power consumption)
Overview

• Machine Learning in Practise
  • Probabilities
  • Finite Resource

• Machine Learning @ Amazon
  • Forecasting
  • Machine Translation
  • Visual Systems

• Conclusions and Challenges
Locations

- ML Seattle
- A9
- ML Los Angeles
- A2Z
- ML Cambridge
- S9
- ML Berlin
- Evi
- Ivona
- ML Bangalore

6/29/17
Machine Learning Opportunities @ Amazon

**Retail**
- Demand Forecasting
- Vendor Lead Time Prediction
- Pricing
- Packaging
- Substitute Prediction

**Customers**
- Product Recommendation
- Product Search
- Visual Search
- Product Ads
- Shopping Advice
- Customer Problem Detection

**Seller**
- Fraud Detection
- Predictive Help
- Seller Search & Crawling

**Catalog**
- Browse-Node Classification
- Meta-data validation
- Review Analysis
- Hazmat Prediction

**Digital**
- Named-Entity Extraction
- XRay
- Plagiarism Detection
- Echo Speech Recognition
- Knowledge Acquisition
Overview

• Machine Learning in Practise
  • Probabilities
  • Finite Resource

• Machine Learning @ Amazon
  • Forecasting
  • Machine Translation
  • Visual Systems

• Conclusions and Challenges
Training Range: Non-fashion items have longer training ranges that we can leverage. Need to information share across new and old products.

Seasonality: This item has Christmas seasonality with higher growth over time. This is where we need growth features in addition to date features.

Missing Features or Input: Unexplained spikes in demand are likely caused by missing features or incomplete input data.
Learning and Prediction

$P(z_{it} | \theta) \sim$
Typical midsize dataset:
• About 5M items
• About 4.5B item-days
• About 98% zero demand
Sampling Predictions

\[ P(z_{it} \mid \theta) \sim \]

- 0 or ≥1?
  Binary classification #1
- 1 or ≥2?
  Binary classification #2
- If ≥2:
  Count regression z-2
Latent State Multistage Likelihood
Modelling Out of Stock

GLM

Bridge
Overview

• Machine Learning in Practise
  • Probabilities
  • Finite Resource
• Machine Learning @ Amazon
  • Forecasting
  • Machine Translation
  • Visual Systems
• Conclusions and Challenges

1432 Girali (Grasping toy) - Selecta Wooden Toys/Selecta Spielzeug
by Selecta Spielzeug

Price: £12.03 & FREE Delivery in the UK. Details

Only 7 left in stock.
Sold by Alle-Spielwaren and Fulfilled by Amazon. Gift-wrap available.

Want it delivered to Germany by tomorrow, 18 March? Order within 5 hrs 41 mins and choose One-Day Delivery to Germany at checkout. Details

18 new from £7.11

- 10 cm / 4 in.
- This classic series of grasping toys has been perfected by Selecta for over 30 years.
- See more product details
Sockeye

• Sequence-to-sequence Neural Machine Translation package build on MXNet: https://github.com/awslabs/sockeye

• Support both CPU and GPU encoding and decoding

• Training

```
> python -m sockeye.train --source sentences.de \\
   --target sentences.en \\
   --validation-source sentences.dev.de \\
   --validation-target sentences.dev.en \\
   --use-cpu \\
   --output <model_dir>
```

• Translating

```
> python -m sockeye.translate --models <model_dir> --use-cpu
```
Overview

• Machine Learning in Practise
  • Probabilities
  • Finite Resource

• Machine Learning @ Amazon
  • Forecasting
  • Machine Translation
  • Visual Systems

• Conclusions and Challenges
Automated Produce Inspection: The Goal

Current Inspection

New Automated Inspection

Computer Vision

<table>
<thead>
<tr>
<th>Defect</th>
<th>Decay</th>
<th>Bruising</th>
<th>Bruising</th>
<th>Overripe/Soft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample #</td>
<td>Defect Cat.</td>
<td>Serious</td>
<td>Damage</td>
<td>Serious</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>3</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>0</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>1</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>240</td>
<td>9</td>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td>% of Total</td>
<td>100%</td>
<td>4%</td>
<td>17%</td>
<td>7%</td>
</tr>
</tbody>
</table>
Challenges

• Illumination
• Clutter/Occlusions
• Viewpoint
• Scale
• Intra-class variability
Age Aligned Strawberries (Test Set)
Overview

• Machine Learning in Practise
  • Probabilities
  • Finite Resource

• Machine Learning @ Amazon
  • Forecasting
  • Machine Translation
  • Visual Systems

• Conclusions and Challenges
Conclusions

• Machine Learning “translates” data from the past into accurate predictions about the future!

• In practice, probabilistic models and finite resources matter.

• Machine Learning helps to improve customer experience at Amazon!
Thanks!