# Machine Learning @ Amazon

Ralf Herbrich
Amazon

# 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
  - $p(A|C') > p(A|C) \quad \wedge \quad p(B|AC') = p(B|AC)$
    $\Rightarrow \quad p(AB|C') \geq P(AB|C)$

**P must be a probability measure!**

# Why Probability?

1. **Mathematics of Uncertainty (Cox' axioms)**

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

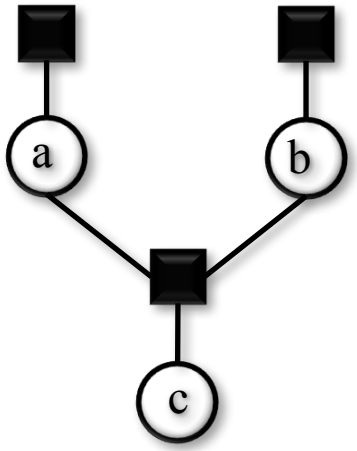
# Factor Graphs

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

- **Semantic:**

$$p(\mathbf{x}) = \prod_f f\left(\mathbf{x}_{V(f)}\right)$$

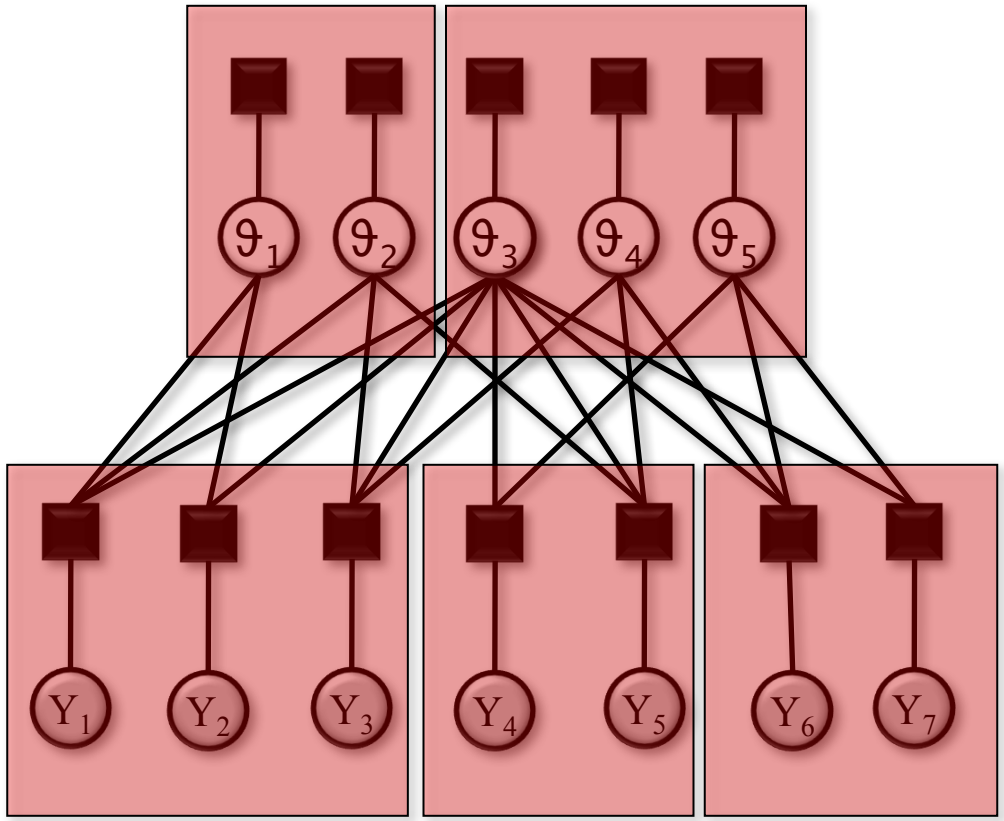
  - Local variable dependency of factors




$$p(a,b,c) = f_1(a) \cdot f_2(b) \cdot f_3(a,b,c)$$

# Factor Graphs and Cloud Computing

$$p(\boldsymbol{\theta}|\mathbf{X},\mathbf{Y}) \propto \prod_i p(y_i|\boldsymbol{\theta},\mathbf{x}_i) \cdot \prod_j p(\theta_j)$$




**Belief Store ("Memory")**

**Message Passing ("Communicate")**

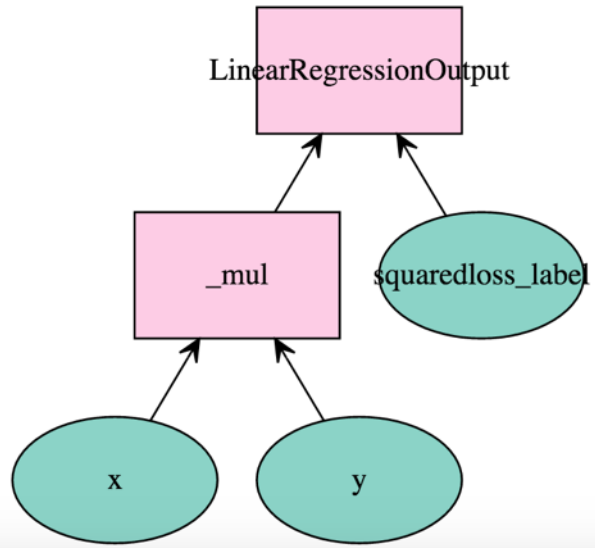
**Data Messages ("Compute")**

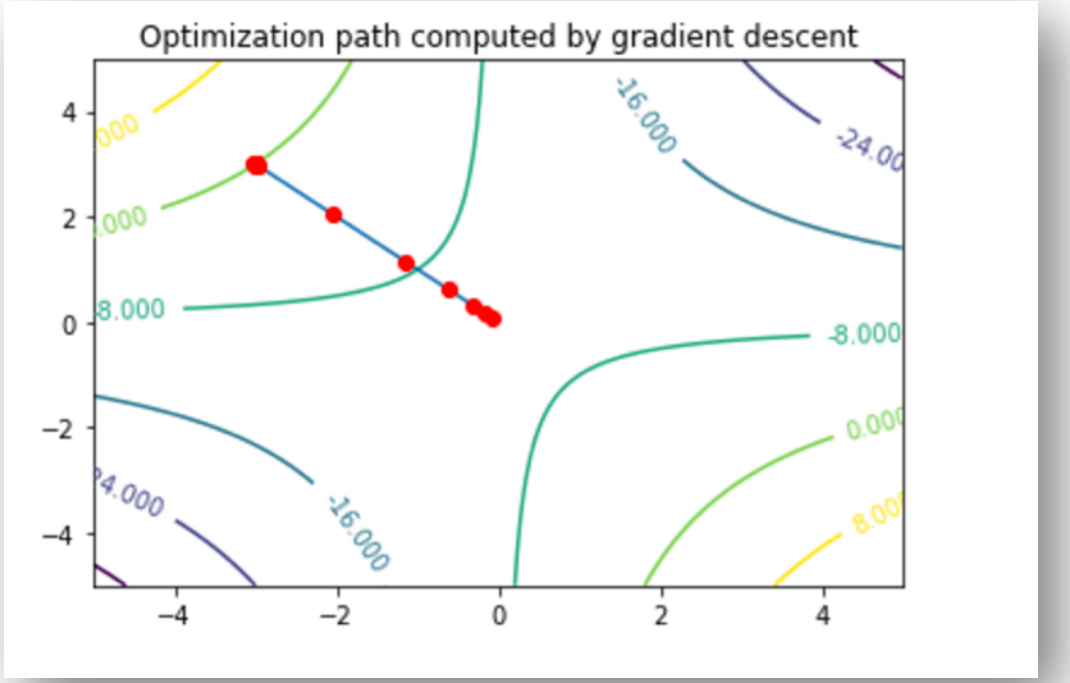
# Factor Graphs and MXNet

```
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]:



# 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!

| Facebook | 2015 |
|---|---|
| Annual Revenue | $17,928,000,000.00* |
| Daily Revenue | $49,117,808.22 |
| Number of DAU | 1,038,000,000** |
| Number of Story Candidates | 1,500*** |
| Number of Daily Stories | 1.557E+12 |
| Maximum Cost per Story Candidate | $0.0000315 |

*http://www.statista.com/statistics/277229/facebooks-annual-revenue-and-net-income/
**http://www.statista.com/statistics/346167/facebook-global-dau/
***https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed

**It's power, stupid!**

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

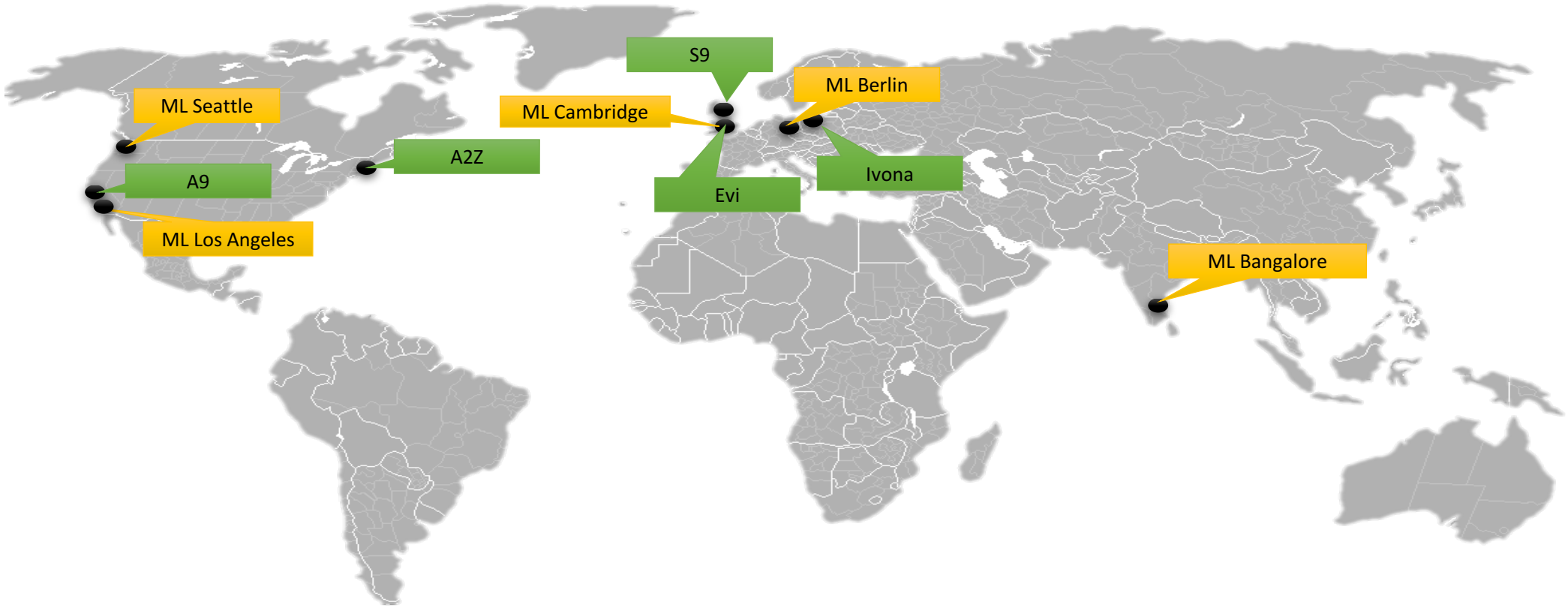
Example: 1 GPU box = 20 CPU boxes
(in terms of power consumption)

amazon

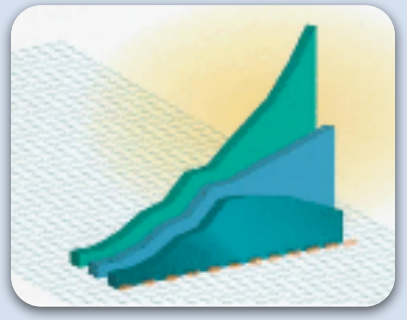
# Overview

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

# Locations


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

# 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

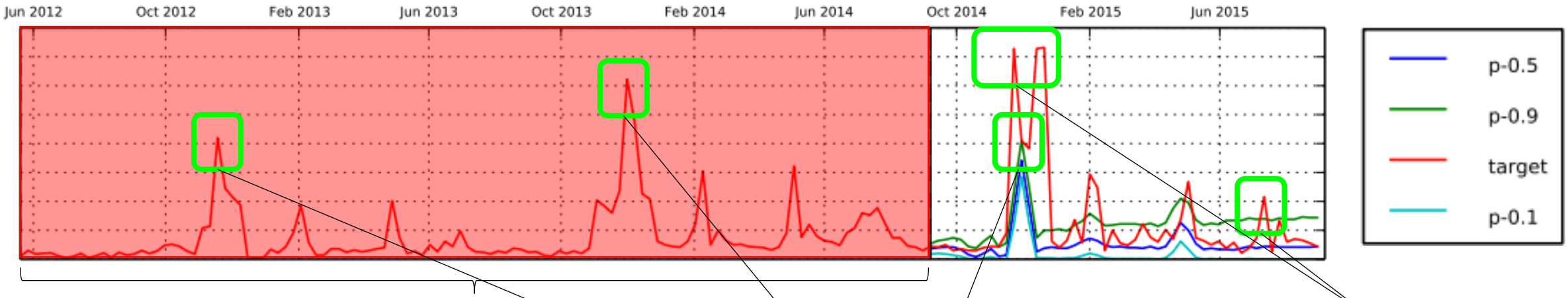
- Named-Entity Extraction
- XRay
- Plagiarism Detection
- Echo Speech Recognition
- Knowledge Acquisiion

# Overview

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

# Demand Forecasting

*Example fashion product to illustrate the challenges of forecasting.*

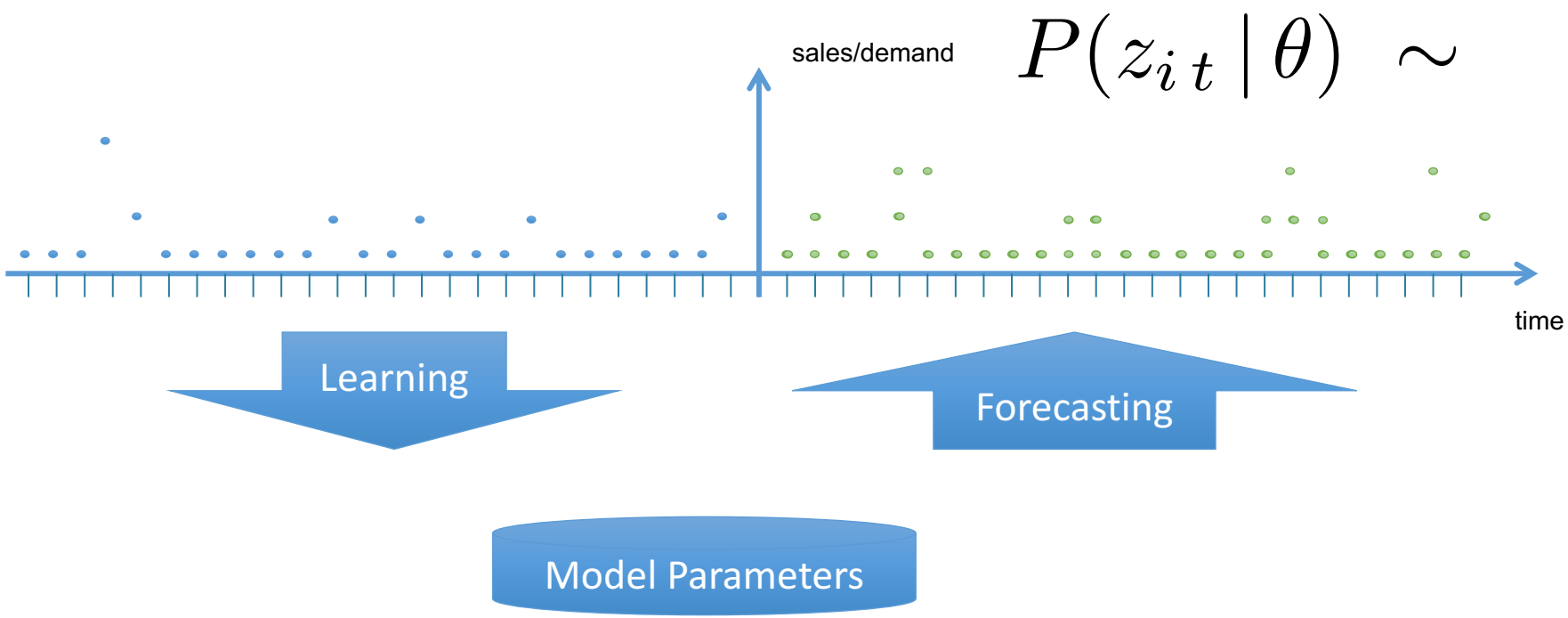



**Training Range:** Non-fashion items have longer training ranges that we can leverage. Need to information share across new and old products.

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

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

# Learning and Prediction

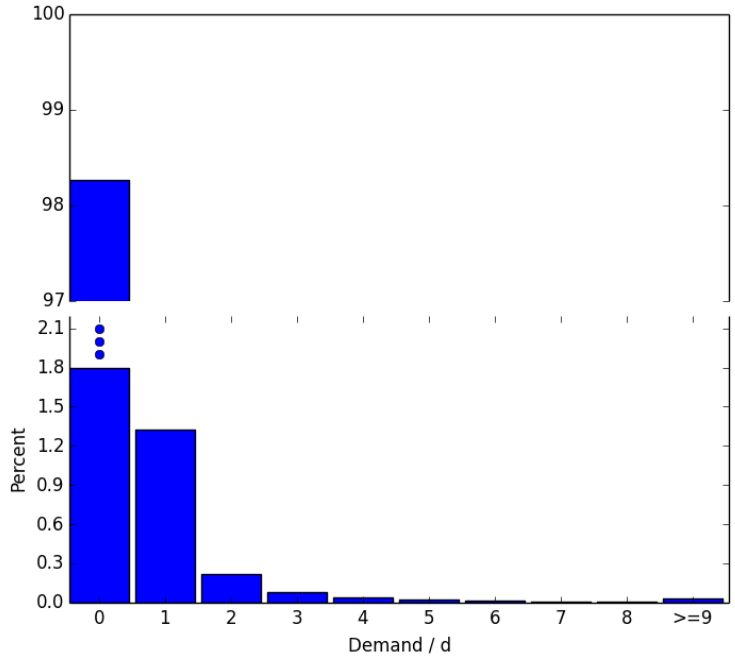



amazon

# Slow Moving Inventory

Typical midsize dataset:

- About 5M items
- About 4.5B item-days
- About 98% zero demand

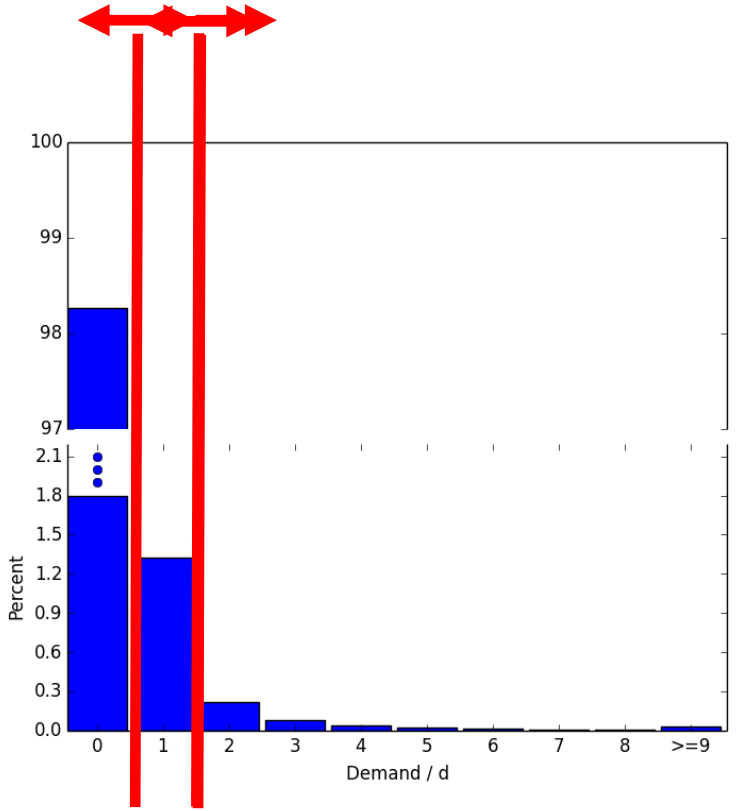



amazon

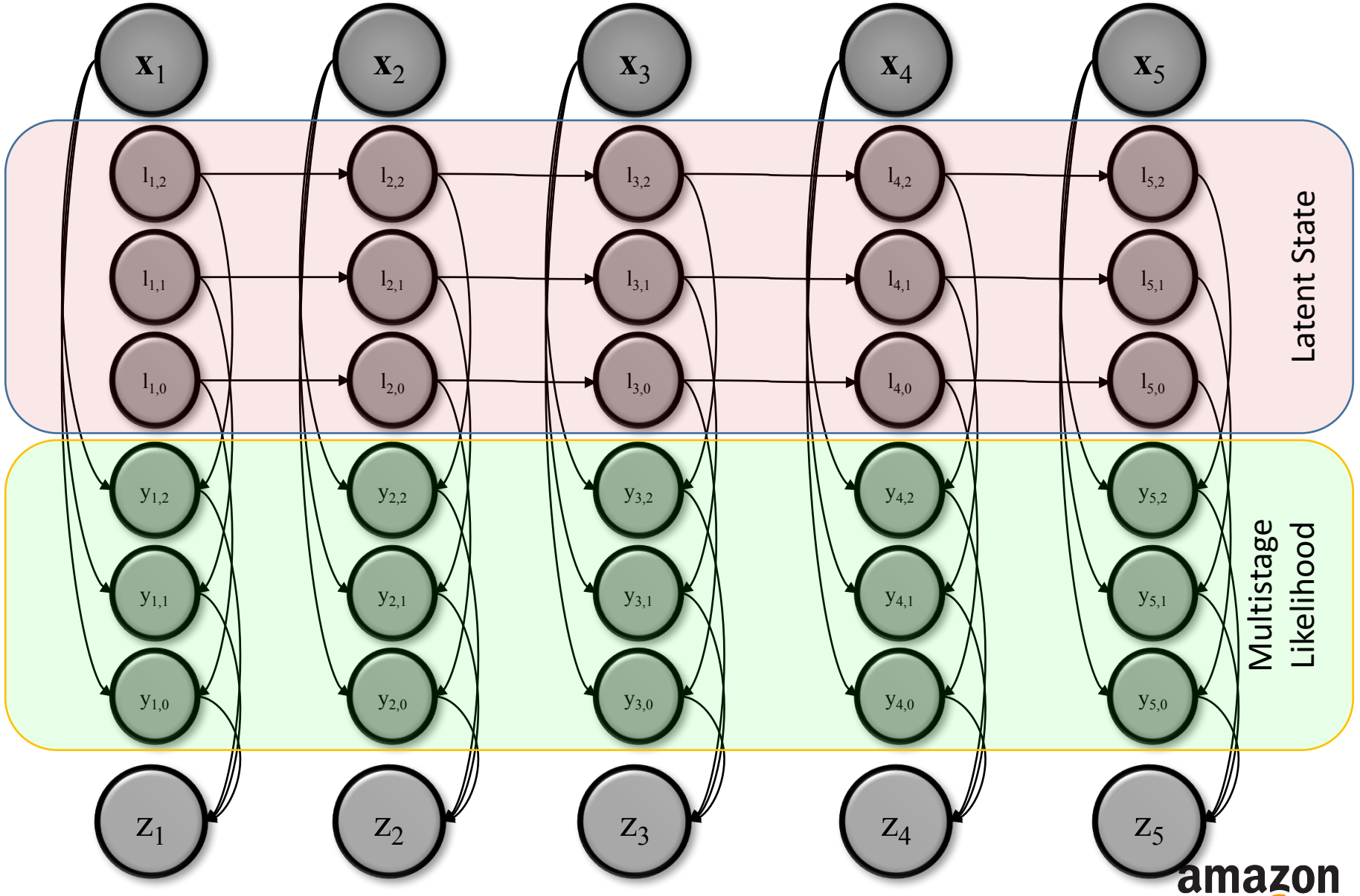
# Sampling Predictions

$$P(z_{i\,t} \mid \theta) \sim$$

- 0 or ≥1 ?
  Binary classification #1
- 1 or ≥2 ?
  Binary classification #2
- If ≥2:
  Count regression z-2

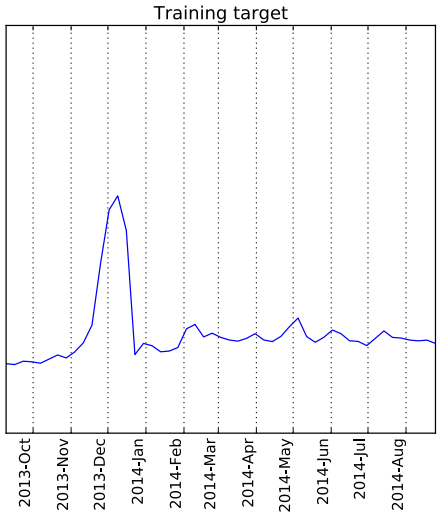


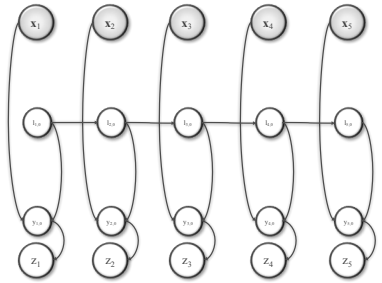

amazon

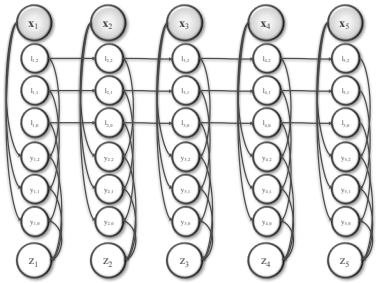

x1
x2
x3
x4
x5
l1,2
l1,1
l1,0
l2,2
l2,1
l2,0
l3,2
l3,1
l3,0
l4,2
l4,1
l4,0
l5,2
l5,1
l5,0
Latent State
y1,2
y1,1
y1,0
y2,2
y2,1
y2,0
y3,2
y3,1
y3,0
y4,2
y4,1
y4,0
y5,2
y5,1
y5,0
Multistage
Likelihood
z1
z2
z3
z4
z5

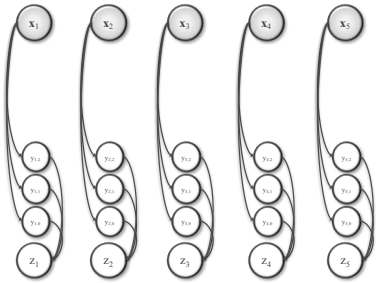
amazon

# In Practice



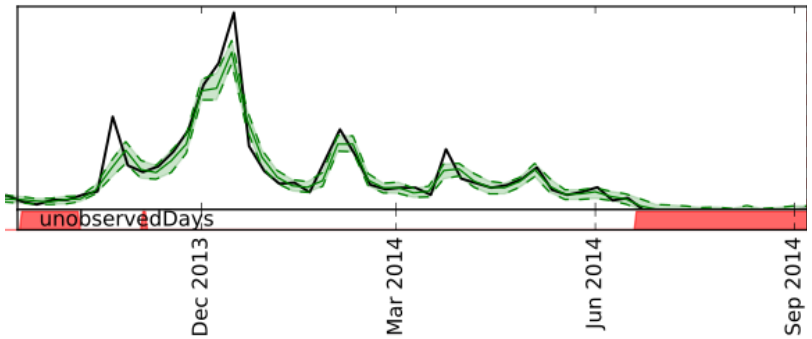




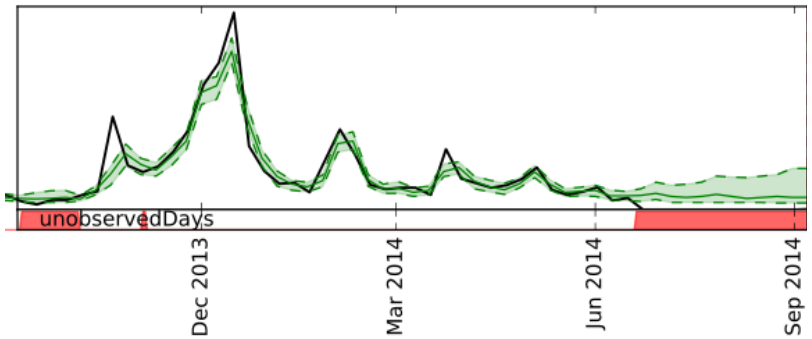
amazon

# Modelling Out of Stock

GLM




Bridge

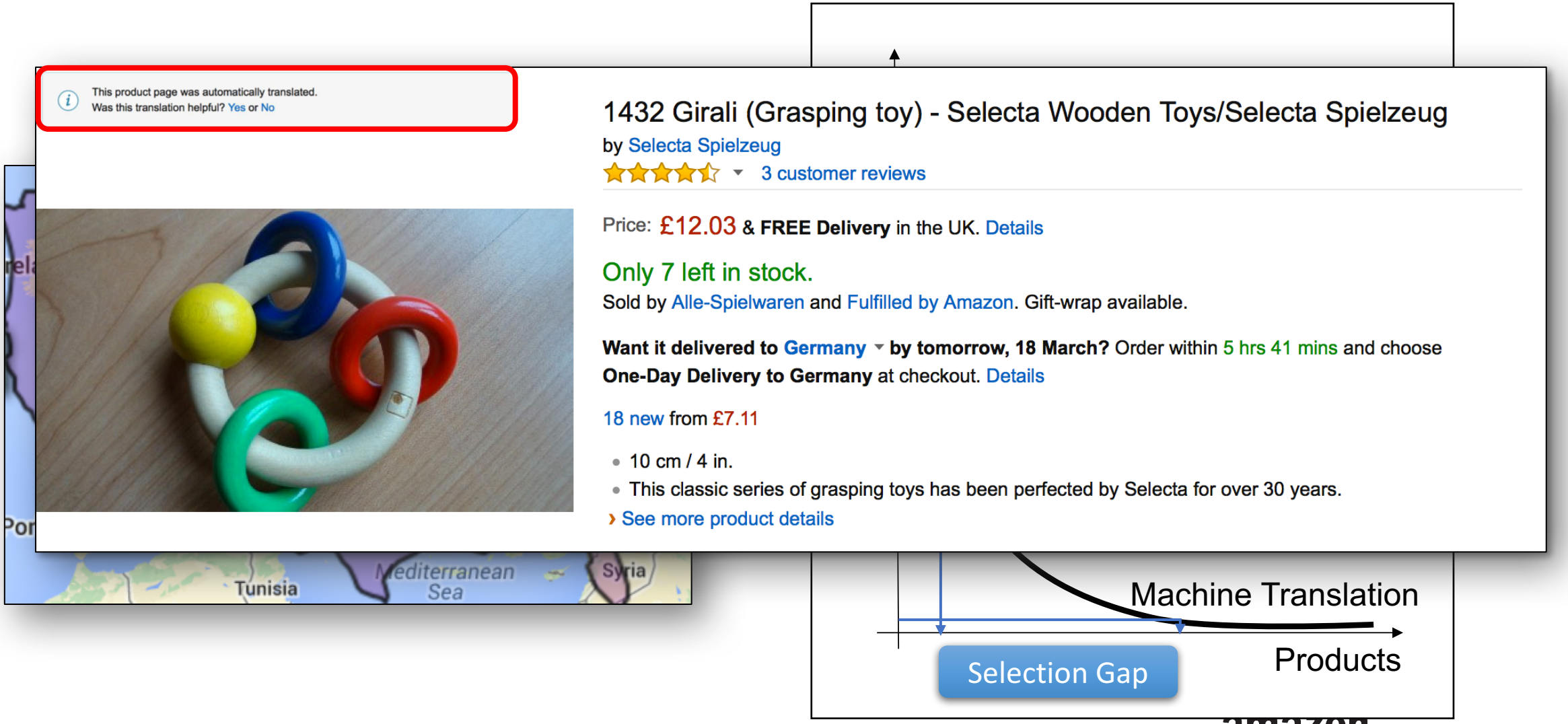



amazon

# Overview

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

# Product Machine Translation (2013 – 2015)



# 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

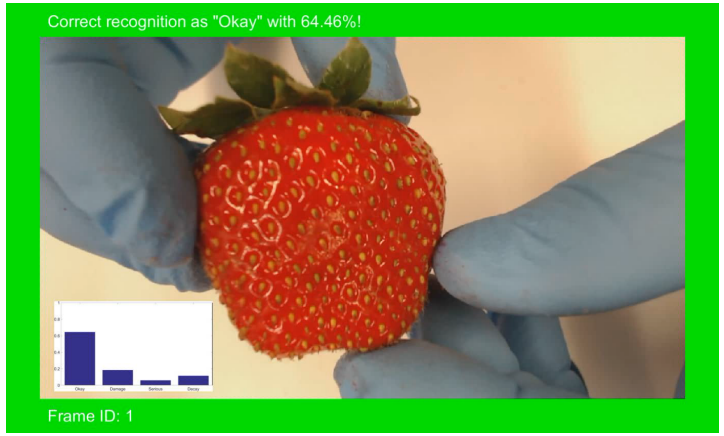
# Automated Produce Inspection: The Goal

## Current Inspection

| | Defect | Decay | Bruising | Bruising | Overripe/Soft |
|---|---|---|---|---|---|
| Sample # | Defect Cat. | Serious | Damage | Serious | Serious |
| 1 | 60 | 3 | 10 | 7 | 1 |
| 2 | 60 | 5 | 5 | 1 | 3 |
| 3 | 60 | 0 | 10 | 7 | 0 |
| 4 | 60 | 1 | 15 | 1 | 0 |
| Total | 240 | 9 | 40 | 16 | 4 |
| % of Total | 100% | 4% | 17% | 7% | 2% |

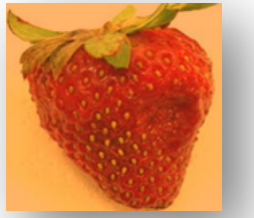
## New Automated Inspection

## Computer Vision



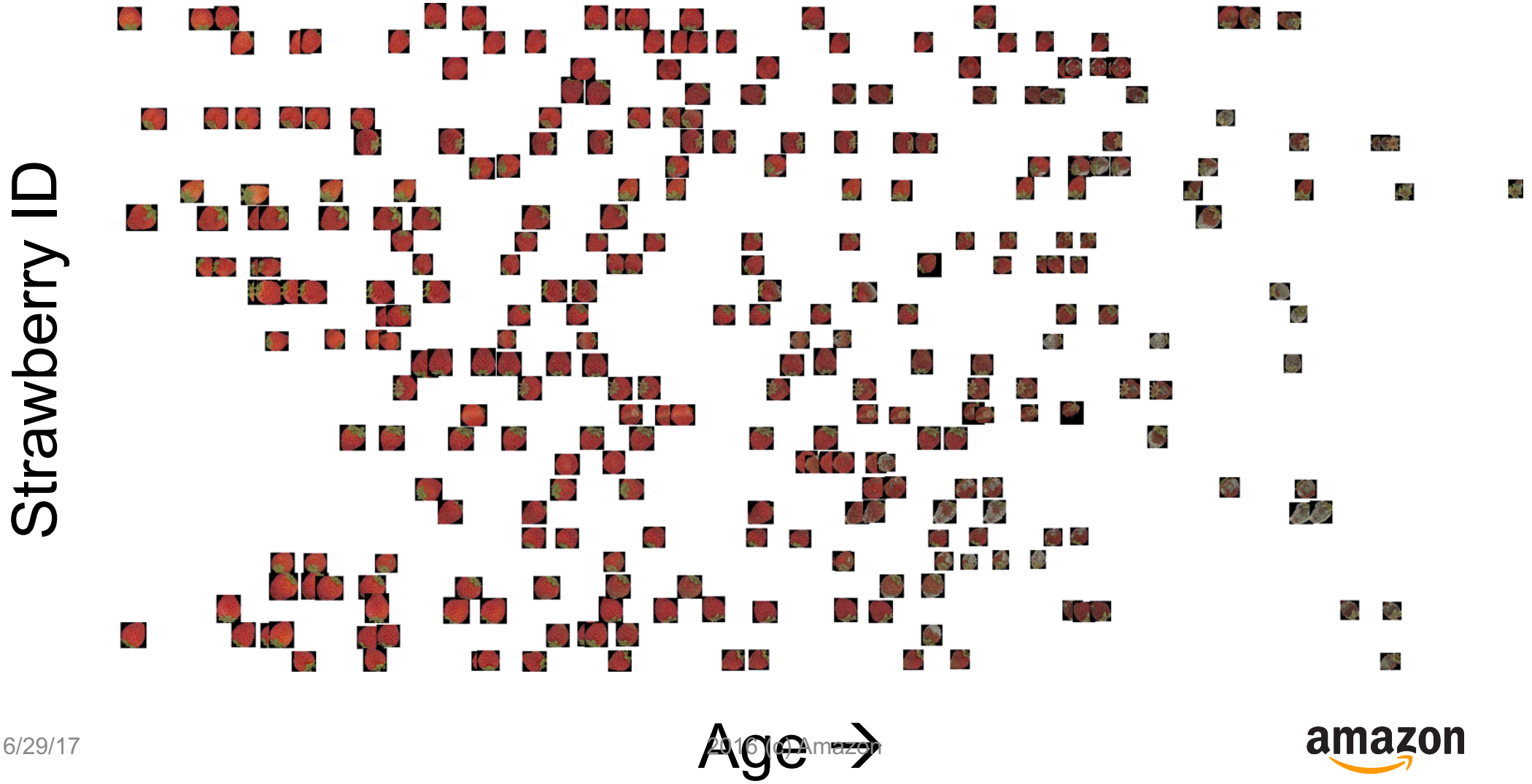
# Challenges

- Illumination
- Clutter/Occlusions
- Viewpoint
- Scale
- Intra-class variability



amazon

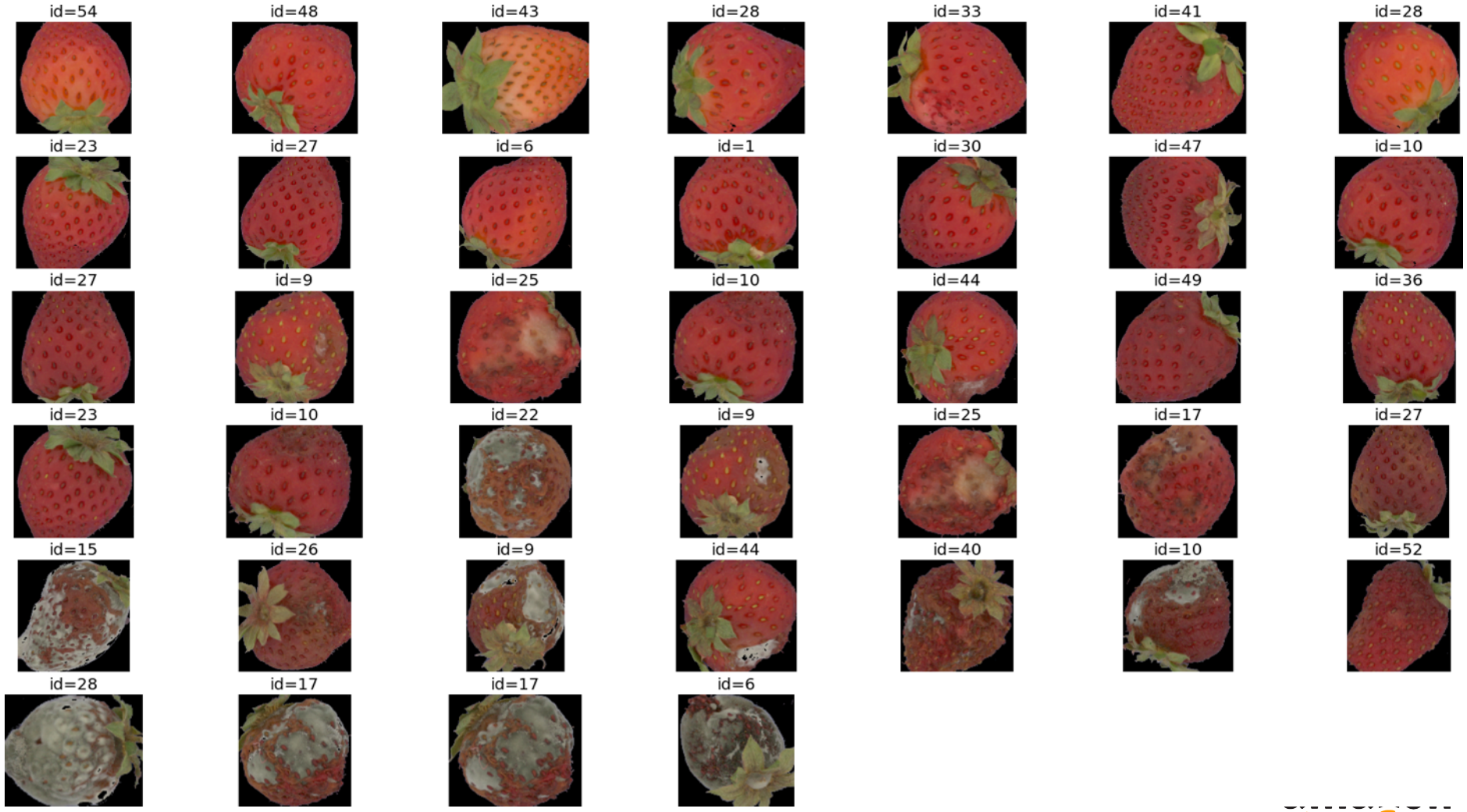
# Predicting Longevity






amazon

# 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!