SPARKBEYOND

PRIOR ON MODEL SPACE

What makes a model simple?

MEIR MAOR
Chief Architect @ SparkBeyond

Outline

Why are simple models desirable

Traditional approaches for simplicity

Alternative approaches

Why Simple models?

PAC model

- No Free Lunch
- Better generalization

In Reality

- Transfer learning and non stationary distributions
- Robust against correlated samples, etc.

Leslie Valiant, 1984

Why Simple Models? Cont.

Understandable

Trustworthy

Explainable (also for Regulatory reasons)

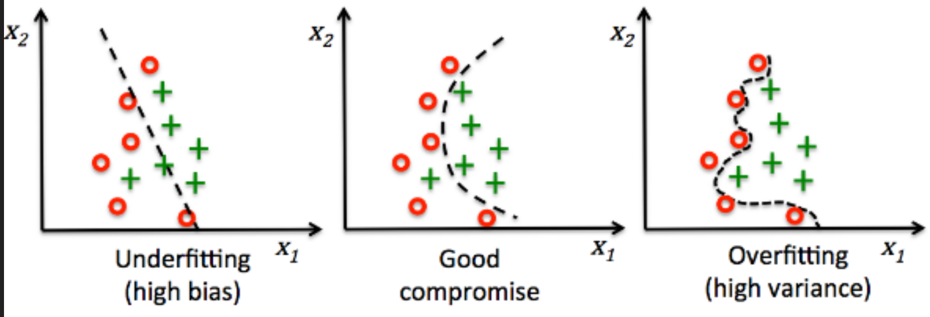
Understandable models are ultimately more accurate

Traditional complexity control

Bias / Variance tradeoff → We must limit our search space

Shrink the hypothesis space:

- limit boosting iterations
  - tree size
  - min sample in leaf
  - number of hidden nodes
  - impose sparsity constraint
  - ...



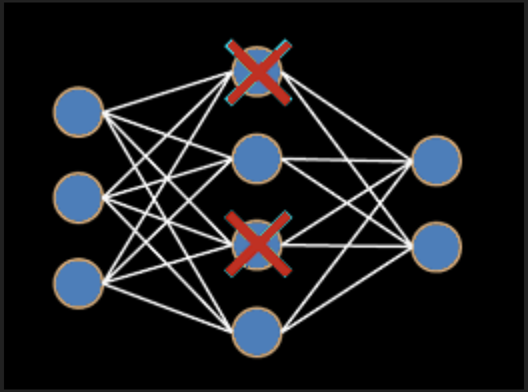
Traditional complexity control cont.

Penalize “less favourable” models:

Lasso / ridge regularization

Bagging / Boot strap sampling

Drop out



Which is more likely?

Coefficients from two feed forward ReLu NN
single-output single hidden layer

NETWORK A	NETWORK B
1.787	-0.590
1.771	0.000
1.735	0.000
1.937	0.806
1.445	0.000
1.601	0.000
1.773	0.092
2.017	0.000
1.418	0.000
1.888	0.000
1.399	0.000
1.733	0.000
2.031	0.000
1.723	-0.932
1.801	0.525
1.598	0.000
1.666	0.097
1.484	0.000

Which feature is a more likely?

Both have $\mathbb{R}^2 = 0.1$

Math.ulp(x) - The positive distance between this floating-point value and the double value next larger in magnitude

vs.

Math.log(x) - Natural logarithm

Which is a more likely feature? #2

The distance to the nearest railway station?

vs.

arctan(latitude * longitude)

Everyone is a domain expert!

We are experts in the world we live in.

Currently, humans have a much better prior than machines.

Many ideas repeat themselves across domains.

For example, you don't have to be a rocket scientist to be familiar with second derivatives.

Transfer Learning to the rescue

We can and must learn from previous problems

How can a child learn to identify a Ring Tailed Lemur from a single photo?

Becoming common

Pre-Trained Neural networks

Pre-Trained embeddings

A lot of work on Images and Text

Much less research on other data:

- TimeNet - RNN for embedding time series data
- Most real-life problems tend to have a more complicated shape

A different approach

Use already codified human knowledge

Explicitly look for patterns similar to things you have seen before

Extraordinary claims require extraordinary evidence

At SparkBeyond

Find the best hypotheses, using simple compositions of tried and true building blocks

The building block may require a lot of code to implement. Yet, will be useful across domains

Incorporate pre-trained embeddings

Use external knowledge

Prioritize simple hypotheses

Always meta-learn how to learn

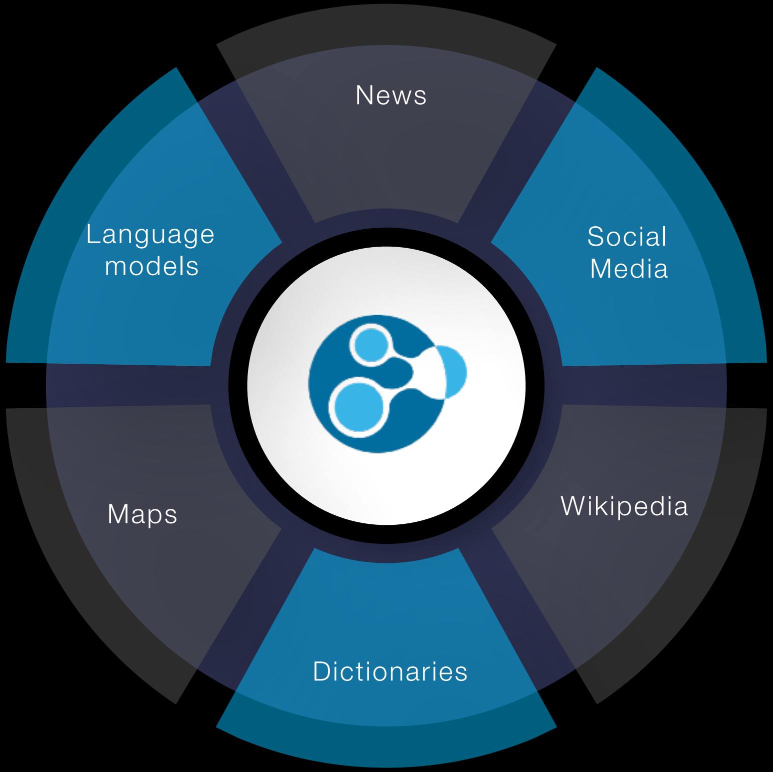
Shops near ***recreational parks*** *are more successful*

- *Domain expert can review such a finding*
  - *True phenomenon*
  - *Insightful and actionable*

Colorful gadgets tend to be cheaper

Becomes intuitive when you see concrete examples

EXTERNAL DATA



Simple = compressible = common

A simple model or feature is one which we can be expressed briefly.

MDL - minimum description length is optimal compression.

Better compression leads to better model performance.

But should we be using a vanilla Turing machine for MDL?

Benefits of a better prior

Learn with less data

More robust to change

More robust to data issues

Understandable and explainable

Actionability without a complete model

Open questions and challenges

What is simple?

What makes an insight insightful?

What makes a feature likely to generalize?

Efficient search over insightful hypothesis space

VISIT OUR BOOTH

SPARKBEYOND