Nuts and Bolts of Applying Deep Learning

Recap of Andrew Ng's talk at the Bay Area Deep Learning School

Bay Area Deep Learning School

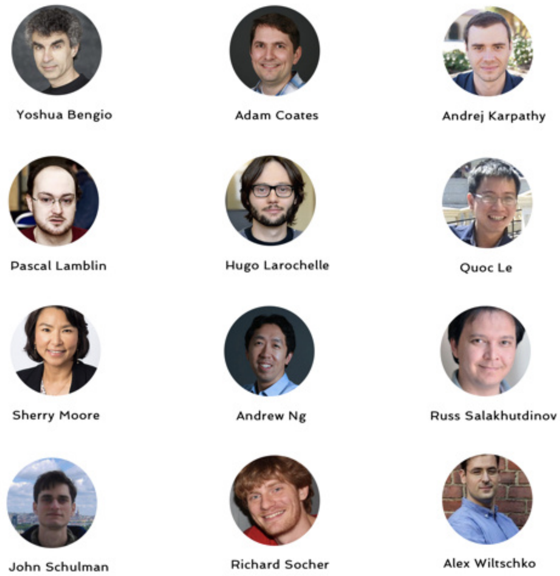
September 24-25, 2016
Stanford, CA

September 24, 2016

9:00 AM	Foundations of Deep Learning, Hugo Larochelle
10:00 AM	Break
10:15 AM	Deep Learning for Computer Vision, Andrej Karpathy
11:45 AM	Lunch
12:45 PM	Deep Learning for NLP, Richard Socher
2:15 PM	Break
2:45 PM	Tensorflow tutorial, Sherry Moore
3:45 PM	Break
4:00 PM	Foundations of Deep Unsupervised Learning, Ruslan Salakhutdinov
5:30 PM	Break
6:00 PM	$\boxed{\text{Nuts and bolts of applying deep learning, Andrew Ng}}$

September 25, 2016

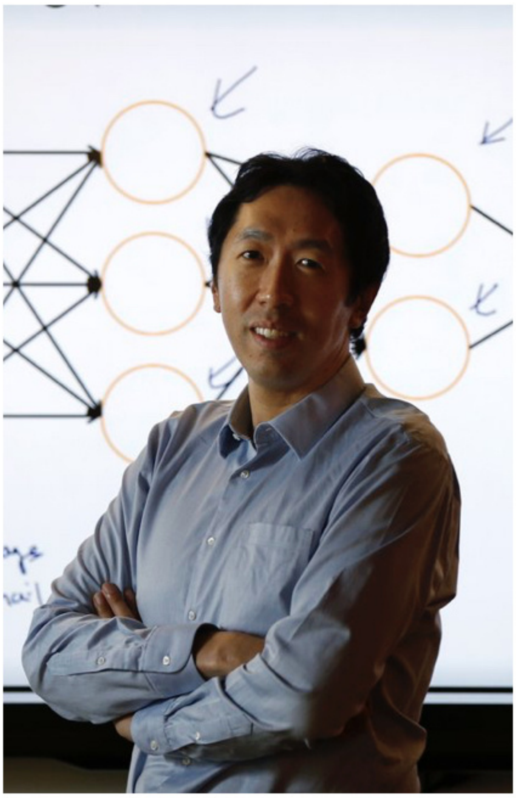
9:00 AM	Foundations of Deep Reinforcement Learning, John Schulman
10:30 AM	Break
10:45 AM	Theano tutorial, Pascal Lamblin
11:45 AM	Lunch
12:45 PM	Deep Learning for Speech, Adam Coates
2:15 PM	Break
2:45 PM	Torch tutorial, Alex Wiltschko
3:45 PM	Break
4:00 PM	Sequence to Sequence Learning for NLP and Speech, Quoc Le
5:30 PM	Break
6:00 PM	Foundations and Challenges of Deep Learning, Yoshua Bengio



ANDREW NG

Andrew Ng is Chief Scientist at Baidu; Co-Chairman and Co-Founder of Coursera; and an Adjunct Professor at Stanford University.

In 2011 he led the development of Stanford University's main MOOC (Massive Open Online Courses) platform and also taught an online Machine Learning class to over 100,000 students, leading to the founding of Coursera. Ng's goal is to give everyone in the world access to a great education, for free.



Content

Major Deep Learning Trends

End-to-End Deep Learning

Bias-Variance Tradeoff

Human-level Performance

Personal Advice

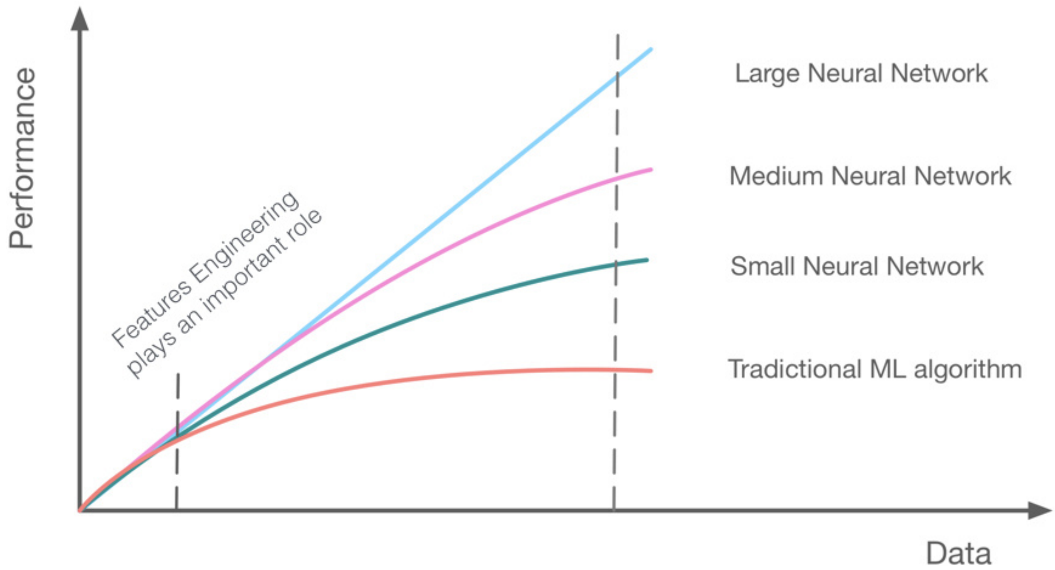
Major Deep Learning Trends

Why do DL algorithms work so well?

With the rise of the Internet, Mobile and IOT era, the amount of data accessible to us has greatly increased.

Neural network models, especially the larger ones have the capacity to absorb all this data.

The performance between difference algorithms



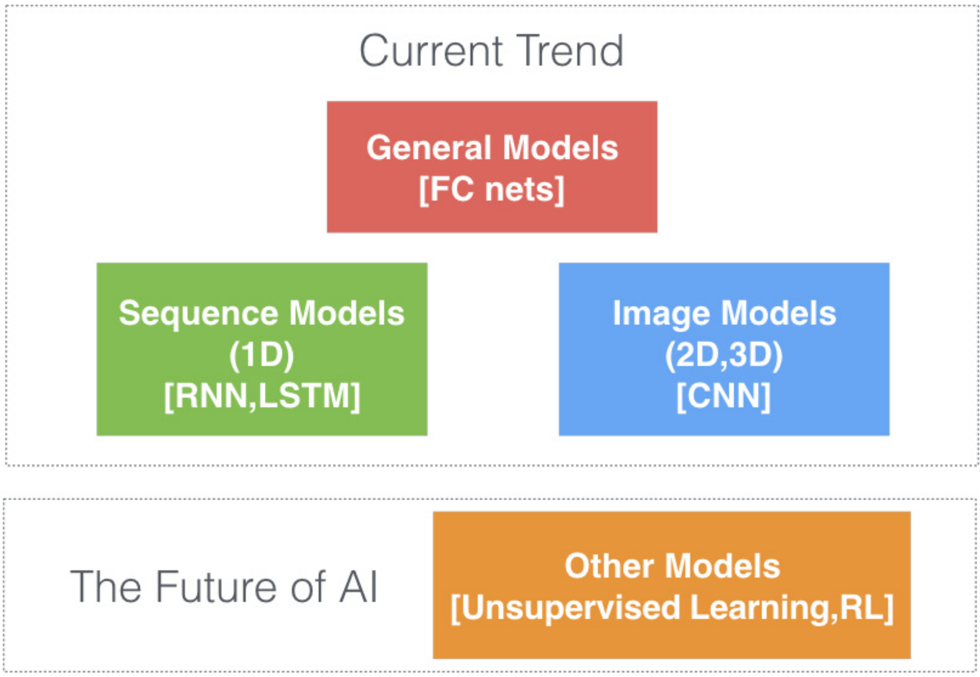
Hand engineering effectively gets replaced by end-to-end approaches

Bigger neural nets combined with a lot of data tend to outperform all other models.

Machine Learning and HPC team

Cooperation from both Machine Learning and HPC team is the key to boosting performance in AI companies.

Categorizing DL models



End-to-End Deep Learning

The rise of End-to-End DL

The fact that model's outputs are becoming more and more complicated

- Images
- Full captions with RNN's
- Audio like in DeepMind's WaveNet

What does End-to-End training mean?

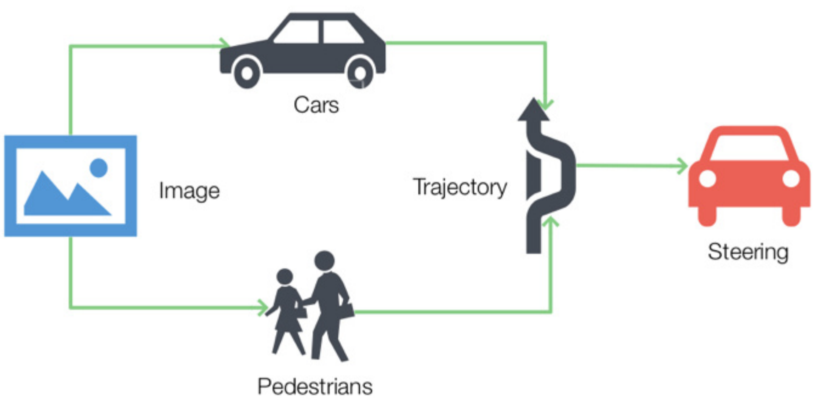
AI practitioners are shying away from intermediate representations

Going directly from one end (raw input) to the other end (output)

Example of End-to-End DL

- Movie Review → Sentiment
- Image → Caption
- Audio → Transcript
- Parameter → Image
- Image → steering

Image → Steering



End-to-End

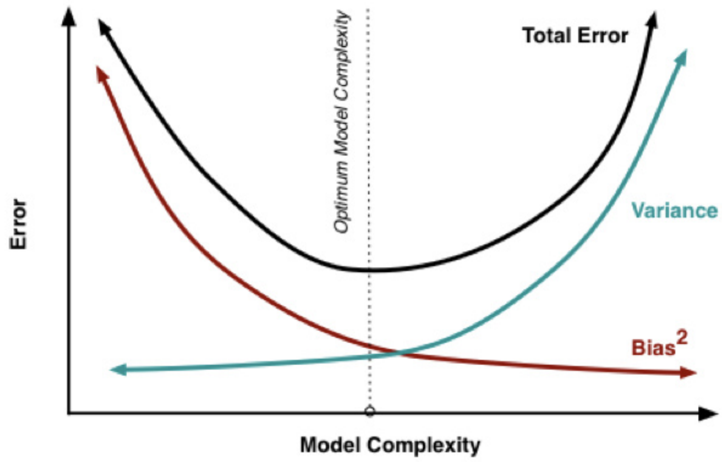
Disadvantages to this approach

End-to-End approaches are data hungry

In practice, not all applications have the luxury of large labelled datasets

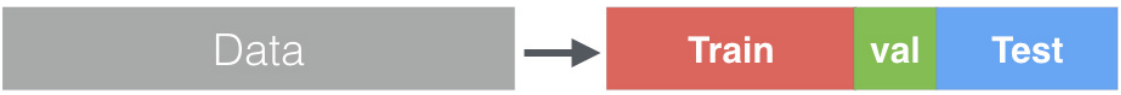
Bias-Variance Tradeoff

Basic Concept



Split Data

Traditional Way

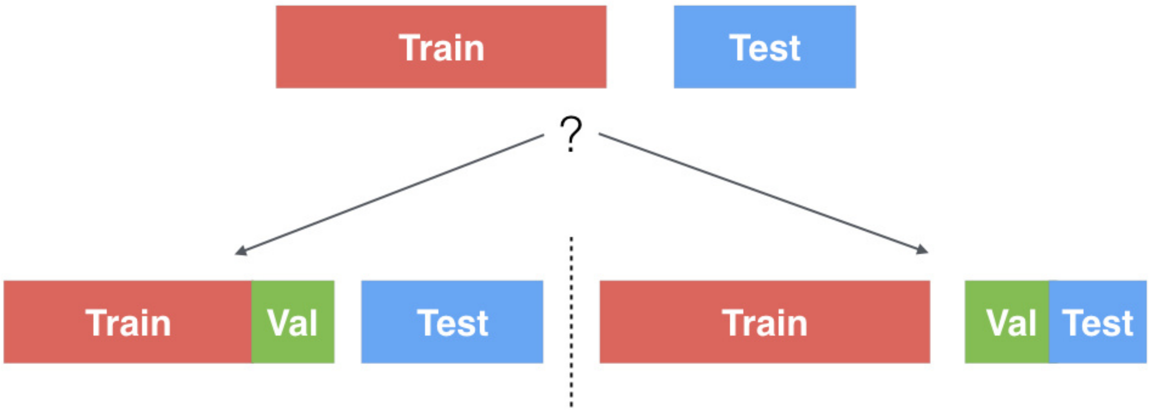


Note: Validation abbr. as Val

- Training set is used to build up our prediction algorithm
- Validation set is used to compare the performances of the prediction algorithms that were created based on the training set
- Test set is used to see how our final model is going to deal in the wild

However

In most deep learning problems, train and test come from different distributions.

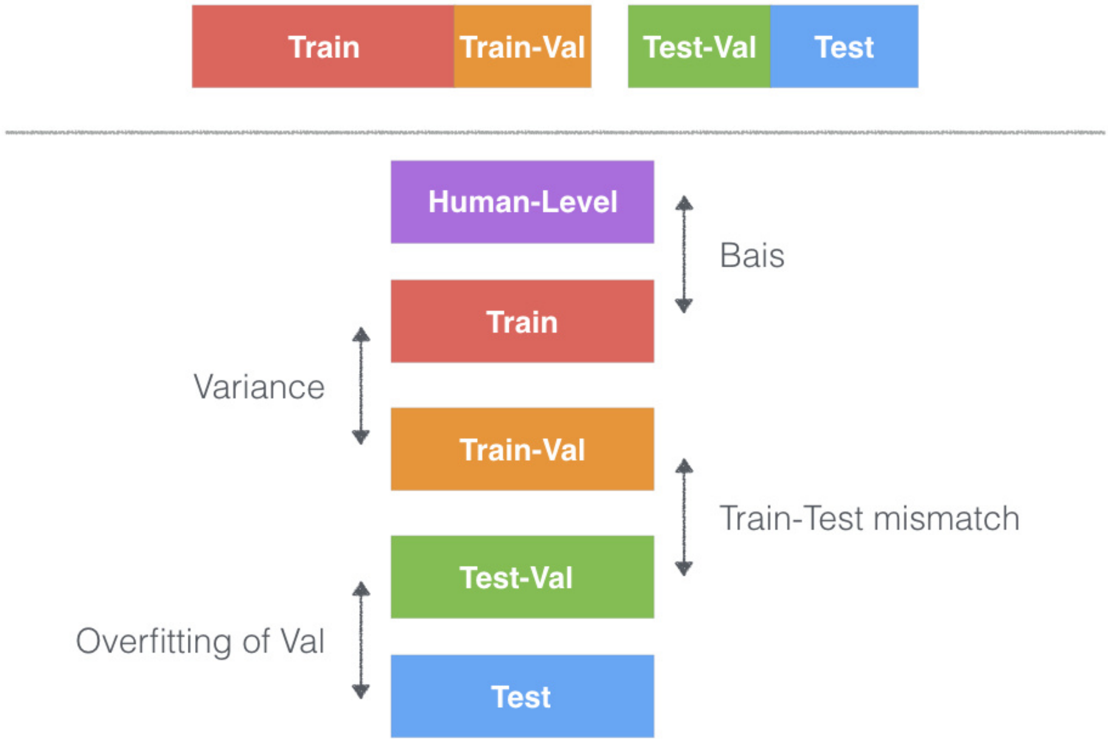


Why

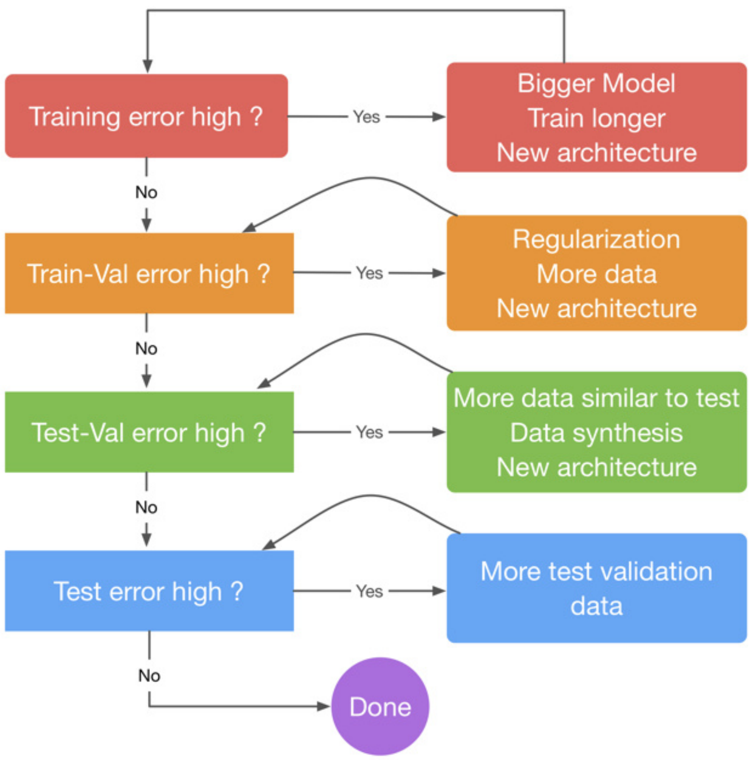
we usually want our validation set and test set to come from the same distribution.

The reason for this is that because a part of the team will be spending a lot of time tuning the model to work well on the validation set, if the test set were to turn out very different from the validation set, then pretty much all the work would have been wasted effort.

Recommendation

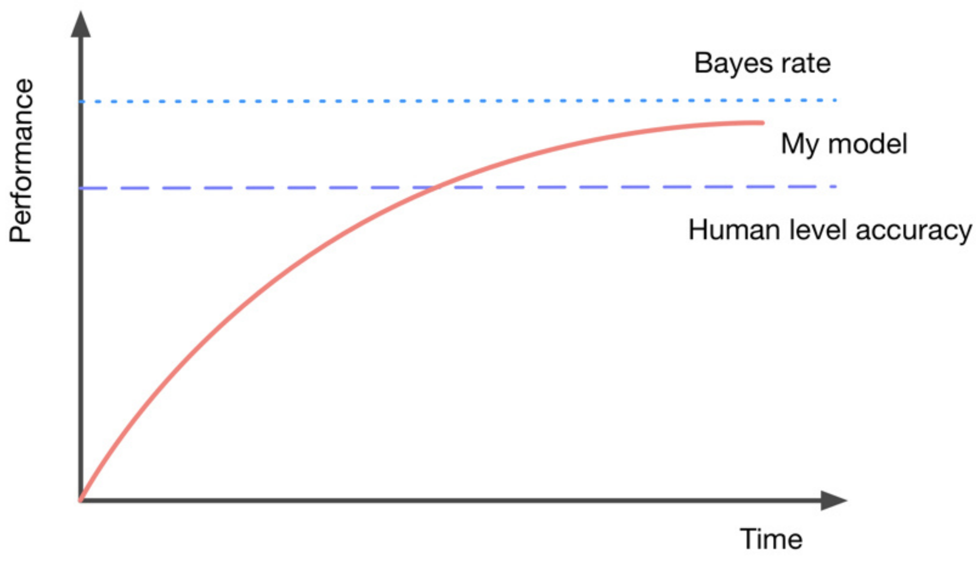


Flowchart for working with a model



Human-level Performance

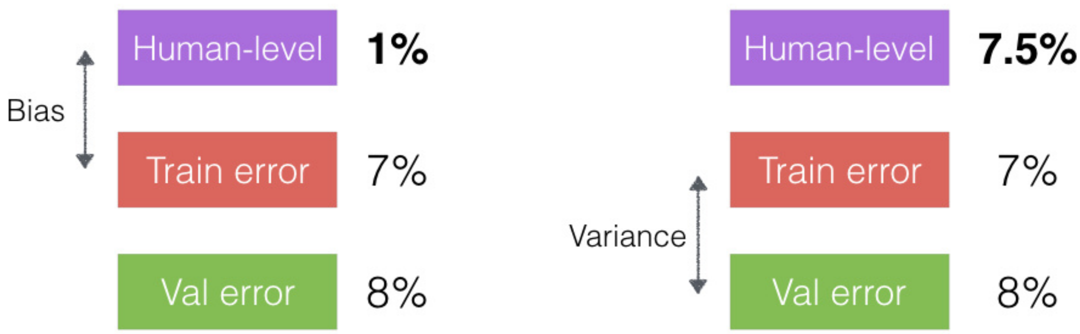
Model Performance



Why better than human ?

- Get labels from humans
- Error Analysis
- Estimate bias / variance effects

The usefulness of human-level accuracy



How to choose human-level accuracy ?

The answer is always the best accuracy possible.

This is because human-level performance is a proxy for the bayes optimal error rate, so providing a more accurate upper bound to your performance can help you strategize your next move.

There's always room for improvement

Even if you are close to human-level accuracy overall, there could be subsets of the data where you perform poorly and working on those can boost production performance greatly.

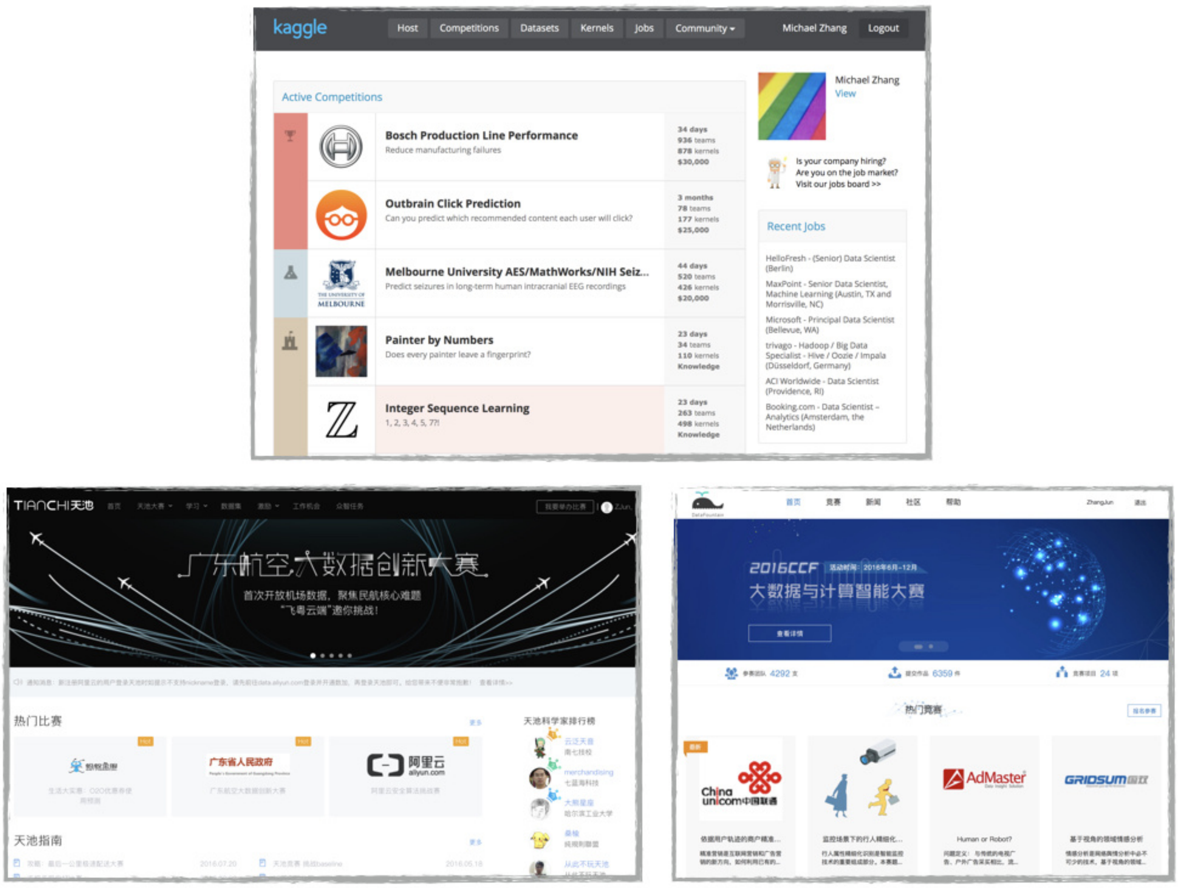
Personal Advice

How to build career in DL

- ML Course
- DL school
- Phd Student Process
- Do Dirty Work

Practice, Practice, Practice

Compete in Kaggle competitions and read associated blog posts and forum discussions.



Papers

Read a lot of papers and try to replicate the results. Soon enough, you'll get your own ideas and build your own models.

Do the Dirty Work

Thanks!