

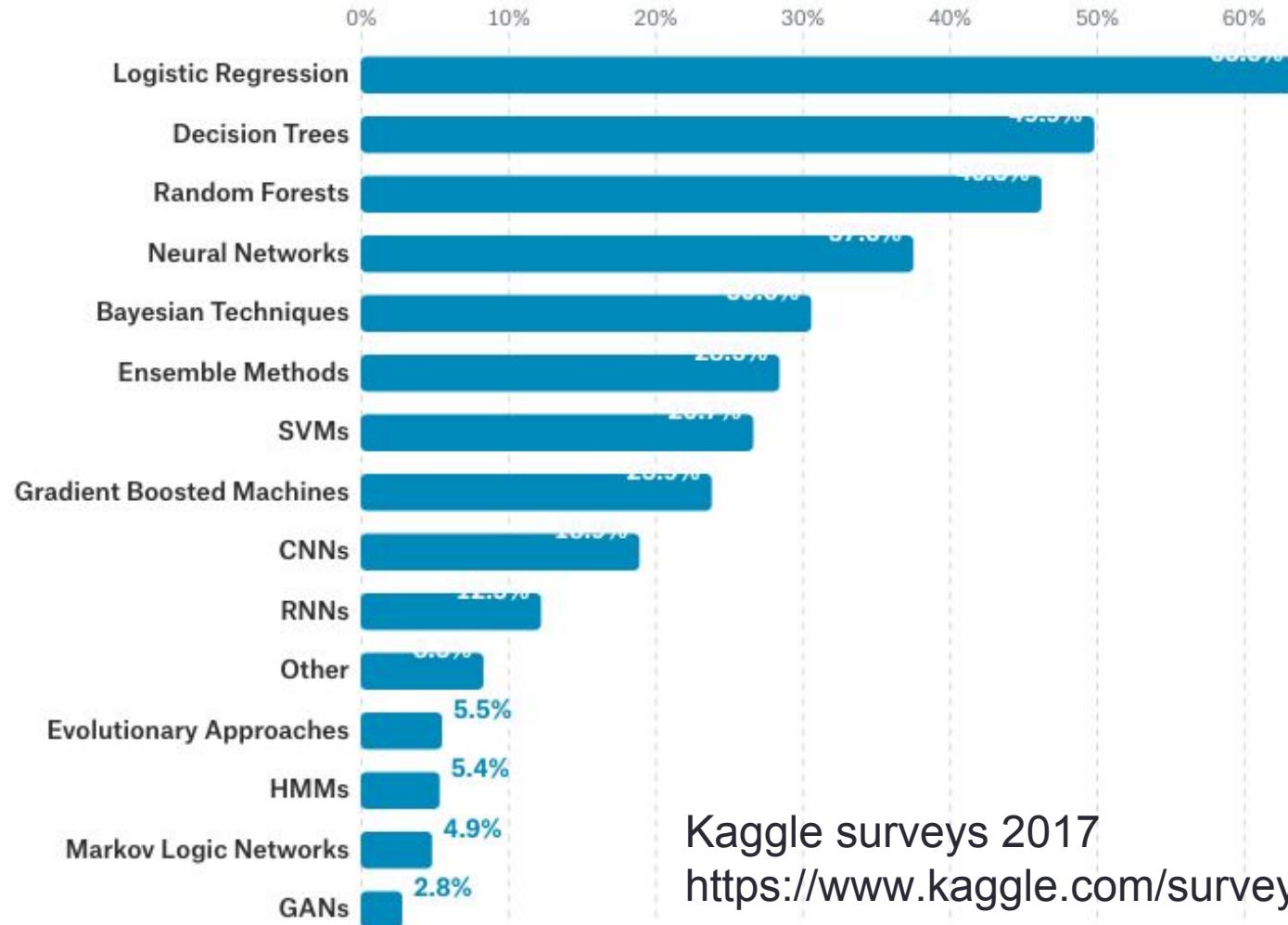
TRICKS OF THE TRADE:

Machine learning in the real world

Burning questions

- Which model to use?
- How should I improve my models?
 - Diagnosis and error analysis
- How to build a machine learning startup?

Which model?



Kaggle surveys 2017
<https://www.kaggle.com/surveys/2017>

Cautionary notes

- “There is no free lunch.”
- The “**No Free Lunch**” theorem states that there is no one model that works best for every problem.
- Depends on
 - Nature of the task
 - Nature of the data
 - Amount of data
- Which model is the best?
 - Try it on your problem.

“Deep learning is not magical.”

That's not so helpful...

Tech Support

...RESTART MY COMPUTER?
I KNOW YOU HAVE A SCRIPT
TO FOLLOW, BUT THE UPLINK
LIGHT ON THE MODEM IS GOING
OFF EVERY FEW HOURS. THE
PROBLEM IS BETWEEN YOUR
OFFICE AND THE MODEM.



MY COMPUTER HAS NOTHING
TO DO WITH ... OK, WHATEVER,
I "RESTARTED MY COMPUTER".

IT'S STILL DOWN, AND EVEN
IF IT COMES BACK, IT'S
GOING TO DIE AGAIN IN A
FEW HOURS, BECAUSE YOUR-

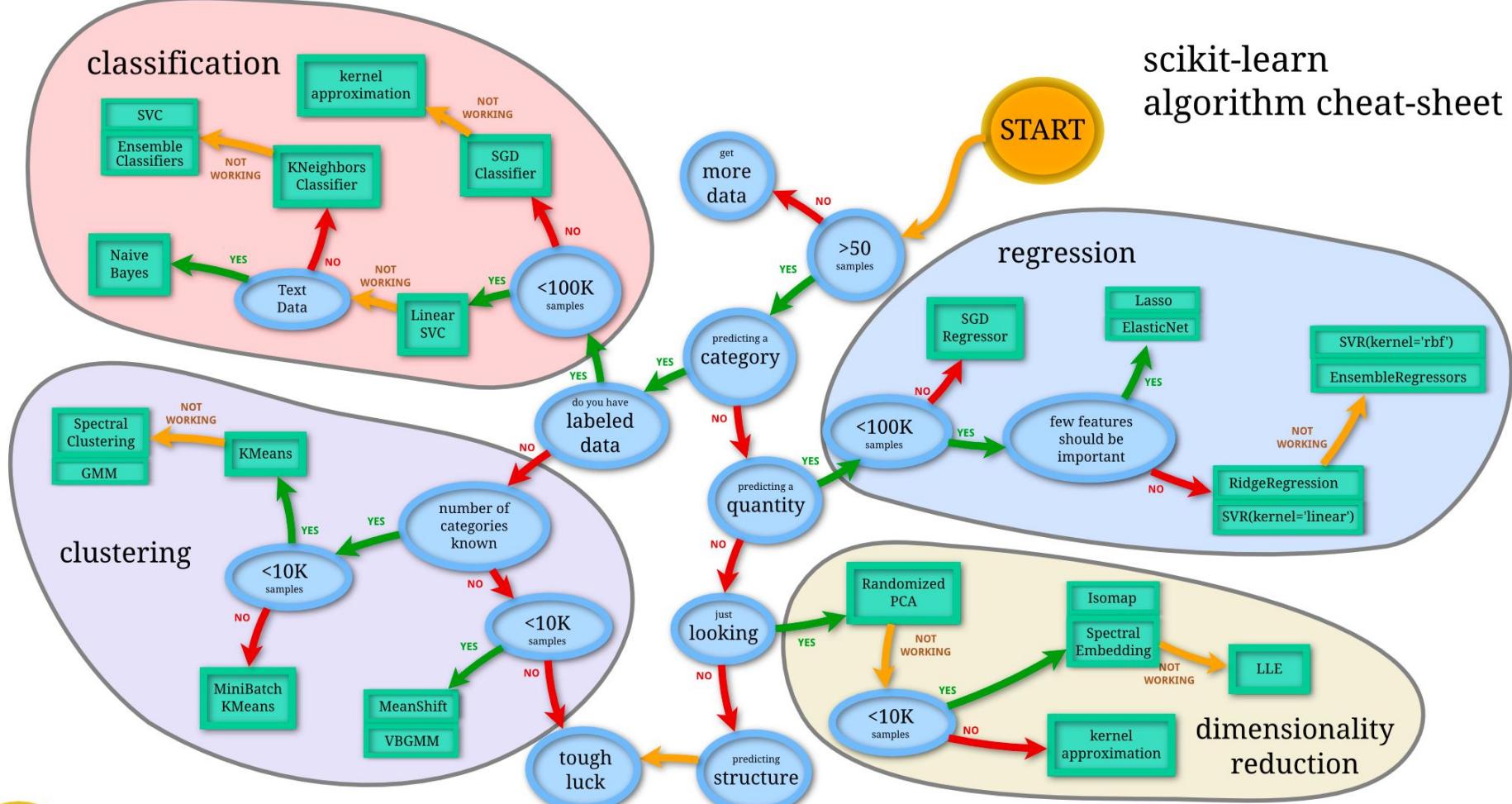


I DON'T HAVE A START MENU.
THIS IS A HAIKU INSTALL,
BUT THAT'S NOT IMPORT-

HAIKU? IT'S AN EXPERIMENTAL
OS THAT I ... OH, NEVER MIND.

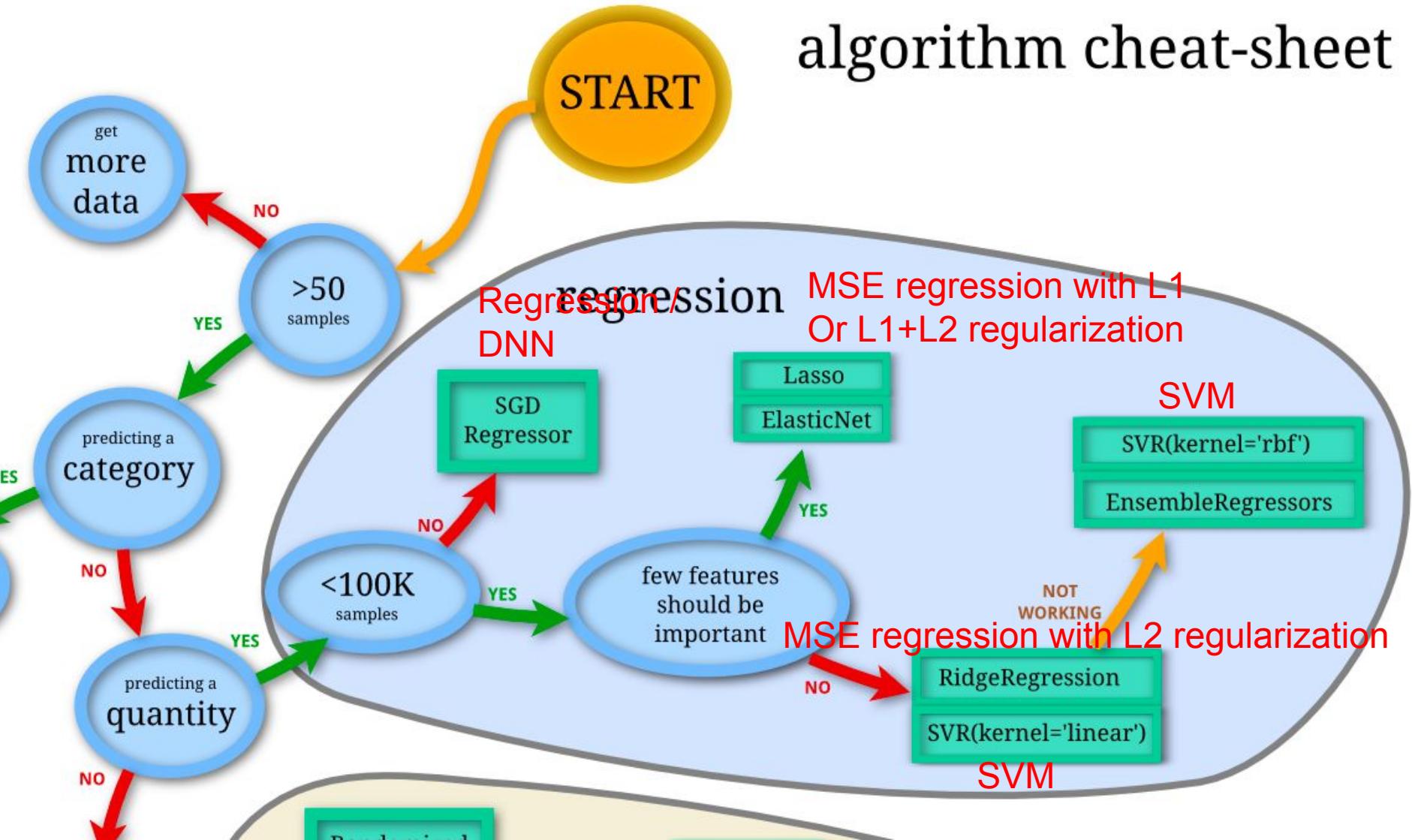


<https://xkcd.com/806/>



Note: treat 100k, 10k samples as a guideline.
 These numbers can go bigger or smaller depending on feature dimension and number of classes.

scikit-learn algorithm cheat-sheet



classification

>>Random Forest/
XGBoost<<

Kernel SVM

SVC
Ensemble
Classifiers

kernel
approximation

KNeighbors
Classifier

NOT
WORKING

SGD
Classifier

Logistic regression/ Linear SVM /
DNN

Naive
Bayes

NOT
WORKING

Text
Data

NOT
WORKING

Linear
SVC

<100K
samples

YES

NO

Linear SVM

Spectral
Clustering
GMM

NOT
WORKING

KMeans

clustering

do you have
labeled
data

predicting a
category

NO

predicting a
quantity

NO

just
looking

YES

<10K
samples

NO

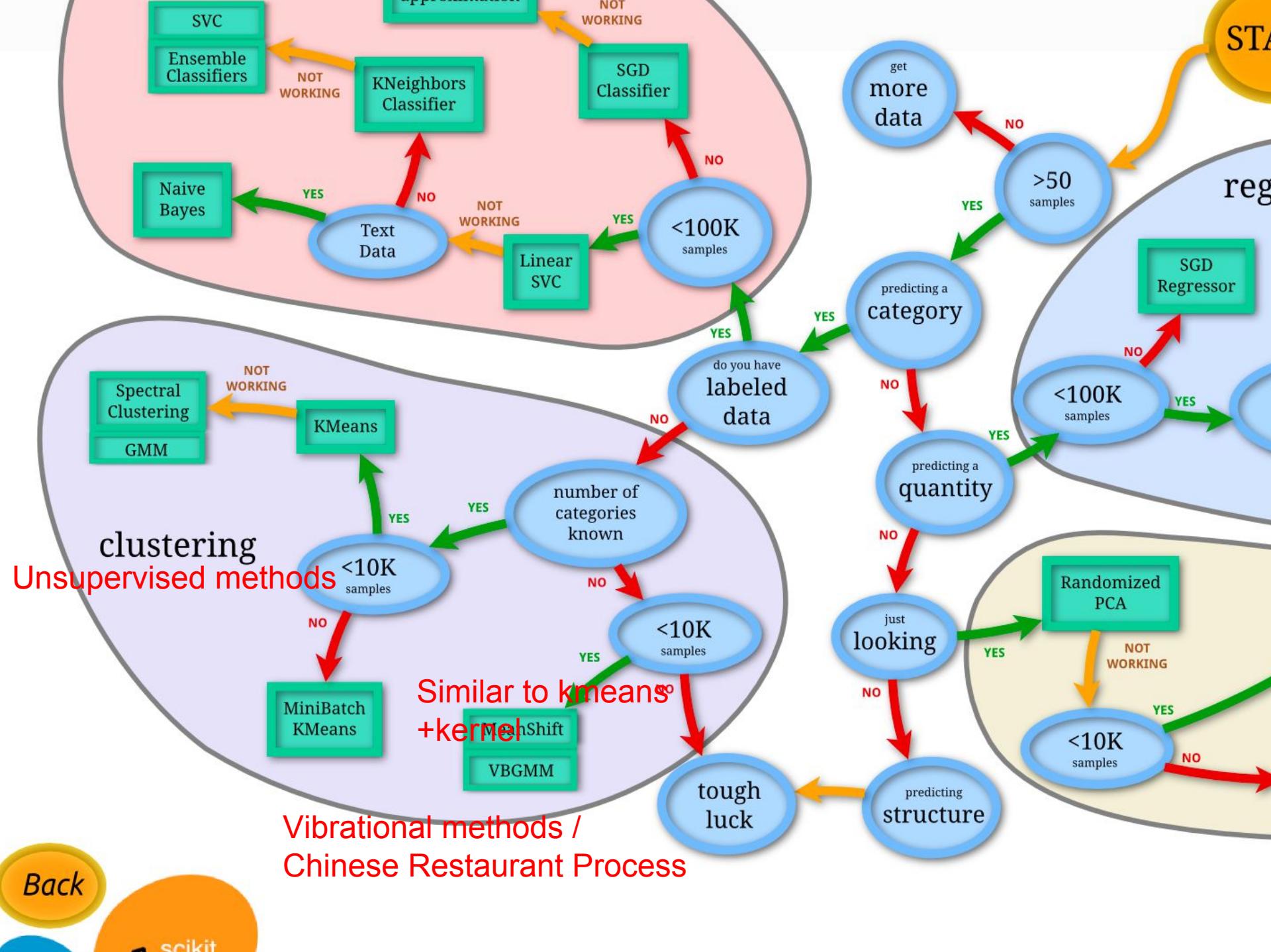
number of
categories
known

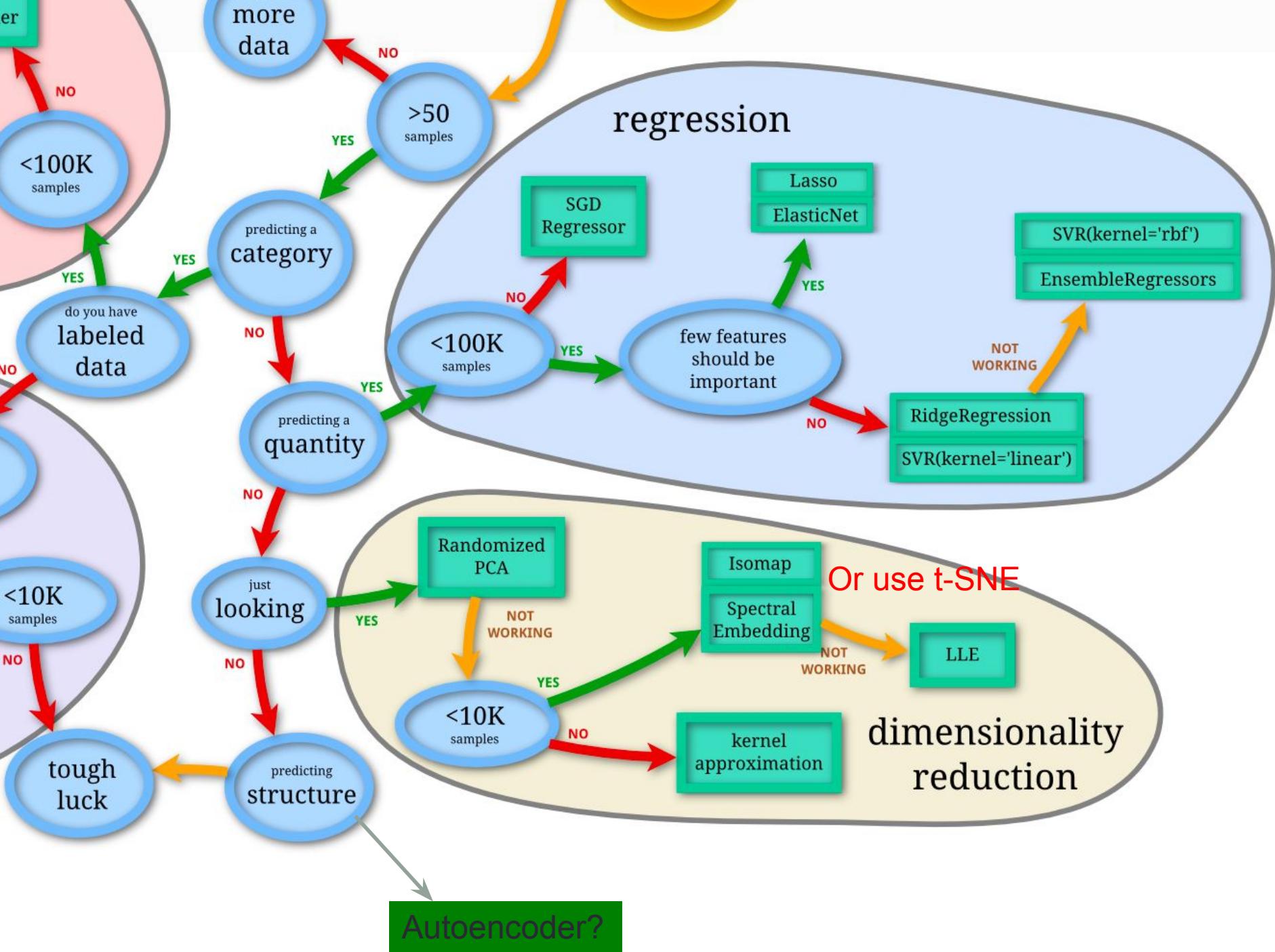
NO

<10K
samples

YES

Random
PCA





“Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?”

We evaluate **179 classifiers** arising from **17 families** (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests and other ensembles, generalized linear models, nearest-neighbors, partial least squares and principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other methods), implemented in Weka, R (with and without the caret package), C and Matlab, including all the relevant classifiers available today. We use **121 data sets**, which represent the whole UCI data base (excluding the large-scale problems) and other own real problems, in order to achieve significant conclusions about the classifier behavior, not dependent on the data set collection. **The classifiers most likely to be the bests are the random forest (RF)** versions, the best of which (implemented in R and accessed via caret) achieves 94.1% of the maximum accuracy overcoming 90% in the 84.3% of the data sets. **However, the difference is not statistically significant with the second best, the SVM with Gaussian kernel implemented in C using LibSVM, which achieves 92.3% of the maximum accuracy.** A few models are clearly better than the remaining ones: random forest, SVM with Gaussian and polynomial kernels, extreme learning machine with Gaussian kernel, C5.0 and avNNet (a committee of multi-layer perceptrons implemented in R with the caret package). The random forest is clearly the best family of classifiers (3 out of 5 bests classifiers are RF), followed by SVM (4 classifiers in the top-10), neural networks and boosting ensembles (5 and 3 members in the top-20, respectively).

Most data sets are small (<1000) in this paper

Do your literature review

- Check for the closest task
 - Does not need to be in the same domain
 - Ex: Strings of DNA -> NLP!
 - Most advances in many fields came from cross-fertilization
- Reading checklist
 - Amount of data
 - Number of classes
 - Data/classes
 - Features
 - Models
- Look at multiple papers

cross-fertilization

noun [U] • UK USUALLY cross-fertilisation **UK**

/,kros.fə:.tr.lar'zeɪʃn/ **US**
/,kra:s.fə:.tə.le'zeɪʃn/

the mixing of the ideas, customs, etc. of different places or groups of people, to produce a better result

Literature review tricks

- Search forward and backward
 - Citations
 - Cited by

[BOOK] **Neural network design**

HB Demuth, MH Beale, O De Jess, MT Hagan - 2014 - dl.acm.org

Abstract This book, by the authors of the **Neural Network** Toolbox for MATLAB, provides a clear and detailed coverage of fundamental **neural network** architectures and learning rules. In it, the authors emphasize a coherent presentation of the principal **neural** networks,

☆ 99 [Cited by 7914](#) Related articles All 3 versions >>

Reading between the lines

- If you don't understand something, re-read
 - Average of 5+ times to understand a paper completely
- Print it out, keep a pen/pencil at hand, and break down equations
- Try to explain it with what you already know

Elastic net Loss function $\min_w \frac{1}{2n_{samples}} \|Xw - y\|_2^2 + \alpha\rho\|w\|_1 + \frac{\alpha(1-\rho)}{2}\|w\|_2^2$

Reading between the lines

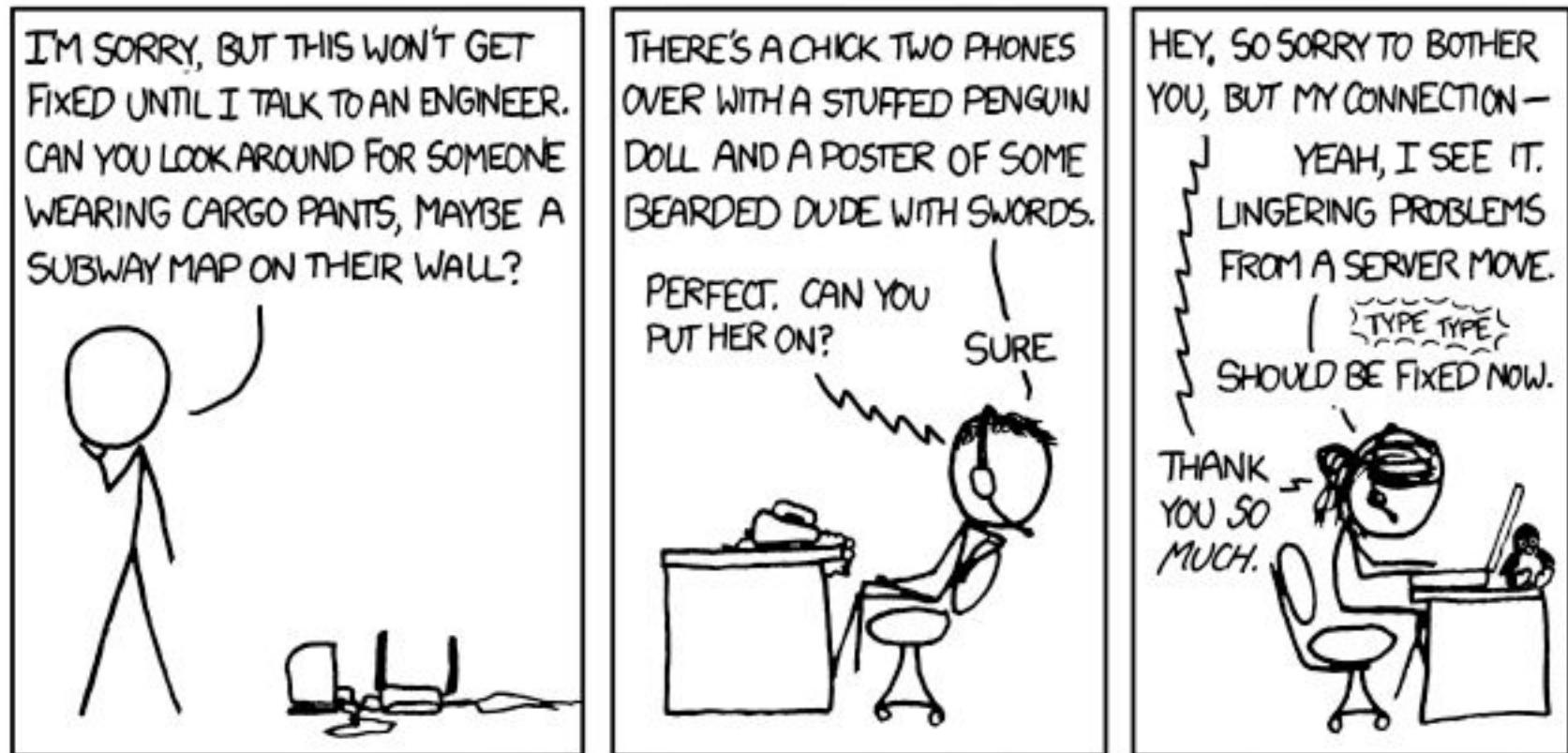
- If you don't understand something, re-read
 - Average of 5+ times to understand a paper completely
- Print it out, keep a pen/pencil at hand, and break down equations
- Try to explain it with what you already know

Elastic net Loss function $\min_w \frac{1}{2n_{samples}} \|Xw - y\|_2^2 + \alpha\rho\|w\|_1 + \frac{\alpha(1-\rho)}{2}\|w\|_2^2$

MSE Loss L1 L2

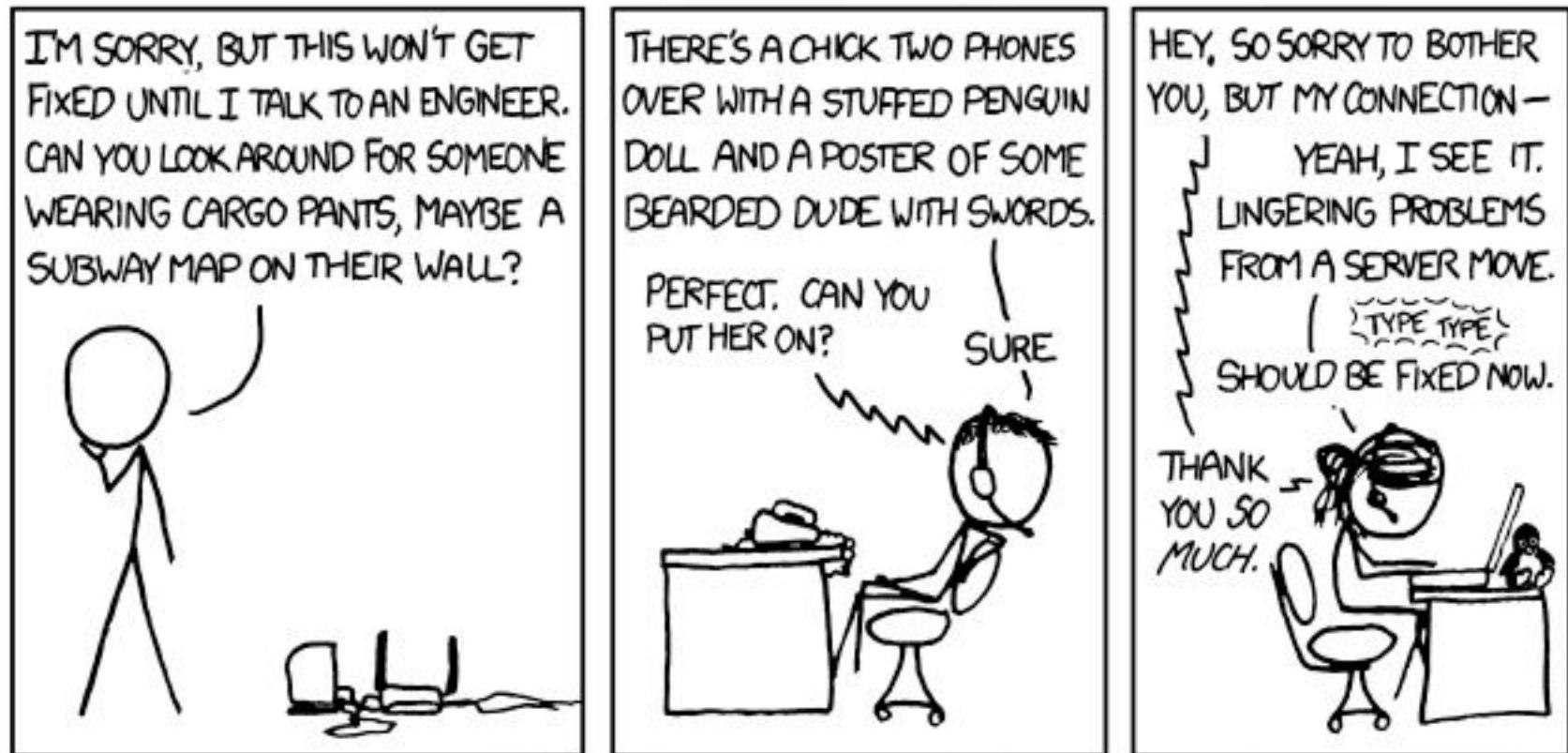
How to improve my results?

- More tech support



How to improve my results?

- More tech support

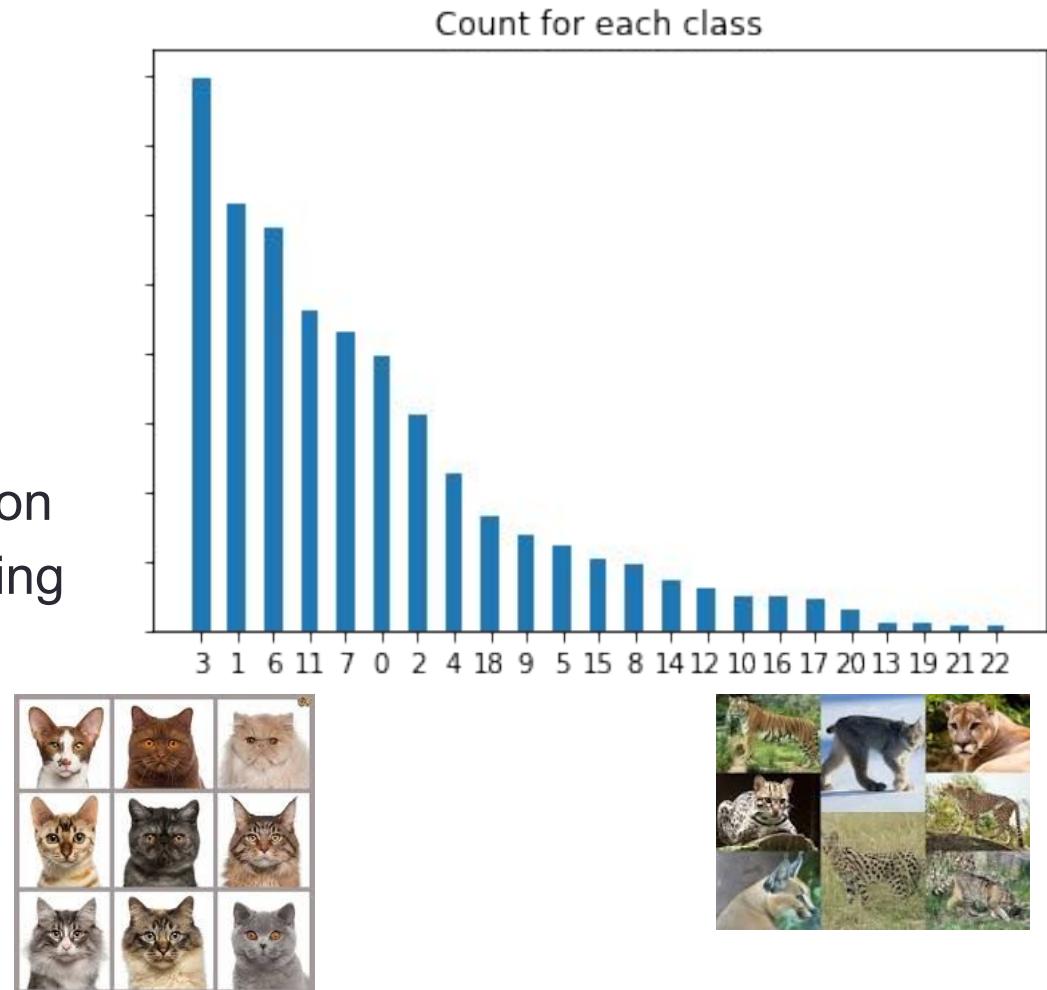


Diagnosis and prescription

- Is my task proper?
- Should I add/remove feature XXX?
- Is my loss function appropriate?
- Is my model overfitting/underfitting?
- Should I try XXX?

Is my task proper? Do you suffer from class imbalance?

- Throwing away
- Refactoring
 - Split
 - Merge
- Data augmentation
- Biasing
 - Weighting the loss function
 - Bias in mini-batch sampling



Diagnosis

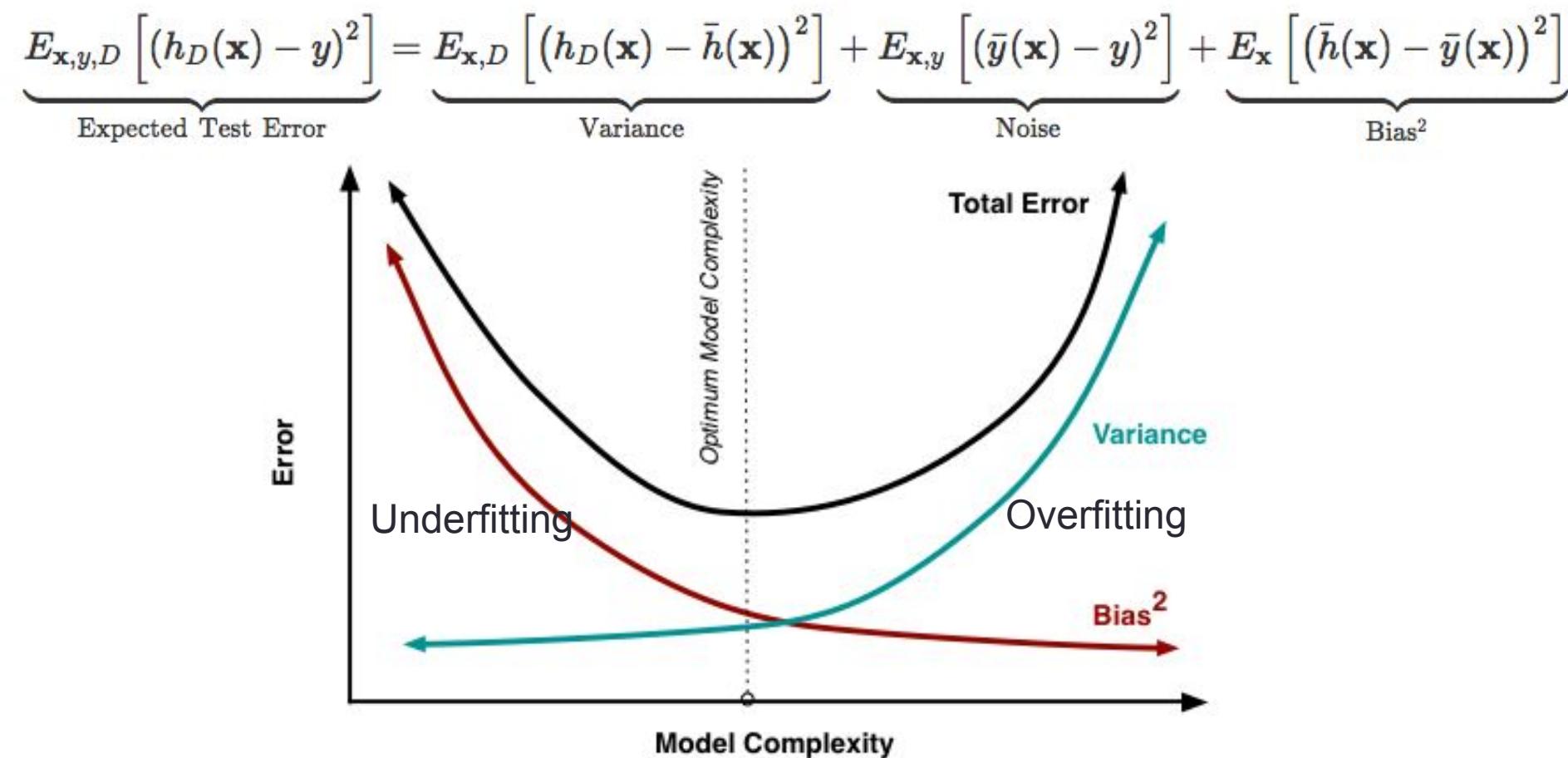
- Is my task proper?
- Should I add/remove feature XXX?
- Is my loss function appropriate?
- Is my model overfitting/underfitting?
- Should I try XXX?

Short answer: Cross Validation

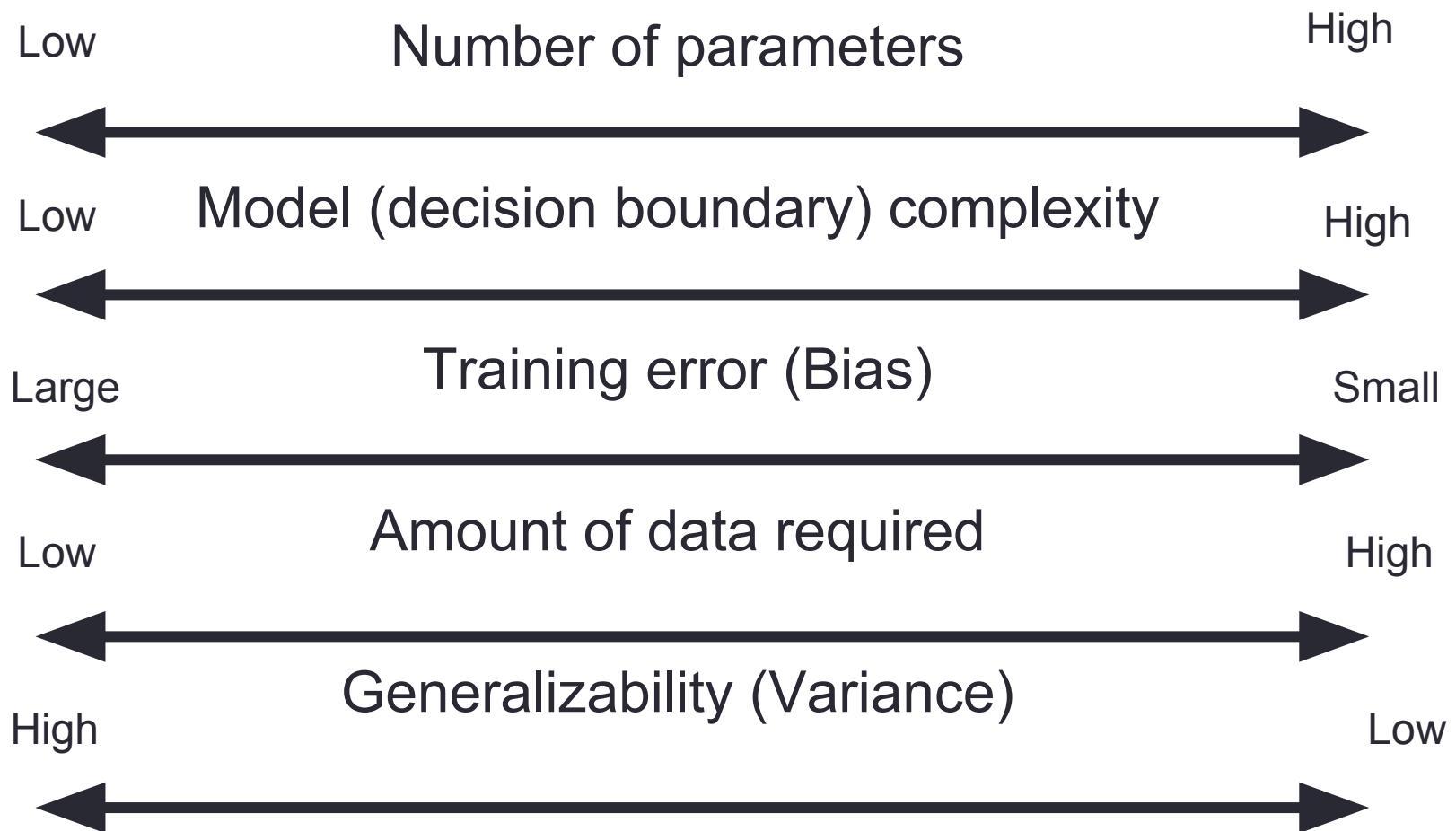
Can we do better?
Find the problem and fix it.

Understanding the bias variance trade-off

- Bias variance analysis can be helpful for diagnosis



Bias-Variance overview

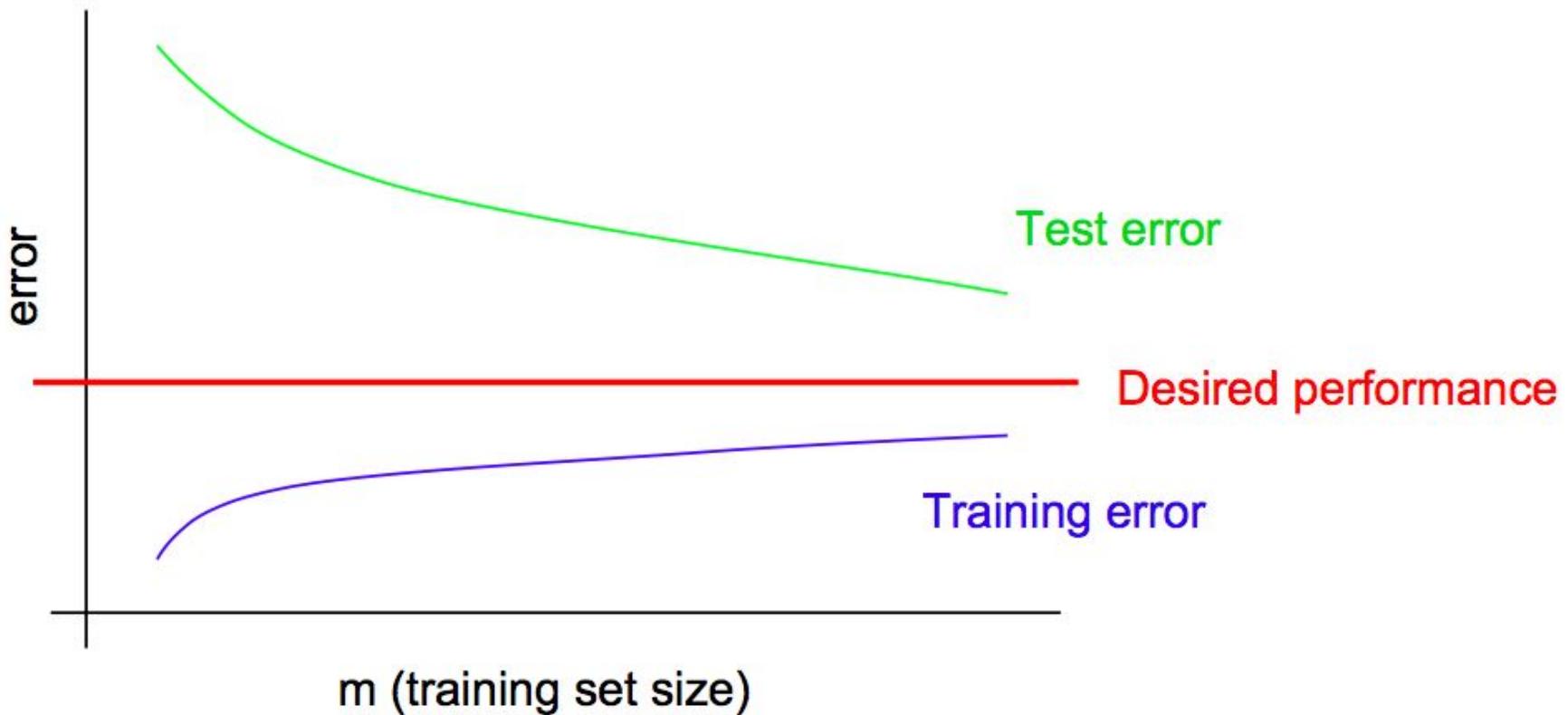


Bias-variance diagnosis

- Suppose the problem is either
 - Overfitting (high variance)
 - Too few features to classify properly (high bias)
- Symptoms
 - Variance: Training error is much lower than test/dev error.
 - Bias: Training error is also high

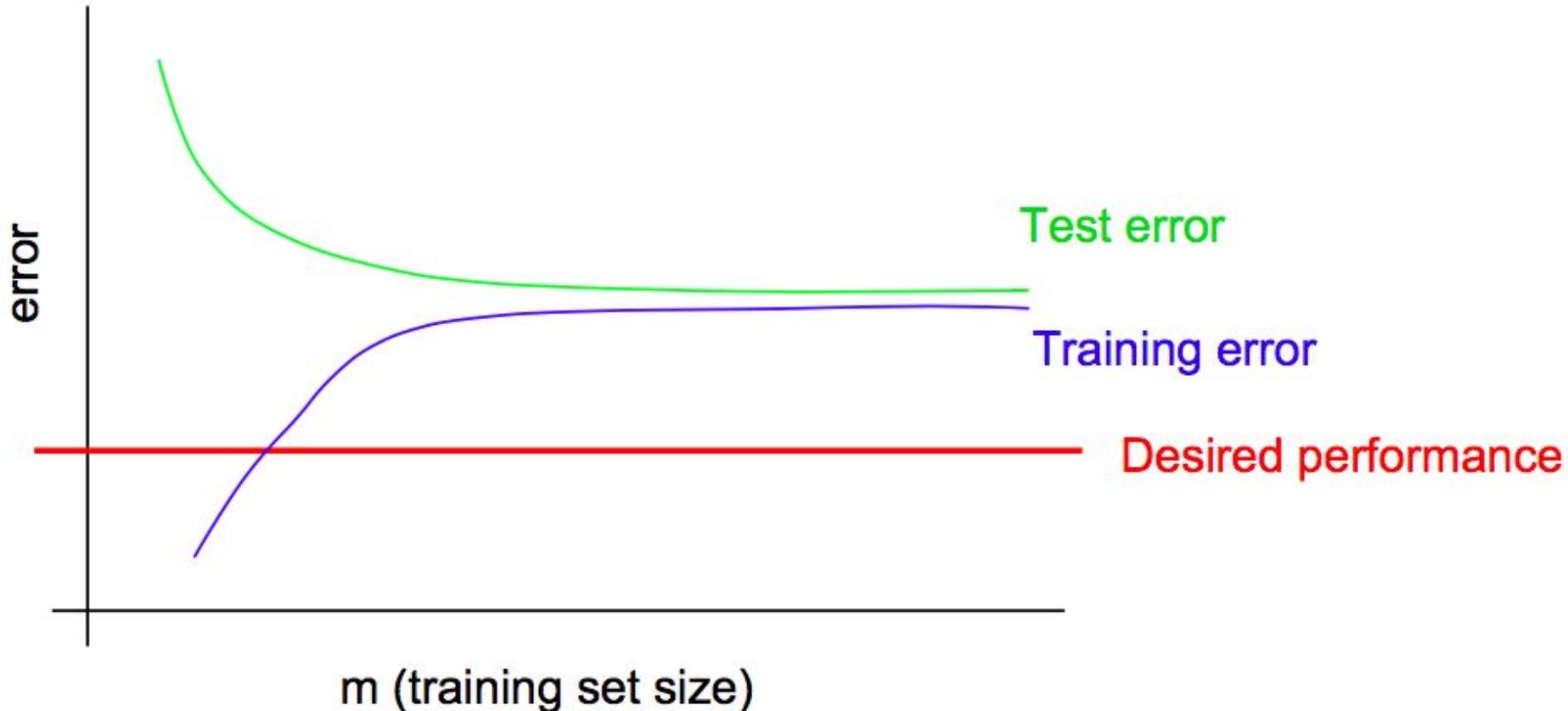
High variance case

- Solution:
 - Reduce overfitting: regularization, reduce features, etc
 - Get more training set



High bias case

- Solution: more features, more layers, bigger models



Desired performance?

- The goal of the application
 - Possibilities
 - Target performance to launch the product
 - Human performance on the task
 - Have A and B do the same task, measure the difference.
- Knowing human performance gives several advantages
 - Knows when to stop
 - Beating human is sometimes a goal, but
 - Some errors might be labeling errors or judgment calls



More diagnosis example

- Flying a drone using RL
- You're poor and can't repeat CMU's drone crashing experiments



- What you did:
- Make a simulation
- Define a reward function R
- Learn the policy, p , using some RL algorithm.
 - Maximize $R(p_{\text{drone}})$

Drone diagnosis example

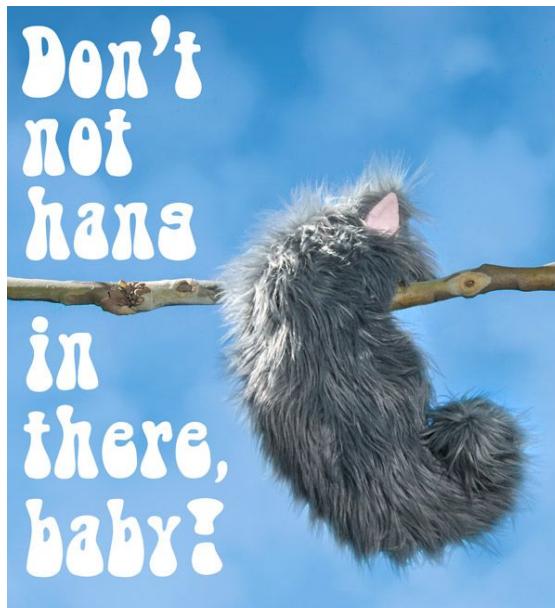
- In real testing, your drone crashes and burns because your policy is bad.



- Diagnosis:
 - If the drone fly well in simulation but not in real life testing, you simulation is bad.
 - Let be p_{rule} be some rule-base control policy developed by drone flying expert. If $R(p_{drone}) < R(p_{human})$, RL algorithm fails to maximize the rewards. Fix the RL algorithm
 - If $R(p_{drone}) > R(p_{human})$, the RL maximization is doing its job properly. Fix the reward function.

Error diagnosis

- You're making a cat classifier. (cat/not cat)
- It sucks.
- You heard of this super new hype algorithm (for example: capsule network). Should you spend months to try it out?



This is not a cat



Looking at the errors

- Spend an hour or two looking at your errors. Identify why.
Keep a table.

	Blurred	Weird angle	Notes
Pic1	x		Stuffed toy
Pic2	x		
Pic3	x		
Pic4		x	Top view
...
	68%	2%	

Solution: Use a method to sharpen the image. Train on blurry images.

The table categories can expand as you look through more pictures and see frequently occurring error cases. So keep notes.

Diagnosis summary

- Simple analysis of the data can help you notice underlying problems
- Bias-variance diagnosis is a general method that can be applied to most tasks
- Other diagnosis depends on the application and need some understanding of the algorithms
- Error analysis can help guide your model improvements

Taking the time to do diagnosis can help you save months of trying random things.

Burning questions

- Which model to use?
- How should I improve my models?
 - Diagnosis and error analysis
- How to build a machine learning startup?

Burning questions

- Which model to use?
- How should I improve my models?
 - Diagnosis and error analysis
- What does a data scientist do at work?
- How to build a machine learning startup?

What does a data scientist do at work?

- Kaggle competition winner
 - <http://blog.kaggle.com/2015/12/21/rossmann-store-sales-winners-in-terview-1st-place-gert/>

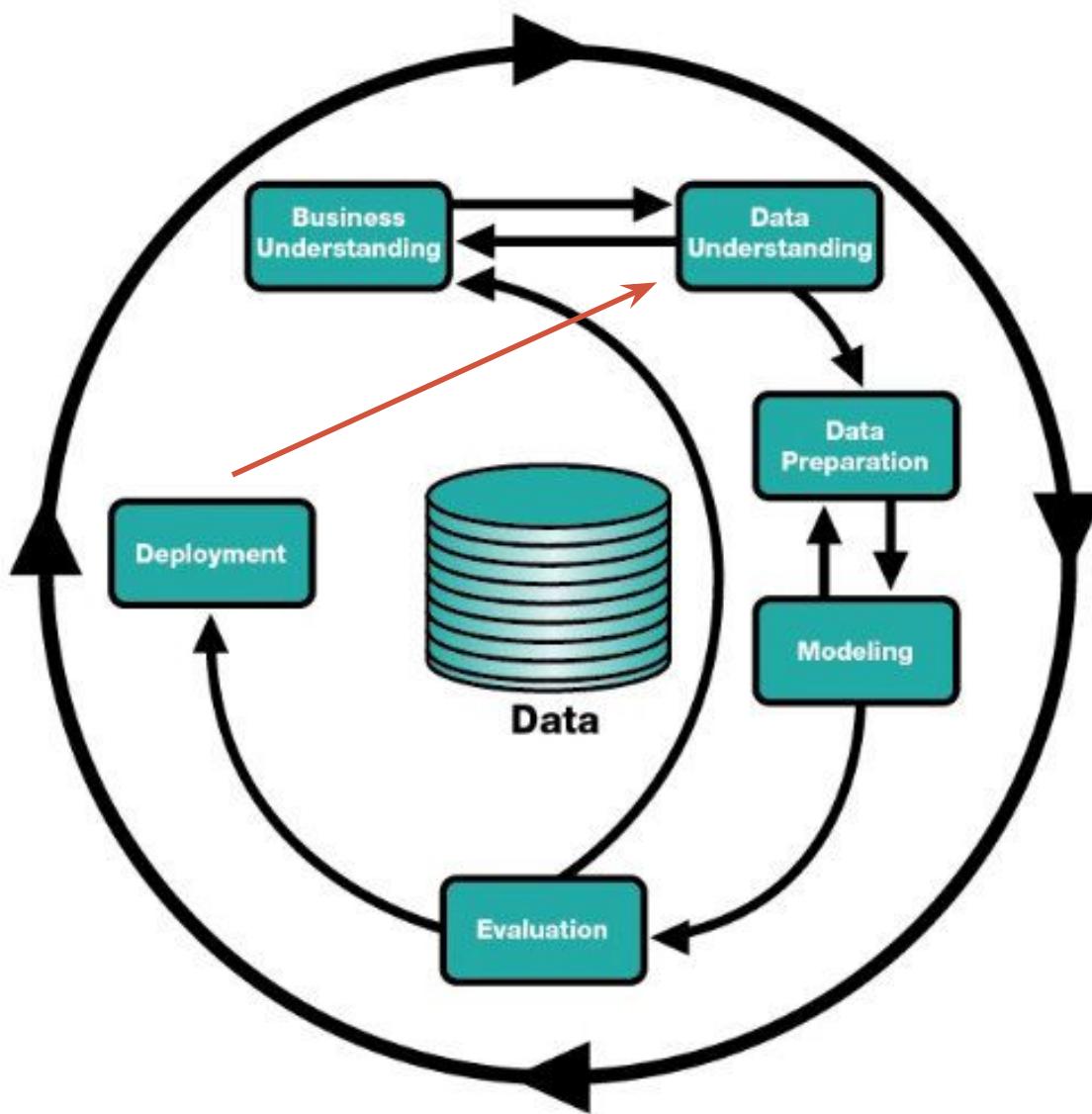
How did you spend your time on this competition?

I spent 50% on feature engineering, 40% on feature selection plus model ensembling, and less than 10% on model selection and tuning.

Note: this is highly automatable (See Datarobot, auto-sklearn, auto-ml, etc.). Will data scientist lose our jobs?

But this is not all of data science

What does a data scientist do at work?



- Data science loop
- In industry:
“data preparation and modeling (10%) and the rest (90%)”

Kaggle competition pitfalls and automation

- Given task
- Given (closed-set) of inputs
- Given performance metrics
- Easy replaced by automation.

Intelligent Machines

Automating the Data Scientists

Software that can discover patterns in data and write a report on its findings could make it easier for companies to analyze it.

by Tom Simonite February 13, 2015

Many organizations have more data than they're able to interpret.

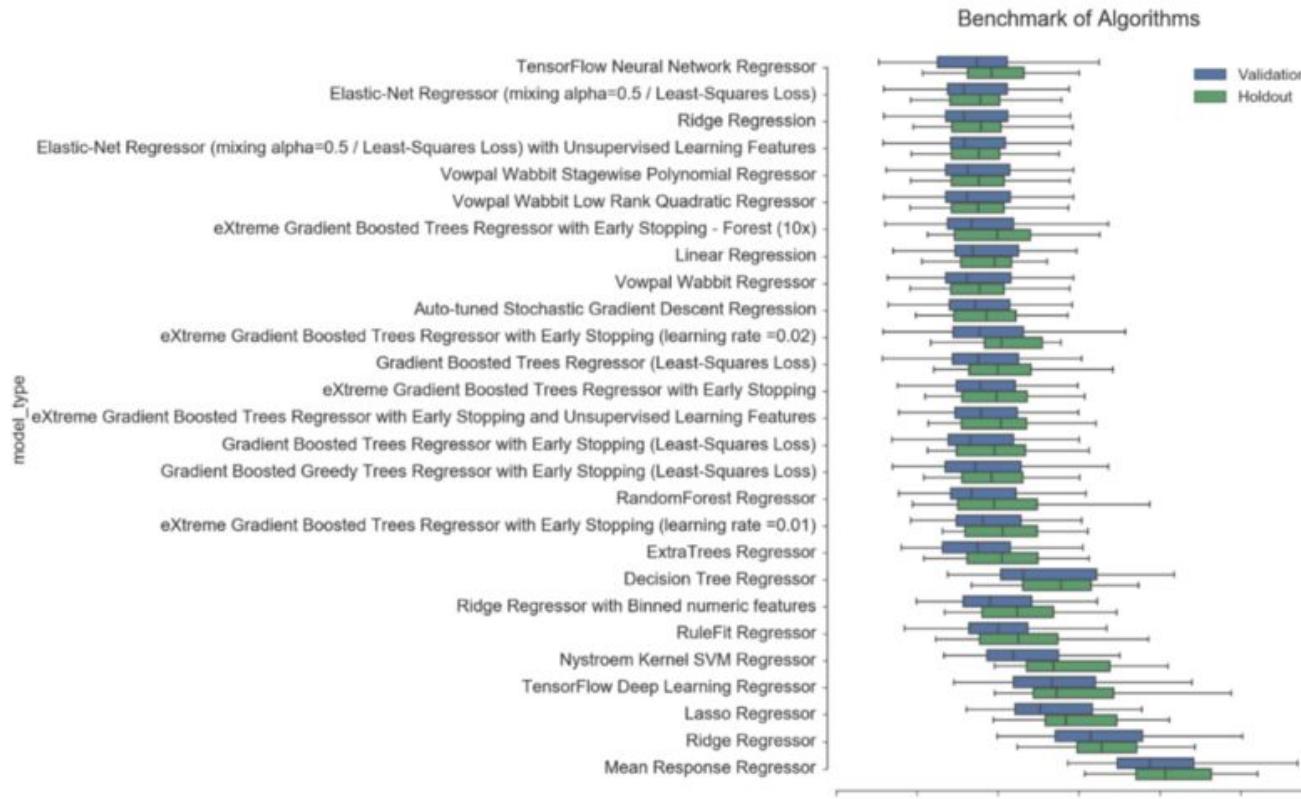
<https://www.technologyreview.com/s/535041/automating-the-data-scientists/>

The 90%

- In real life, you will have to look for and decide your tasks, your inputs, your metrics.
- Understand what's possible and what's not
 - And be able to describe it to non-data scientists
- Justify business usage and make sure about deployment.

Data scientists + automation tools

- Automate tools can remove boring tasks in data science
 - Give powerful benchmarks
 - Explore the space faster

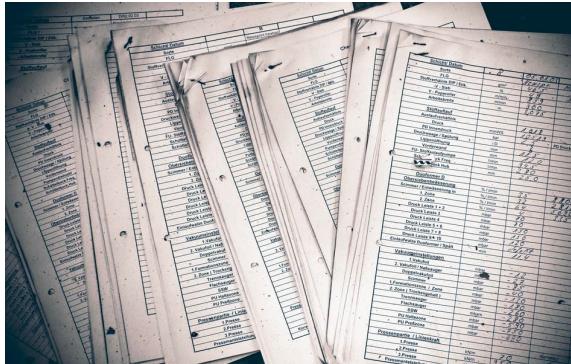


Burning questions

- Which model to use?
- How should I improve my models?
 - Diagnosis and error analysis
- What does a data scientist do at work?
- How to build a machine learning startup?

Build a machine learning application

- What are needed for Machine Learning in business applications?



Data



Business goal



Team



Business goal

Pick the task

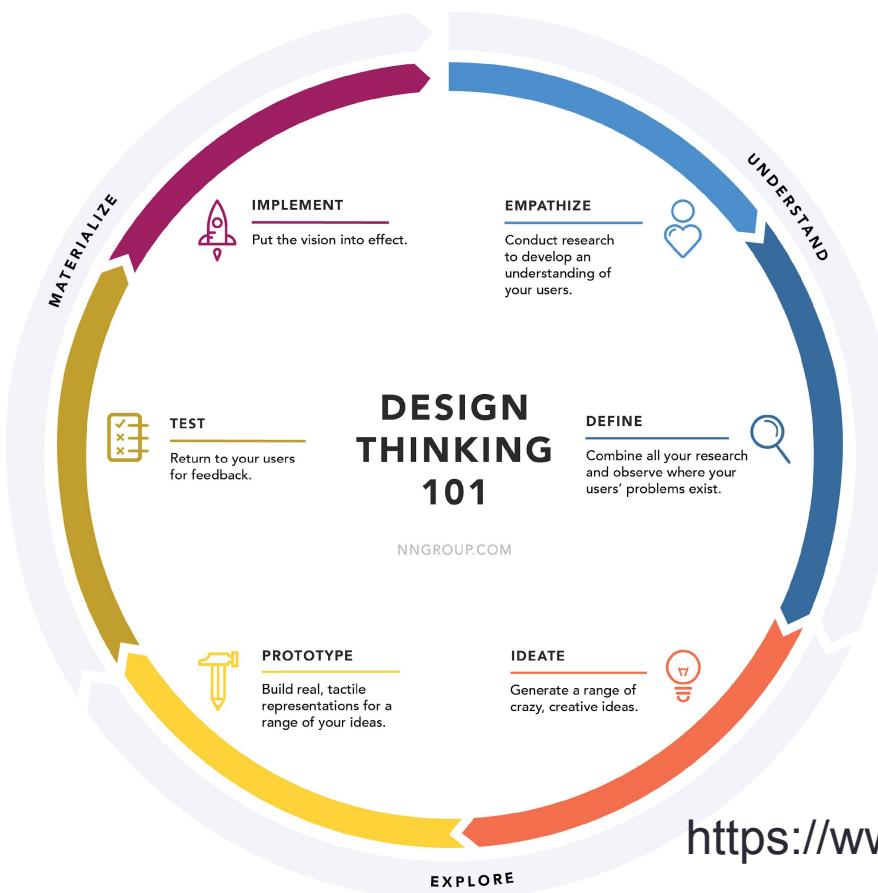
Pick the problem

- The goal can dictate the collection of data and the machine learning methods
- Treat it just like any other business plan
 - Machine Learning tasks are harder to evaluate the potential benefits
 - As machine learning gets automated, the bottleneck now is in **management, implementation, and business imagination.**

Machine learning is not a magic bullet

I hate to use this buzzword

- Design thinking
 - User centric thinking framework



<https://www.nngroup.com/articles/design-thinking/>

Example: Recommendation systems

- How can the recommendation system generates value?
- How can it be integrated into the existing UX flow?
- How can this feature differentiate you from competitors?
- What is a minimum requirement for the feature?

Customers who bought this item also bought

Page 1 of 13



Dragonpad USA Pop filter
Studio Microphone Mic
Wind Screen Pop Filter
★★★★★ 3,564
\$8.39 ✓prime



NEEWER Adjustable
Microphone Suspension
Boom Scissor Arm Stand,
Compact Mic Stand...
★★★★★ 3,050
#1 Best Seller in
Microphone Stands
\$12.50 ✓prime



Elgato Chat Link, Party
Chat Adapter for Xbox One
and PlayStation 4
★★★★★ 330
\$9.95 ✓prime



Neewer NW(B-3) 6 inch
Studio Microphone Mic
Round Shape Wind Pop
Filter Mask Shield with...
★★★★★ 882
#1 Best Seller in
Microphone Windscreens &
Pop...
\$6.99



Elgato Game Capture
HD60 S - stream, record
and share your gameplay
in 1080p60, superior low...
★★★★★ 1,544
\$179.94 ✓prime



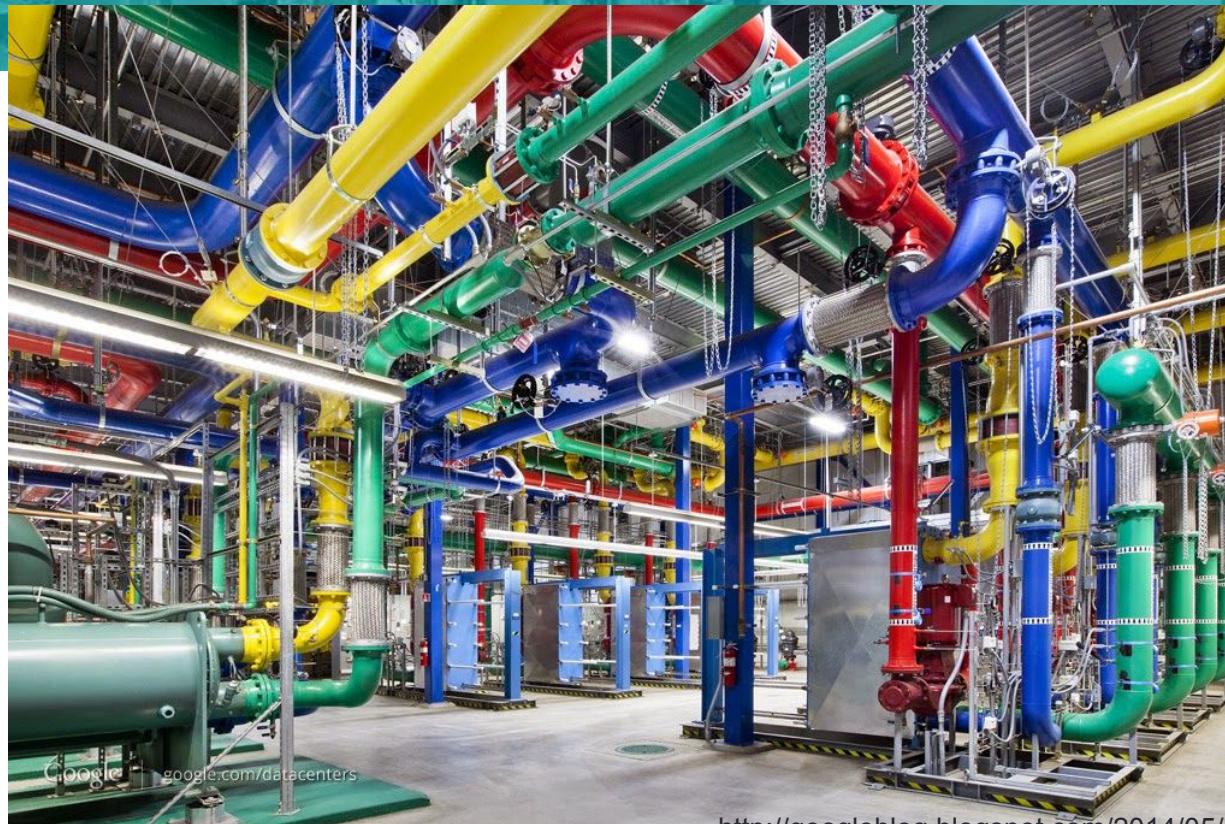
Elgato Game Capture
HD60, for PlayStation 4,
Xbox One and Xbox 360,
or Wii U gameplay, Full...
★★★★★ 1,544
#1 Best Seller in Internal
TV Tuner & Video...
\$151.74 ✓prime

What tasks are good for machine learning?

- Tasks that requires less than 1 second to think
 - ASR, Object recognition, Face recognition
- Augmenting human performance
 - Healthcare diagnosis
 - Self-driving cars
- Tasks at a large scale



DeepMind AI Reduces Google Data Centre Cooling Bill by 40%



<http://googleblog.blogspot.com/2014/05/better-data-centers-through-machine.html>

Aspects to consider

- Metrics
- Timings



Defining a metric

- Machine learning methods are evaluated based on metrics

Speech Recognition - Word Error Rate

Search – Recall rate

Machine Translation - BLEU score

Recommendation system – Click through rate

- Tasks can have multiple metrics
- Understand the relationship between the metric and the business goal

Example: Chatbots

- Task completion rate
- Number of turns to complete a task
- Satisfaction
- Word Error Rate (if speech input)



Understand the trade-offs

- Performance metrics
- Amount of computation power required
- Type of computation required
- Latency



Picking one guiding metric

- In order to develop fast, pick one summarizing metric to optimize
 - Precision + recall -> F-score
 - Weighted average of multiple metrics
- Task driven
- Instead of combining all metrics. Some metrics are **optimizing metrics** while some are **constraints metrics**
 - Must use less than X ram
 - Produce output in less than x milliseconds
 - Less than 1 false alarm per day
 - Note performance, metrics can sometimes be relaxed.
 - Easier to make something fast, than make something good.

Understand the timings

- Most data are not ready for machine learning
- Data preparation takes time

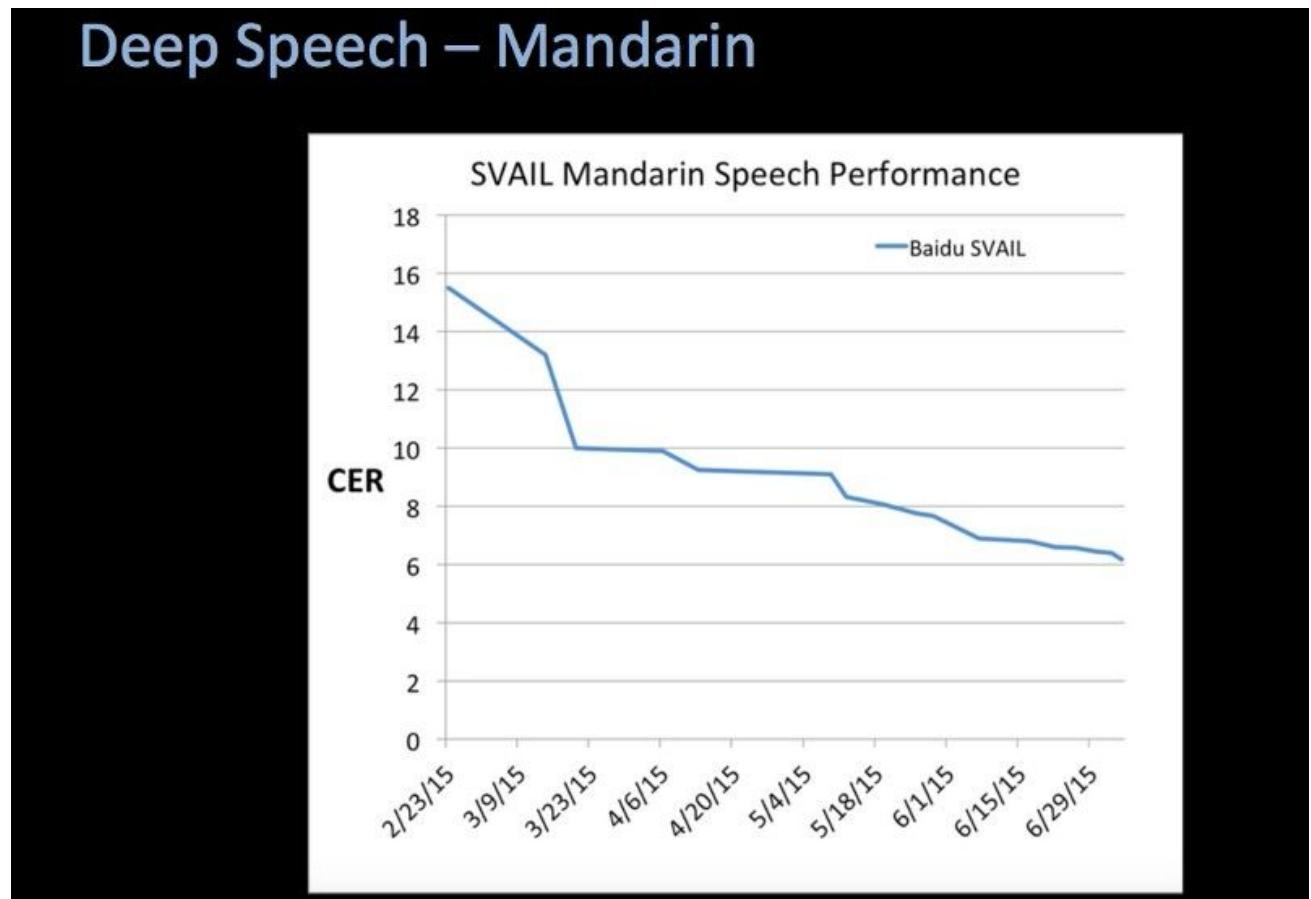


- Improving performance takes time

<http://www.nvidia.com/object/drive-px.html>

Performance and time spent on development

- 80/20 rule



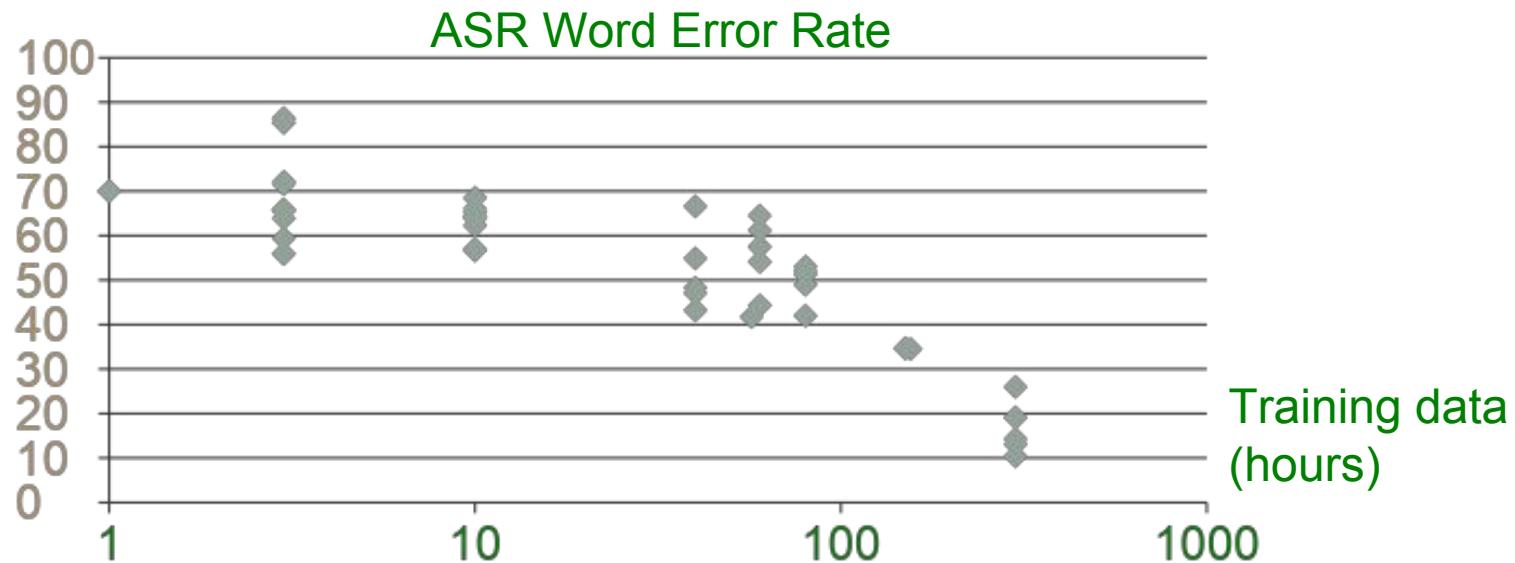
<https://medium.com/s-c-a-l-e/how-baidu-mastered-mandarin-with-deep-learning-and-lots-of-data-1d94032564a5>

Data		
1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	30
31	32	33
34	35	36
37	38	39
40	41	42
43	44	45
46	47	48
49	50	51
52	53	54
55	56	57
58	59	60
61	62	63
64	65	66
67	68	69
70	71	72
73	74	75
76	77	78
79	80	81
82	83	84
85	86	87
88	89	90
91	92	93
94	95	96
97	98	99
100	101	102
103	104	105
106	107	108
109	110	111
112	113	114
115	116	117
118	119	120
121	122	123
124	125	126
127	128	129
130	131	132
133	134	135
136	137	138
139	140	141
142	143	144
145	146	147
148	149	150
151	152	153
154	155	156
157	158	159
160	161	162
163	164	165
166	167	168
169	170	171
172	173	174
175	176	177
178	179	180
181	182	183
184	185	186
187	188	189
190	191	192
193	194	195
196	197	198
199	200	201
202	203	204
205	206	207
208	209	210
211	212	213
214	215	216
217	218	219
220	221	222
223	224	225
226	227	228
229	230	231
232	233	234
235	236	237
238	239	240
241	242	243
244	245	246
247	248	249
250	251	252
253	254	255
256	257	258
259	260	261
262	263	264
265	266	267
268	269	270
271	272	273
274	275	276
277	278	279
280	281	282
283	284	285
286	287	288
289	290	291
292	293	294
295	296	297
298	299	300
301	302	303
304	305	306
307	308	309
310	311	312
313	314	315
316	317	318
319	320	321
322	323	324
325	326	327
328	329	330
331	332	333
334	335	336
337	338	339
340	341	342
343	344	345
346	347	348
349	350	351
352	353	354
355	356	357
358	359	360
361	362	363
364	365	366
367	368	369
370	371	372
373	374	375
376	377	378
379	380	381
382	383	384
385	386	387
388	389	390
391	392	393
394	395	396
397	398	399
400	401	402
403	404	405
406	407	408
409	410	411
412	413	414
415	416	417
418	419	420
421	422	423
424	425	426
427	428	429
430	431	432
433	434	435
436	437	438
439	440	441
442	443	444
445	446	447
448	449	450
451	452	453
454	455	456
457	458	459
460	461	462
463	464	465
466	467	468
469	470	471
472	473	474
475	476	477
478	479	480
481	482	483
484	485	486
487	488	489
490	491	492
493	494	495
496	497	498
499	500	501
502	503	504
505	506	507
508	509	510
511	512	513
514	515	516
517	518	519
520	521	522
523	524	525
526	527	528
529	530	531
532	533	534
535	536	537
538	539	540
541	542	543
544	545	546
547	548	549
550	551	552
553	554	555
556	557	558
559	560	561
562	563	564
565	566	567
568	569	570
571	572	573
574	575	576
577	578	579
580	581	582
583	584	585
586	587	588
589	590	591
592	593	594
595	596	597
598	599	600
601	602	603
604	605	606
607	608	609
610	611	612
613	614	615
616	617	618
619	620	621
622	623	624
625	626	627
628	629	630
631	632	633
634	635	636
637	638	639
640	641	642
643	644	645
646	647	648
649	650	651
652	653	654
655	656	657
658	659	660
661	662	663
664	665	666
667	668	669
670	671	672
673	674	675
676	677	678
679	680	681
682	683	684
685	686	687
688	689	690
691	692	693
694	695	696
697	698	699
700	701	702
703	704	705
706	707	708
709	710	711
712	713	714
715	716	717
718	719	720
721	722	723
724	725	726
727	728	729
730	731	732
733	734	735
736	737	738
739	740	741
742	743	744
745	746	747
748	749	750
751	752	753
754	755	756
757	758	759
760	761	762
763	764	765
766	767	768
769	770	771
772	773	774
775	776	777
778	779	780
781	782	783
784	785	786
787	788	789
790	791	792
793	794	795
796	797	798
799	800	801
802	803	804
805	806	807
808	809	810
811	812	813
814	815	816
817	818	819
820	821	822
823	824	825
826	827	828
829	830	831
832	833	834
835	836	837
838	839	840
841	842	843
844	845	846
847	848	849
850	851	852
853	854	855
856	857	858
859	860	861
862	863	864
865	866	867
868	869	870
871	872	873
874	875	876
877	878	879
880	881	882
883	884	885
886	887	888
889	890	891
892	893	894
895	896	897
898	899	900
901	902	903
904	905	906
907	908	909
910	911	912
913	914	915
916	917	918
919	920	921
922	923	924
925	926	927
928	929	930
931	932	933
934	935	936
937	938	939
940	941	942
943	944	945
946	947	948
949	950	951
952	953	954
955	956	957
958	959	960
961	962	963
964	965	966
967	968	969
970	971	972
973	974	975
976	977	978
979	980	981
982	983	984
985	986	987
988	989	990
991	992	993
994	995	996
997	998	999
999	999	999

Accumulate data

Machine learning uses data

- More data means better models



- You can start a machine learning project without data
 - If the business plan says it's feasible

Making use of your data

- Train/dev/test split
 - 80/10/10, 90/5/5, 5-fold CV, leave one out CV, etc. for academia
- For real applications, get dev and test sets that represent your users.
 - Reflects the data you want to do well on.
 - There can be a mis-match between train and dev data. But avoid mis-match between dev and test data.
 - If no users, recruit friends to pretend to be the users.
- Example: Cat classifier.
 - Should you use ImageNet cat pictures as train/dev/test?
 - Go pretend you're a user and take cat pictures for the dev/test set.

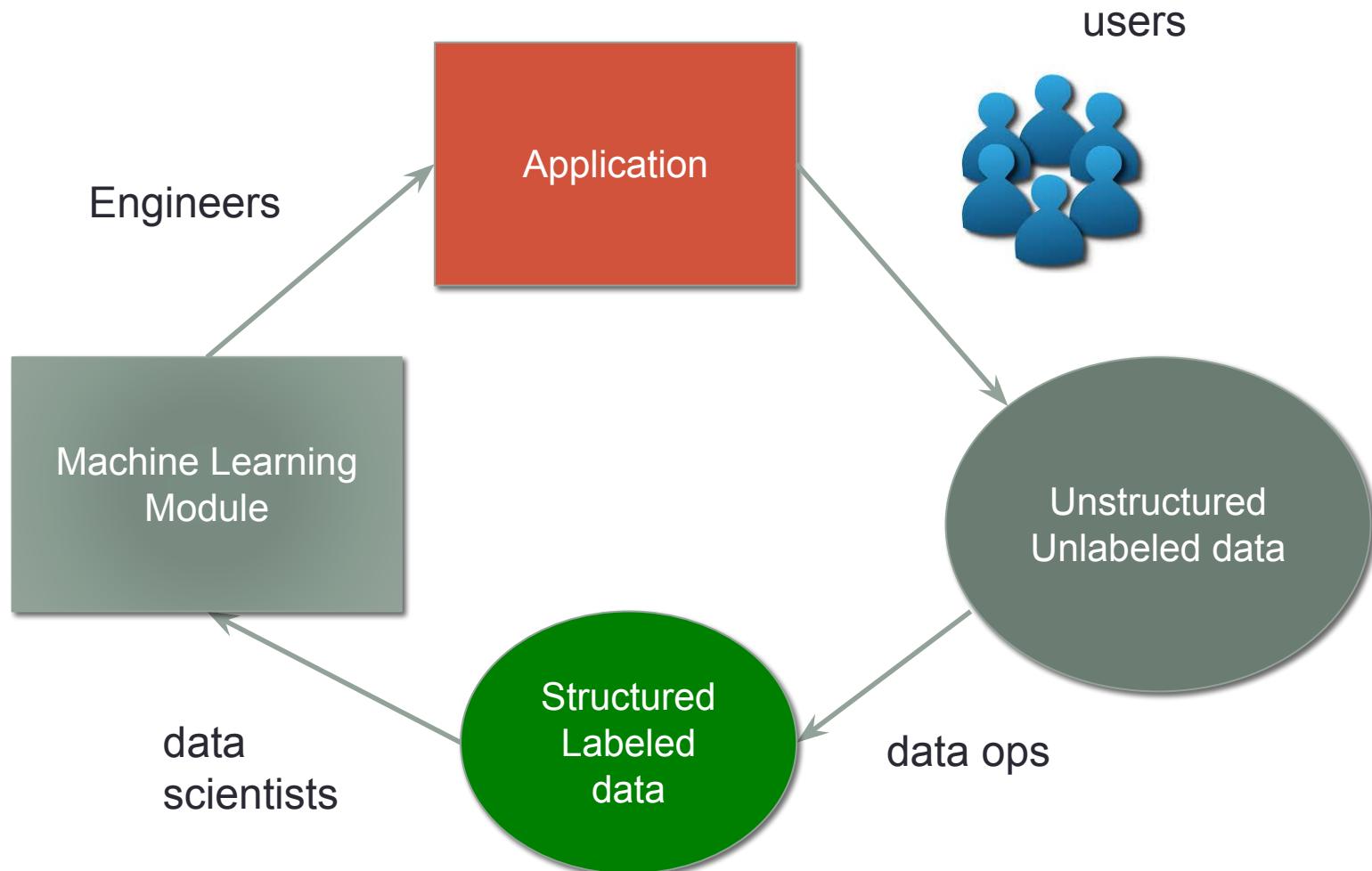
Why match dev/test data

- You tune on the dev set.
- Scenario: do well on dev set, but bad on test set
- If dev and test set comes from the same distributions
 - Diagnosis: you overfit on the dev set. Get more dev data
- If dev and test set comes from different distributions
 - 3 possibilities
 - Overfit on the dev set
 - Test set is harder than the dev set
 - Dev and test set are just different. And what works on dev might not work on the test set. Wasted effort.

Dev and test set and size

- Dev - tune hyperparameters, select features, and make other decisions regarding the learning algorithm.
 - Test - evaluate the performance of the algorithm, but not to make any decisions about regarding what learning algorithm or parameters to use.
-
- Dev – big enough to notice difference between algorithms (if you care about 0.1% difference, make sure you have enough dev set to spot it).
 - Test – large enough to give confidence that your model will do well in real task

Getting into the data cycle



Build and fix

- Aim high, but launch fast, and iterate
 - Aim high: business goal must be sound, dictates what data to collect
 - Launch fast: get users and more training data. Gets error feedback fast, so you can diagnose and plan your efforts
 - Has a proper pipeline from start to finish to help fast iteration
- Train on public domain datasets for fast training data acquisition (but dev/test must be real cases)
- Sometimes simple models are usually enough to launch
 - Logistic regression
 - Rule-based
 - Decision trees

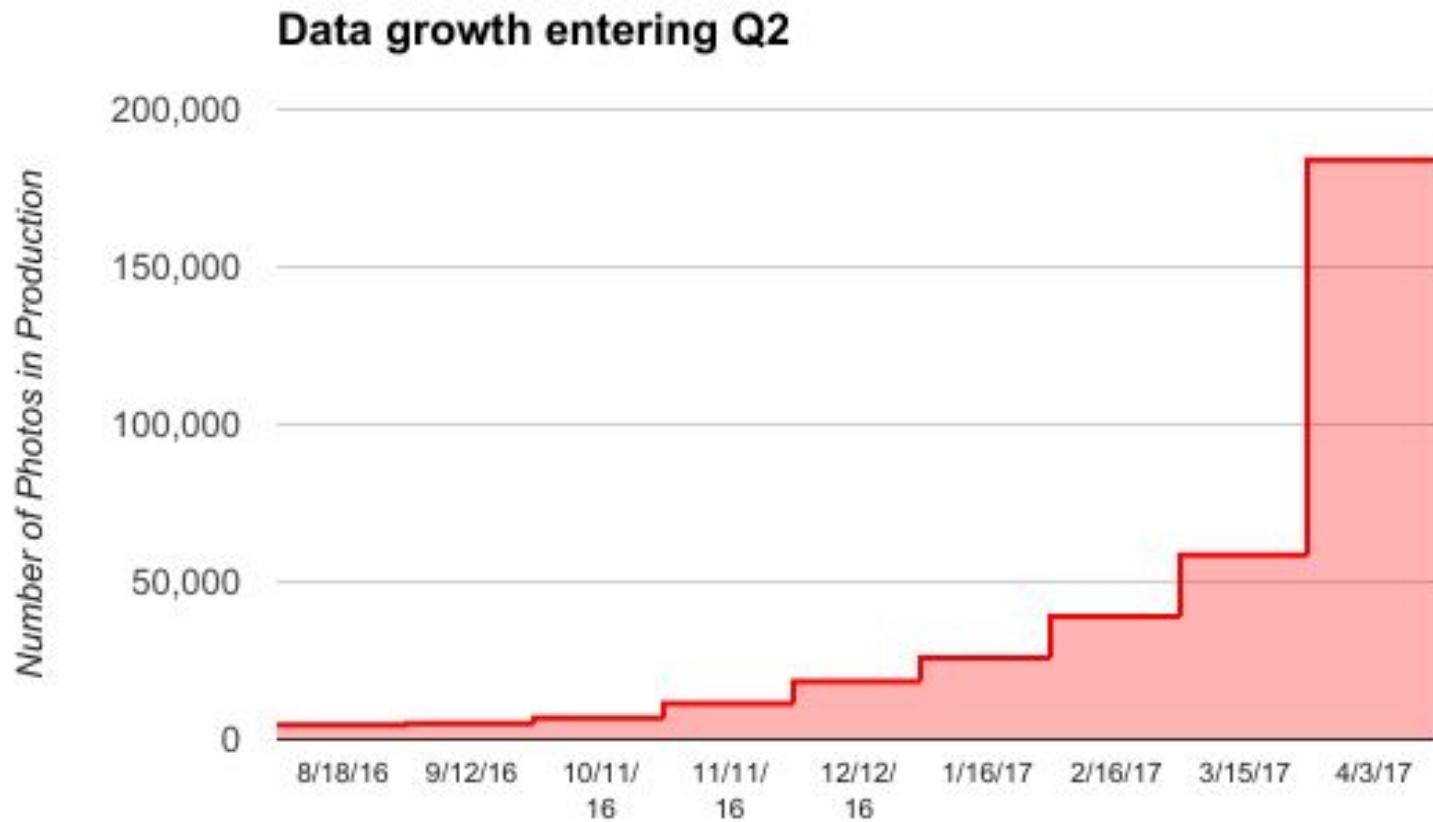
Example: gmail attachment

Starts with simple rule-based



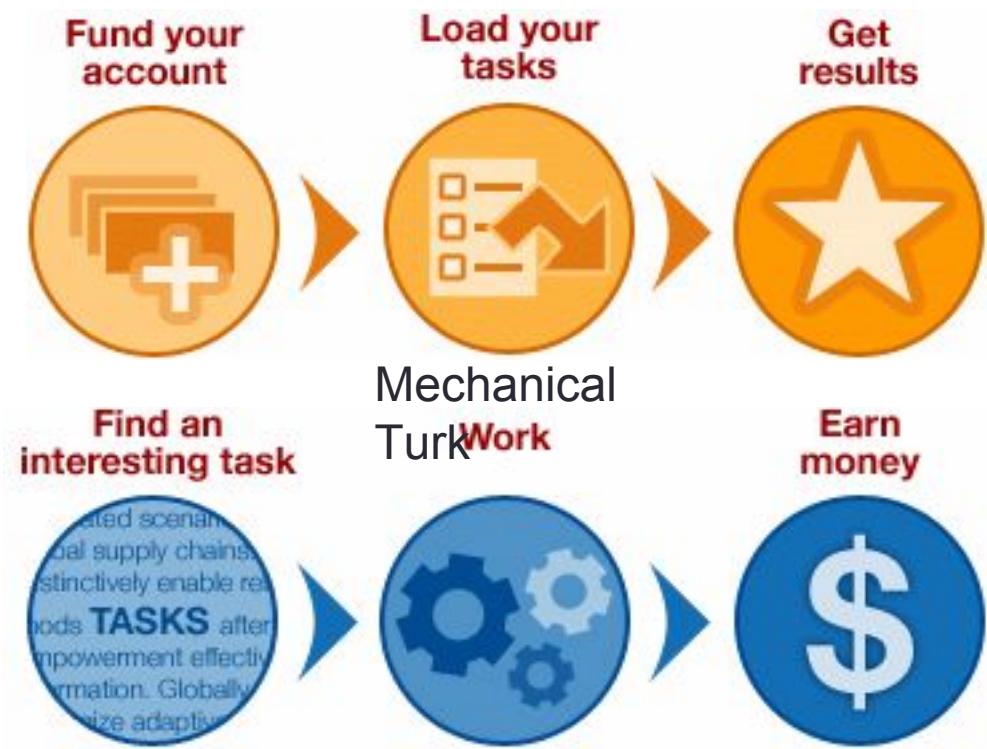
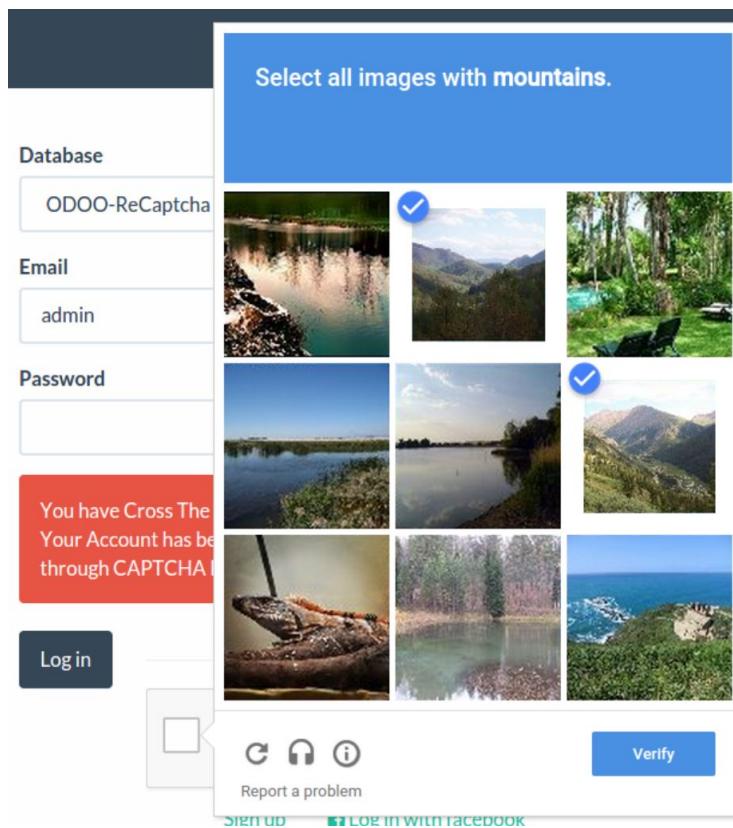
<https://www.technospot.net/blogs/gmail-doesnt-allow-you-to-forget-attachment/>

Getting into the data cycle



Crowdsourcing

- Distribute data labeling to the general public
- Need some incentives



Mturk Example



Is receipt valid?

Yes

Business Name:

Payment Method:

✓ Choose below...

CASH

DEBIT

CREDIT - AMERICAN EXPRESS (AMEX)

CREDIT - DISCOVER

CREDIT - MASTERCARD (MCARD)

CREDIT - VISA (VS)

CREDIT - NOT SPECIFIED

FOOD STAMP

EBT

SNAP

WIC

CHECK CARD

CHECK

GIFT CARD

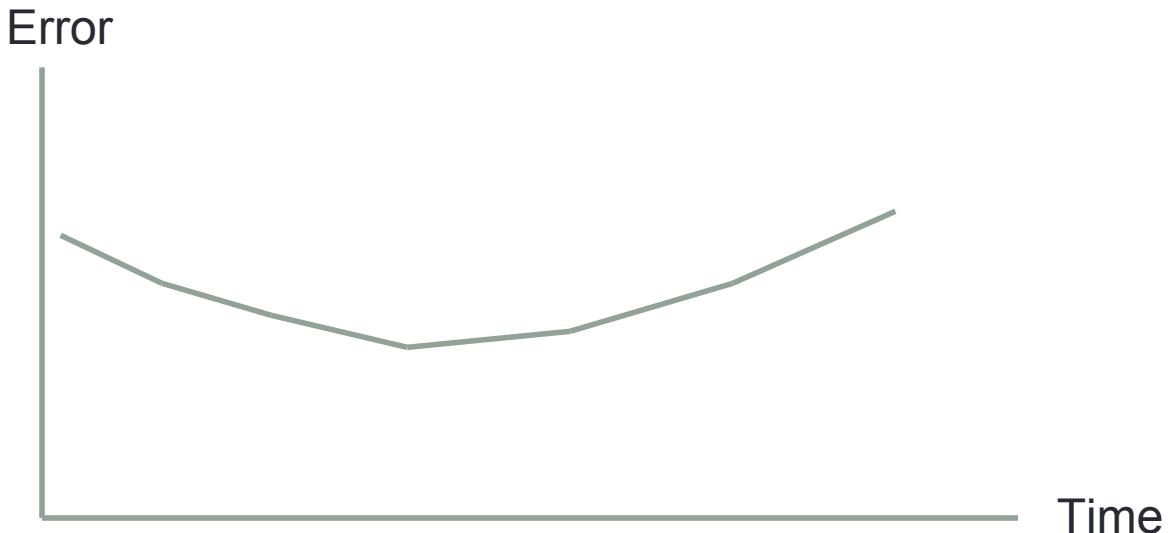
MONEY CARD

- Payment method not in above list

- No payment method is on receipt

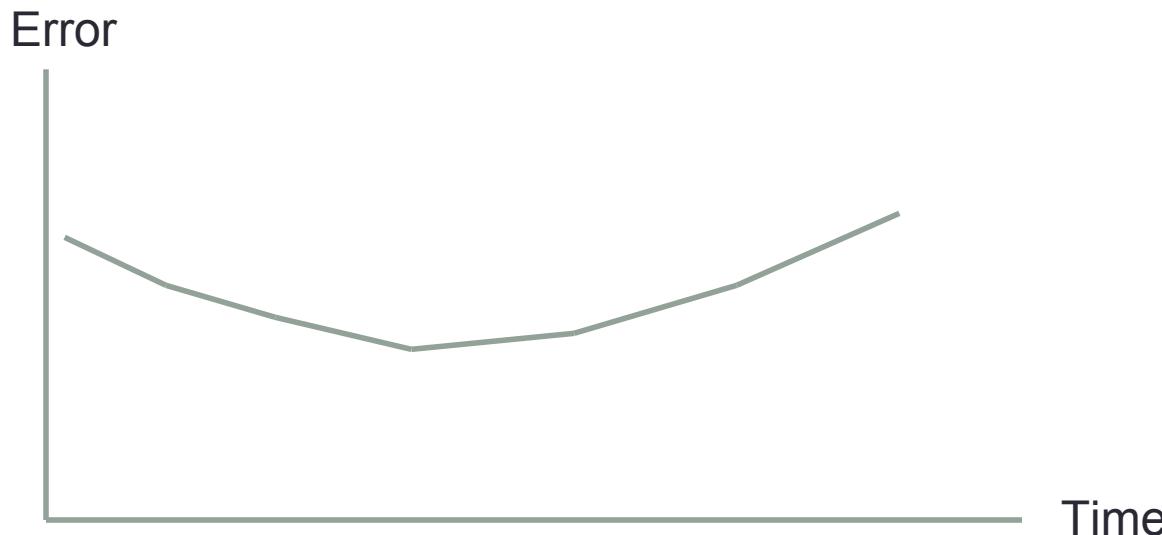
The rolling metric and datasets

- Example: You launched your cat classifier model
 - got into the data loop
 - used the new training data to train new models every month
 - then you noticed the test error keeps going up
 - What happened?



The rolling metric and datasets

- Your test no longer aligns with current usage
 - New phone cameras
 - New breed of cats(???)
 - New camera filters



The rolling dataset and metrics

- Train/dev/set must change over time to accommodate the change in usage
 - Example:
 - Test/dev: last month usage
 - Train: last year usage
- The metrics can also change. Don't cling on the same metrics if your usage no longer aligns. This will waste your team efforts
 - New hardware, new requirements



Team

Build the team

The team

- Data science now require a diverse skill sets. Sometimes, multiple roles can be covered by one person
- Data scientists (modeling)
- Data ops/engineers (deployment, data ETL)
- Performance engineers (optimization)
- Domain expert (help guide modeling)
- Labeling team
- Compute

Summary

- Modeling: Picking a starting point
 - Lots of model. Pick depending on the task.
 - Always good to consult experts/papers
 - Go with simpler models first then do diagnosis
- Diagnosis: Guide your efforts efficiently when iterating
 - Bias-variance
 - Error analysis

Summary

- Be a good communicator
 - Find potential tasks that can benefit from data science
 - You will work as a team with non-data scientists
 - Practice by writing blogs/facebook posts explaining data science.

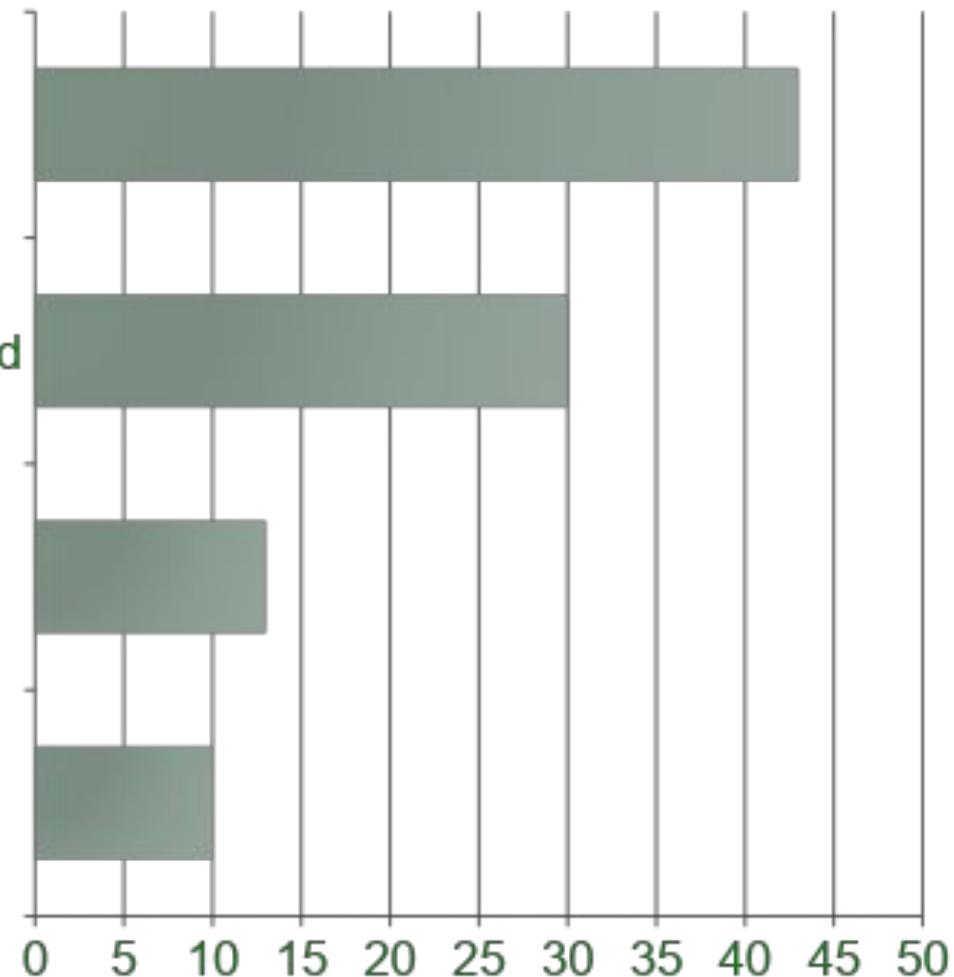
Common misconceptions business people have

Wrong expectation of capabilities

Underestimating resources needed

Technical misunderstanding

Not understanding what AI is



Further readings

Rules of Machine Learning: Best Practices for ML Engineering

http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdf

Rules of Machine Learning: Best Practices for ML Engineering

Martin Zinkevich

This document is intended to help those with a basic knowledge of machine learning get the benefit of best practices in machine learning from around Google. It presents a style for machine learning, similar to the Google C++ Style Guide and other popular guides to practical programming. If you have taken a class in machine learning, or built or worked on a machine-learned model, then you have the necessary background to read this document.

[Rule #1: Don't be afraid to launch a product without machine learning.](#)

Ethics issues

- Abuse
- Bias

Different Levels In a Self Driving Car



LEVEL 0



There are no autonomous features.

LEVEL 1



These cars can handle one task at a time, like automatic braking.

LEVEL 2



These cars would have at least two automated functions.

LEVEL 3



These cars handle "dynamic driving tasks" but might still need intervention.

LEVEL 4



These cars are officially driverless in certain environments.

LEVEL 5



These cars can operate entirely on their own without any driver presence.

Tesla is updating Autopilot's 'Hold Steering Wheel' alert after complaints, says Elon Musk

Fred Lambert - Jun. 13th 2018 6:12 am ET  @FredericLambert



<https://electrek.co/2018/06/13/tesla-autopilot-hold-steering-wheel-alerts-complaints/>

'The Business of War': Google Employees Protest Work for the Pentagon



Thousands of Google employees have signed a letter to Sundar Pichai, the company's chief executive, protesting Google's role in a program that could be used to improve drone strike targeting.

Michael Short/Bloomberg

<https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html>

Bias

- Our society is biased which leads to biased data and biased models
 - Understanding how models make prediction help us remove the bias
- Our morality is also biased
 - Understanding the cost-benefit tradeoff

Harms of Bias

Allocation

Models allocate resource to certain groups of people more - credit prediction

Representation

Models have more data of certain demographic and does better



<http://content.time.com/time/business/article/0,8599,1954643,00.html>

Study finds gender and skin-type bias in commercial artificial-intelligence systems

Examination of facial-analysis software shows error rate of 0.8 percent for light-skinned men, 34.7 percent for dark-skinned women.

 [Watch Video](#)

Larry Hardesty | MIT News Office
February 11, 2018

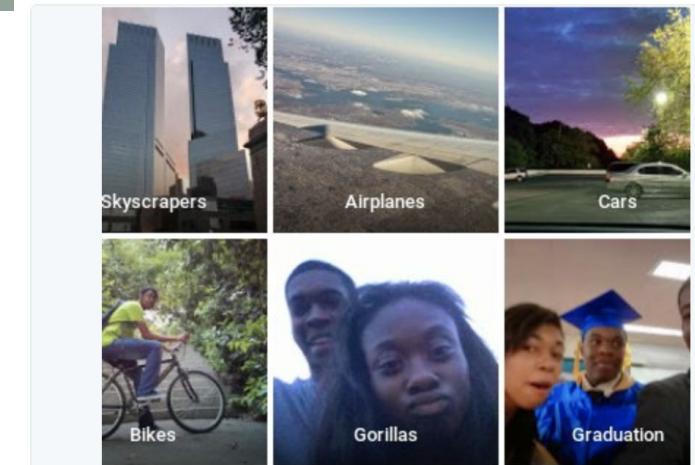
▼ Press Inquiries

PRESS MENTIONS

<http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>

Ways to combat bias

- Improve accuracy <- hard
- Blacklist
- Equal representation
- Awareness
- Test the system on general users



Jacky lives on @jalcine@playvicious.social now.
@jackyalcine



Google Photos, y'all fucked up. My friend's not a gorilla.
♡ 2,303 8:22 AM - Jun 29, 2015

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech

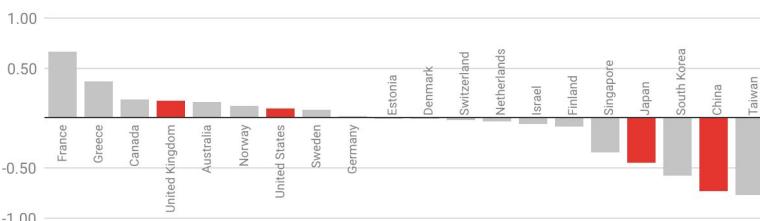
Nearly three years after the company was called out, it hasn't gone beyond a quick workaround

By James Vincent | @jjvincent | Jan 12, 2018, 10:35am EST

<https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai>

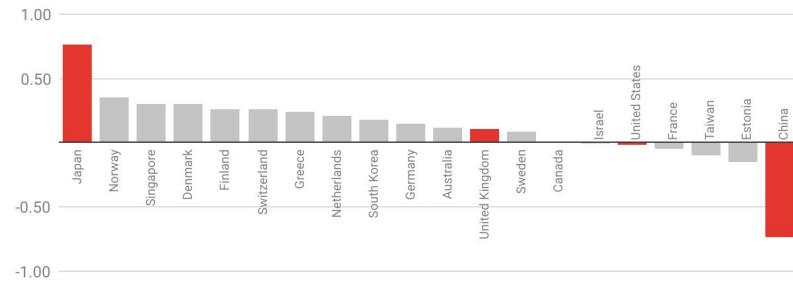
Moral machine

Countries with more individualistic cultures are more likely to spare the young



A comparison of countries piloting self-driving cars: If the bar is closer to 1, respondents placed a greater emphasis on sparing the young; if the bar is closer to -1, respondents placed a greater emphasis on sparing the old; 0 is the global average.

How countries compare in sparing pedestrians over passengers



If the bar is closer to 1, respondents placed a greater emphasis on sparing pedestrians; if the bar is closer to -1, respondents placed a greater emphasis on sparing passengers; 0 is the global average.

Problems with bias is both a social issue and a technological issue

- “If I were to guess like what **our biggest existential threat** is, it’s probably that. So we need to be very careful with the artificial intelligence. There should be some regulatory oversight maybe at the national and international level, just to make sure that we don’t do something very foolish.”



- “I think people who are naysayers and try to drum up these doomsday scenarios — I just, I don’t understand it. It’s really negative and in some ways I actually think it is pretty irresponsible”



Poll



Beyond this course

- Things you might want to pursue more
 - Graphical models
 - Stochastic process
 - Graph theory
 - Random forests and ensemble methods
 - Optimization

Automatic Speech Recognition

- Basics of human speech generation
- Hidden Markov Models
- Signal processing, Fourier Transforms (basis transforms)
- Connectionist Temporal Classification
- How to build and deploy your own speech application

ASR

DEMO



Natural language processing

- Conditional Random Fields
 - Basics of language understanding
 - Machine translation and Question Answering
 - Chatbots
-
- Take both for the best experience!

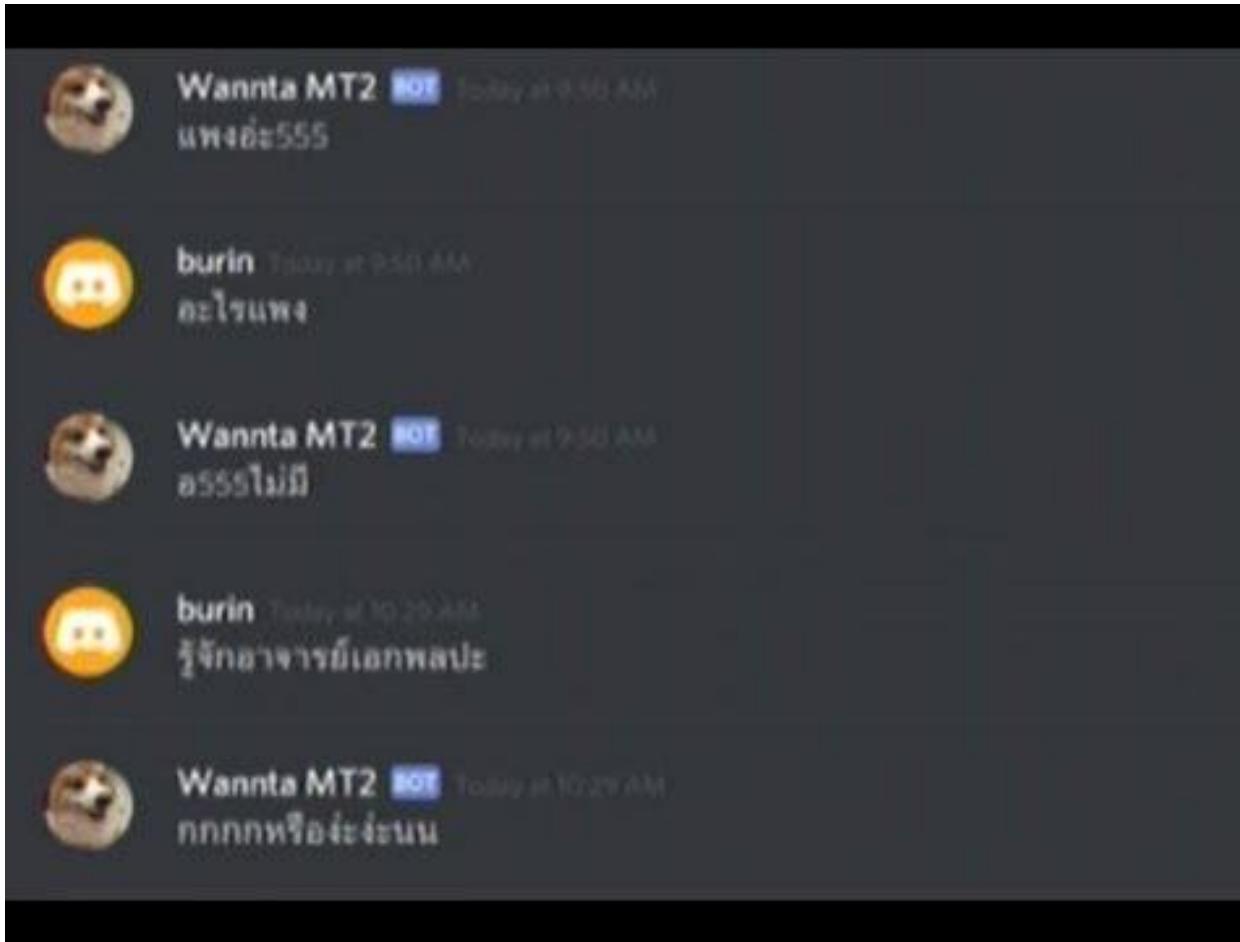
■ GitHub, Inc. [US] | <https://github.com/google-research/bert/blob/master/multilingual.md>

Data Source and Sampling

The languages chosen were the [top 100 languages with the largest Wikipedias](#). The entire Wikipedia dump for each language (excluding user and talk pages) was taken as the training data for each language

The only language which we had to unfortunately exclude was Thai, since it is the only language (other than Chinese) that does not use whitespace to delimit words, and it has too many characters-per-word to use character-based tokenization. Our WordPiece algorithm is quadratic with respect to the size of the input token so very long character strings do not work with it.

NLP



Course philosophy

- Pattern Recognition: understand, **build**, and use machine learning models
- ASR: understand, and use machine learning models for ASR
- NLP: understand, and use machine learning models for NLP