Using Machine Learning to Automatically Predict and Identify Defects in Automotive Assembly Processes

David Johnson

Foxcon 2017 Software Developer’s Conference

Outline

- Brief introduction to Machine Learning
  - Frequentist statistics
  - Bayesian statistics
- Machine Learning in practice
  - Precision vs. Recall
  - Classification vs. Regression
- Automotive Assembly Defects
  - Torque tool operations
  - Common defects and errors
  - Case study: simulated data from 49,000 vehicles

What is Machine Learning?

- **Machine Learning** (ML) is the subfield of computer science that gives computers the ability to learn without being explicitly programmed[1]
  - Using statistical analyses
  - Processing large amounts of data
  - Adapting without new programming

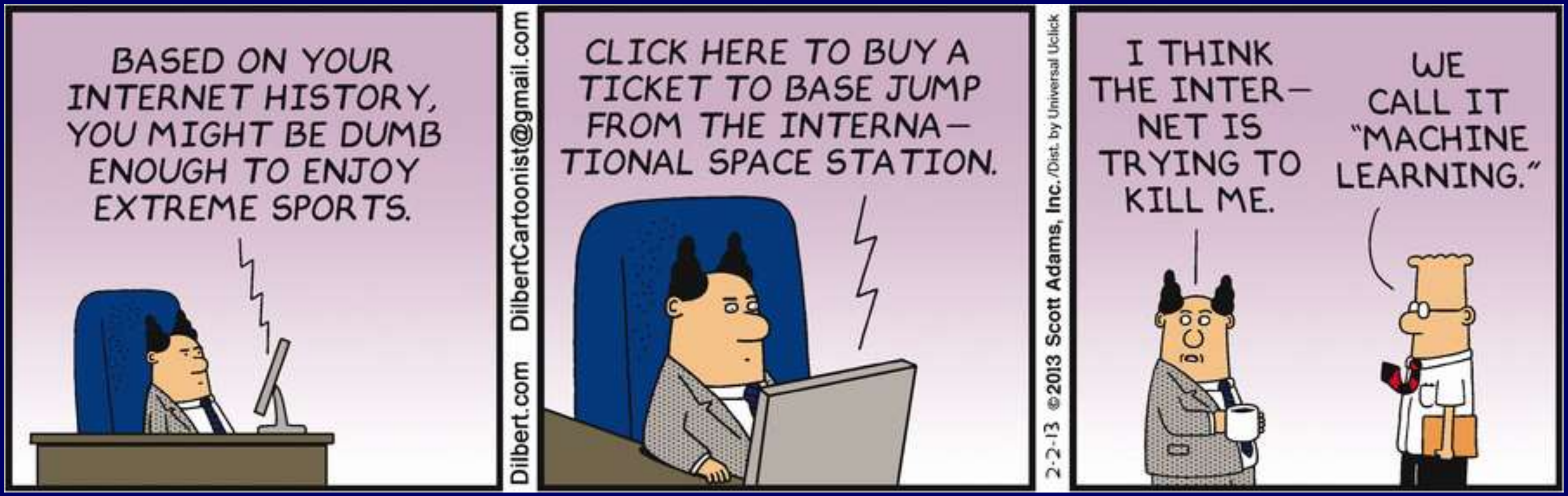
[1]Arthur Samuel, 1959. from Wikipedia

Machine Learning

- Using statistical analyses
  - Statistics 101 still applies:
    - Need a model, data, and an objective function
  - But prediction is more important than model validation for ML
- Processing large amounts of data
  - Since analysis is automated or semi-automated, more data is usually helpful
- Adapting without new programming
  - Unlike general purpose artificial intelligence, ML is data-driven

Examples of Machine Learning

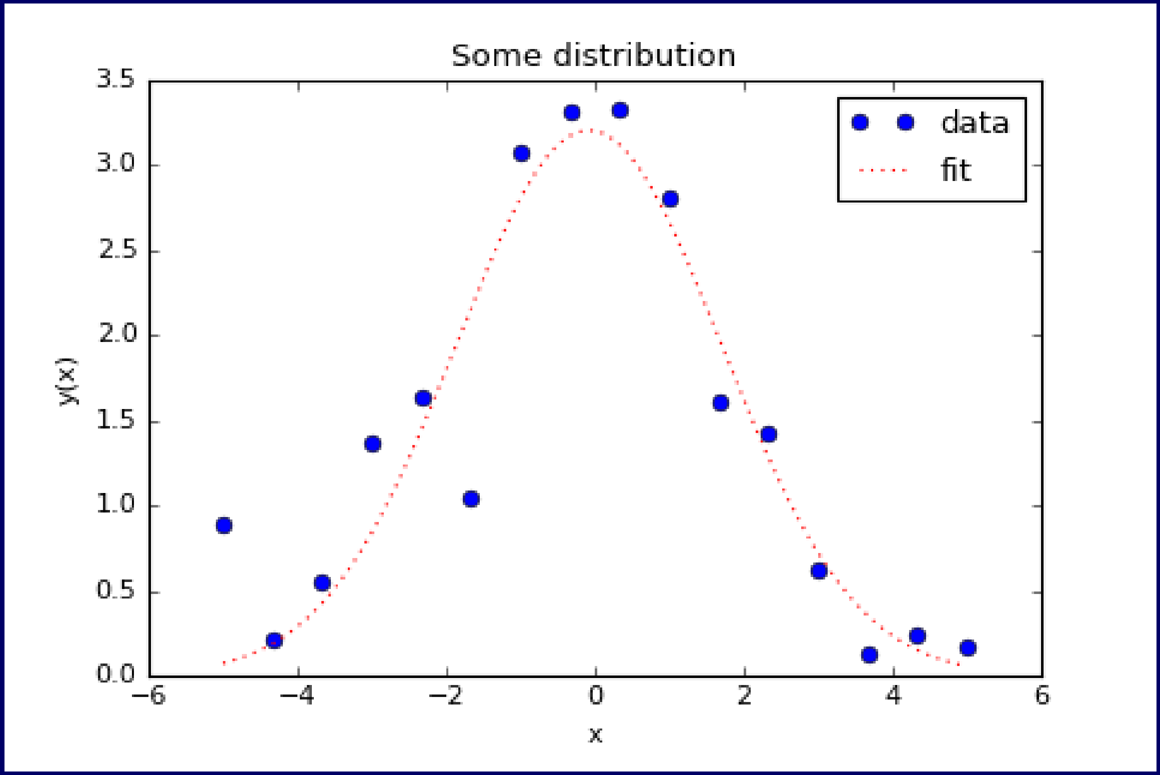
- Product suggestions
  - Amazon suggested products; Netflix similar films
- Cybersecurity
  - Automatically identifying malware based on actions and/or file signatures
- Job ads / HR recruiting
  - Linked In suggested jobs; automated resume processing
- Google AlphaGo
  - World champion of Go beaten 4-1.
- Criminal sentencing
  - Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)
- Tesla's autopilot system
  - Camera, radar, GPS, ultrasonic sensor => follow lanes, adjust speed



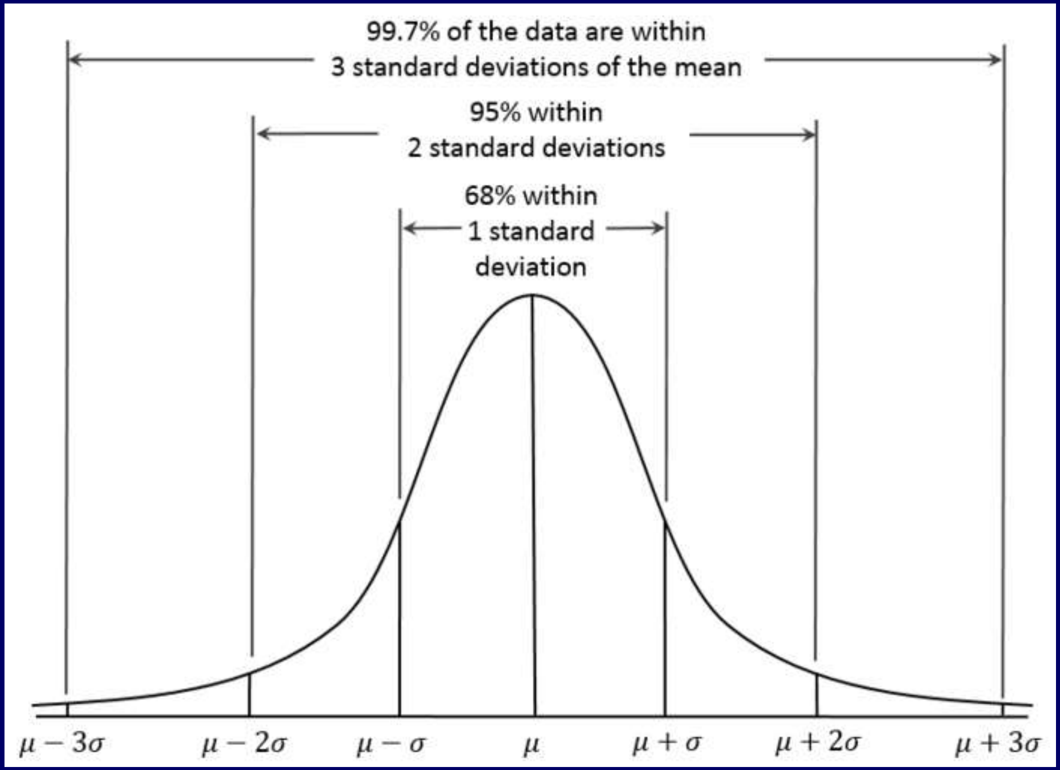
- Next: Frequentist statistics review

http://dilbert.com/strip/2013-02-02

Frequentist statistics



Gaussian

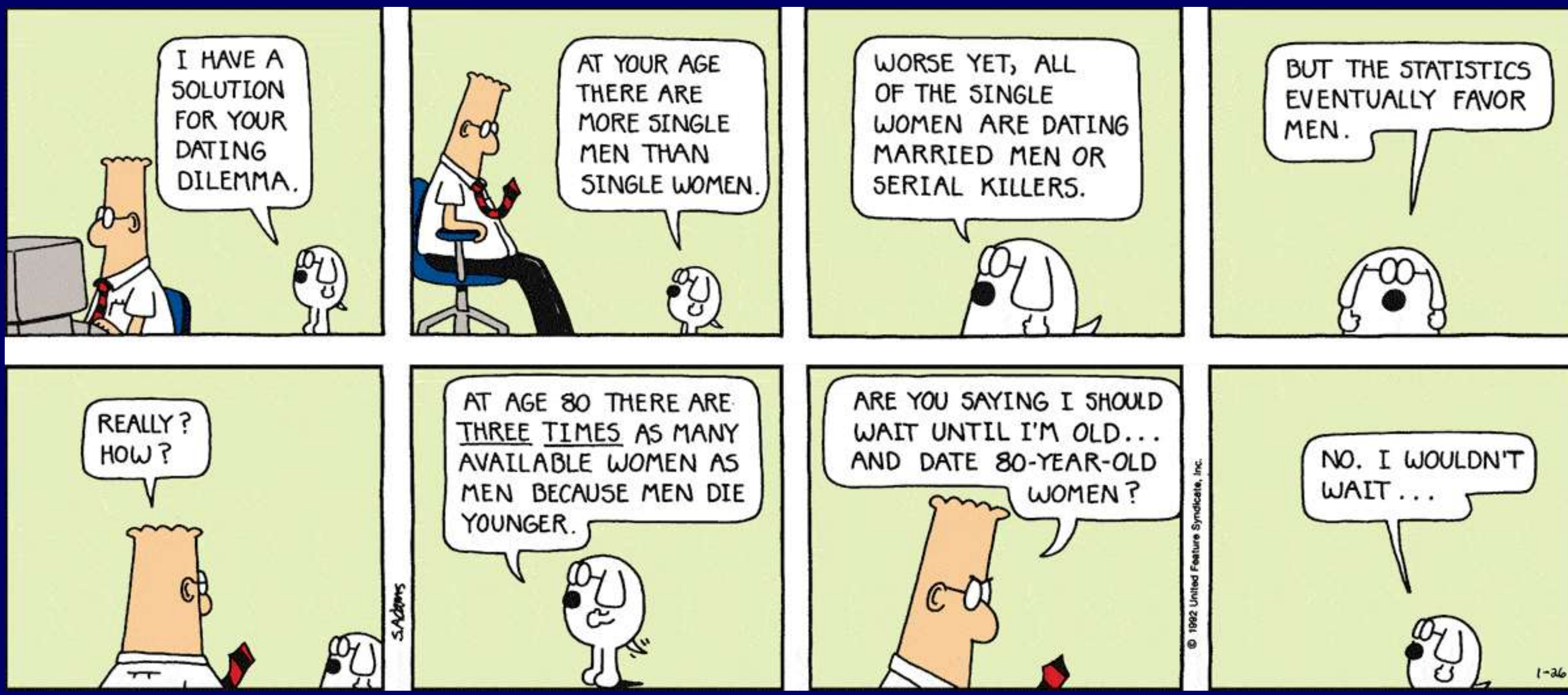


- Gaussian model: $\exp(-(x-\mu)^2/(2\sigma^2))$
  - Model: Only two parameters
  - Data: requires relatively few points for a fit (~ 10)
  - Objective function: goodness of fit (χ^2 test)

Figure from Wikipedia

Questions for the Frequentist

- Model validation
  - Why would you believe that this data was from a Gaussian distribution?
  - What would refute that belief?
  - How certain are the fitted parameters?
- Model prediction
  - How certain are new data points from the model?



http://dilbert.com/strip/1992-01-26

Bayesian statistics

Bayes' Theorem:

$$\underbrace{P(\Theta|X)}_{\text{posterior}} = \frac{\overset{\text{Joint distribution}}{P(X, \Theta)}}{\underbrace{P(X)}_{\text{data}}} = \frac{\overset{\text{likelihood}}{P(X|\Theta)}\ \overset{\text{prior}}{P(\Theta)}}{\underbrace{P(X)}_{\text{data}}}$$

- Bayes' Theorem combines prior beliefs and observed data to infer the posterior distribution
- Frequentist models are still used in the likelihood, but the joint distribution is new
- This allows us to answer the questions on the previous slide ("how certain …")

Bayesian Example - Lunch

lunch orders

File Edit View Insert Format Data Tools Add-ons Help Last edit was yesterday at 10:14 PM

	A	B	C	D	E	F	G	H
1	_restaurants	times chosen	last picked by	last picked on	days since picked	P(Restaurant)	P(Jon|Restaura	P(Restaurant|Jon)
2	_ Sort	Sort	Sort	Sort	Sort	Sort	jo Sort	Sort 3%
3	Ideal Hotdog	1	jon	2/18/2016	344	0.2%	100.0%	0.7%
4	penn station	16	jon	1/9/2017	18	3.2%	87.5%	10.4%
5	chubbys	24	jon	11/28/2016	60	4.7%	75.0%	13.3%
6	rice blvd	6	jon	8/25/2016	155	1.2%	66.7%	3.0%
7	Pita Pit	3	jon	12/28/2016	30	0.6%	66.7%	1.5%
8	ya halla	7	jon	4/21/2016	281	1.4%	57.1%	3.0%

- What should we get for lunch?
- Where are we likely to choose?

Bayes' Lunch

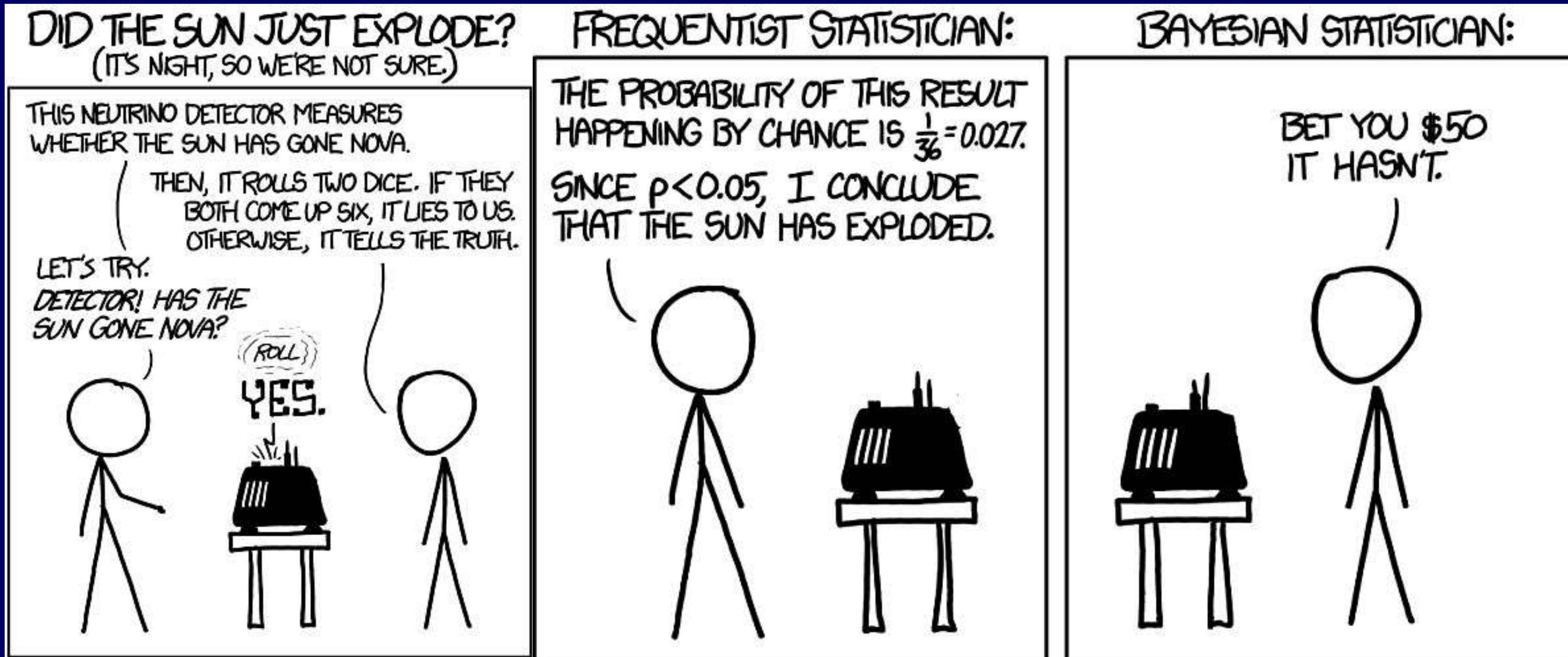
$$\underset{\text{posterior}}{P(R=\text{Restaurant} \mid N=\text{Name})} = \frac{\overset{\text{Joint distribution}}{P(R, N)}}{P(N)} = \frac{\overset{\text{likelihood}}{P(N|R)}\ \overset{\text{prior}}{P(R)}}{P(N)}$$

- Using Bayes' Theorem, we can predict the Restaurant (R) given the Name (N) of the person whose turn it is
- Maximizing P(R|N) is a common algorithm
- Non-parametric; derived entirely from spreadsheet.

Statistics Summary

- Frequentist statistics focuses on model evaluation, assuming parameters are deterministic
- Bayesian statistics uses prior and posterior probabilities to quantify the uncertainties of both the model and the data
- Both are still relevant, but they require a statistician to formulate and evaluate models

Frequentist vs. Bayesian XKCD



https://xkcd.com/1132/

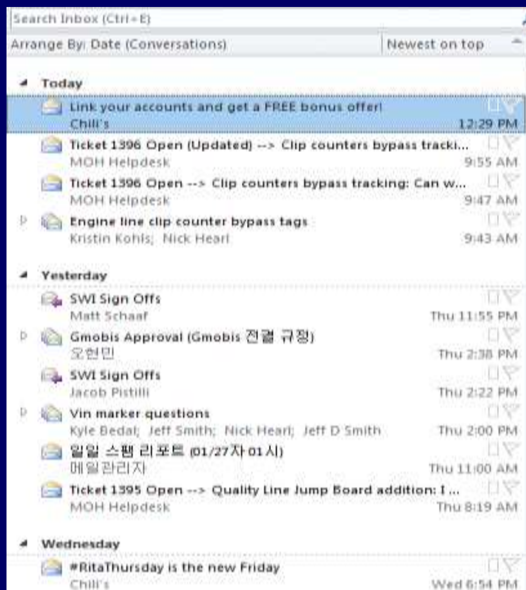
Machine Learning in practice

- Email spam filtering
  - What is the probability of each word in a dictionary appearing in a spam email vs. a non-spam email?
  - Using Bayes' Theorem, infer posterior probability, mark spam if P(spam) > cutoff (e.g., 90%)
- What goes wrong if the wrong decision is made?
  - Spam marked as non-spam
  - Non-spam marked as spam

Email spam

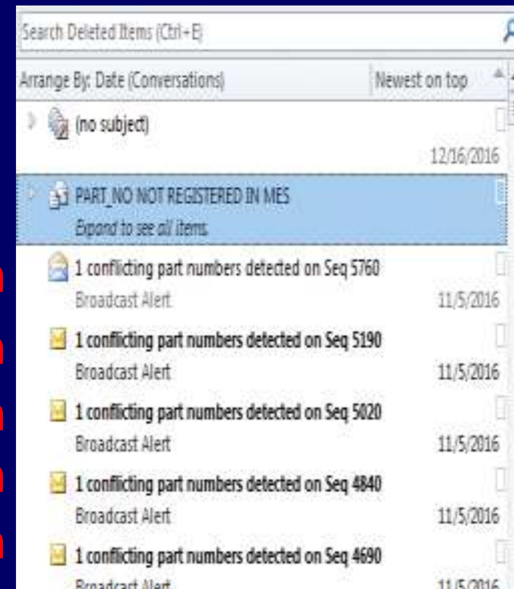
Inbox

Spam
Not spam
Not spam
Not spam
Not spam
Not spam
Not spam
Spam



Spam folder

Not spam
Not spam
Spam
Spam
Spam
Spam
Spam



- **Keywords that identify non-spam: P(Engine| non-spam) = 0.70, P(VIN| non-spam)=0.58, …**
- **Keywords that identify spam: P(Broadcast Alert|spam)=0.89, …**
- **Naive Bayesian classifier:**

$$\mathbf{P(Spam \mid K_1, \ldots K_n) = P(Spam)\ \Pi\ P(K_i \mid Spam)}$$

Email spam

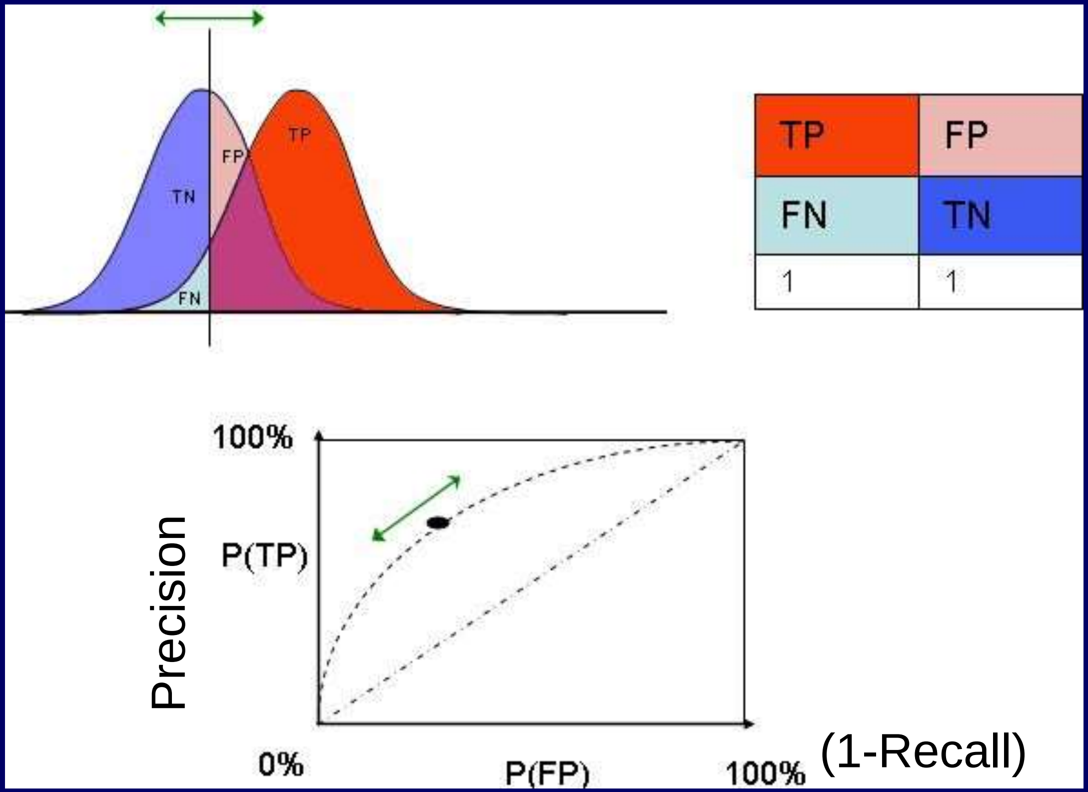
	Predicted: Not spam	**Predicted: Spam**	**Totals**
Inbox	95 (TP)	5 (FN)	100
Spam	1 (FP)	99 (TN)	100
Totals	96	104	200

Recall (sensitivity) = TP/(TP+FN) = 0.95

Precision (positive predictive value) = TP/(TP+FP) = 0.99

- Classification algorithms aren’t perfect
- Is FP worse than FN? Always?

Precision vs. Recall tradeoff



- ROC curve: the relative errors can be compared by adjusting the parameters of the algorithm
- E.g., consider more words to be spam -> better recall, worse precision

https://en.wikipedia.org/wiki/Receiver_operating_characteristic

Classification vs. Regression

- Model output type makes important differences to the algorithms available
- **<u>Classification</u>**: the model output is a categorical variable with discrete values
  - E.g., labels, attributes, colors, statuses, 1st, 2nd, 3rd, etc.
- **<u>Regression</u>**: the model output is a continuous variable
  - E.g., measurements, sizes, physical values

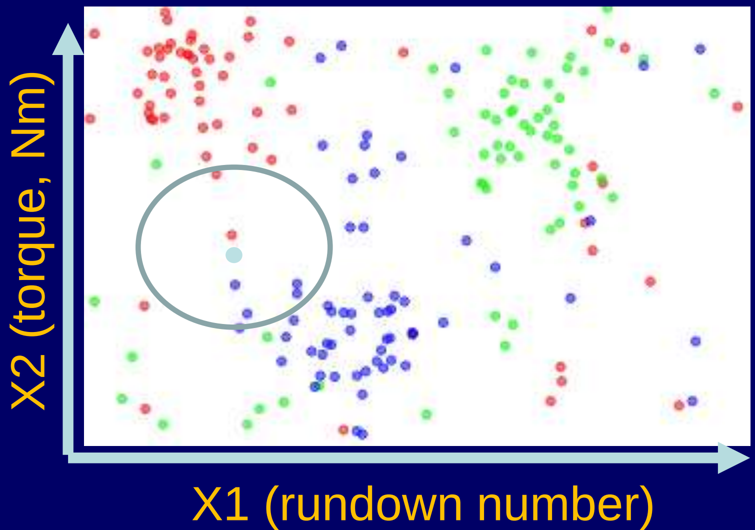
Examples of outputs

- Classification problems:
  - Predict products that a consumer might want to buy
  - Predict who will vote for a given candidate
  - Identify ZIP codes from handwritten envelopes
- Regression problems:
  - Predict stock prices based on company performance
  - Predict chances of a patient having a second heart attack
  - Identify sources of cancer risk from clinical prostate samples
  - Estimate time to failure for a piece of industrial equipment

Machine Learning Algorithms

- k-Nearest Neighbors (k-NN):
  - The oldest **classification** algorithm
  - Successful due to simplicity
- Linear regression:
  - The oldest **regression** algorithm
  - Surprisingly flexible with generalized linear models
- Many other algorithms exist

k-Nearest Neighbors

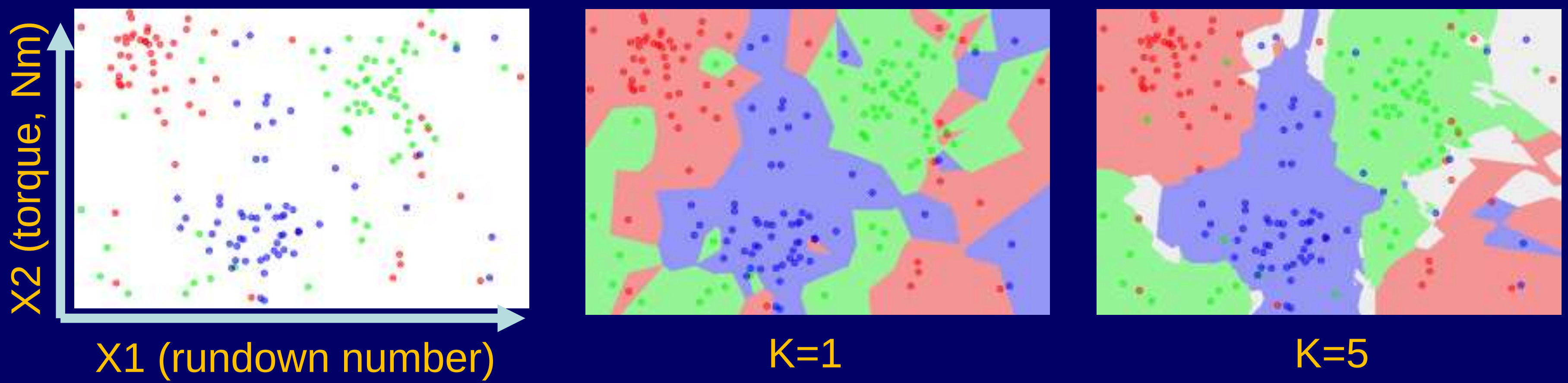


Error Types:
- No error
- Trigger loss
- Cross-threaded

- Suppose you wanted to predict what type of error will occur from the features of rundown number (1, 2, 3…) and torque value (e.g., 10 Nm)
- When you get a new point at •, which error is most likely?
- Suppose k=3. 3 nearest points are: Trigger loss, No error, No error
- Majority vote: No error

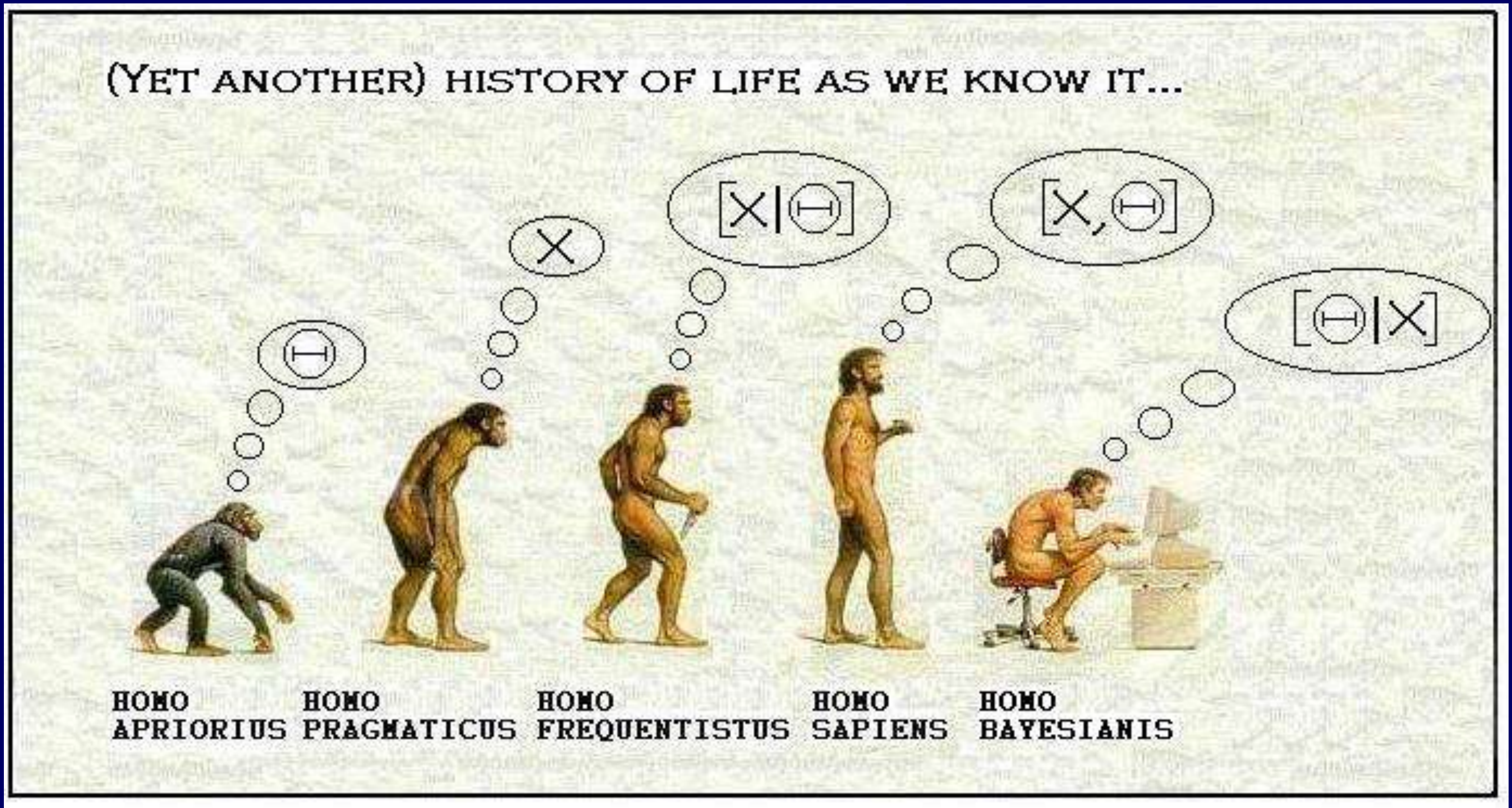
https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

k-Nearest Neighbors (2)



- Predict all the points!
- Practical limitations: can’t use all the data due to curse of dimensionality, so use dimensionality reduction preprocessing or representative data sub-sampling
- How do you pick k? What does it mean?

https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm



(Θ is all of the “known” parameters; x is all of the observed data)

http://www2.stat.duke.edu/~mw/fineart.html

http://www.wranglerforum.com/
Jan 2013 Winner “There I Fixed It”

Automotive Assembly - Torque

- Video – Atlas Copco Electric nutrunner

https://youtu.be/4an9H6VTxVc

Torque tool operations

- Normal mode
  - Torque is inside engineering range (min, max)
  - Angle is inside engineering range (min, max)
  - Duration is acceptable
- Failure modes
  - Failed to reach min torque or angle
  - Exceeded maximum torque or angle
  - Operator running behind

Common defects and errors

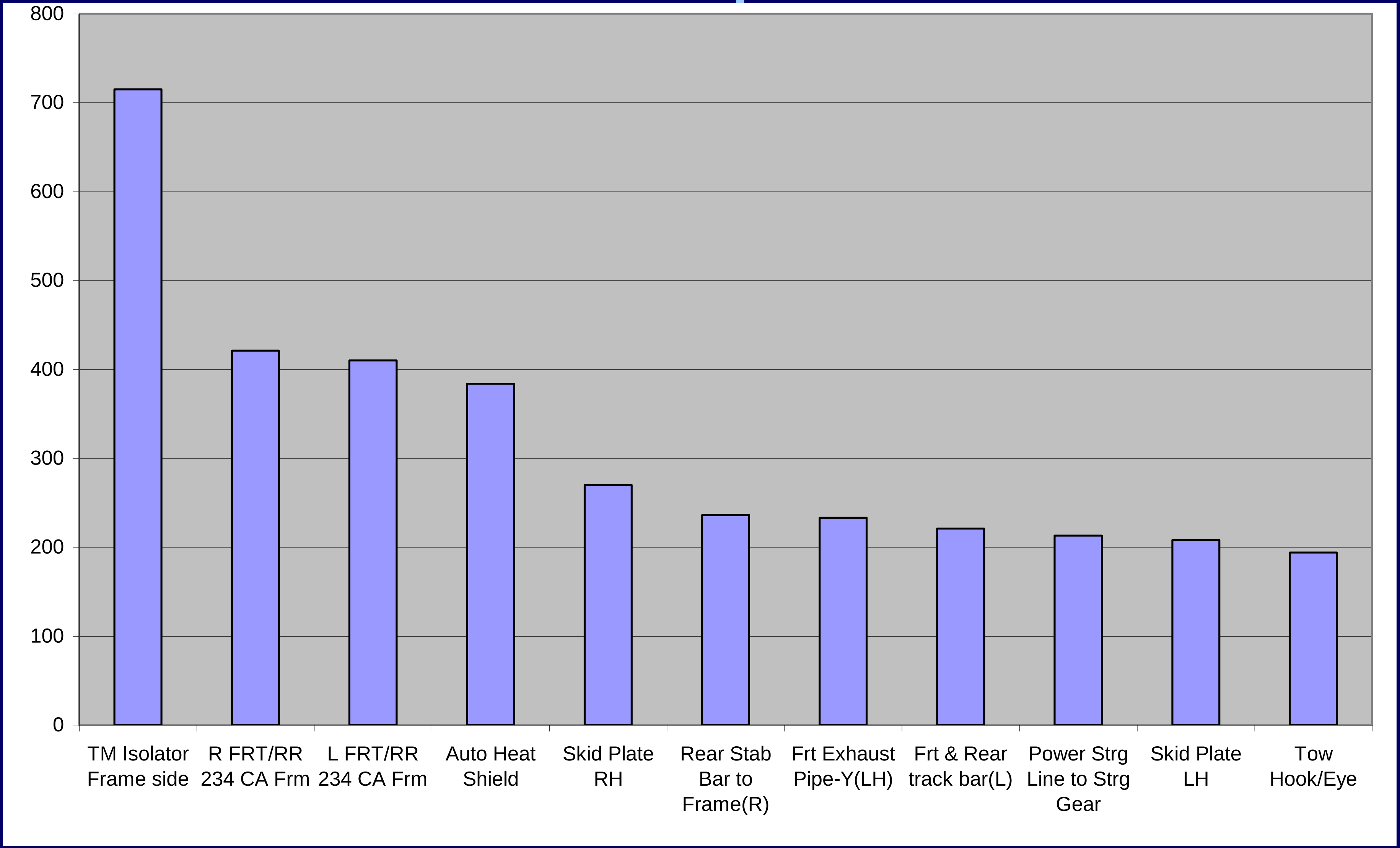
- Trigger loss
  - The operator let go of the trigger too soon
- Wrong number of torques:
  - E.g., Fuel tank has 4 bolts, so 4 torques required
  - Operator only got 3 done before running out of time
- Part is wrong or defective
- Cross threading
  - The nut slipped or was incorrectly loaded
- Electrical issues
  - Power failure
  - Ethernet failure
- Tool breakdown (calibration or mechanical)

- 8 Production lines, ~250 operators, ~400 vehicles per shift

Case study

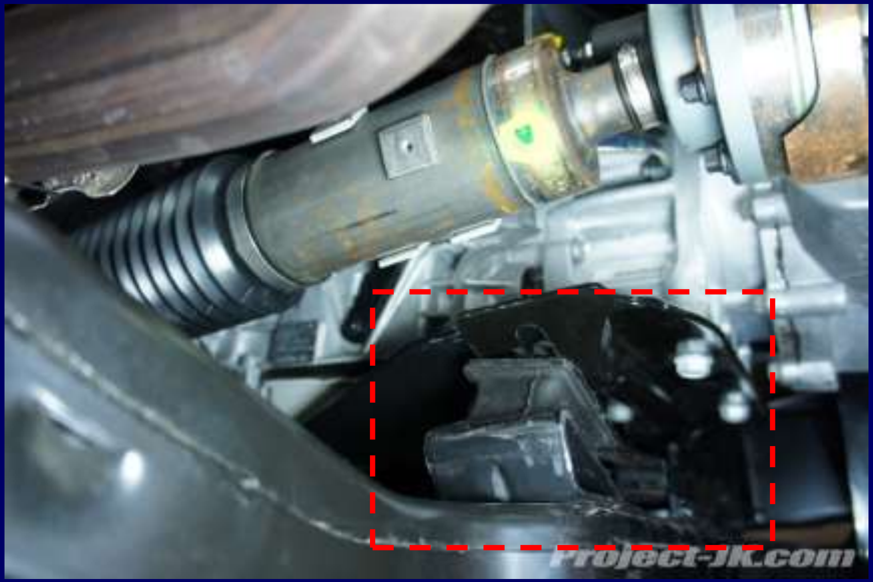
- Data from a preliminary 3 month study:
  - 49,000 vehicles
  - 180 torque tools
  - 4.37M rundowns (4.35M first time successes)
  - 8,500 failures on 7,000 distinct VINs
- Approximate failure rate: 0.0019 failed torques per required rundown
- Due to confidentiality concerns, the data has been generated from a simulation

Worst torque tools



- 4.37M rundowns; 8,500 errors total

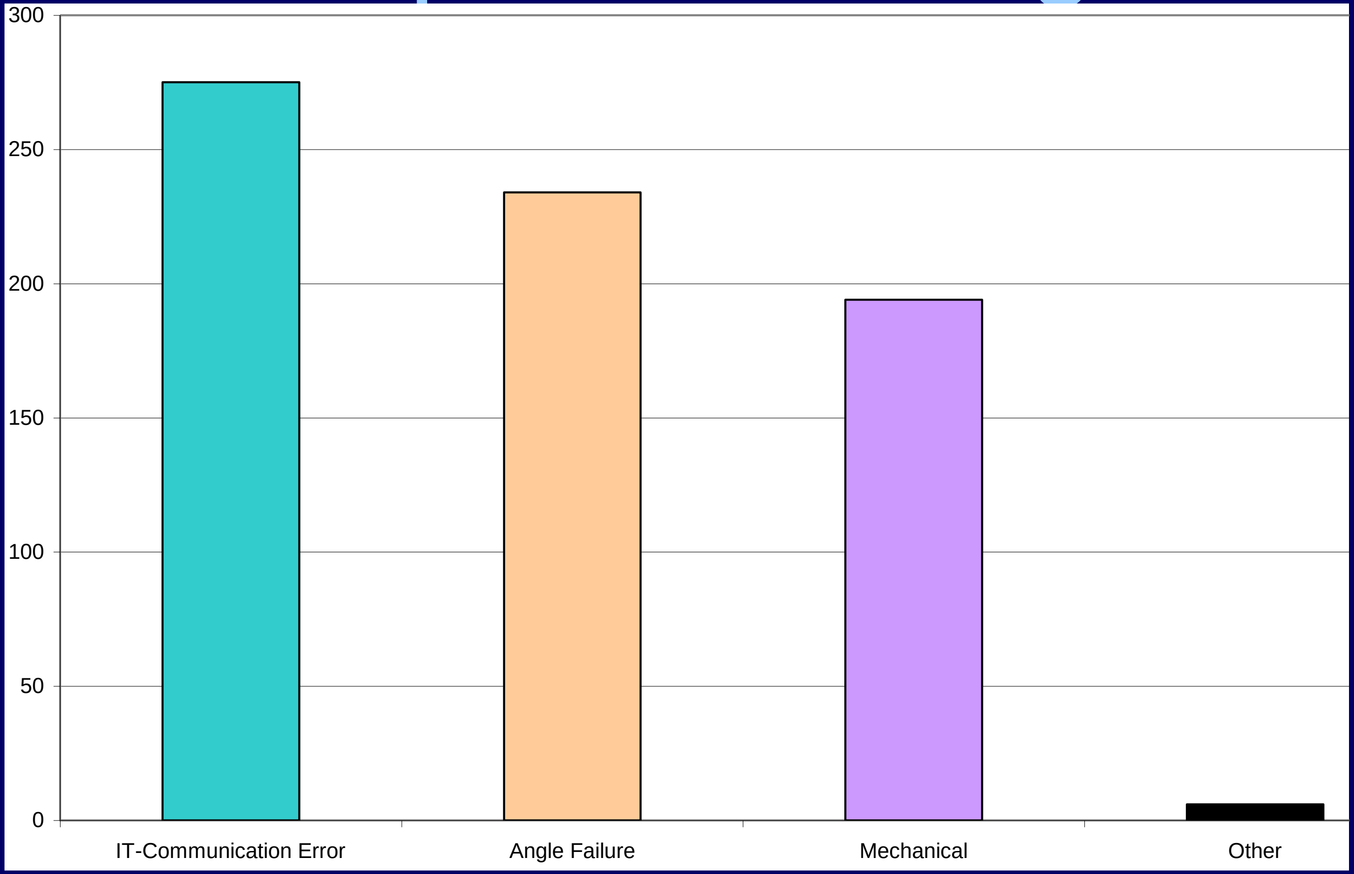
What's wrong with TM Isolator?



- Transmission isolator – fully automated torque robot.
- Only 3 torques, 40 – 70 Nm (from repair manual)
- Why does this torque tool fail so often?

http://project-jk.com/jeep-jk-write-ups/

RepairTech Log



- Upon further investigation, the ethernet communication between the robot and the torque tool was found to be faulty (replacement pending)
- Angle failures are due to rubber / steel nut interface

Predicting failures

- Available features in the model:
  - All part numbers
  - All torque values (torque, angle, OK/NG)
  - All sales codes (export nations)
  - RHD vs LHD, manual vs. auto trans., gas vs diesel
  - Number of rundowns, last calibration, etc.
- Desired outputs:
  - Time to failure on torque tools
  - Probability of requiring jumps for each vehicle
  - Predict type of repairs given vehicle information
- Still a work in progress (unbalanced data)

Preventative Maintenance

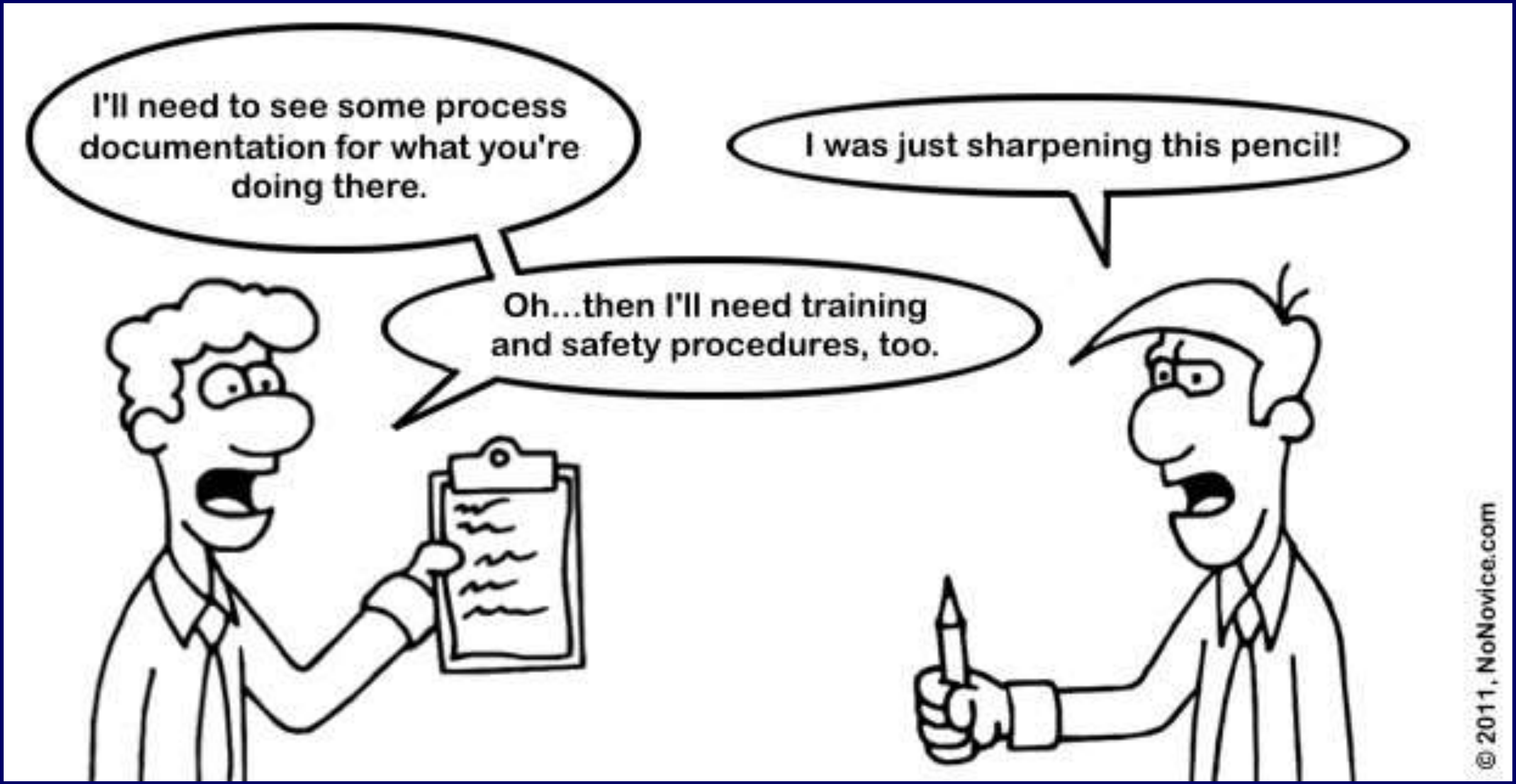
- Current maintenance schedule is fixed
  - E.g., every month, tools X, Y, and Z must be calibrated
- Proposed:
  - Predict time to fail based on actual usage
  - Schedule maintenance based on failures
- Probable predictors:
  - Last date of calibration
  - Total rundowns since calibration
  - Min, max torque
  - Drifting residuals

Unacceptable Maintenance Schedule

http://piximus.net/fun/there-i-fixed-it-9

Auditing

- Manual audits are used to intentionally introduce errors and verify that the production line stops and produces alarms as intended
  - LPA (Layered process audit)
  - EPV (Error proofing validation)
- Scheduling is fixed
  - Every week, stations A-P are audited, then Q-Z, etc.
- Current problem: pencil whipping
- Proposed solution: schedule audits based on failures



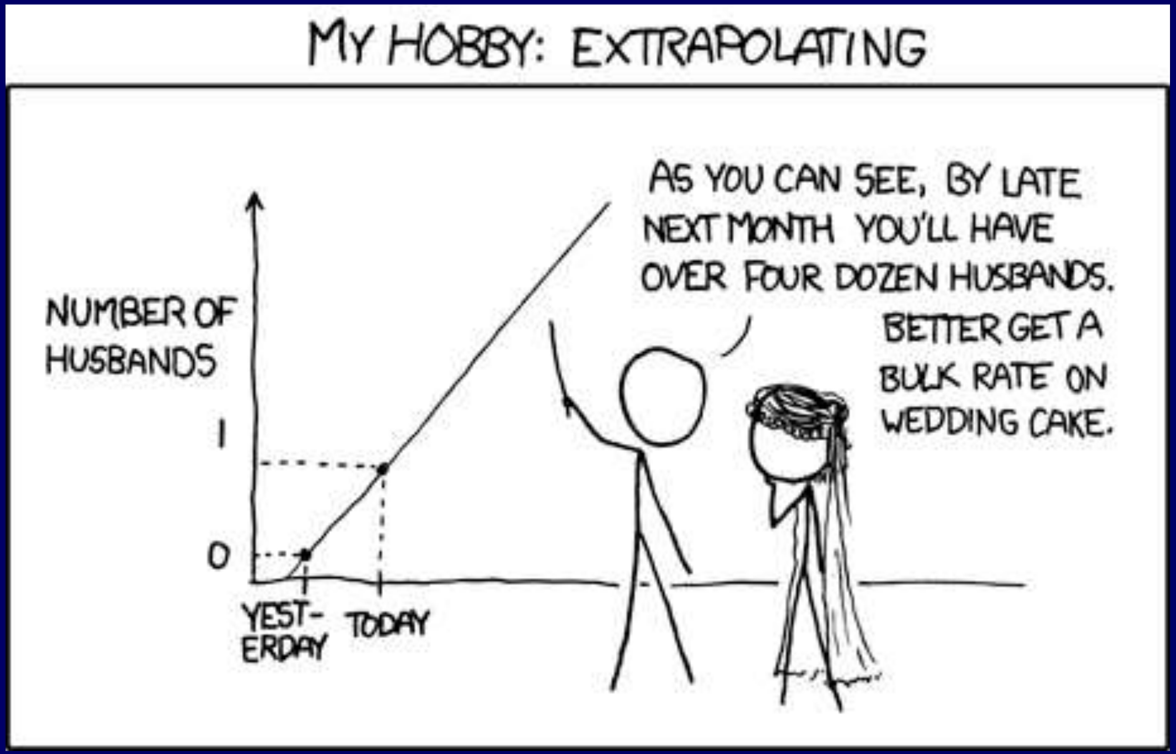
http://auditstuff.com/audit-jokes/

Future work

- Goals for the future:
  - Get python sklearn to work in production
  - Automate the analysis
  - Do a trial run with live data
  - Schedule audits and maintenance based on model, then compare failure rates to similar interval
- Problems right now:
  - Volume of non-predictive data
  - False positives
  - Overfitting and unbalanced data

Overall Conclusions

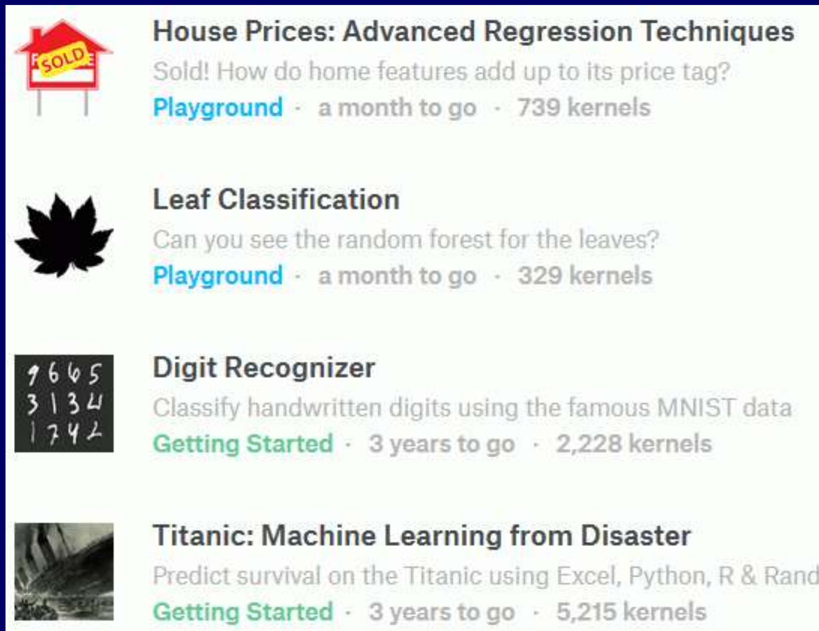
- Machine learning is powerful
  - Convert existing large datasets into predictions
  - Semi-automated or automated analysis
  - Wide range of applications
- Limitations
  - Only works if the future looks like the past
  - Not a general purpose AI
  - Not always better than traditional statistics



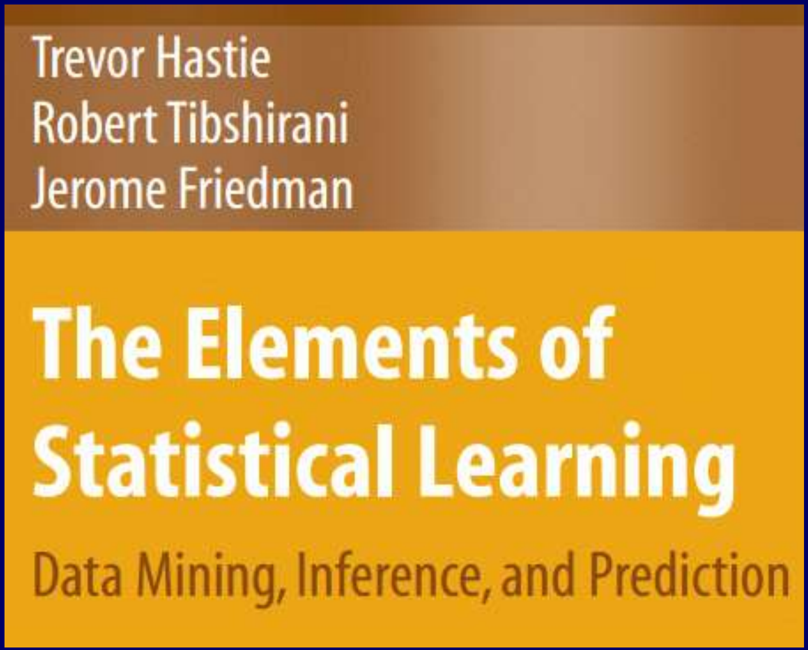
By the third trimester, there will be hundreds of babies inside you.

https://xkcd.com/605/

References

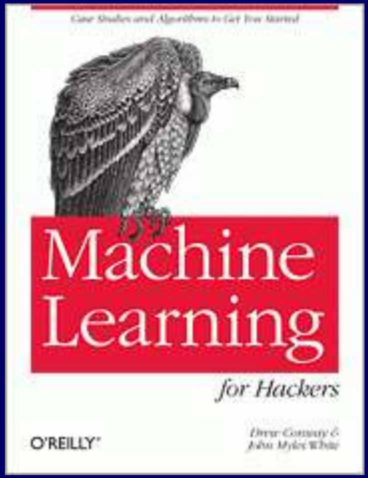


www.kaggle.com



David Donoho

Python scikit-learn.org



djohn89@gmail.com
www.djohn89.com