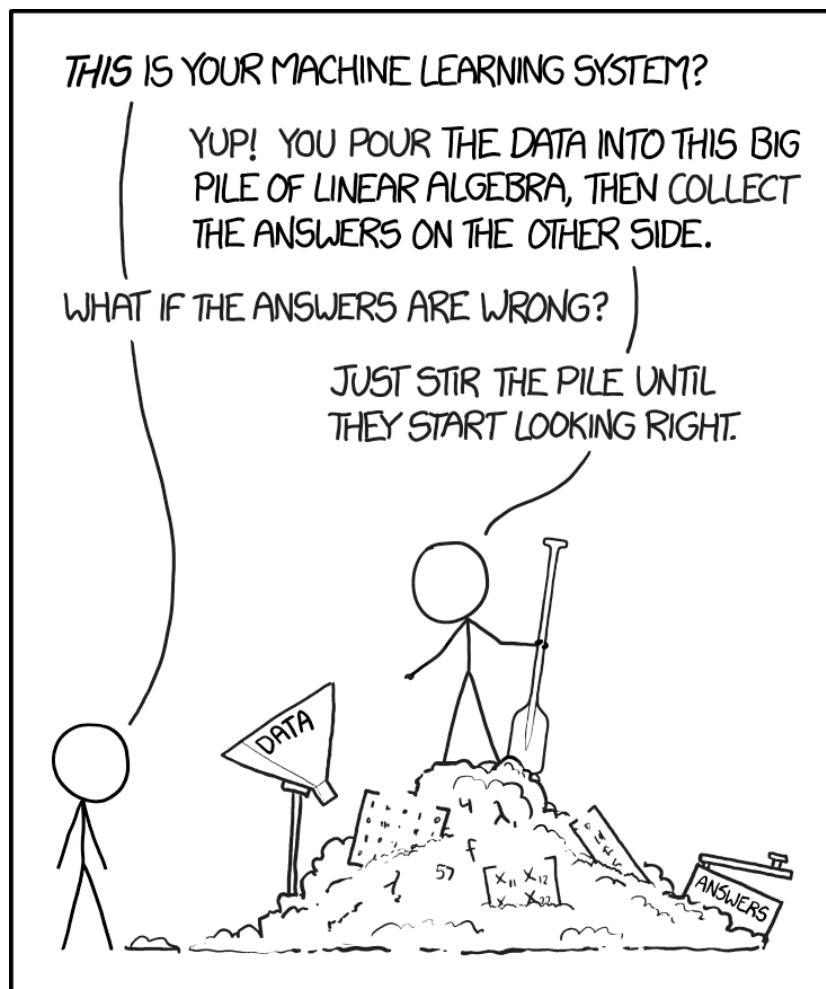


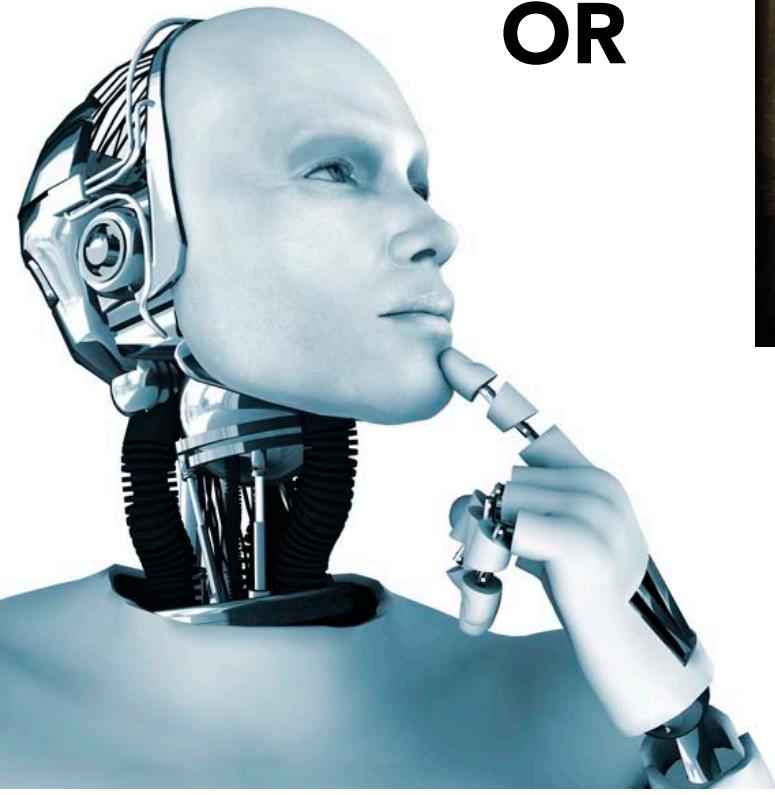
Deep Learning in Chemistry

Matt Robinson



What we think about when we think about deep learning

OR



Source:

<http://nerdist.com/ouija-boards-are-even-creepier-when-you-know-how-they-work/>, Popularspeed.com

What Silicon Valley Thinks About Deep Learning



DEEP VISION

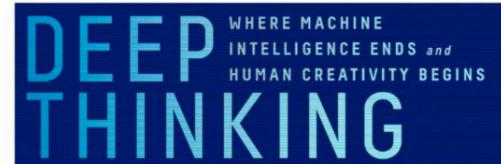
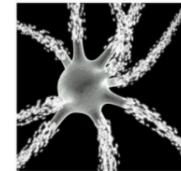
deepScience.ai



Deep Dive



deeplearning.ai



DEEP
LEARNING
INSTITUTE

DEEPVR

Source: Eric Topol, Twitter

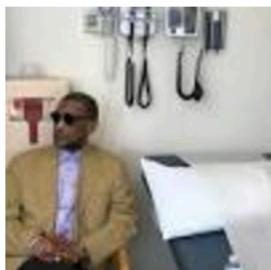
What Pharma thinks about it



Big pharma turns to AI to speed drug discovery, **GSK signs deal** 

Reuters - Jul 1, 2017

LONDON (**Reuters**) - The world's leading drug companies are turning to artificial intelligence to improve the hit-and-miss business of finding ...



How AI Is Transforming Drug Creation 

Wall Street Journal (subscription) - Jun 25, 2017

For example, **Merck & Co.** is using **Atomwise's** deep-learning technology to identify compounds that could be developed into medications for ...

Numerate to Use AI Platform in Developing Drugs for **Takeda**

Genetic Engineering & Biotechnology News - Jun 12, 2017

Numerate agreed to lead the companies' **discovery** programs, which are aimed at identifying clinical candidates in **Takeda's** core therapeutic ...

So What is Deep Learning?

Deep Learning is Everywhere

Siri / Alexa/ Cortana



Fraud Detection



Facial Recognition



Recommendation Systems



<http://www.businessinsider.com/netflix-new-logo-site-redesign-2014-6>,
<http://money.cnn.com/>,
<http://www.sickchirpse.com/most-disturbing-siri-responses>,
<http://sociable.co/social-media/how-to-disable-facebook-facial-recognition/>

What is Machine Learning?

“Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.”

- Arthur Samuel

What is Machine Learning?

More Rigorously:

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E .”

- Tom Mitchell

What is Machine Learning?

What all that means:

Machine Learning algorithms *learn from data*.

When is ML Useful?

- A pattern exists.
- We cannot pin it down mathematically.
- We have data on it.

List taken from *Learning From Data* by Abu-Mostafa et al.

Why is ML Useful?

- We can use a set of observations to uncover an underlying process
 - Often much easier to get examples of input and output, then to develop equation for the process.
- It can be automated

Why is ML Useful: An Example

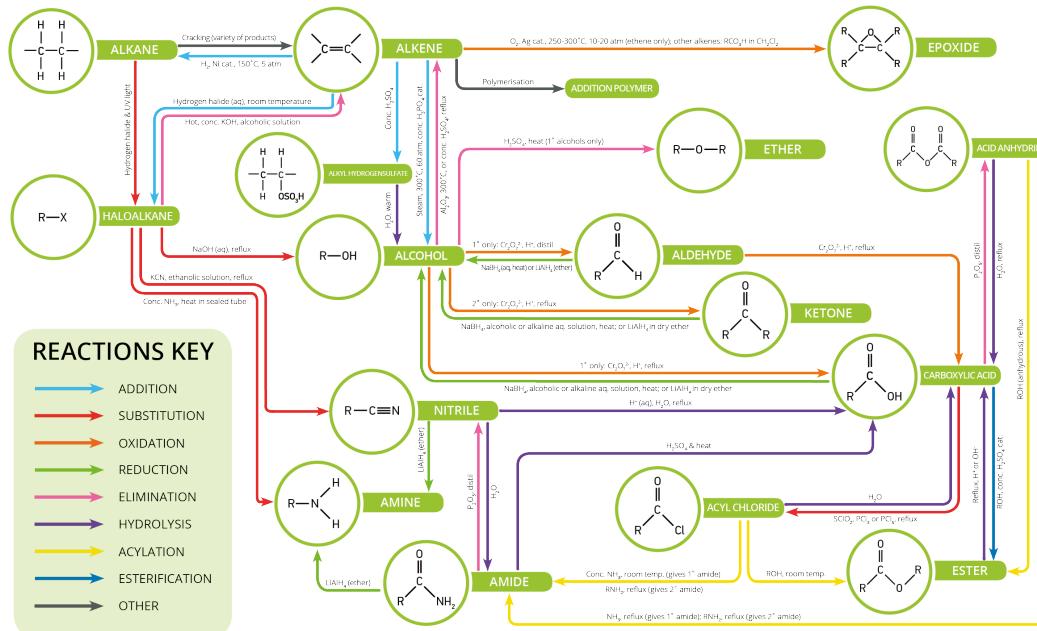
Spam Filter:

- Learns to flag spam emails given a *training set* of examples of spam and ham emails
- No need to write out long list of rules/exceptions

Why is ML Useful: An Example

What's something else with a lot of explicit rules and exceptions?

ORGANIC REACTIONS MAP



2014 COMPOUND INTEREST - WWW.COMPOUNDCHM.COM

Source: MIT News, "Computer system predicts of chemical reactions."

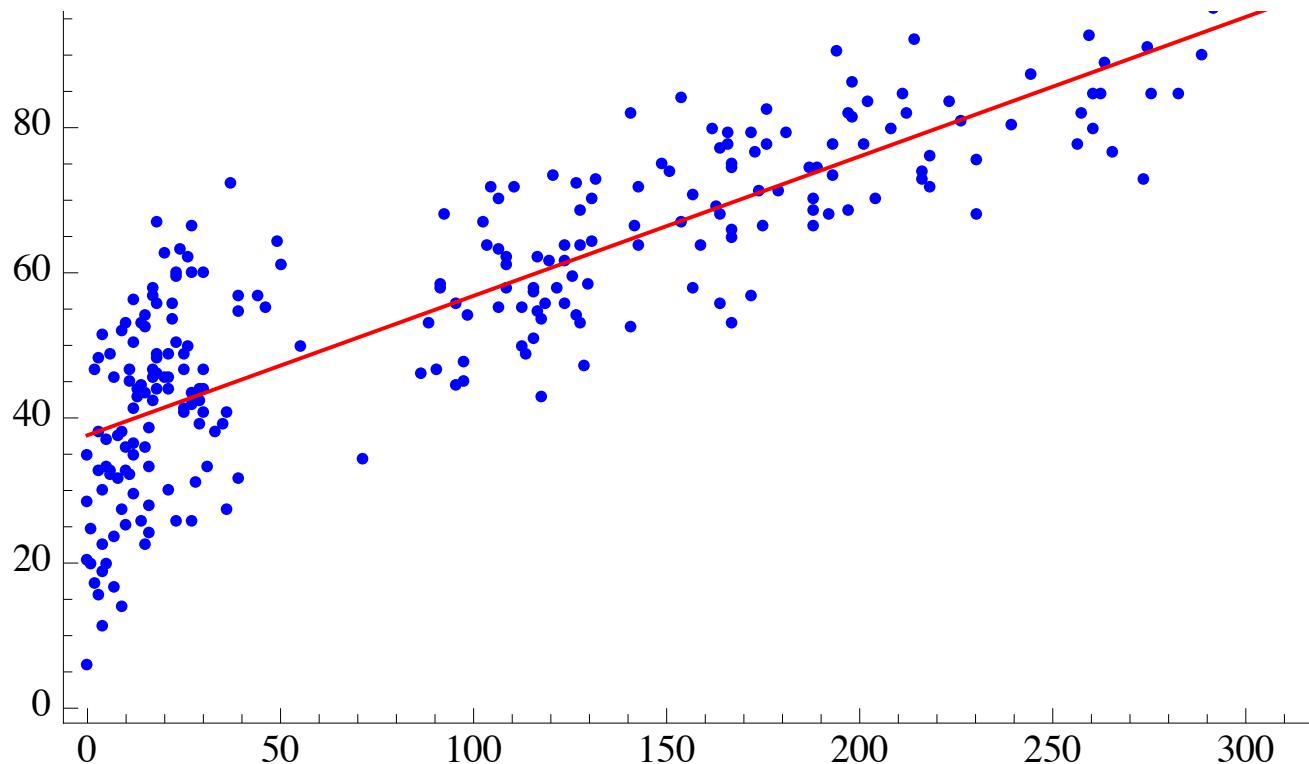
Why is ML Useful: An Example

What's something else with a lot of explicit rules and exceptions?

- “If molecule A has reaction site X, and molecule B has reaction site Y, then X and Y will react to form group Z – unless molecule A also has reaction sites P, Q, R, S, T, U, or V.”

Source: MIT News, “Computer system predicts of chemical reactions.”

A Simple ML Model: Linear Regression



A 200 year old machine learning algorithm

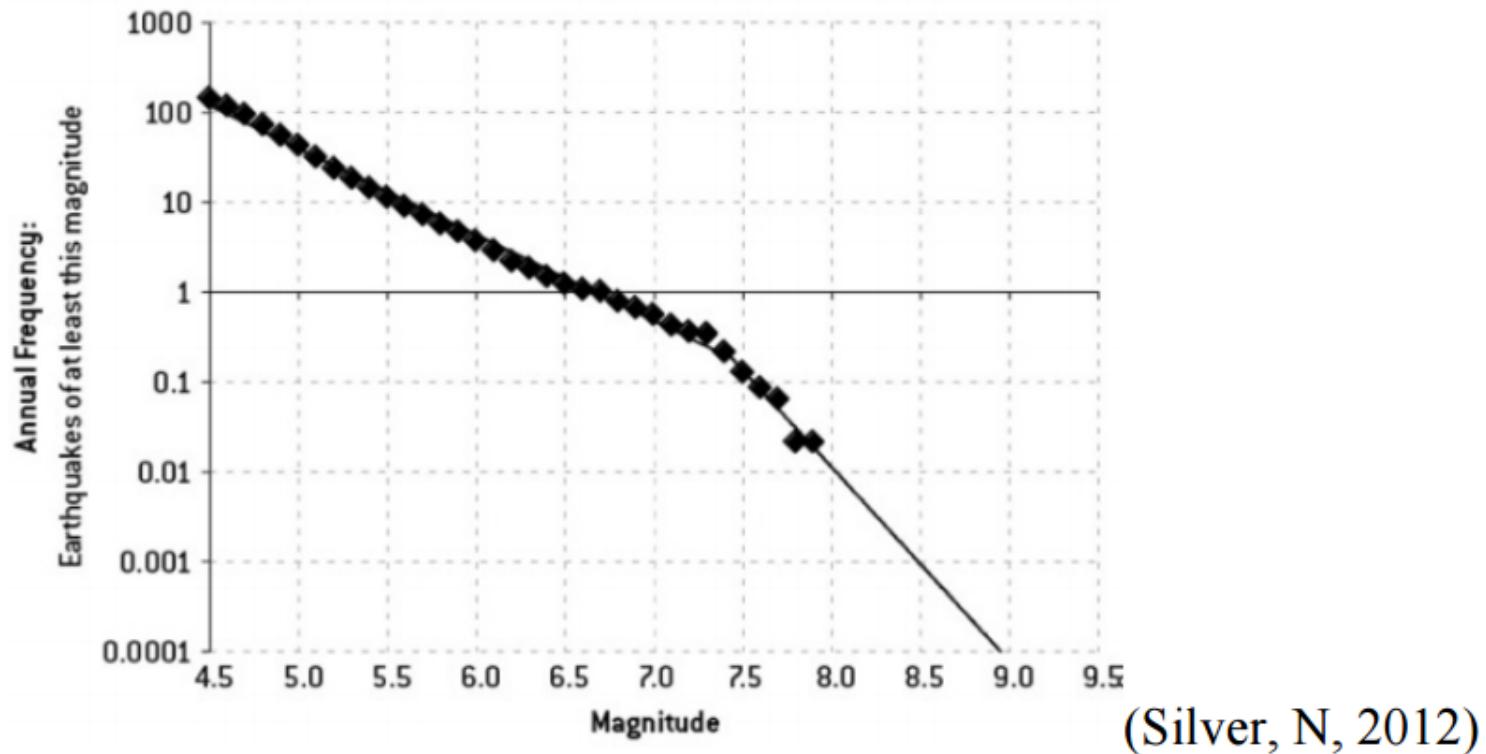
A Simple ML Model: Linear Regression

- 'Learns' a general function from **training set** examples of input-output (x, y) pairs.
- Can then use general function to **predict** output values (y) for new input values (x)
- The goal: a function that **generalizes** well to inputs outside the training set

The 'Nemesis' of Generalization: Overfitting

An example, how often do earthquakes occur?

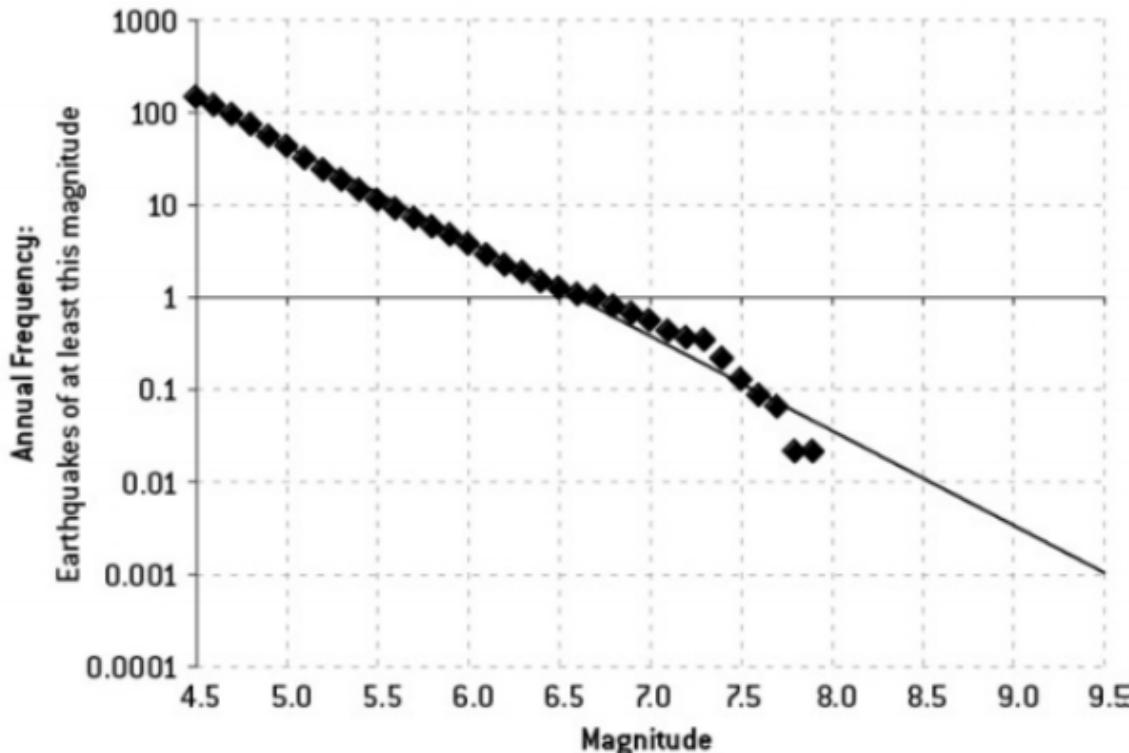
FIGURE 5-7C: TŌHOKU, JAPAN EARTHQUAKE FREQUENCIES
CHARACTERISTIC FIT



The 'Nemesis' of Generalization: Overfitting

An example, how often do earthquakes occur?

FIGURE 5-7B: TŌHOKU, JAPAN EARTHQUAKE FREQUENCIES
GUTENBERG-RICHTER FIT

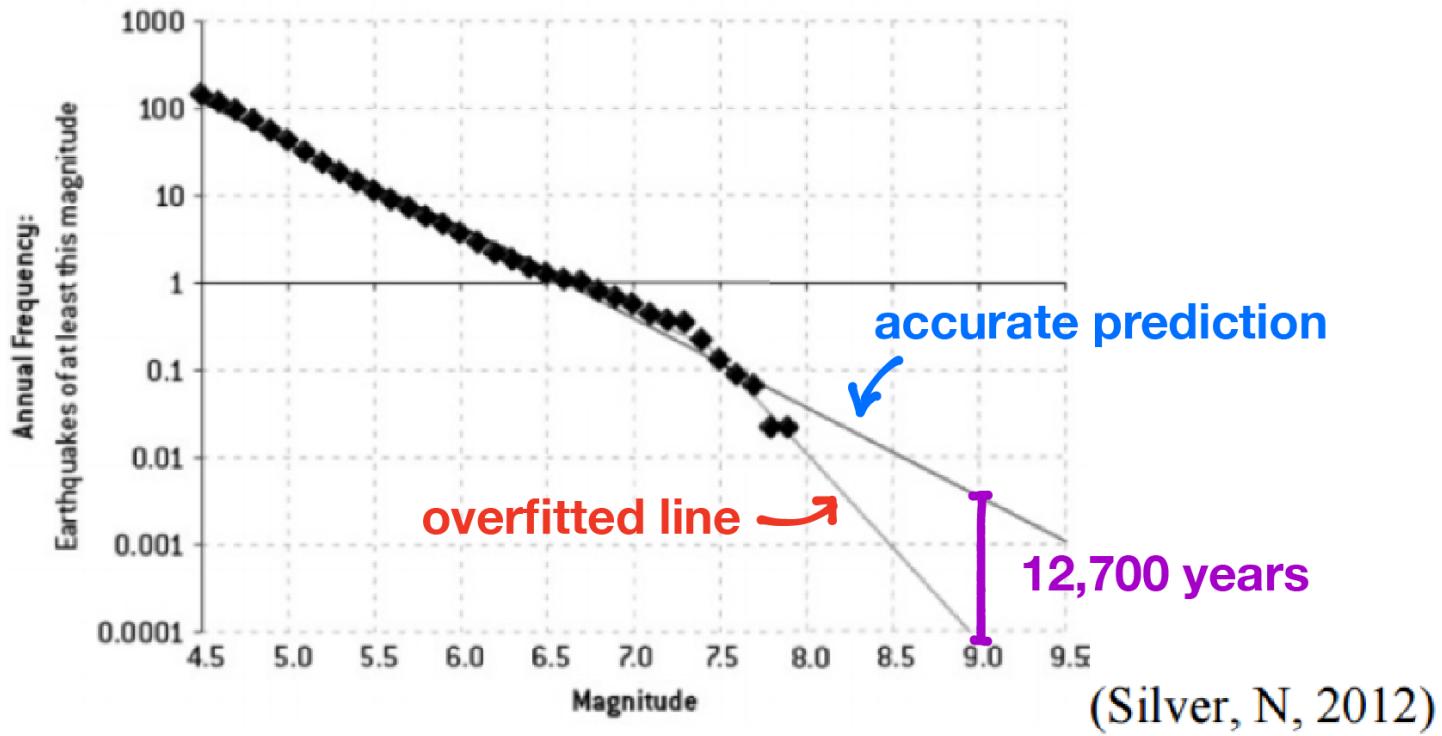


(Silver, N, 2012)

The 'Nemesis' of Generalization: Overfitting

An example, how often do earthquakes occur?

FIGURE 5-7C: TŌHOKU, JAPAN EARTHQUAKE FREQUENCIES
CHARACTERISTIC FIT

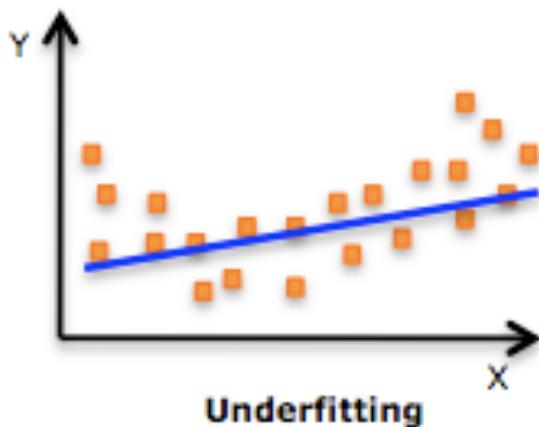


Source: Brian Stacey, 'Fukushima: The Failure of Predictive Models'

A Related Problem: Underfitting

- But, also need to adequately capture the underlying structure of the data.

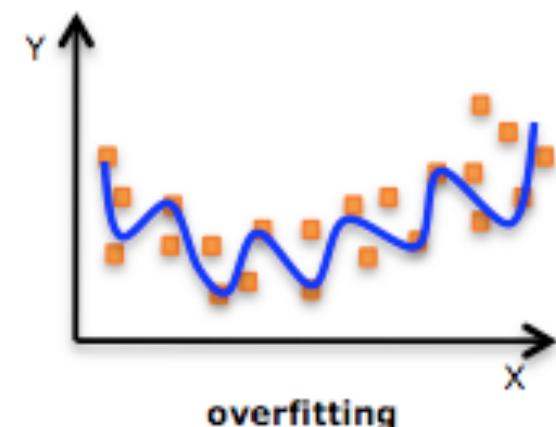
Need to find that Goldilocks fit!



Underfitting

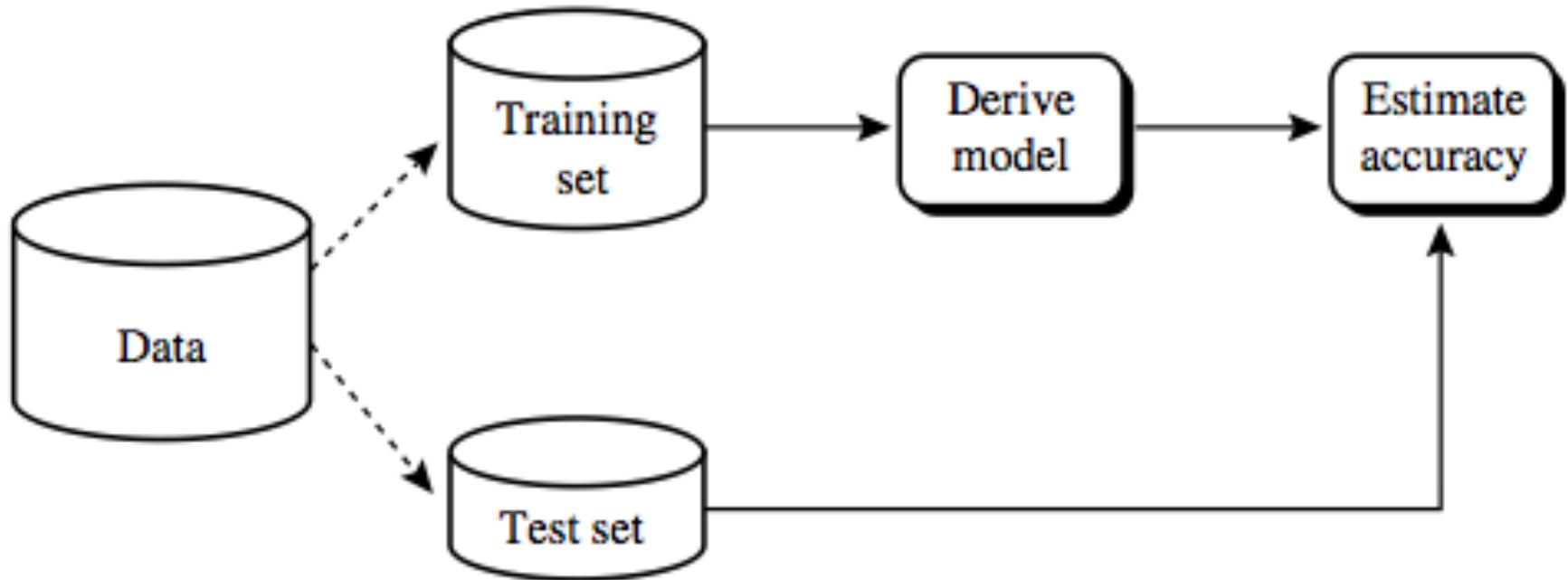


Just right!

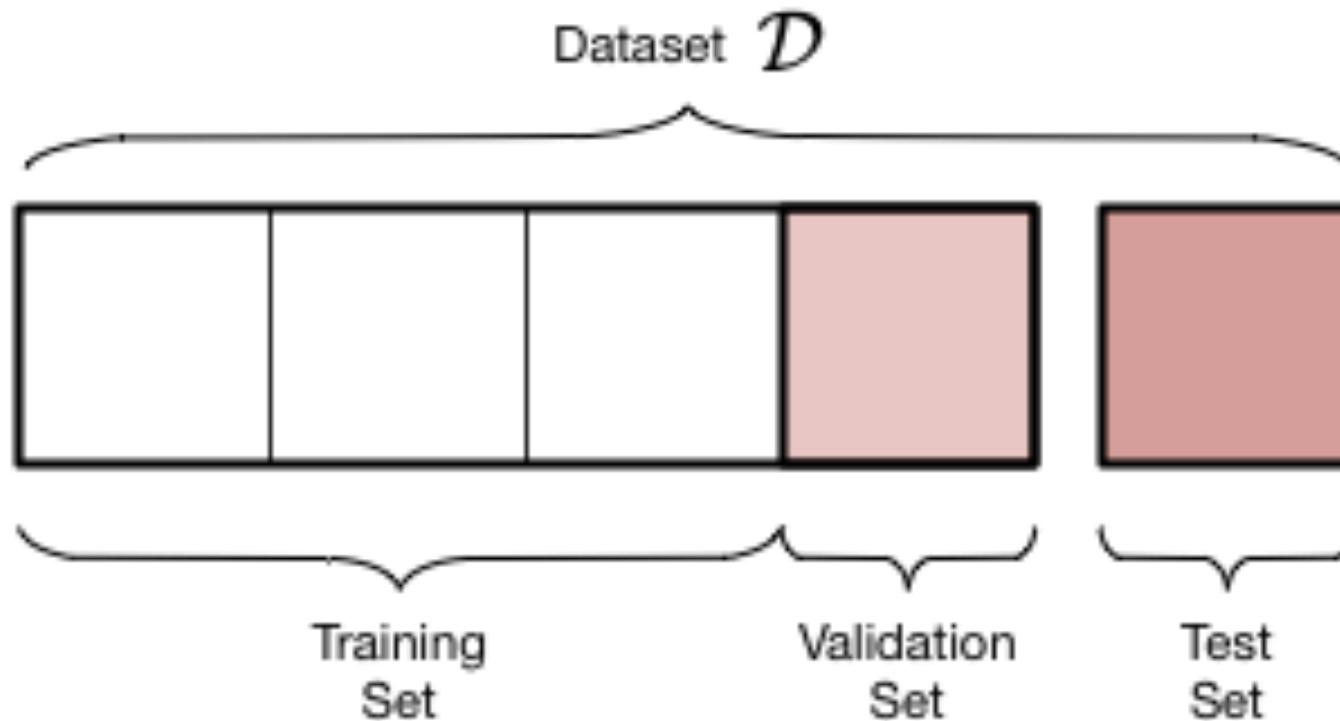


overfitting

Testing Generalizability: The Training/Test Set Split

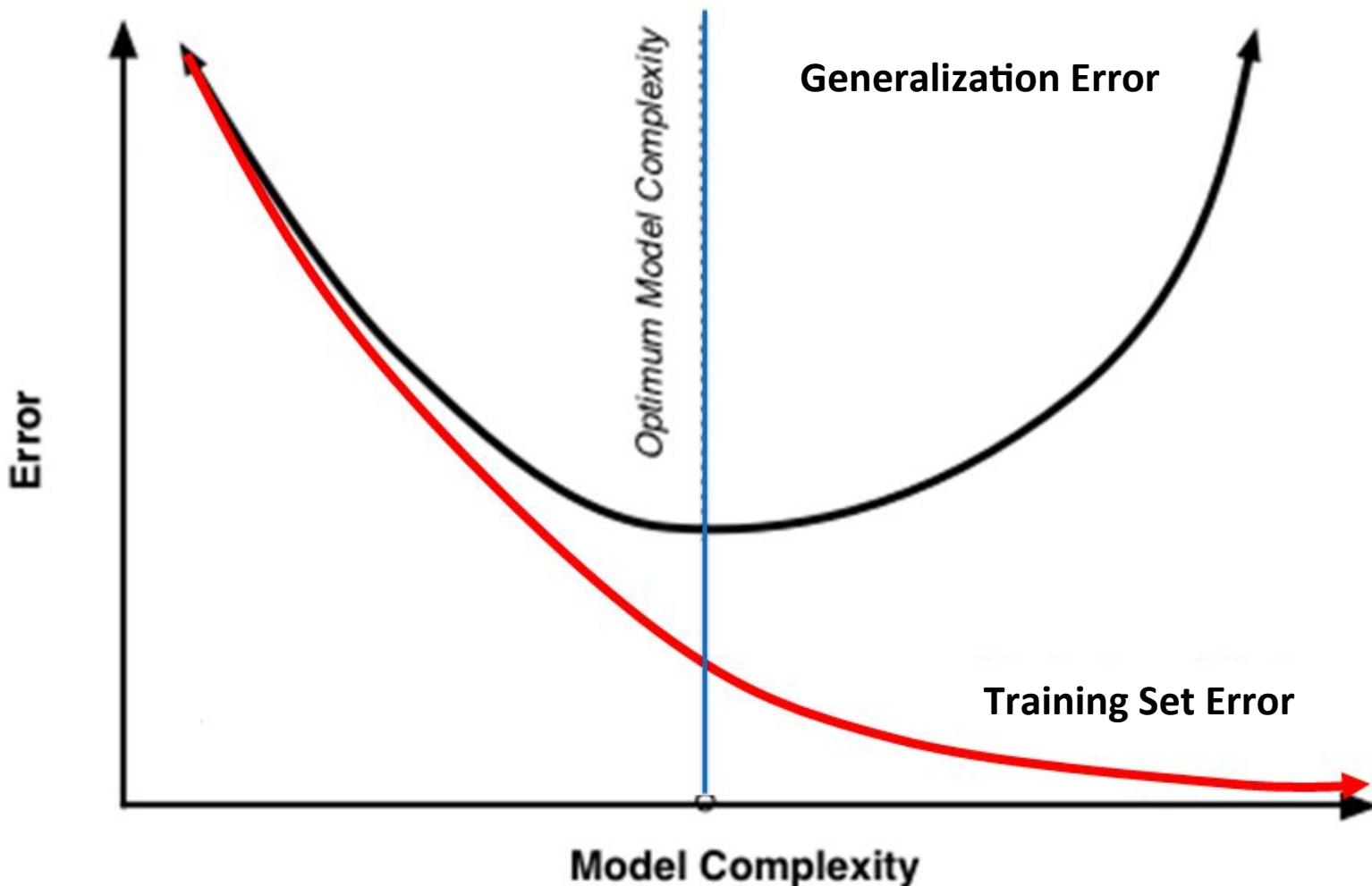


Testing Generalizability: The Training/Test Set Split



Source: AM207 Lecture Notes, Harvard University

The Training/Test Set Split



Source: Bridget Heath, "Deep Learning and Neural Nets"

A Different Model: Artificial Neural Networks (ANNs)

“Birds inspired us to fly, burdock plants inspired velcro, and nature has inspired many other inventions. It seems only logical, then, to look at the brain’s architecture for inspiration on how to build an intelligent machine.”

- Aurelien Geron

A Different Model: Artificial Neural Networks (ANNs)

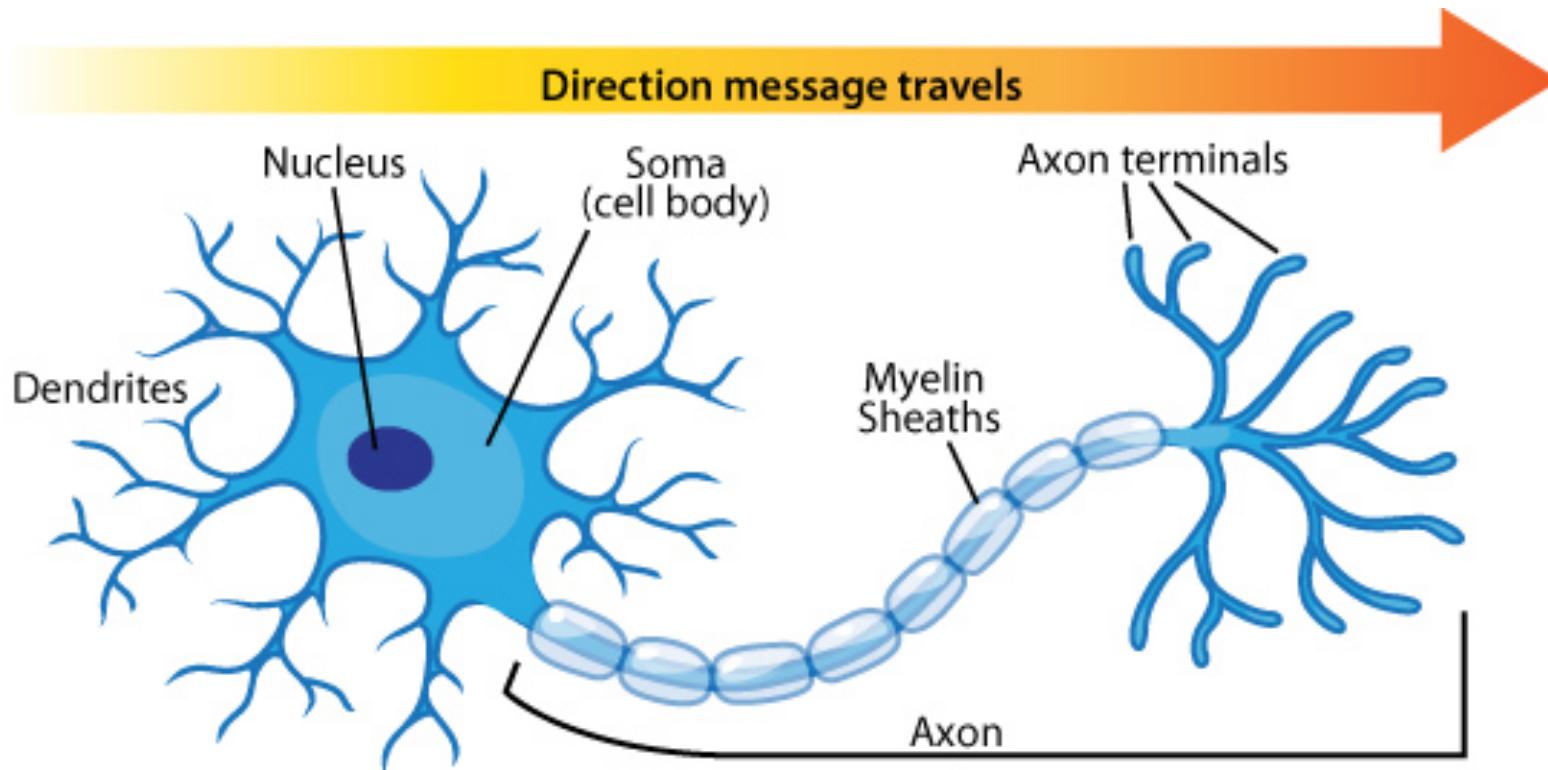
BUT:

“The quest for ‘artificial flight’ succeeded when the Wright brothers and others stopped imitating birds and started...learning about aerodynamics.”

- Stuart Russell and Peter Norvig

An Early Model: The Perceptron (Rosenblatt, 1957)

Simplified mathematical model of a neuron:



Source: Arizona State University, Ask a Biologist, "Neuron Anatomy"

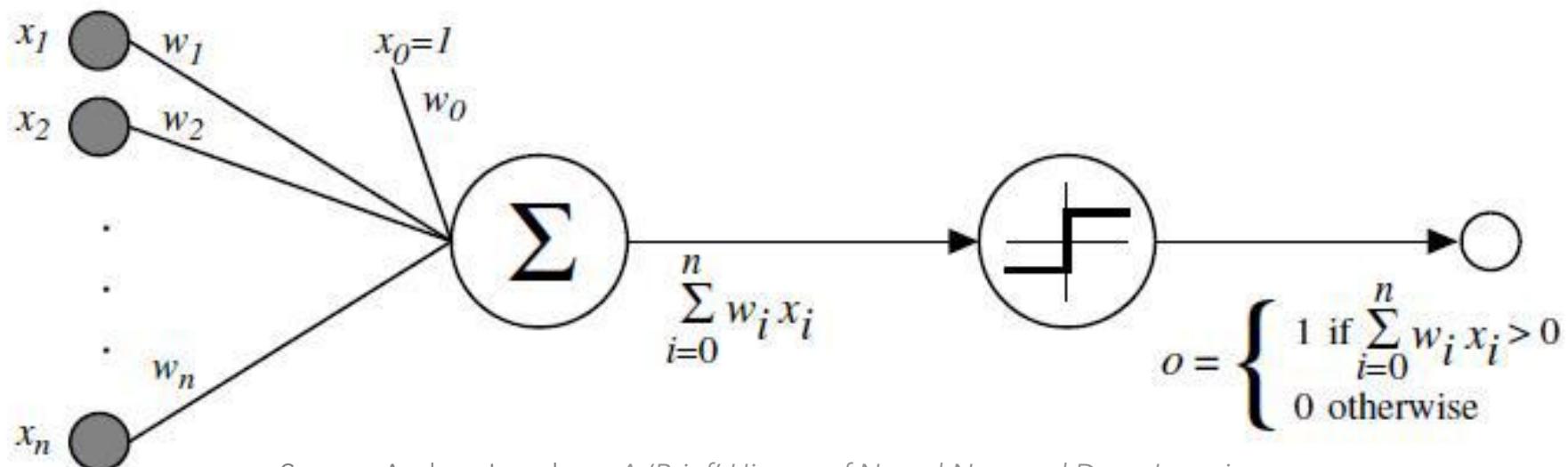
An Early Model: The Perceptron (Rosenblatt, 1957)

Simplified mathematical model of a neuron:

Inputs (x_i) = nearby neurons

Weights (w_i) = synapse strength

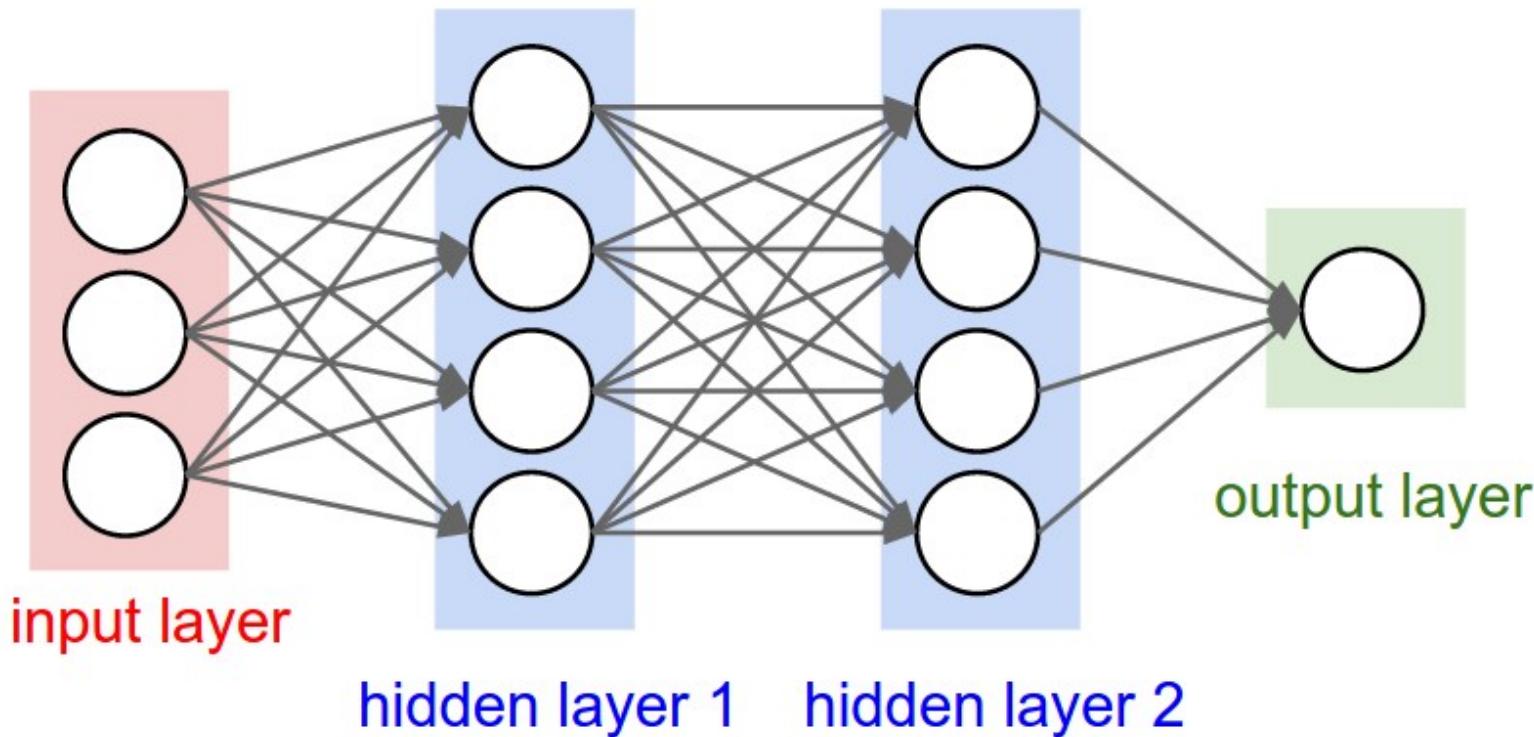
Activation function = threshold to fire



Source: Andrey Jurenkov,, A 'Brief' History of Neural Nets and Deep Learning

From the Perceptron to Artificial Neural Networks (ANN)

output of one neuron = input of another neuron



ANN with ≥ 2 hidden layers = *Deep Neural Network (DNN)*

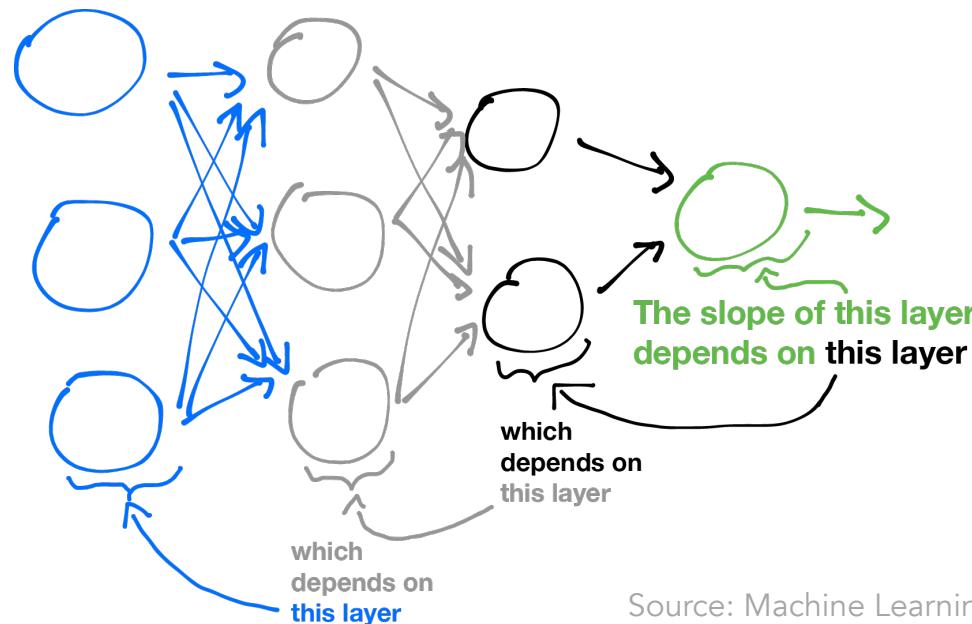
- Can model complex non-linear relationships

Training Neural Networks: The Backpropagation Algorithm

- The general idea:
 - Adjust parameters until NN outputs the answer.
- The algorithm (non-mathy version)
 - Use NN to make prediction on training example
 - Measure the error
 - Go through each layer in reverse to measure how each connection contributes to error
 - Tweak connections weights to reduce error

Training Neural Networks: The Backpropagation Algorithm

- If you prefer some math: We are minimizing the cost function, which is a function of the weights, using gradient descent.
 - It's just the chain rule!



Source: Machine Learning at Berkeley Blog

A Second AI Winter Dawns

Again Neural Nets ran into some issues:

- The vanishing gradient problem
- Overfitting
- Training was slow
- People switched to the then more effective SVMs (Support Vector Machines) and RFs (Random Forests)

The Current Spring

What brought neural nets back:

- New algorithms and activation functions
- New methods to avoid overfitting
- GPUs
- Much more data

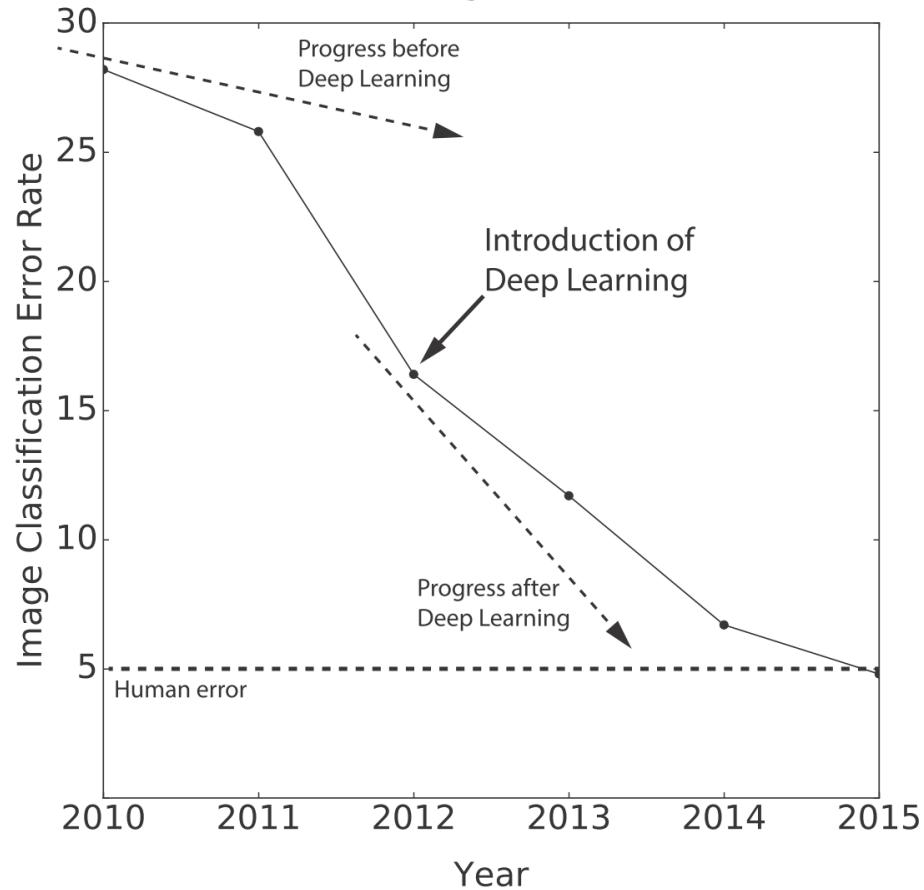
Some Amazing Results: Perhaps Deserving of Some Hype

Go



[https://en.wikipedia.org/wiki/Go_\(game\)](https://en.wikipedia.org/wiki/Go_(game))

ImageNet



Goh, G. B.; Hadas, N. O.; Vishnu, A., *Journal of Computational Chemistry* 2017, 38 (16), 1291-1307.

And Finally to Chemistry: Worth the Hype?

- “We observed a consistent outperformance against non-neural networks state-of-the-art models across disparate research topics .” (Goh et al., 2017)
- A note of caution: often hard to compare machine learning models to each other.

Deep Learning for QSAR

- Desired prediction: biological activity of compound
- “Shallow” ML models used for a long time
 - Input data = molecular descriptors (physiochemical properties of compound)
 - Informed chemists select appropriate features

Deep Learning for QSAR

- 2012 Merck Activity Prediction Challenge
 - Teams given pre-computed molecular descriptors of compounds, as well as experimentally measured activities for 15 drug targets.
 - DNN won competition and also outperformed Merck internal baseline model.

Deep Learning for QSAR

How they did it: Multi-task DNNs

- Predict multiple outputs simultaneously
- Creates shared features among multiple tasks
- Can deal with correlated features
- Can help when data is sparse for one task
- Less severe outliers

Deep Learning for QSAR

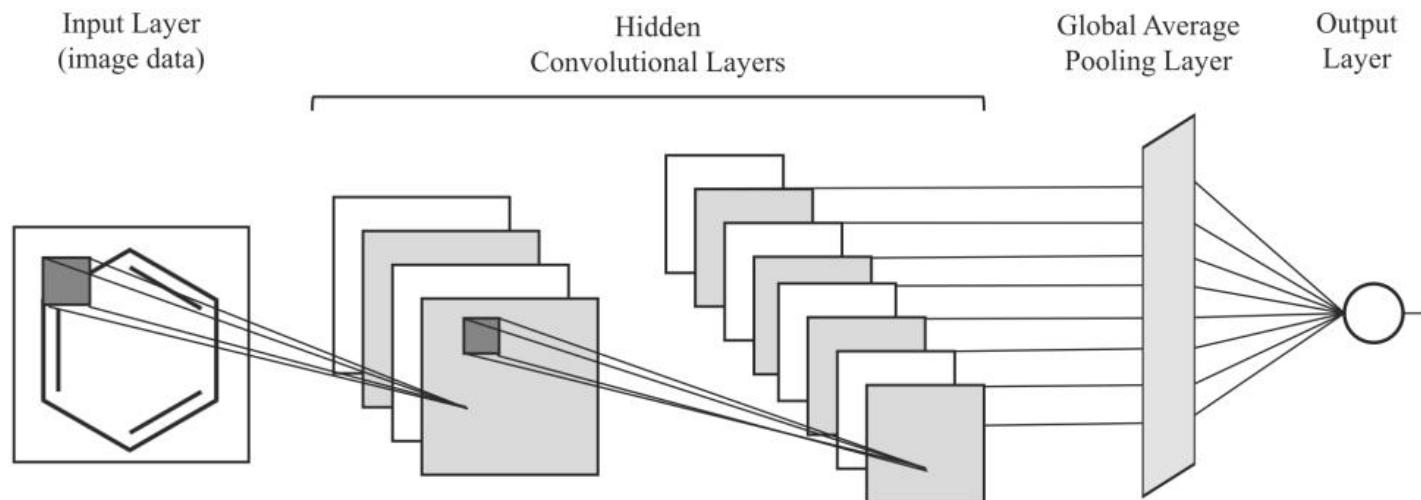
Assay	RF	GBM	NNET	MULTI
1851_1a2	0.915	0.926	0.926	0.938
1851_2c19	0.882	0.894	0.897	0.903
1851_2c9	0.876	0.891	0.889	0.907
1851_2d6	0.839	0.857	0.863	0.861
1851_3a4	0.871	0.896	0.895	0.897
1915	0.754	0.757	0.756	0.752
2358	0.745	0.764	0.738	0.751
463213	0.651	0.659	0.651	0.676
463215	0.614	0.610	0.613	0.654
488912	0.664	0.672	0.664	0.816
488915	0.700	0.713	0.723	0.873
488917	0.818	0.834	0.835	0.894
488918	0.785	0.800	0.784	0.842
492992	0.804	0.827	0.803	0.829
504607	0.673	0.670	0.684	0.670
624504	0.851	0.869	0.871	0.889
651739	0.775	0.793	0.790	0.825
651744	0.870	0.885	0.886	0.900
652065	0.787	0.793	0.793	0.792

Average test set AUC values.

Dahl, G. E.; Jaitly, N.; Salakhutdinov, R. arXiv:1406.1231 2014.

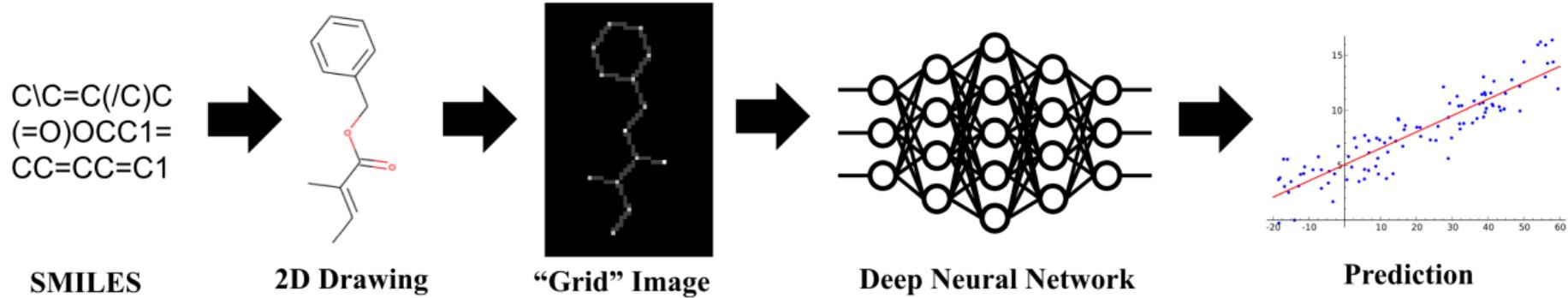
Deep Learning for QSAR

Chemception: A Deep Neural Network with Minimal Chemistry Knowledge Matches the Performance of Expert-developed QSAR/QSPR Models



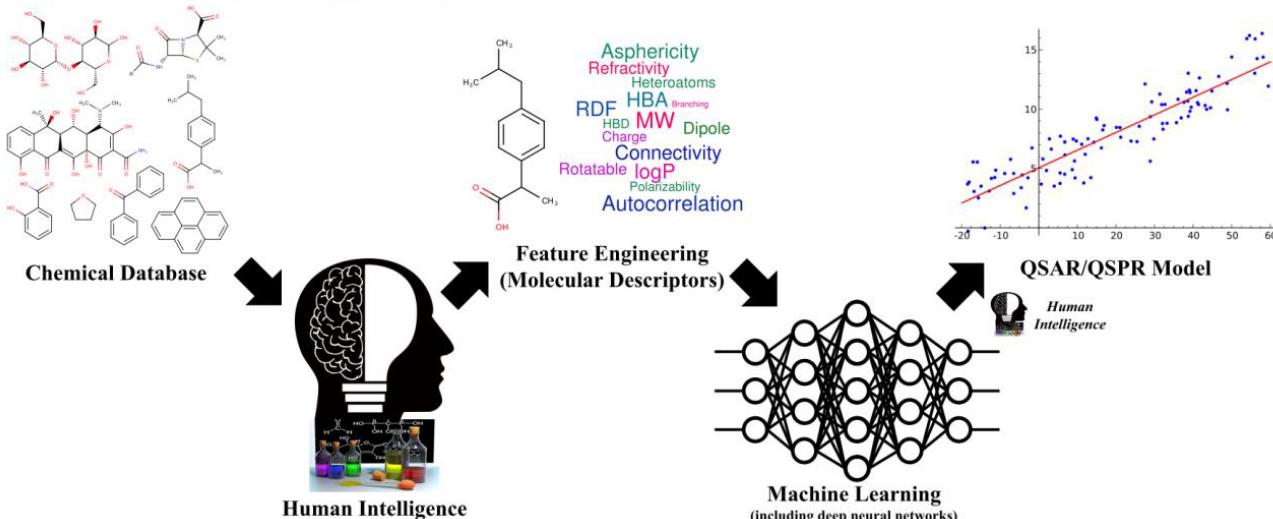
Deep Learning for QSAR

- Only uses images of 2D drawings of molecules.
 - No explicit chemistry knowledge provided to model
 - **Automatic feature engineering!**
 - Predicts toxicity, activity, and solvation properties after training on modest database

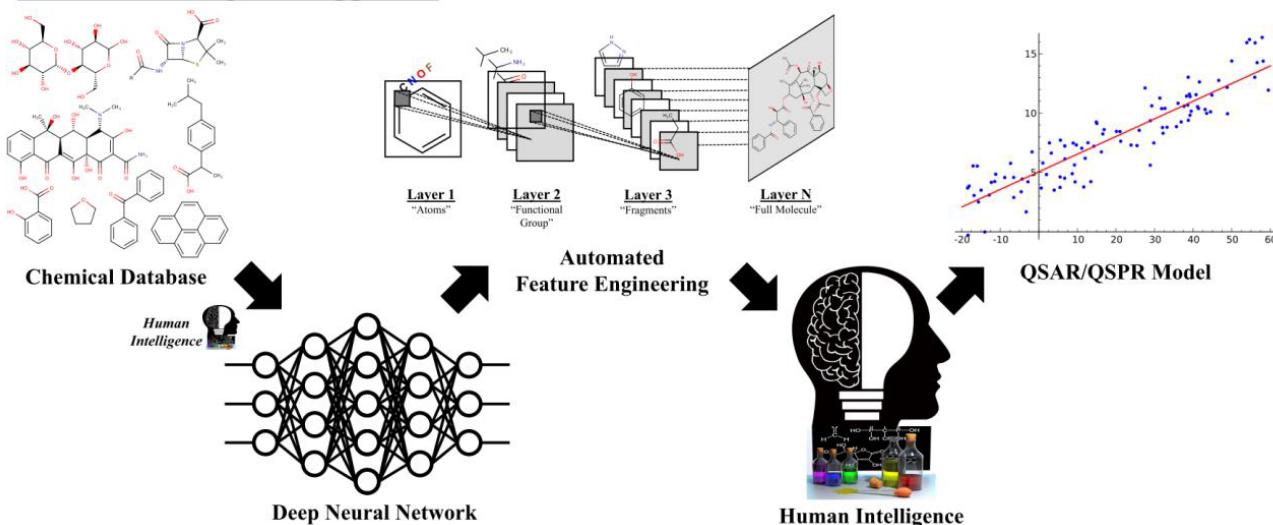


Deep Learning for QSAR

Machine Learning Tool Approach

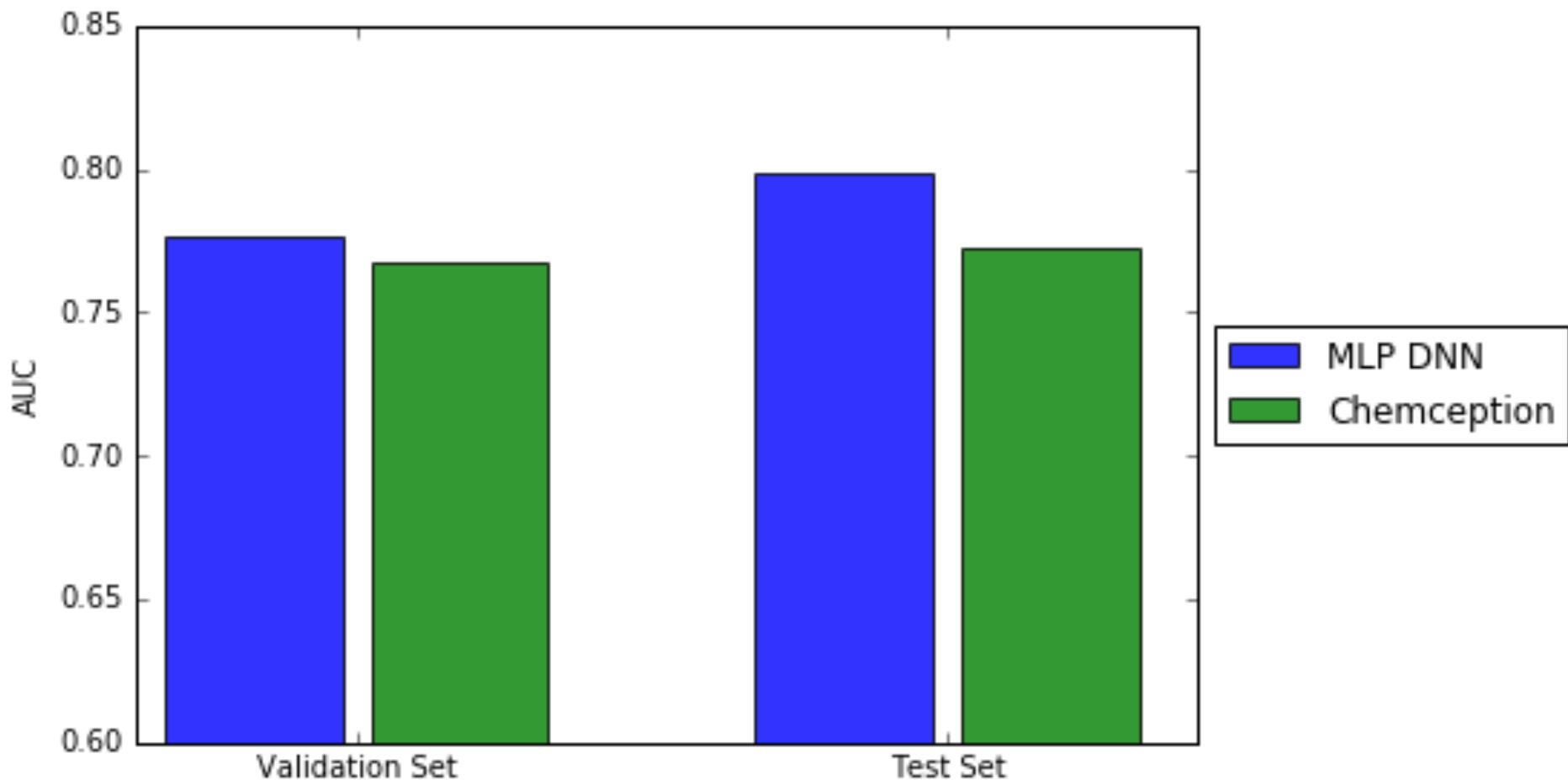


“Machine Intelligence” Approach



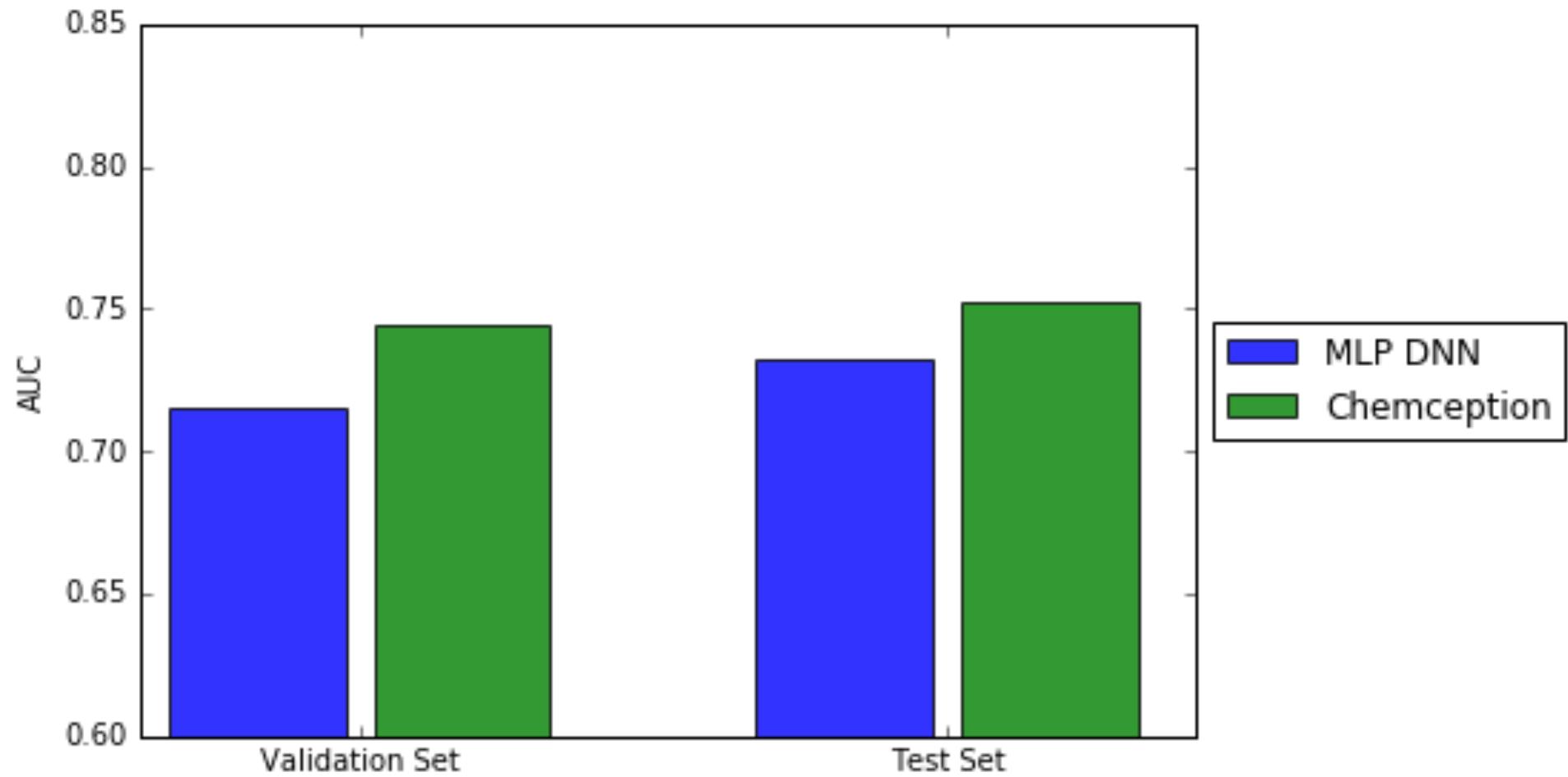
Deep Learning for QSAR

Tox21 Results



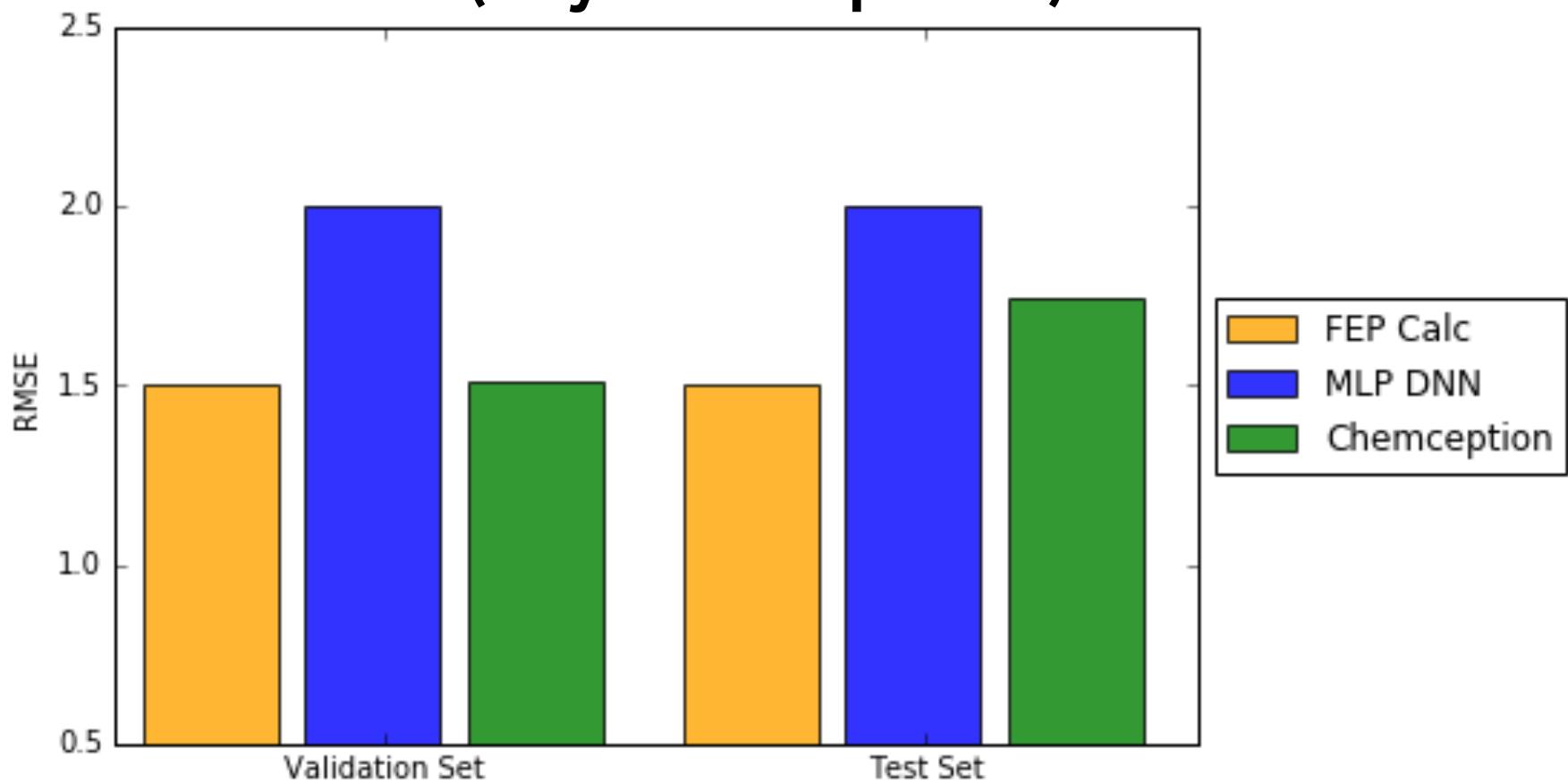
Deep Learning for QSAR

HIV Activity Results



Deep Learning for QSAR

Solvation Free Energy Results (only 600 compounds)



Deep Learning for QM

ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost

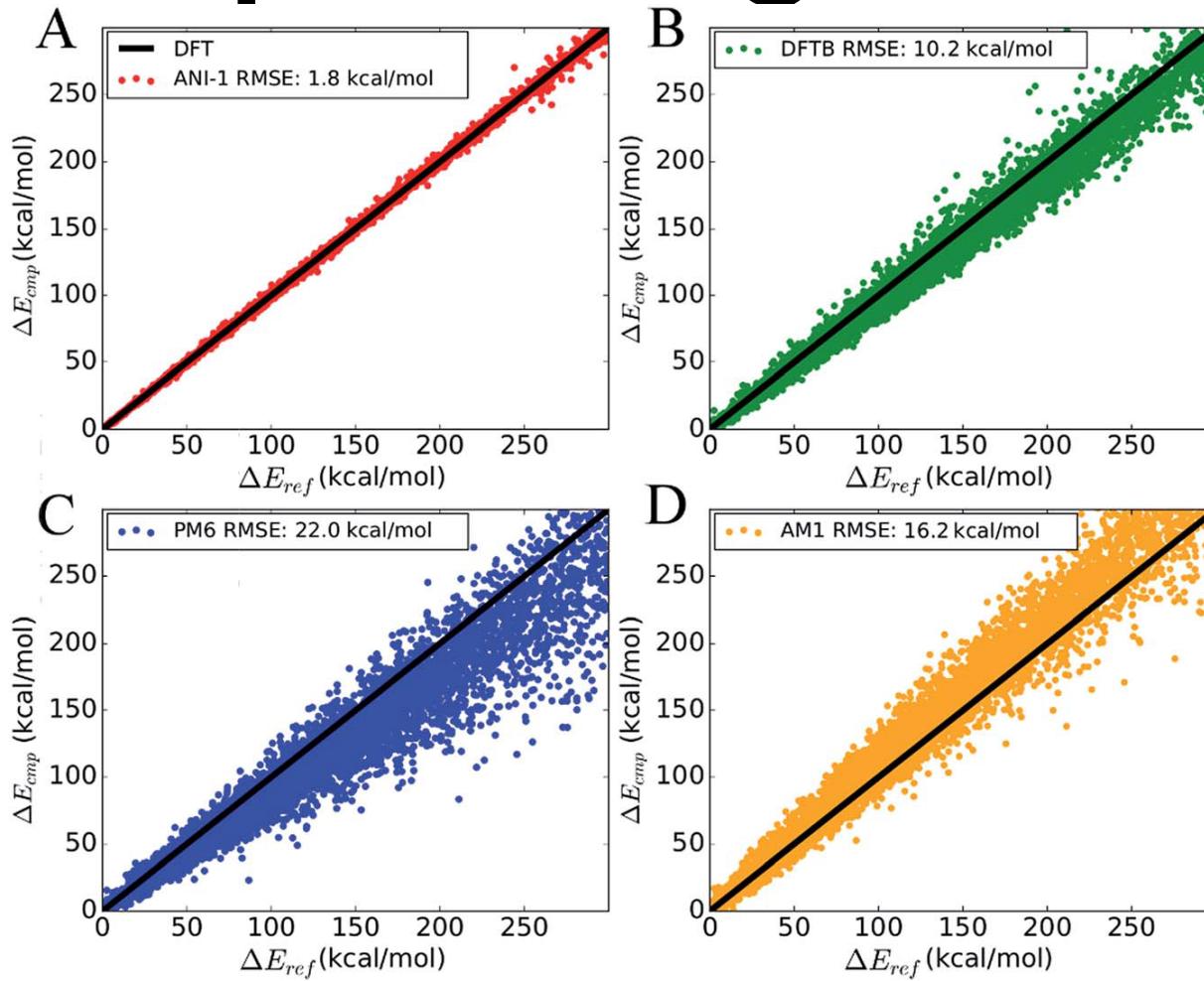
- ANAKIN-ME = (Accurate NeurAl networK englNe for Molecular Energies)
- DNN trained on DFT calculations learns an accurate and transferable potential for organic molecules
- Essentially a neural network force field

Smith, J. S.; Isayev, O.; Roitberg, A. E., *Chemical Science* 2017, 8 (4), 3192-3203.

Deep Learning for QM

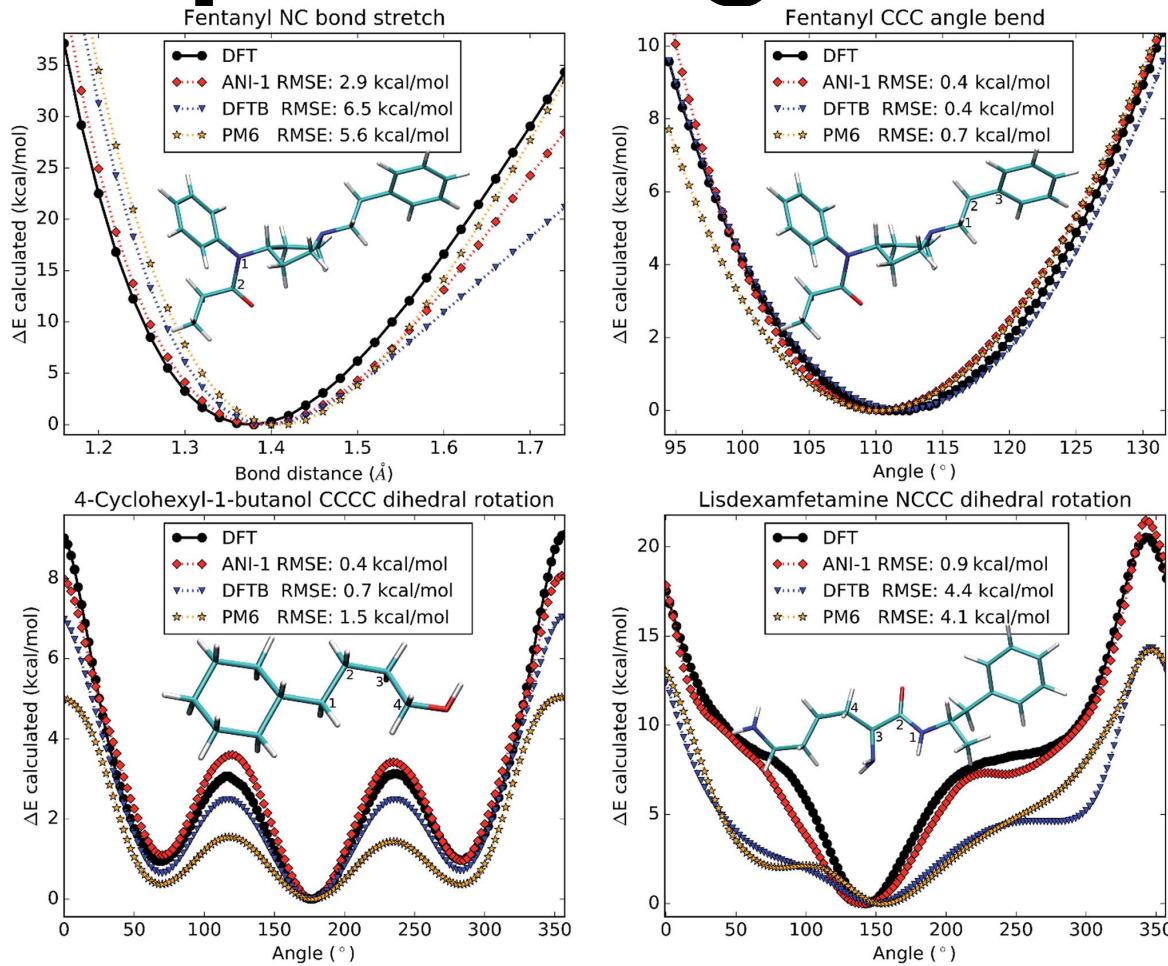
- Model utilizes modified Behler-Parinello symmetry functions to improve transferability
- Trained on 13.8 million ω B97X/6-31G(d) energies for molecules with 8 or less non-H atoms
- Transferable to larger molecules!
- Single point energies and forces can be calculated up to 10^6 X faster than with DFT

Deep Learning for QM



"Relative energy comparisons from random conformations of a random sampling of 134 molecules from GDB-11 all with 10 heavy atoms. None of these molecules are included in the training set."

Deep Learning for QM



“Each subplot shows a one-dimensional potential surface scan generated from DFT, the ANI-1 potential, and two popular semi-empirical methods, DFTB and PM6.” (Note the smooth potential)

The 'essence' of all these papers:

- A pattern exists.
- We cannot pin it down mathematically.
- We have data on it.

Arguments Against Deep Learning

- There is no physics
- DNNs are essentially a “black box”
 - Difficult to understand what NN has ‘learned’ or how it is making predictions

Arguments Against Deep Learning

- There is not enough data in Chemistry for DNNs to excel
 - Data in Chemistry is hard/expensive to get
- Improved performance not worth increased complexity
- “Climbing a tree to get to the moon.”

Arguments For Deep Learning

- DNNs work
- New methods are coming out to improve interpretability
- They are a tool to aid chemists' decisions, not to replace them
 - Use as 'hypothesis generation device'

Arguments For Deep Learning

- DNNs are useful for automatic feature engineering
- “Almost all of computational chemistry is an empirically-determined, and at times even intuitively determined, approximation of the ‘true’ first principles Schrodinger Equation.” (Goh et al., 2017)

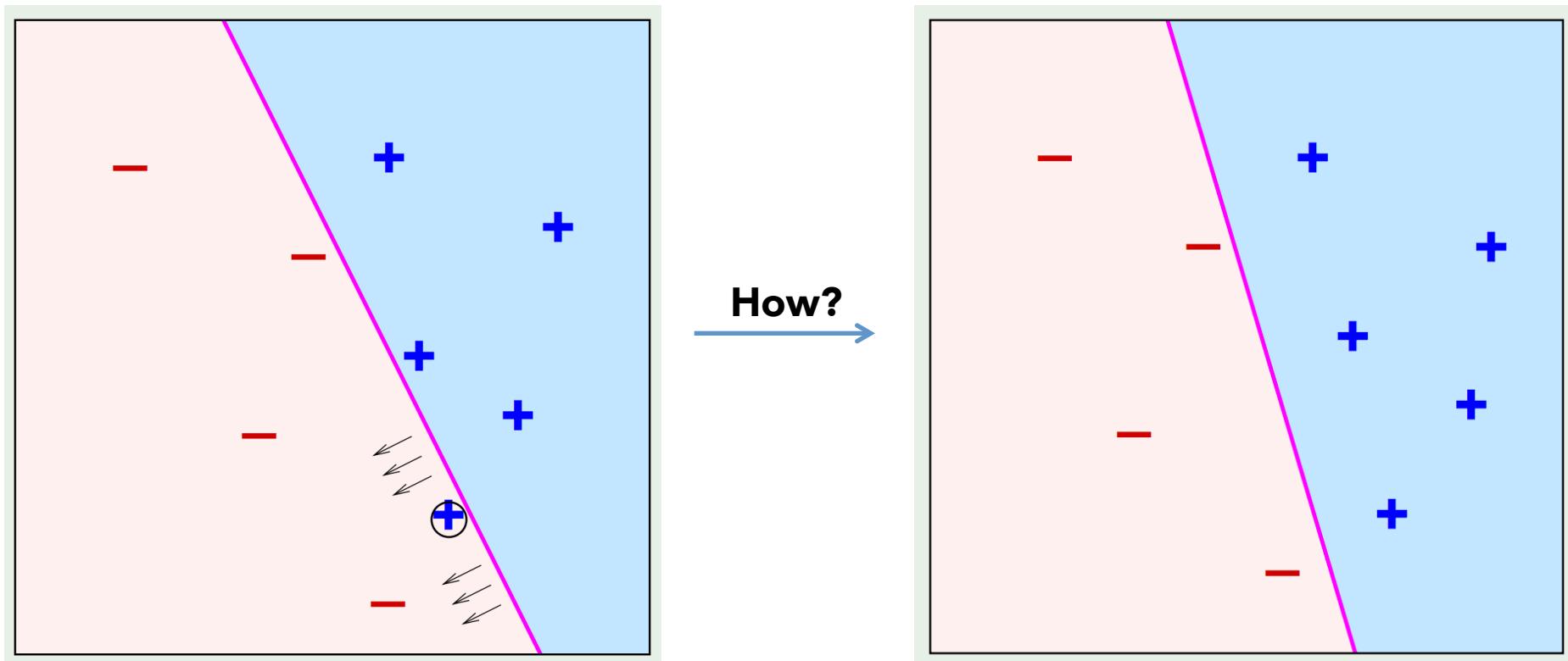
Arguments For Deep Learning

- “Coupled with the maturity of GPU-accelerated computing for training deep neural networks and the exponential growth of chemical data on which to train these networks on, we anticipate that deep learning algorithms will be a valuable tool for computational chemistry.” (Goh et al., 2017)

THANKS FOR LISTENING!

- Resources to learn more:
 - Lots of MOOCS/ great textbooks/ free online material

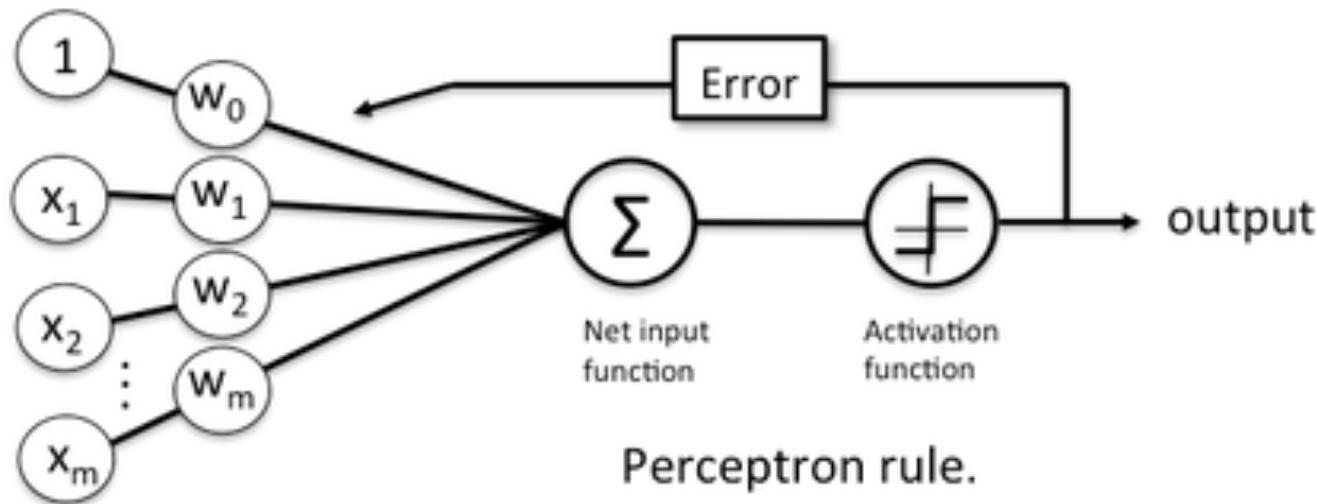
How a Perceptron Learns



Source: *Learning From Data* by Abu-Mostafa et al.

How a Perceptron Learns

“Cells that fire together wire together”



Pick a misclassified point:

Reinforce connections (increase weights) that would have led to the correct classification, and vice versa