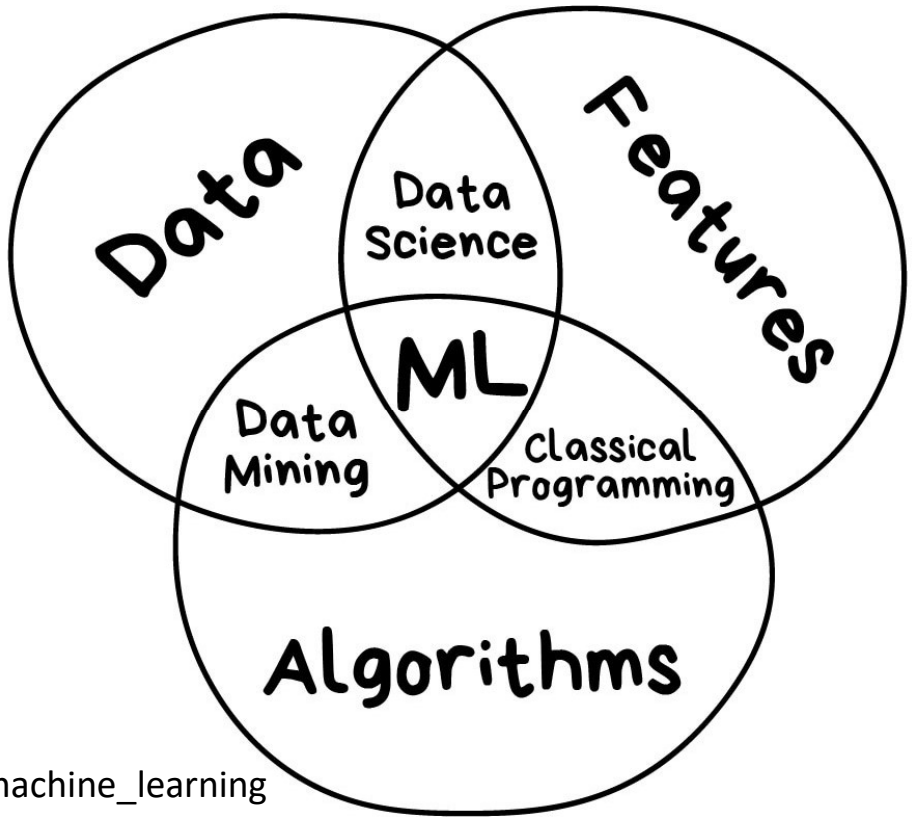
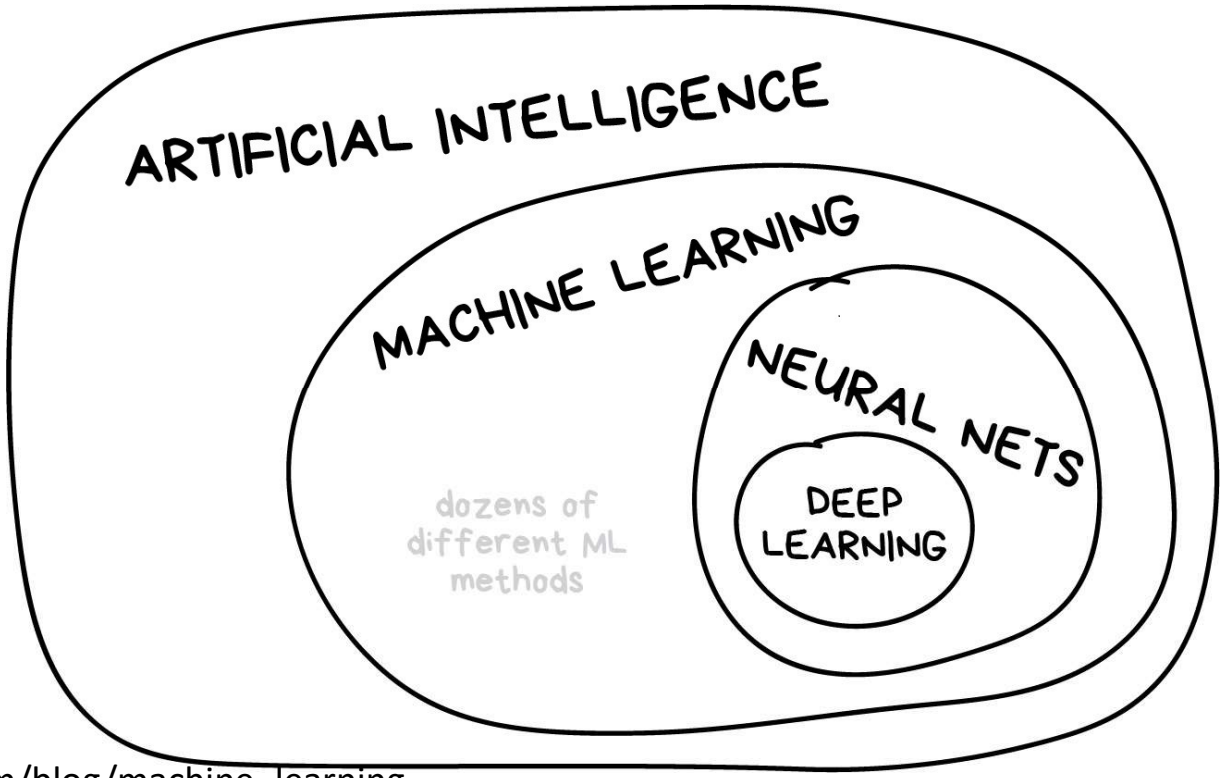
Machine Learning Review

Three Components of ML



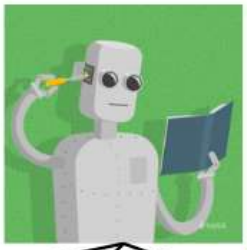
Source: vas3k.com/blog/machine_learning

Machine Learning and Artificial Intelligence



Source: vas3k.com/blog/machine_learning

Taxonomy of Machine Learning



Labeled Data

Reward

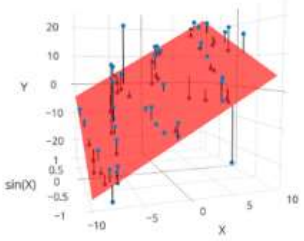
Unlabeled Data

Supervised Learning

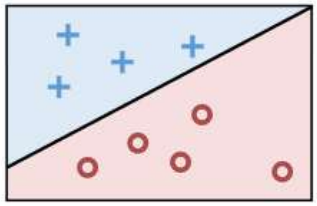
Quantitative Response

Categorical Response

Regression



Classification

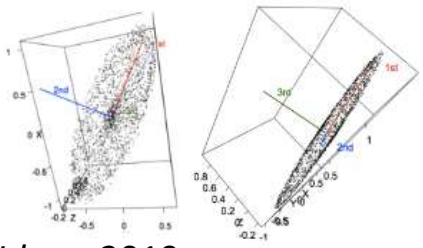


Reinforcement Learning

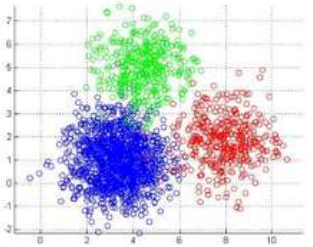
Alpha Go

Unsupervised Learning

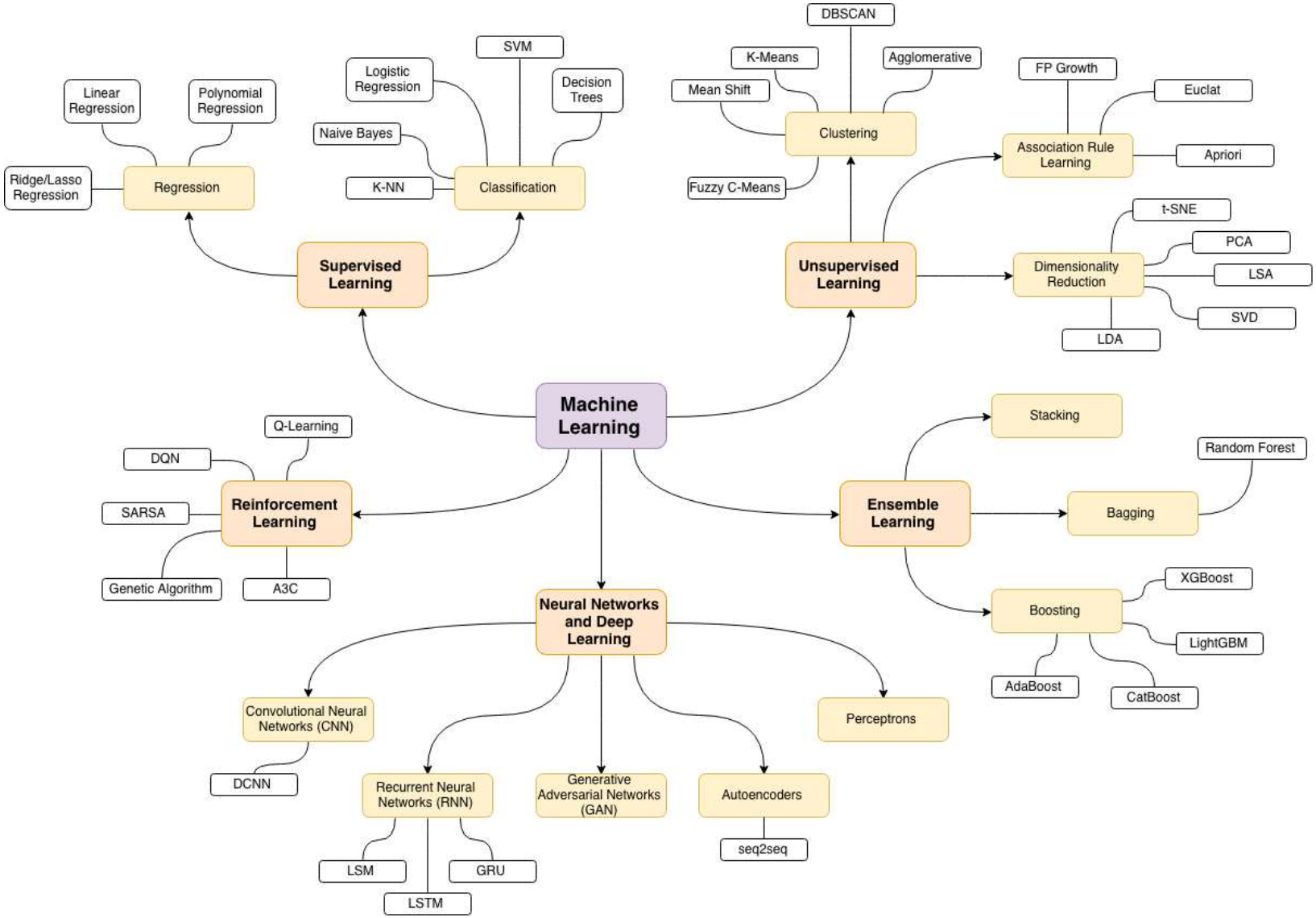
Dimensionality Reduction



Clustering



Source: Joseph E. Gonzalez, *AI-Systems Big Ideas*, 2019.



Source: vas3k.com/blog/machine_learning

Machine Learning ≈ Function Approximation

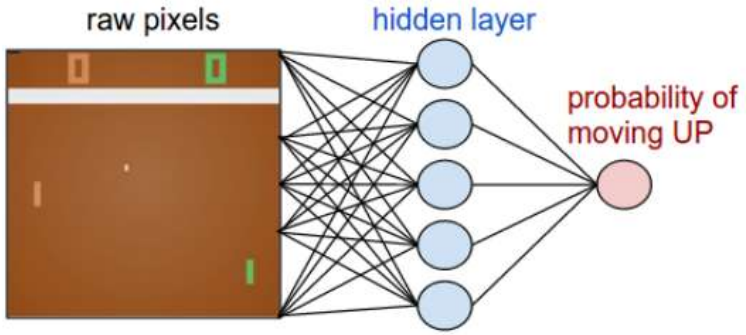
Object Recognition

Label:*Cat*

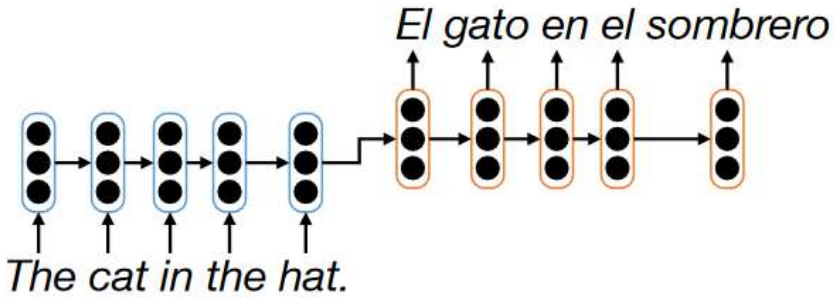
Speech Recognition

"The cat in the hat"

Robotic Control



Machine Translation



Source: Joseph E. Gonzalez, *AI-Systems Big Ideas*, 2019.

Supervised Machine Learning

- Given data containing the function inputs and outputs

Data

Input	Output
X_1 	Y_1 cat
X_2 	Y_2 *baby*
…	…
X_n 	Y_n *baby*

Model

$$f_\theta(x) \rightarrow y$$

Parameters

Goal

Loss

$$\theta^* = \arg\min_\theta \mathbb{E}_D \left[L\left(f_\theta(x), y \right) \right]$$

Over future data

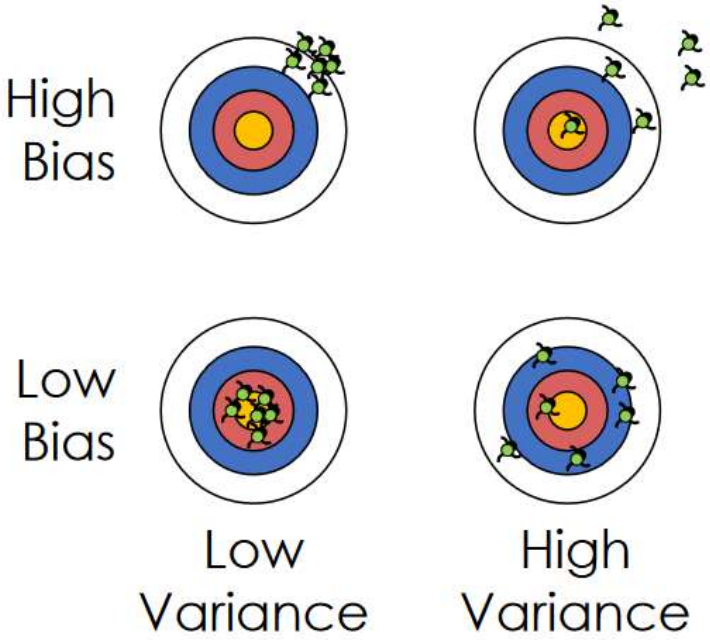
Training (approximates the goal over training data):

$$\hat{\theta} = \arg\min_\theta \frac{1}{n} \sum_{i=1}^{n} L\left(f_\theta(x_i), y_i \right)$$

Source: Joseph E. Gonzalez, *AI-Systems Big Ideas*, 2019.

The Bias Variance Tradeoff

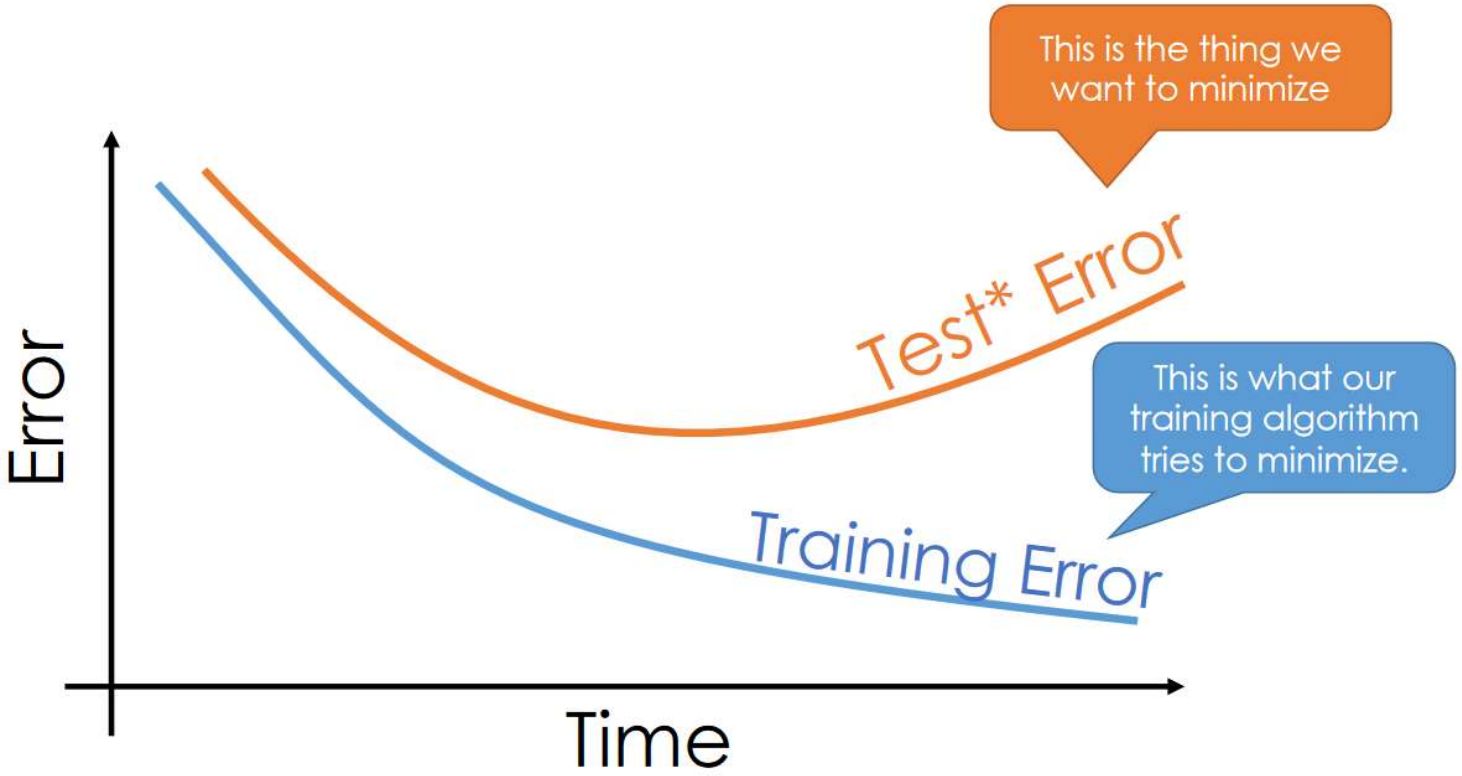
- Fundamental trade-off in ML (classically)



- Low bias learning techniques
  - Typically higher variance ...
- Increasing data supports
  - Higher variance techniques
- Deep neural networks?
  - Focus on training procedure not models to control tradeoff
    - Initialization, SGD, Dropout, learning rates, early stopping, ...

Source: Joseph E. Gonzalez, *AI-Systems Big Ideas*, 2019.

Training and Validation



*If we are making modeling decisions based on this then it should be called validation error.

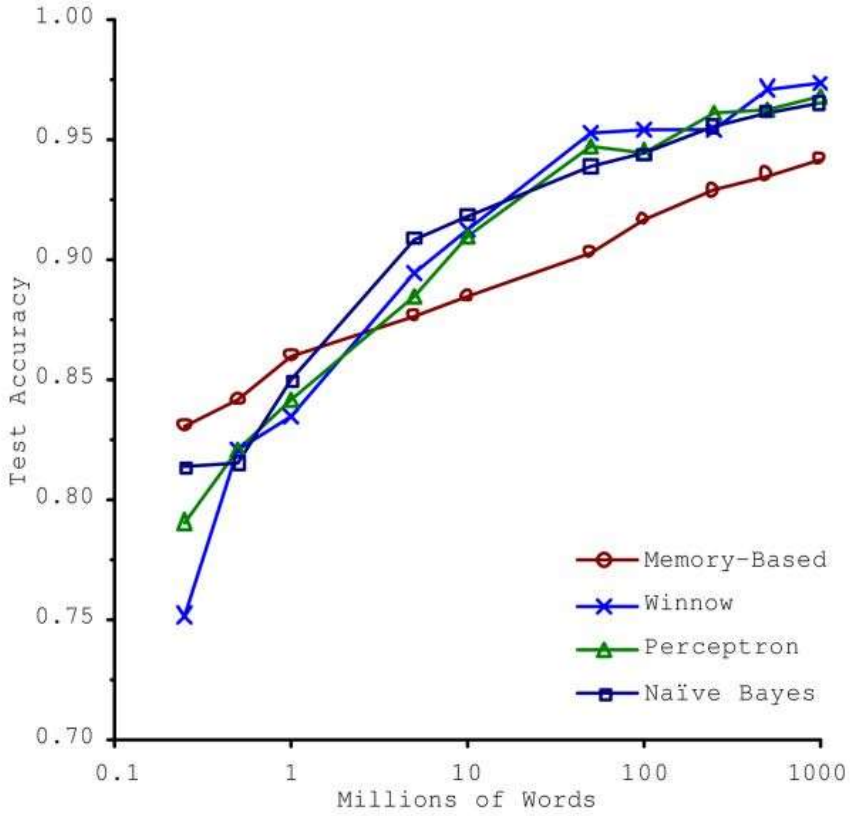
On Dataset Size and Learning

- Data is a a resource! (e.g., like processors and memory)
  - Is having lots of processors a problem?
- You don't have to use all the data!
  - Though using more data can often help
- More data often *dominates models and algorithms

Source: Joseph E. Gonzalez, *AI-Systems Big Ideas*, 2019.

*More data also enables more sophisticated.

On Dataset Size and Learning



References

- Joseph E. Gonzalez, *AI-Systems Big Ideas*, 2019.
- Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, CACM, 2012.