ME 5374-ST

Machine Learning for Materials Science and Discovery

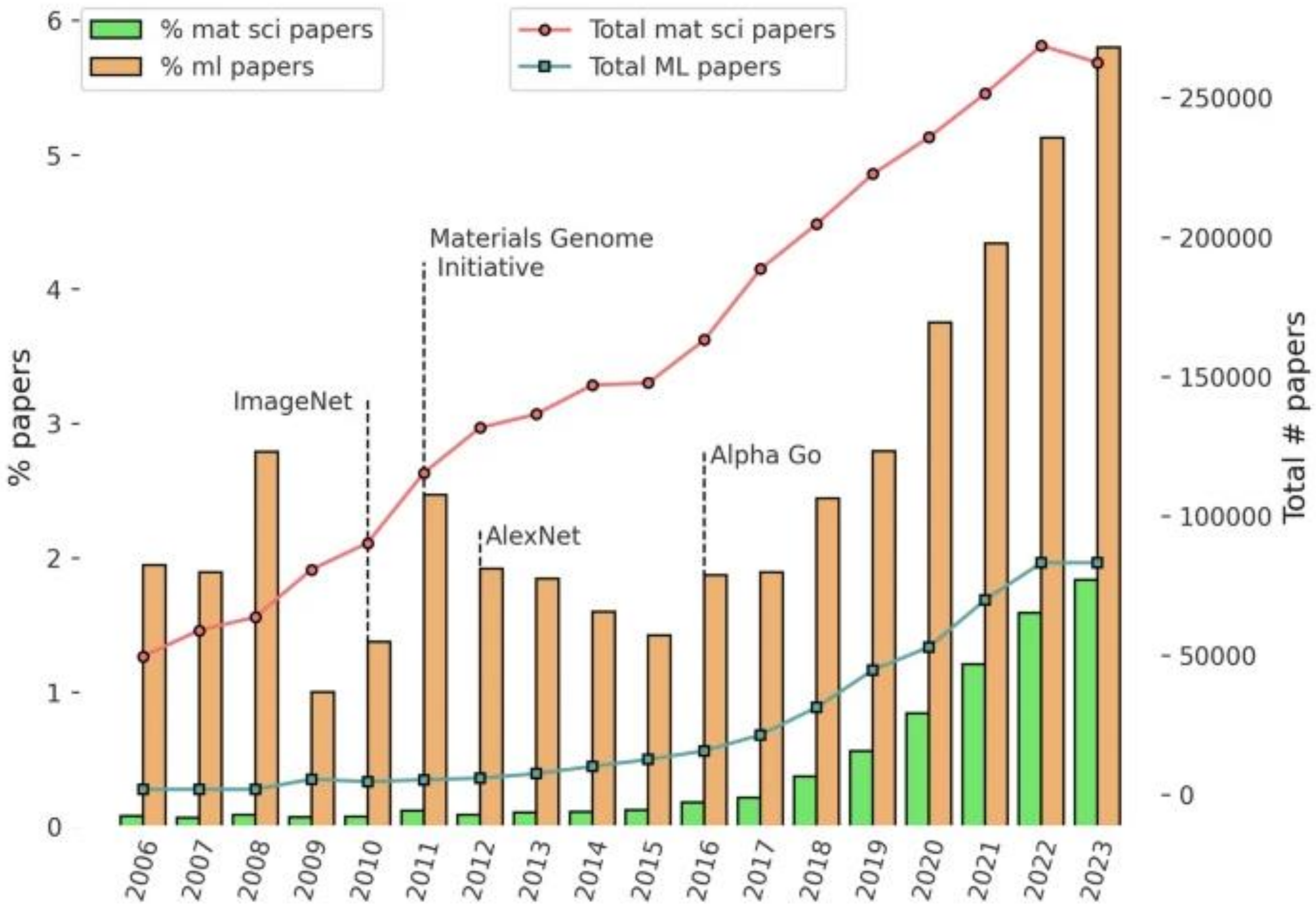
Fall 2025

Asst. Prof. Peter Schindler

Lecture 5 – Machine Learning Basics 1

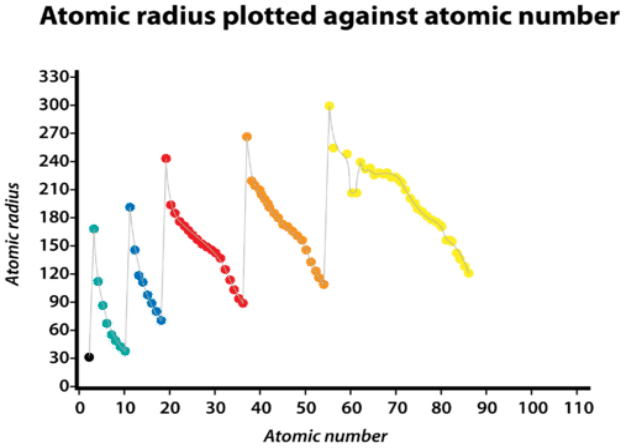
- Overview of Artificial Intelligence (AI) and Machine Learning (ML)
- Types of ML: Supervised vs. Unsupervised Learning
- Typical ML Pipeline
- Optimization and Gradient Descent
- Multi-linear Regression and Normal Equation
- Over/Underfitting and Bias vs. Variance Tradeoff
- Error Metrics and ML Terminology

ML in Materials Science is Hot, Hot, Hot!



https://www.nature.com/articles/s41524-024-01411-6

Data-Centric Pattern Matching Not New!



Period ↓ / Group →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57 La *	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	89 Ac **	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
			*	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	
			**	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	

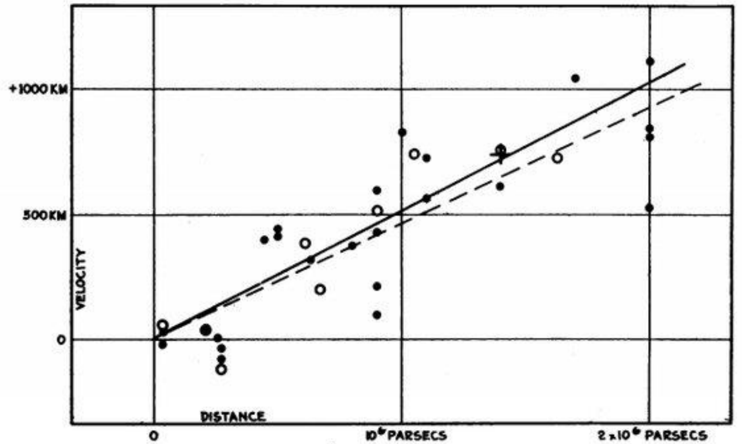
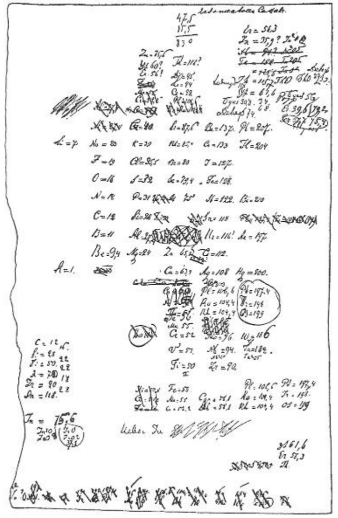


FIGURE 1
Velocity-Distance Relation among Extra-Galactic Nebulae.

Hubble (1929)



Draft of table from D. Mendelejeff, *Zeitschrift für Chemie* **12**, 405-406 (1869) (Work from Dmitrii Mendeleev)

Kepler: the first data scientist

What can one of the greatest geniuses in physics teach us about data analysis and artificial intelligence?

https://francisco-rodrigues.medium.com/kepler-the-first-data-scientist-70a0a0e4a2a8

Artificial Intelligence and Machine Learning

Artificial Intelligence

Computer Vision

Symbolic AI

NLP

Optimization

Machine Learning

Architectures: Shallow vs. Deep Learning

Supervised

Labeled Data

- Regression
- Classification
- Active Learning

Semi-Supervised

Partially Labeled Data

Self-Supervised

Masked/Hidden Data

Unsupervised

Unlabeled Data

- Clustering
- Dim. Reduction

Reinforcement

Reward (Feedback)

Game AI
Robotics
Real-time decisions

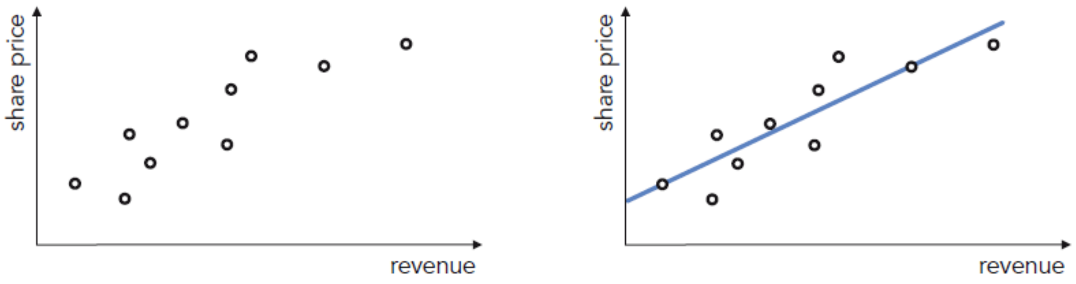
Generative AI

LLMs, Transformers

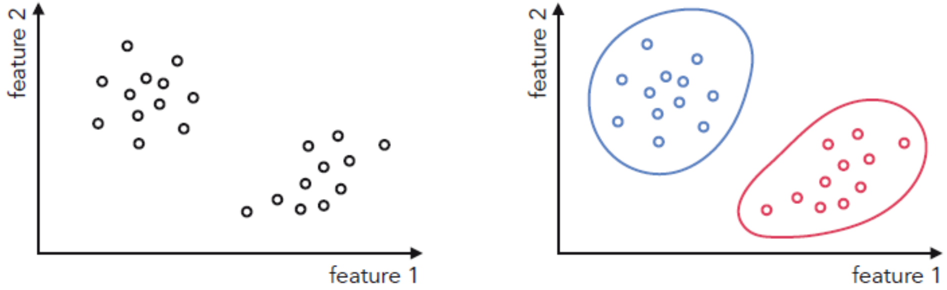
GANs, VAEs, Diffusion models

Supervised vs. Unsupervised Learning

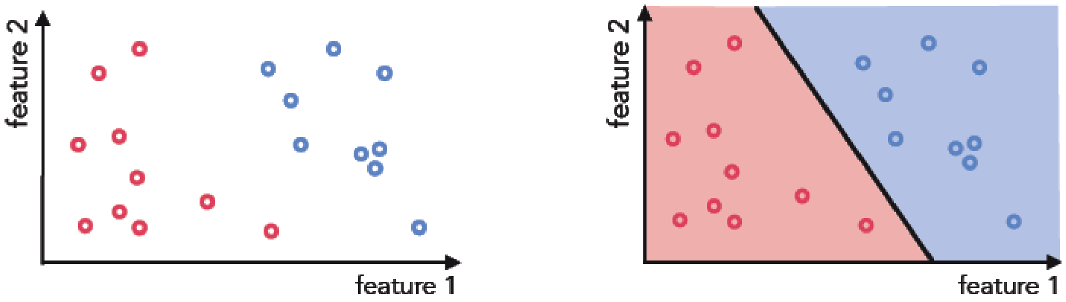
Supervised: Regression



Unsupervised: Clustering



Supervised: Classification

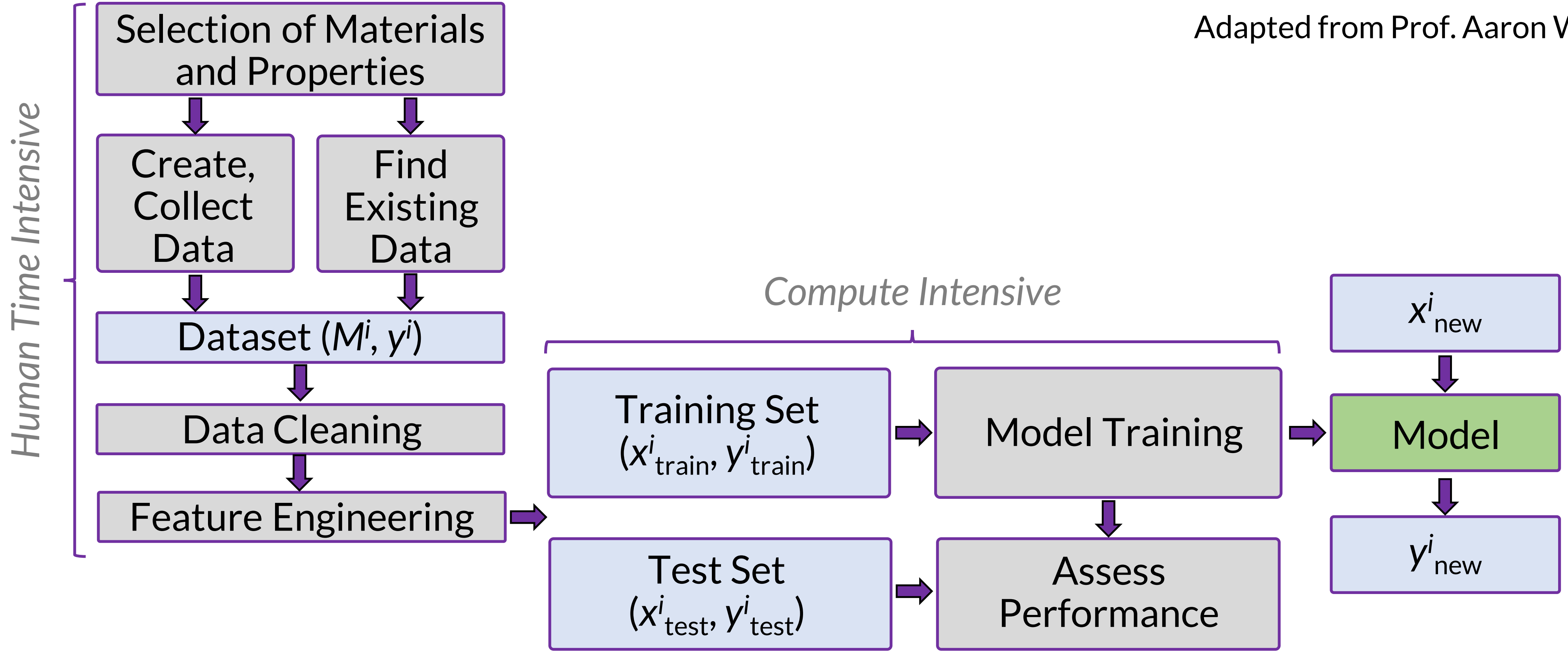


Unsupervised: Dim. Reduction

"*Machine Learning Refined*", 2nd Edition, J. Watt, R. Borhani, A. K. Katsaggelos

Machine Learning Pipeline

Adapted from Prof. Aaron Walsh



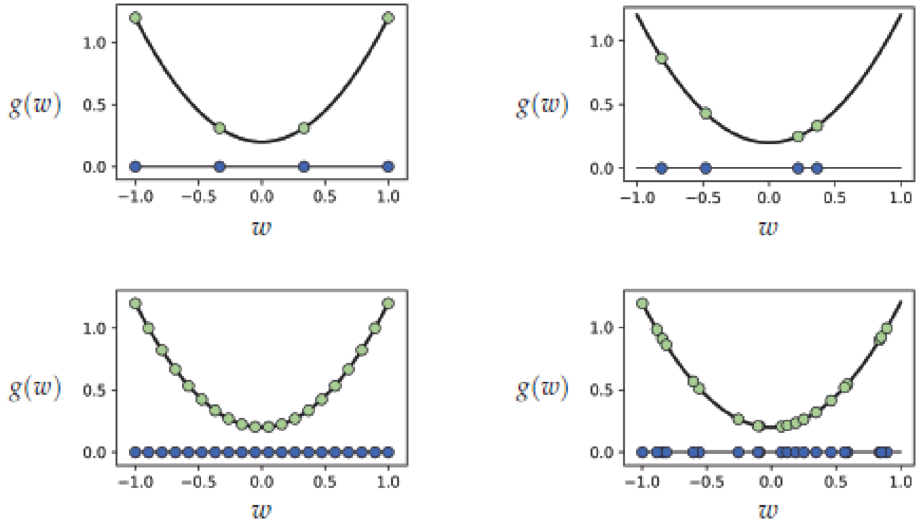
Arranging Data
(Data Engineer/Analyst)

Training and Model Development
(ML Researcher / Data Scientist / DevOps)

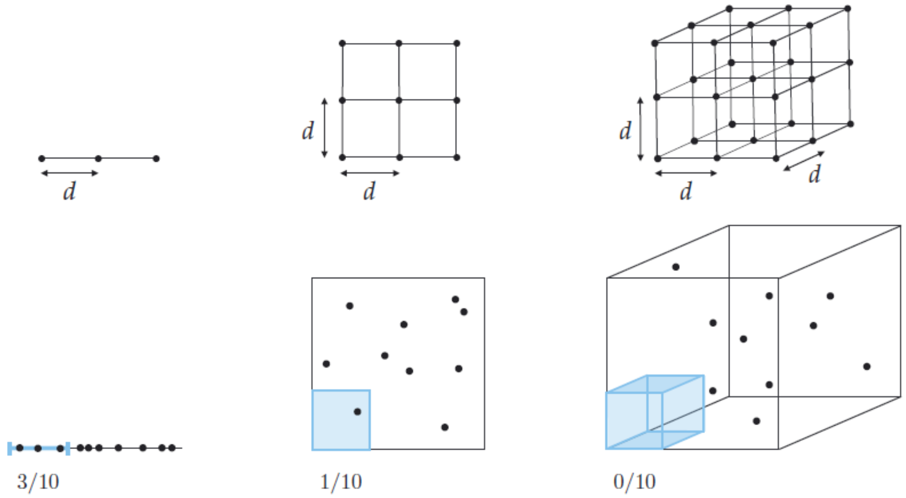
Deployment
(MLOps Engineer)

Zero-Order Optimization: The Curse of Dimensionality

- In ML we will need to optimize (minimize a cost function)
- *Zero-order optimization*:
  Sample points uniformly/randomly
  to find the minimum



- Curse of dimensionality

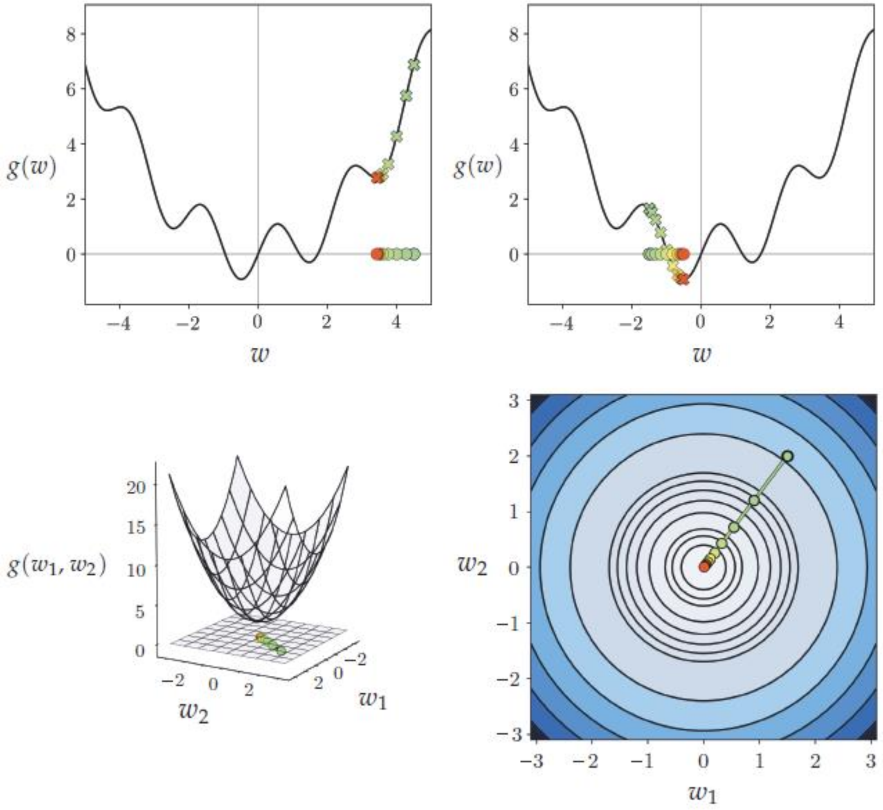


"*Machine Learning Refined*", 2nd Edition, J. Watt, R. Borhani, A. K. Katsaggelos

First-order Optimization: Gradient Descent

- Take steps in the direction of the gradient (“roll down the slope”)
- Need to pick a *learning rate*
- Non-convex functions
  still need sampling

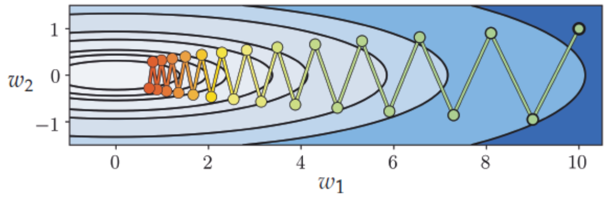
Interactive Demonstration

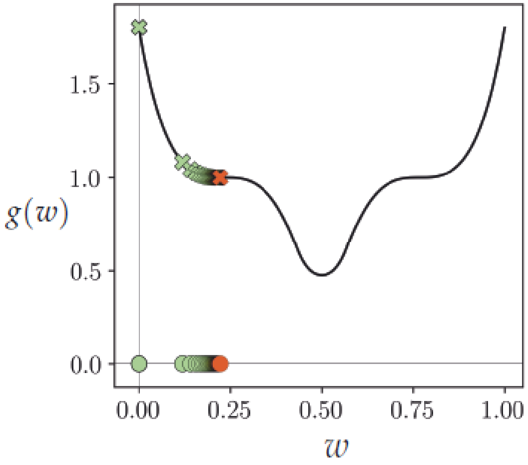


“*Machine Learning Refined*”, 2nd Edition, J. Watt, R. Borhani, A. K. Katsaggelos

Remark: Second-Order Optimization – Hessian / Newton's Method

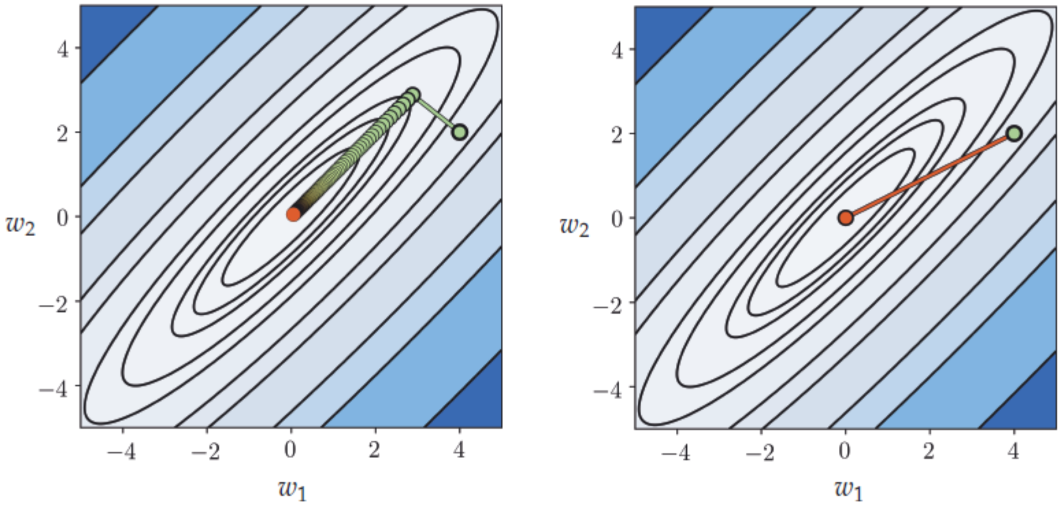
Gradient descent:
- Prone to zig-zagging
- Slow near saddle points





Newton's method:

Taking steps to stationary points of the **second-order** Taylor series of a function.



- Issues with nonconvex functions
- Computing Hessian for large *N* expensive

"*Machine Learning Refined*", 2nd Edition, J. Watt, R. Borhani, A. K. Katsaggelos

Multi-linear Regression: Gradient Descent vs. Normal Equation

On Board

Normal Equation: $\theta = (X^T X)^{-1} X^T y$

- Works for pseudo-inverse, which may be caused by
  - redundant features
  - too many features ($m \leq n$)
- No choosing of learning rate
- No iterations necessary
- Matrix operations slow for large number of feature n (>10,000)

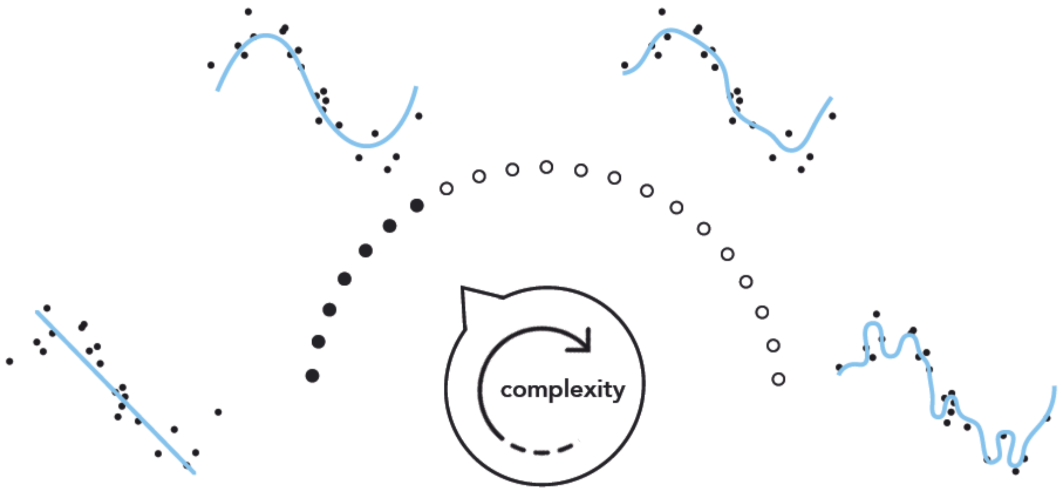
Generalized linear models:

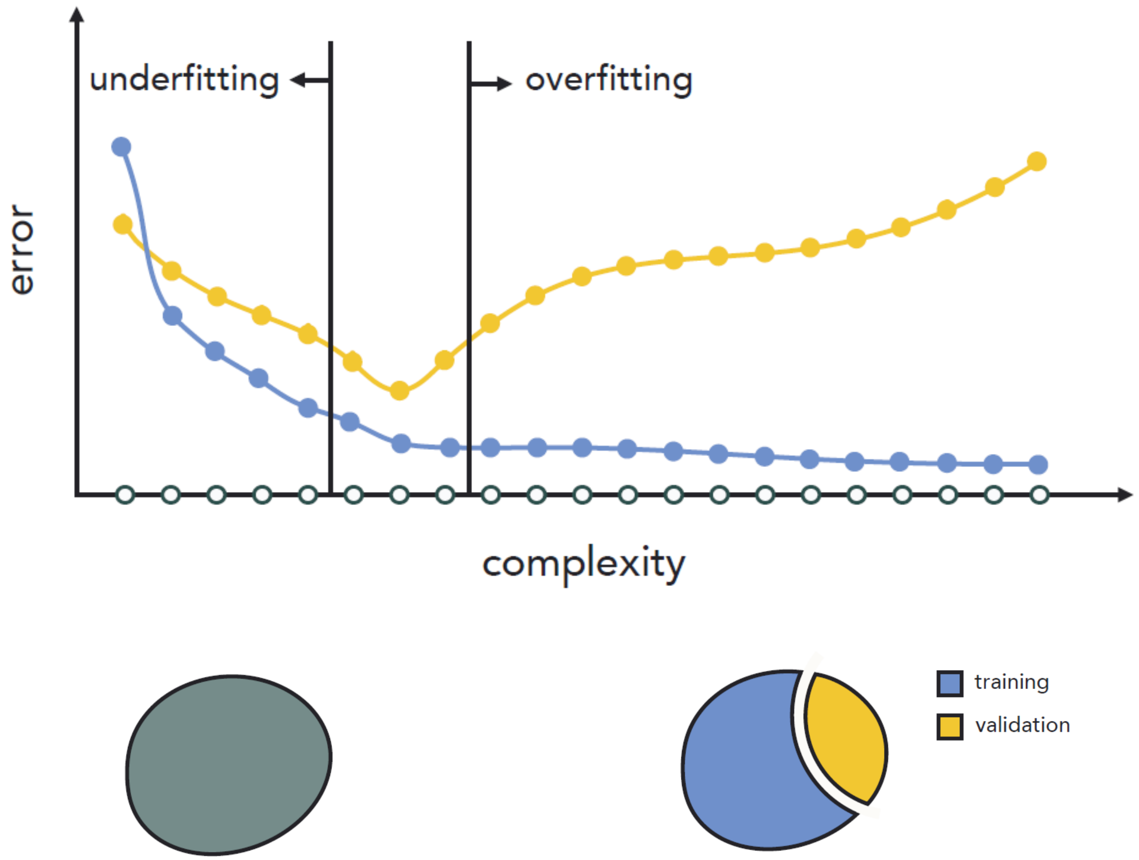
- Can easily incorporate nonlinearity to approx. functions (e.g., polynomials)

$$y = \theta_0 + \theta_1 f(x_1) + \theta_2 f(x_2) + \ldots$$

Interactive Demonstration

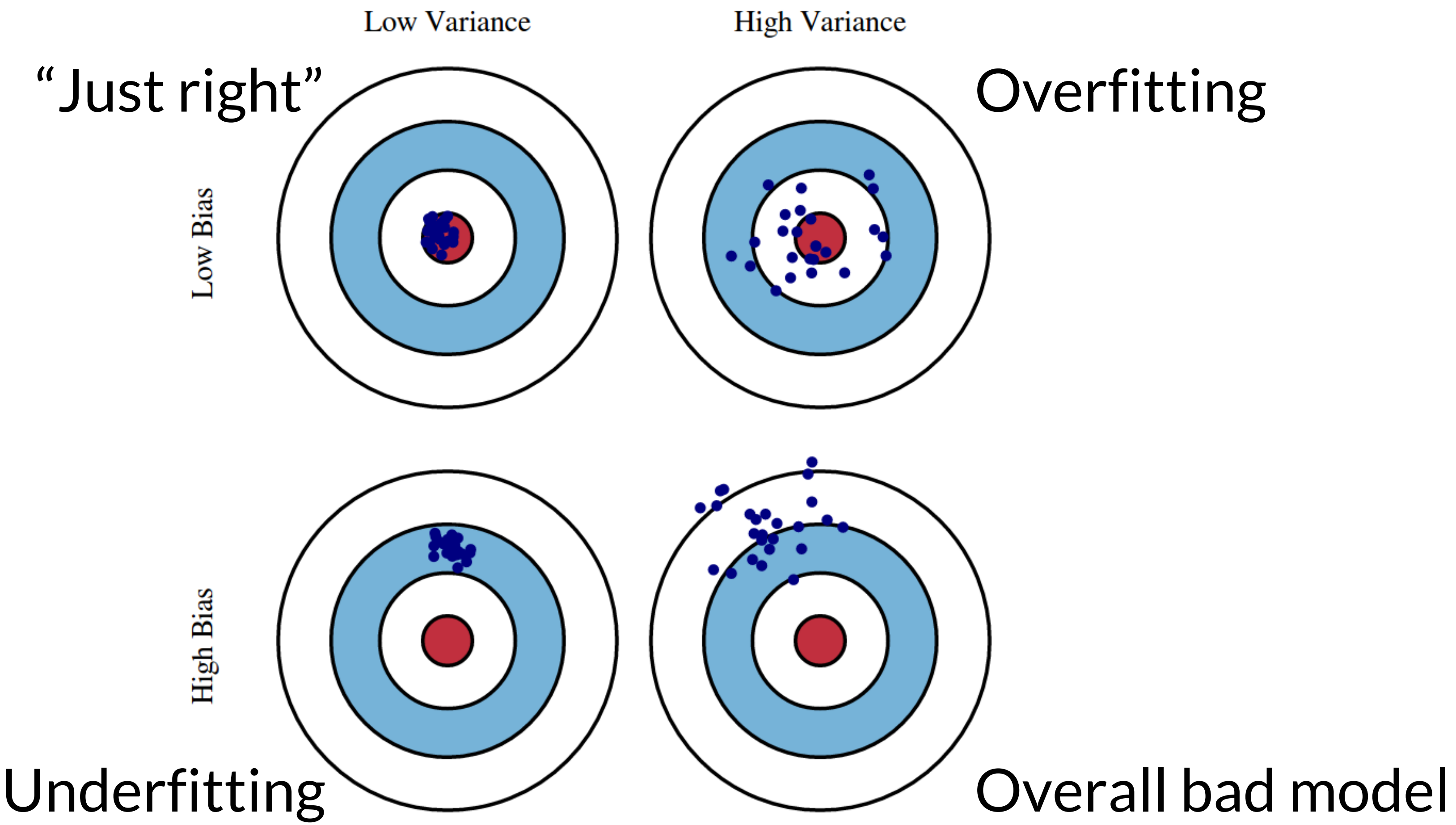
Overfitting and Underfitting





"*Machine Learning Refined*", 2nd Edition, J. Watt, R. Borhani, A. K. Katsaggelos

Bias and Variance



https://www.cs.cornell.edu/courses/cs4780/2015fa/web/lecturenotes/lecturenote12.html

Other Error Metrics and Terminology

- *Residual*: $r^i = y^i - y^i_{\text{pred}}$
- *Mean squared error*: $\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(r^i)^2$
- *Root mean squared error*: $\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(r^i)^2}$
- *Mean absolute error*: $\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}\left|r^i\right|$ (less sensitive to outliers)
- *Feature or Descriptor*: Input variable
- *Labelled example*: True value/answer paired with the input variables
- *Ground truth*: The true label
- *Hyperparameter*: Model parameters that are tuned during training
- More comprehensive glossary (found through Prof. Aaron Walsh): https://developers.google.com/machine-learning/glossary

Lecture Feedback

Please, scan the QR code and take a minute to let me know how the lecture was and mention any **feedback/questions**

This form is **anonymous!**