

IEMS 304 Lecture 1: Introduction to Statistical Learning

Yiping Lu

yiping.lu@northwestern.edu

Industrial Engineering & Management Sciences
Northwestern University



NORTHWESTERN
UNIVERSITY

Logistics

Textbook: James G, Witten D, Hastie T, et al. *An introduction to statistical learning.*

CS 229 Lecture Note: https://cs229.stanford.edu/main_notes.pdf

Time and Location: Monday, Wednesday and Friday, 9.00 A.M.- 9.50 A.M.
Tech L251

Office Hour: Friday: 1 P.M. Tech M237

TA Office Hour:

Pre-requisite and Pre-test

This is a **mathematically intense** course. But that's why it's exciting and rewarding!

Pre-requisite: A previous course in statistics at the level of IEMS 303 plus a course in matrix analysis. Comfort with programming (we will be programming in R) is also necessary.

Pre-test: Passing the pretest is worth 3% of your final course grade. You must achieve a passing score of 70% or higher by

Monday, Apr 15th at 11:59 p.m. This deadline will be firmly enforced.

Honor Code

Do's

- form study groups (with arbitrary number of people); discuss and work on homework problems in groups
 - write down the solutions independently
 - write down the names of people with whom you've discussed the homework
 - use ChatGPT as a TA
-

Don'ts

- It is an honor code violation to copy, refer to, or look at written or code solutions from a previous year, including but not limited to: official solutions from a previous year, solutions posted online, solutions you or someone else may have written up in a previous year, and solutions for related problems.
- Directly copy the answer from ChatGPT/Claude/Any GenAI

Lab Session

Homework

Publish on course website, due Friday (except pretests/midterm weeks)

Submit on Gradescope

Exams

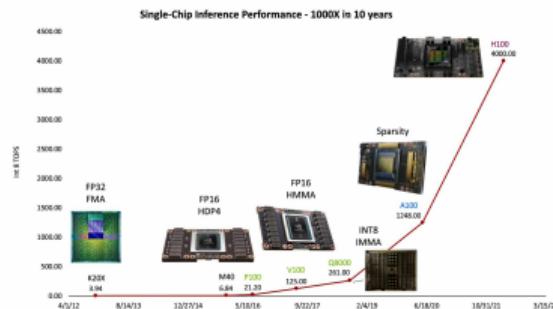
Let's Start

Massive Data

Massive complex data : Images, Acoustic signals, Text, ...

- Wikipedia pages: 13 millions (2014), 57 million (2022)
- Facebook users: 800 million (2014), 2.96 billion (2022)
- Flickr photos: 6 billion (2014), 10 billion (2022)
- Twitter tweets/day: 340 million (2014), 500 million (2022)
- Youtube video/min: 24 hours (2014), 500 hours (2022)
- Google pages: ≥ 1 trillion (2014), ≥ 130 trillions (2016)

Massive Computing : Huang's Law



Broad Applications in Science and Engineering

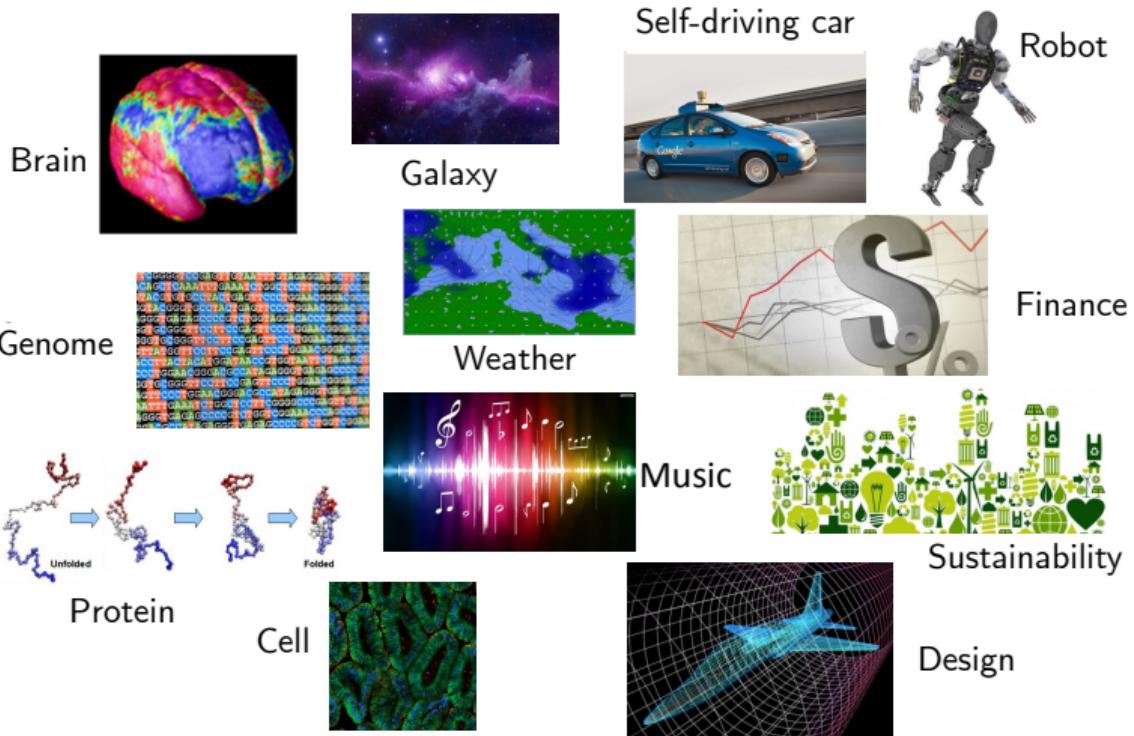
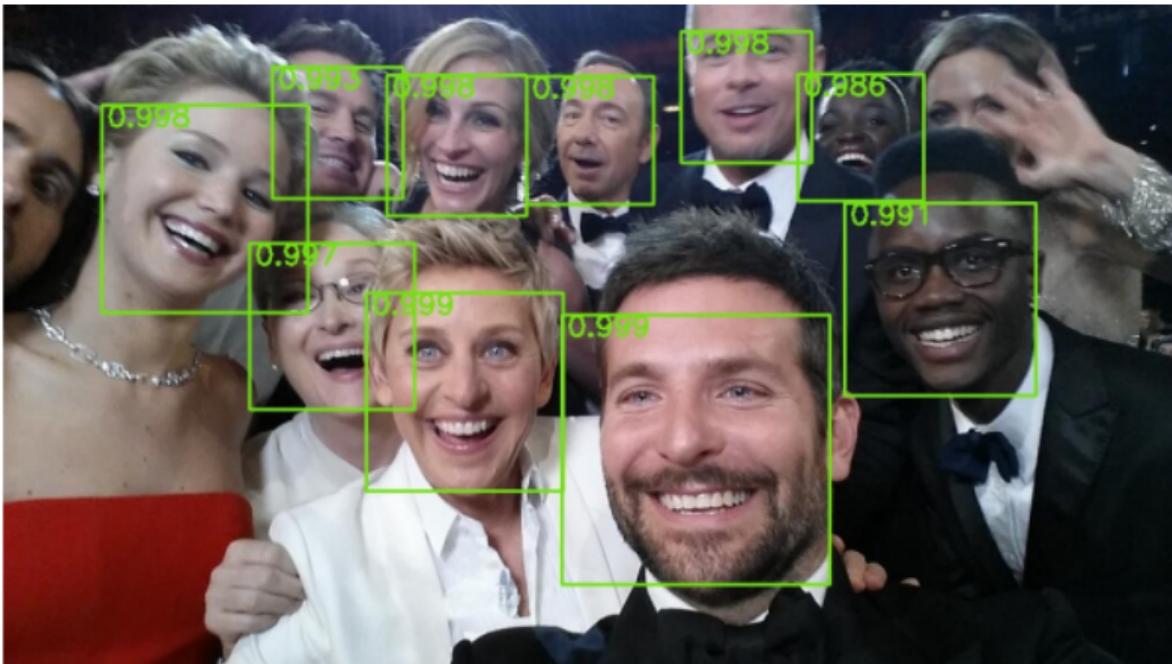


Image Classification

			
mite mite black widow cockroach tick starfish	container ship container ship lifeboat amphibian fireboat drilling platform	motor scooter go-kart moped bumper car golfcart	leopard leopard jaguar cheetah snow leopard Egyptian cat
			
grille convertible grille pickup beach wagon fire engine	mushroom agaric mushroom jelly fungus gill fungus dead-man's-fingers	cherry dalmatian grape elderberry ffordshire bullterrier currant	Madagascar cat squirrel monkey spider monkey titi indri howler monkey

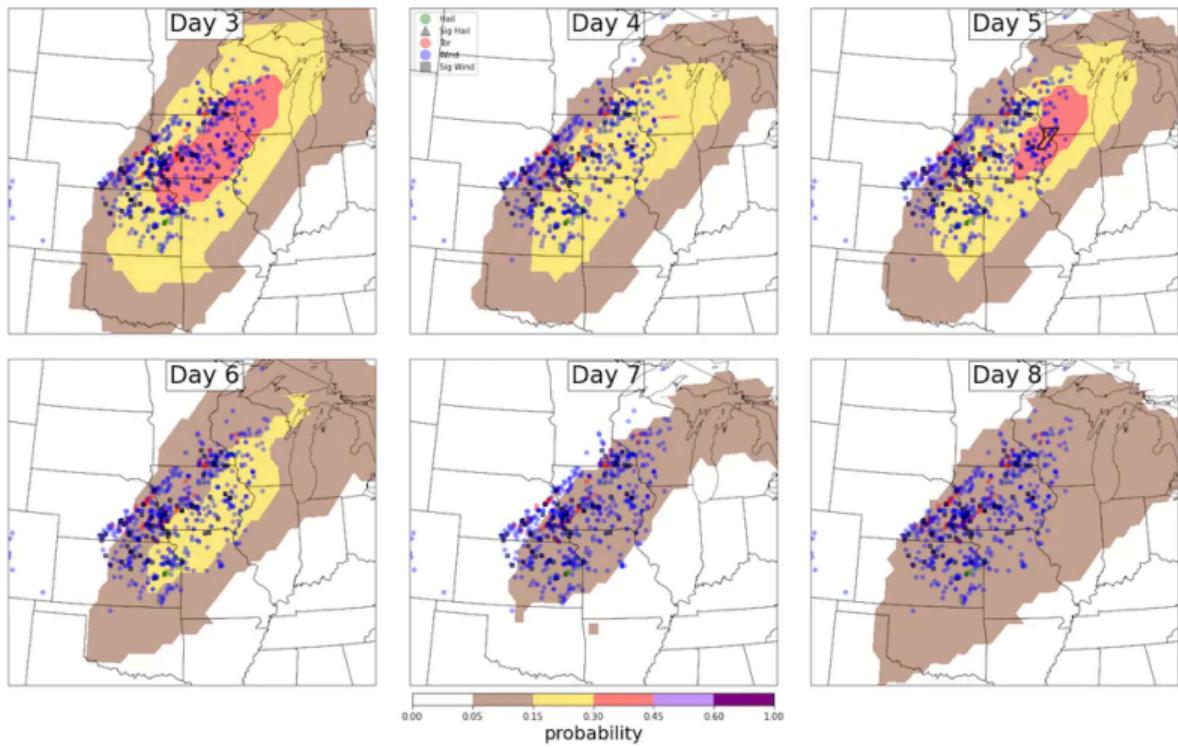
Face Detection



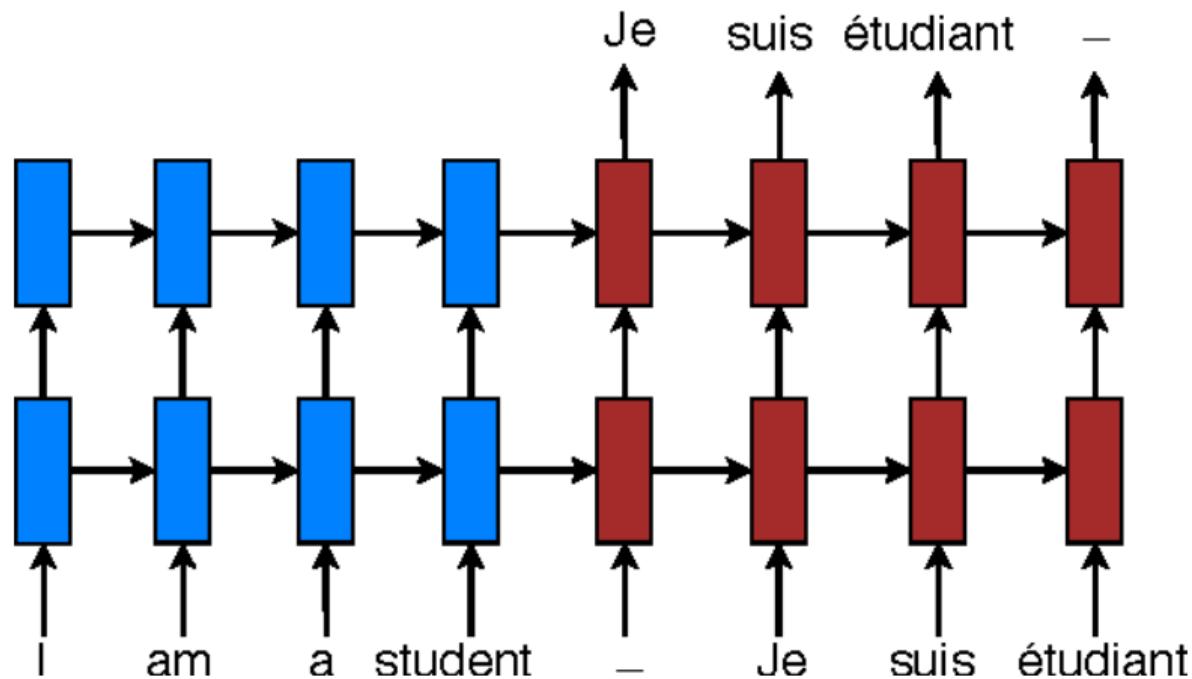
Spam Detection

Subject	Correspondents	Date
URGENT RFQ	• ← AL WALEED EQUIPMENTS	03/13/2017 06:55
New Order Attached **KINDLY SEND INVOICE	• ← starsescorts@gmail.com	03/15/2017 01:27
We're sad to let you know that our delivery was unsuccessful....	• ← Amr Hassan	03/15/2017 19:30
47929 username2	• ← FedEx Expedited Express	03/16/2017 02:53
Delivery Status Notification	• ← pkeith@gejlaw.com	03/16/2017 05:29
Formal Inquiry	• ← webmaster@stroy-exp...	03/16/2017 05:47
We have delivery problems with your parcel #7104543	• ← vowsbyjudy@shaw.ca	03/16/2017 14:38
INQUIRY	• ← "Anaïs VANACKER" <Va...	03/16/2017 21:16
54343 username	• ← webmaster@whfarm2....	03/17/2017 00:57
Item Delivery Notification	• ← Saigon Offshore	03/17/2017 03:47
UPS courier can not deliver parcel #004287245 to you	• ← dava@ac-lyon.fr	03/17/2017 14:25
Parcel Delivery Notification	• ← juanro5554@hotmail.c...	03/17/2017 14:48
Visa Card Award	• ← alifeof8@server.alifeofj...	00:34
Problems with item delivery, n.4930349	• ← webmaster@stroy-exp...	06:23
Package Delivery Notification	• ← abidjanbateau@vps286...	06:52
Delivery Status Notification	• ← info@visa.com	07:21
	• ← Apache	09:54
	• ← Apache	10:06
	• ← contrav8@box980.blue...	17:05

Weather Forecasting



Machine Translation

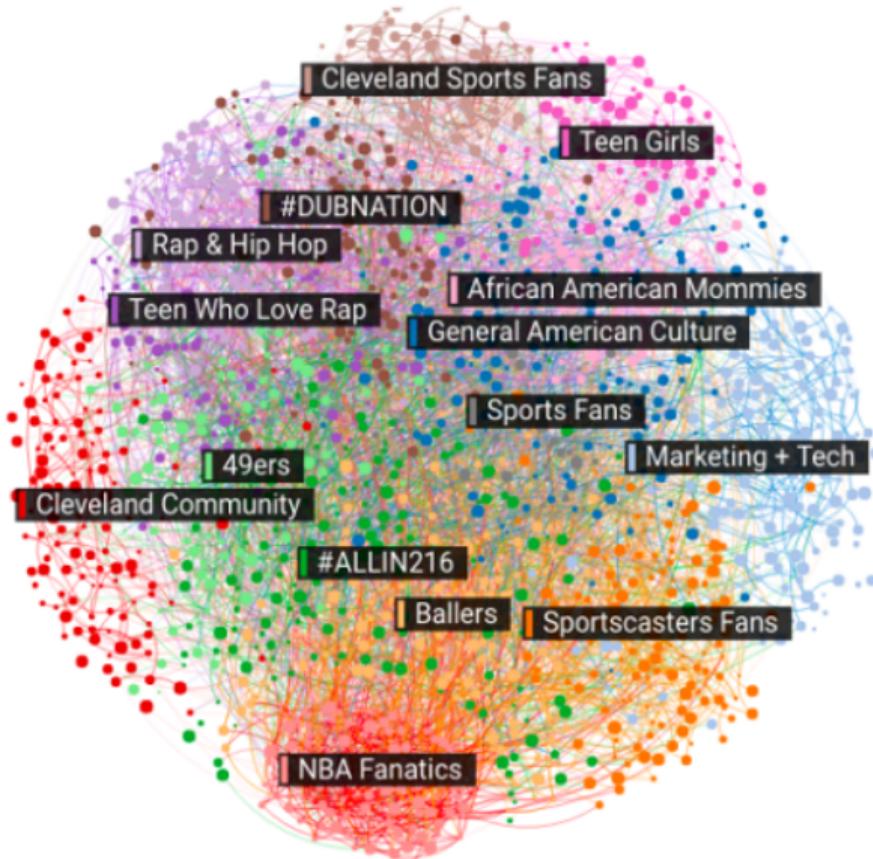


Autonomous Driving

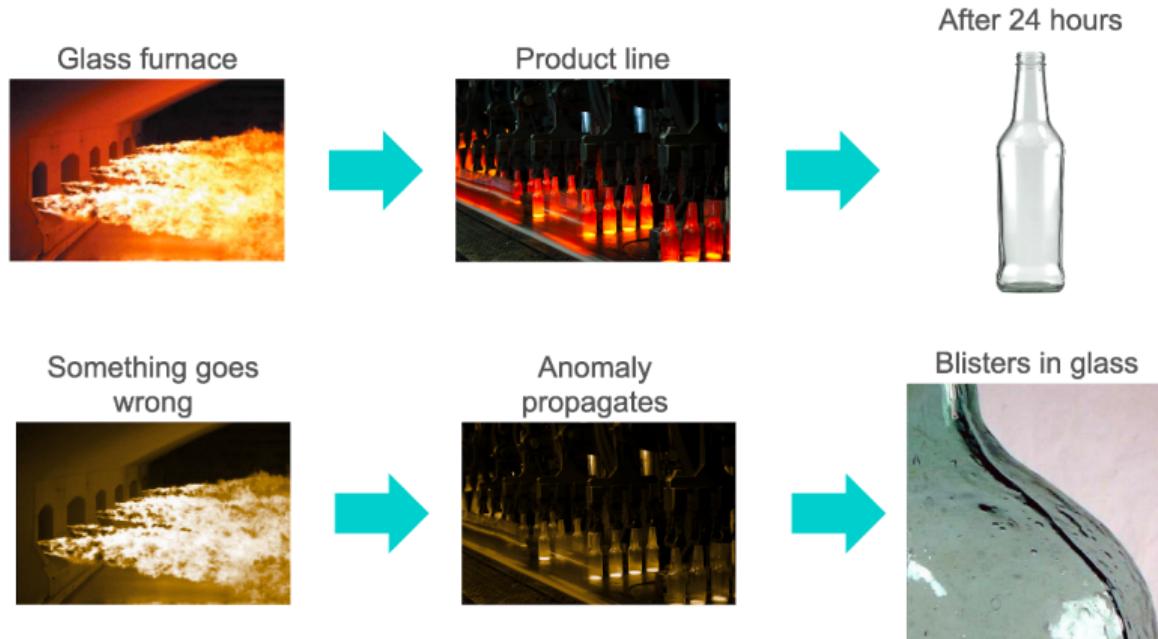


**Do We Always have the input
output pair**

Community Detection

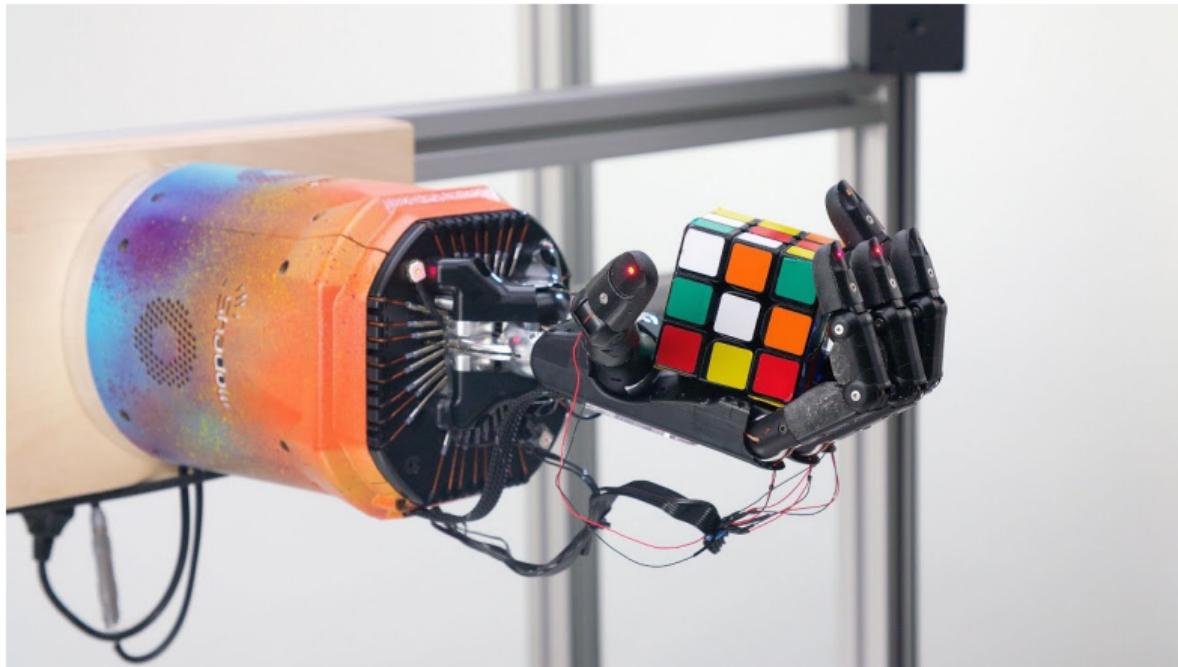


Anomaly Detection

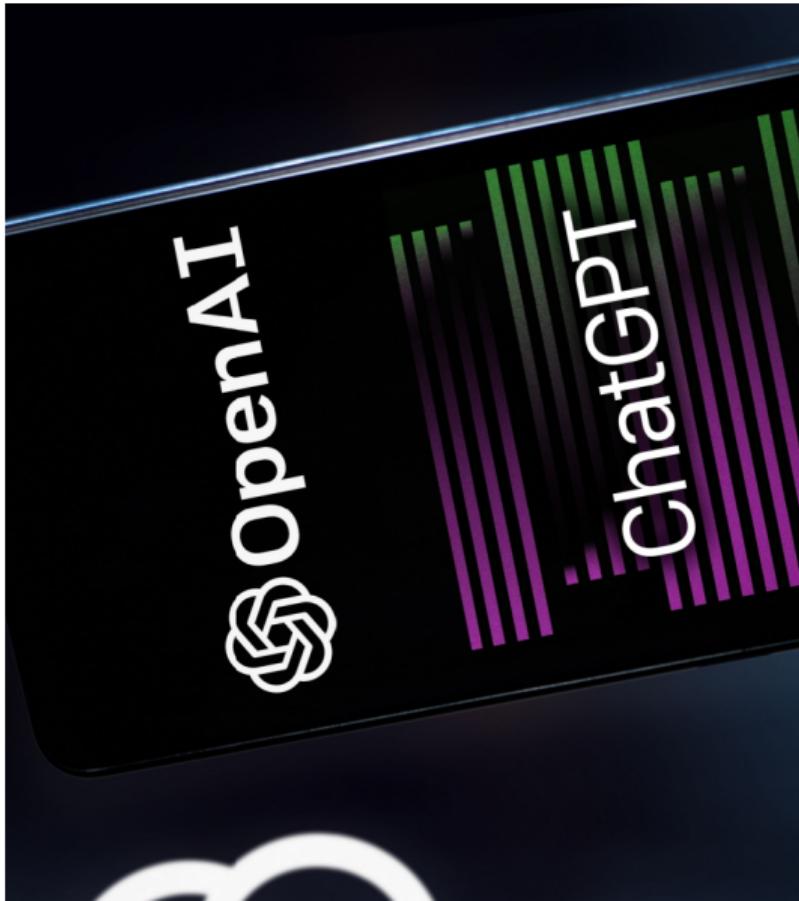


Movie Recommendation

							...
	★★★★★	?	★★★★★	?	?	?	...
	?	★★★★★	?	?	★★★★★	?	...
	?	?	?	★★★★★	★★★★★	?	...
	?	★★★★★	★★★★★	?	?	★★★★★	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



Chatbot



MIDJOURNEY

AI



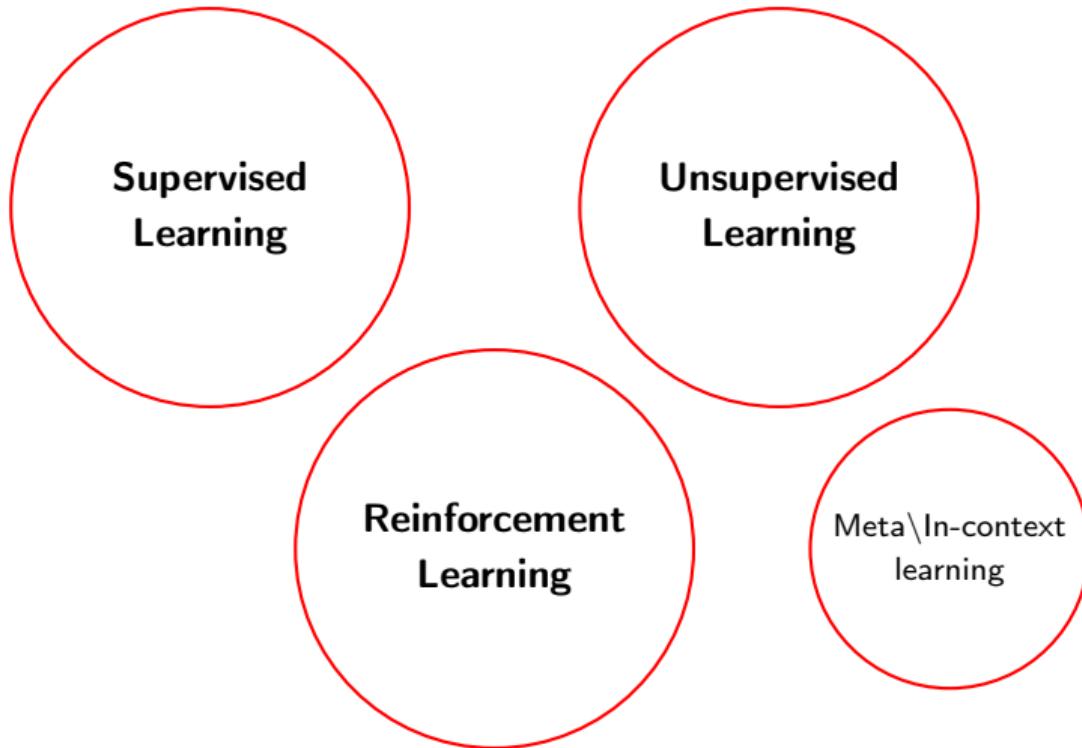
Introduction: Machine Learning

Tom Mitchell (1998): 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.

- Experience (data): games played by the program (with itself)
 - Performance measure: winning rate
-
- (⌚) We want to provide clear, interpretable models. These models allow you to understand the direct influence of each predictor on the outcome, which is essential in fields where insight into relationships (rather than just prediction) is needed.
 - (⌚) No confidence interval estimation
 - (⌚) In cases where data is scarce, simpler parametric models used in statistical learning can perform better. (**Why?**)

Regression: Predict the Unknown

Taxonomy of Machine Learning



Supervised Learning (Regression)

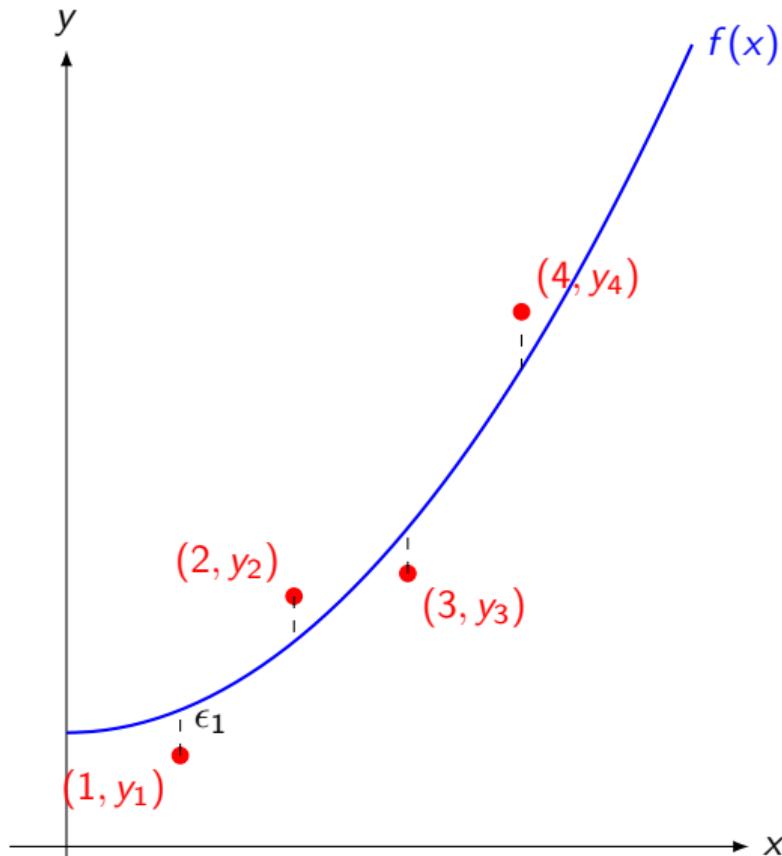
Supervised Learning: a set of observed data points $\{(x_i, y_i)\}_{i=1}^n$, where x_i represents the predictor (or vector of predictors) and y_i represents the response variable. Regression is the process of modeling the relationship between x and y by assuming:

$$y_i = f(x_i) + \epsilon_i,$$

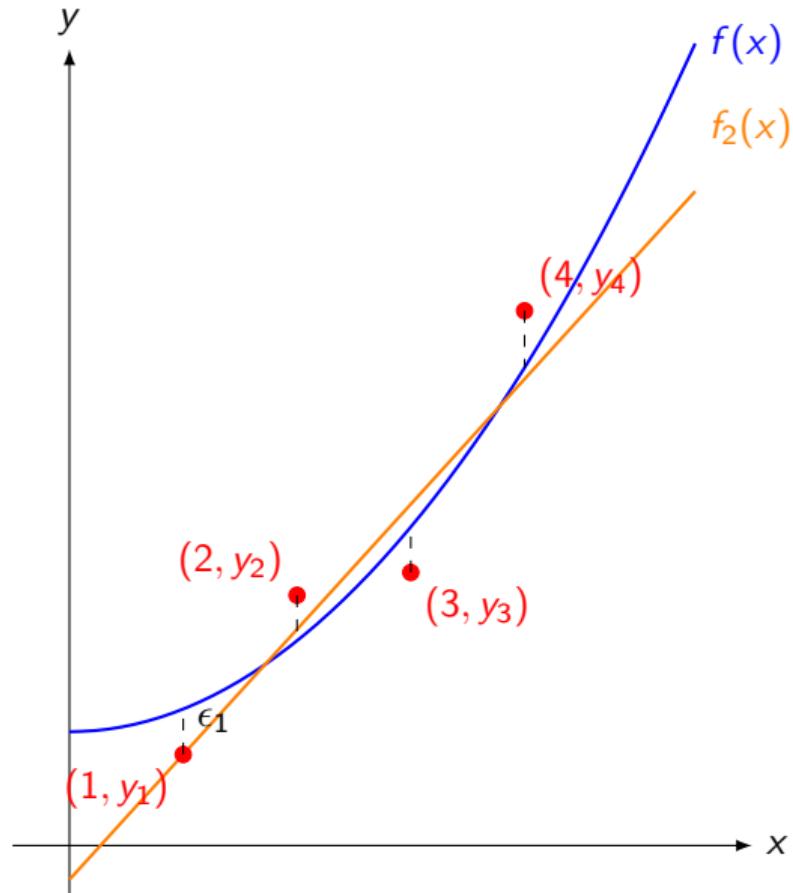
where:

- $f(x_i)$ is an unknown function that describes the systematic component of the relationship
- ϵ_i is a random error term.

Regression

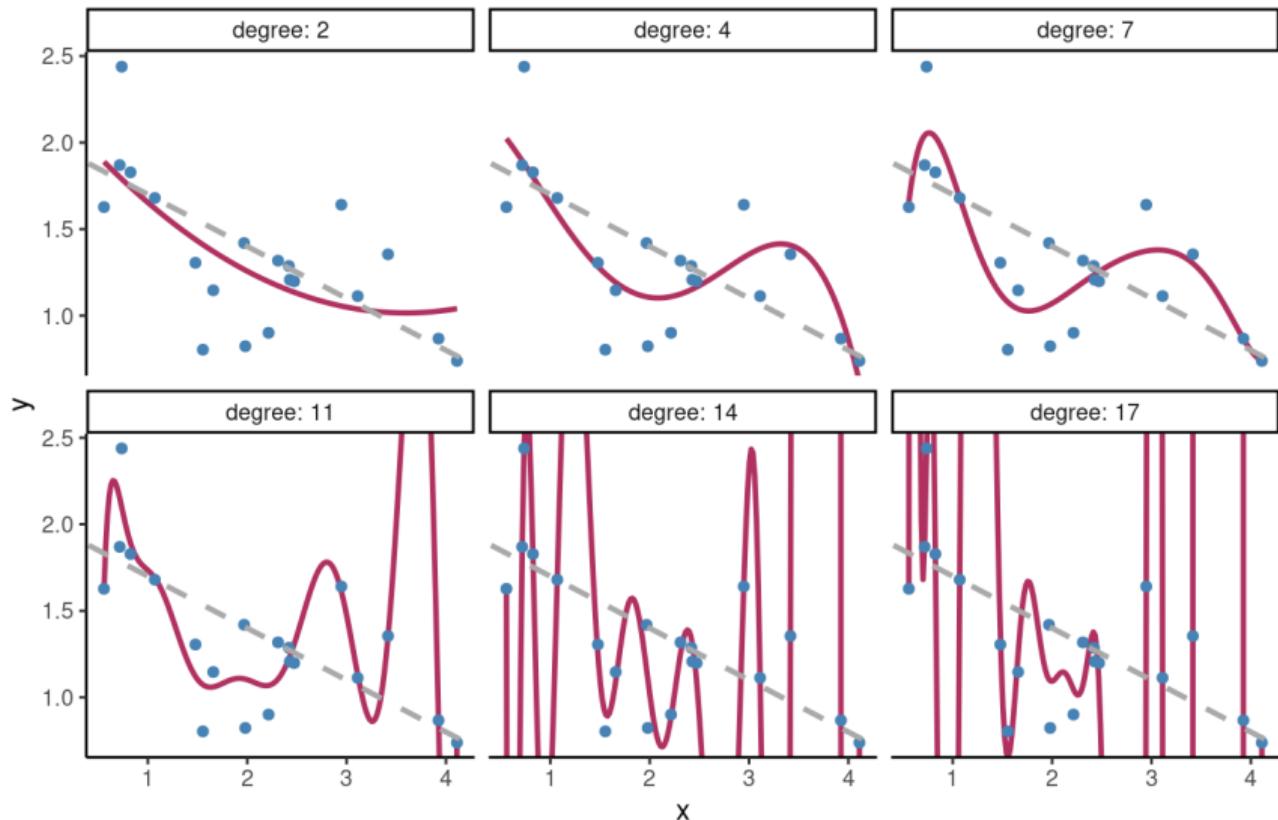


Regression



Runge Phenomenon

High degree polynomial models fit data better



Bias and Variance Trade-off

$$\mathbb{E}[(y - \hat{f}(x))^2] = \underbrace{\left(f(x) - \mathbb{E}[\hat{f}(x)]\right)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left[\left(\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\right)^2\right]}_{\text{Variance}} + \underbrace{\sigma^2}_{\text{Irreducible}}$$

-
- ⌚ An unbiased estimator could still make systematic mistakes – for example, if it overestimates 99% of the time, and underestimates 1% of the time *by a lot*, in expectation it could be unbiased.
 - ⌚ An unbiased estimator is **not** necessarily better than a biased estimator, because the total error depends on both the bias and variance of the estimator.

Bias and Variance Trade-off

	Underfitting	Just right	Overfitting
Symptoms	<ul style="list-style-type: none">• High training error• Training error close to test error• High bias	<ul style="list-style-type: none">• Training error slightly lower than test error	<ul style="list-style-type: none">• Very low training error• Training error much lower than test error• High variance
Regression illustration			
Classification illustration			

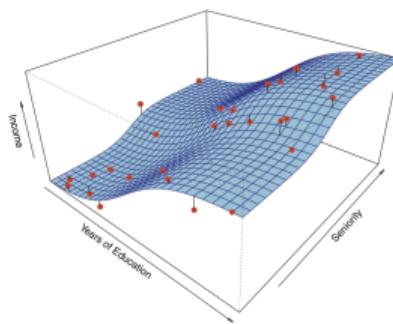
Prediction Accuracy and Model Interpretability

Why would we ever choose to use a more restrictive method instead of a very flexible approach?

High Dimensional Features

□ $x \in \mathbb{R}^d$

$$x = \begin{bmatrix} x_1 & \text{--- living size} \\ x_2 & \text{--- lot size} \\ x_3 & \text{--- \# floors} \\ \vdots & \text{--- condition} \\ x_d & \text{--- zip code} \end{bmatrix} \longrightarrow y \text{ --- price}$$



Data as a Matrix

Linear Algebra Review this **friday!**

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z							
1	tau																															
2	44	21	1	1	0.185	0.843	-0.027	-0.045	-0.315	0.36	0.641	0.784	-0.164	0.718	-0.437	-0.237	0.331	3.19	0.097	-0.859	-0.117	-0.132	-0.116	-0.334	-1.18	-0.134	-0.402					
3	44	21	1	1	1.89	0.484	-0.027	-0.045	-0.315	-0.208	0.909	1.09	-0.222	-0.223	-0.299	-0.282	0.238	-0.468	0.0897	0.0246	-0.509	-0.204	-0.426	-0.235	-0.334	-0.257	-0.076	-0.402				
4	44	21	1	1	-0.185	0.484	-0.027	-0.045	0.484	-0.103	0.908	1.09	-0.787	-0.817	-0.356	-0.382	-0.594	-0.432	-0.495	0.089	0.076	-0.54	-0.771	-0.544	-0.481	1.92	-0.355	-0.735				
5	44	21	1	1	0.487	0.177	-0.027	-0.045	-0.315	0.037	0.641	0.152	-0.408	0.163	0.216	0.388	-0.0586	-0.187	0.441	0.0246	2.46	0.084	-0.454	0.0246	2.46	0.55	-0.859	0.559	-0.773	-0.694	0.158	-0.402
6	44	21	1	1	-0.381	-0.53	-0.027	-0.045	-0.315	-0.389	0.134	0.62	0.169	-0.57	1.09	-0.844	0.804	-0.576	-0.454	0.0246	2.46	0.55	-0.859	-0.559	-0.773	-0.694	0.158	-0.402				
7	44	21	1	1	0.0361	-0.195	-0.027	-0.045	-0.315	-0.32	0.641	-0.784	-0.813	-0.822	-0.415	-0.998	-0.831	-0.143	-1.26	-0.839	-0.509	-0.397	-0.804	-0.602	-1.07	0.179	0.205	1.87				
8	44	21	1	1	-0.538	-0.27	-0.027	-0.045	1.88	-0.847	0.31	0.62	0.169	-0.857	-0.34	-0.41	-1.15	-0.867	-0.509	-0.397	-0.804	-0.602	-1.07	0.179	0.205	1.87						
9	44	21	1	1	1.43	-0.847	-0.027	-0.045	-0.315	-0.324	0.641	-0.784	-0.813	-0.822	-0.415	-0.998	-0.831	-0.143	-1.26	-0.839	-0.509	-0.397	-0.804	-0.602	-1.07	0.179	0.205	1.87				
10	44	21	1	1	-0.0977	-0.413	-0.027	-0.045	-0.315	0.398	0.641	0.462	-0.352	-0.686	0.317	0.851	-0.6	-0.042	-0.487	-0.878	-0.876	-0.718	-0.556	-0.2705	0.837	-0.179	0.733	-0.402				
11	44	21	1	1	0.511	-1.11	-0.027	-0.045	0.484	2	0.134	0.62	-0.0762	1.04	-0.256	-0.227	0.0866	-0.398	0.26	0.0446	-0.509	0.222	0.0352	0.226	0.397	-0.694	0.733	-0.402				
12	44	21	1	1	0.488	-0.348	-0.027	-0.045	0.484	0.783	0.641	-0.784	-0.459	-0.558	-0.437	-0.327	-0.111	0.255	-0.133	-0.839	-0.509	-0.412	-0.342	-0.618	-0.334	-0.257	0.733	-0.402				
13	44	21	1	1	-0.443	-0.0396	-0.027	-0.045	-0.315	-0.265	0.641	0.152	0.0948	-0.508	0.391	1.63	-0.274	0.649	0.0246	1.75	0.076	0.396	0.48	0.403	1.71	-0.107	-0.334	-0.402				
14	44	21	0	0	0.0548	1.23	-0.027	-0.045	-0.315	-0.265	0.641	-0.784	-0.285	-0.189	-0.437	-0.581	-0.507	-0.576	0.0449	0.0246	-0.509	0.6	-0.236	-0.602	-0.354	-0.694	0.158	-0.402				
15	44	21	0	0	0.594	0.812	-0.027	-0.045	0.484	-1.18	0.134	0.169	-0.315	1.04	1.09	1.09	0.096	-0.576	-0.517	0.12	0.0446	0.47	0.212	0.477	0.91	-0.479	0.817	-0.402				
16	44	21	0	0	0.6889	0.0875	-0.027	-0.045	0.484	-0.195	0.134	0.62	-0.0763	-0.138	-0.195	-0.205	-0.284	-0.289	-0.298	-0.309	-0.309	-0.309	-0.309	-0.309	-0.309	-0.309	-0.309	-0.309	-0.309			
17	44	21	0	0	0.117	1.01	-0.027	-0.045	-0.315	-0.494	0.641	0.86	-0.132	-0.517	-0.37	-0.389	0.083	0.432	-0.141	-0.888	-0.509	-0.214	-0.0678	-0.134	0.191	0.179	0.733	-0.402				
18	44	21	0	0	1.18	0.484	-0.027	-0.045	-0.315	-0.745	0.641	-0.784	-0.103	-0.477	-0.543	0.429	-0.143	-0.275	-0.839	-0.509	0.018	0.176	0.0200	0.397	-0.372	0.733	-0.402					
19	44	21	0	0	0.688	-0.107	-0.027	-0.045	-0.315	0.574	0.134	0.152	1.59	1.21	1.84	-0.582	0.285	0.613	0.084	-0.454	-0.509	1.19	0.174	1.2	-0.188	4.11	-0.678	0.735				
20	44	21	0	0	0.13	0.349	-0.027	-0.045	-0.315	-0.799	0.134	-0.326	-0.509	0.641	-0.182	0.0743	-0.571	-0.859	-0.509	-0.253	-0.606	-0.254	-0.775	-0.775	-0.852	-0.402						
21	44	21	0	0	-0.0682	-0.8952	-0.027	-0.045	-0.315	-0.489	0.641	-0.784	-0.759	0.795	0.31	1.77	-0.437	-0.555	3.57	0.848	-0.878	-0.896	1.62	-0.276	1.84	-0.775	-0.994	0.733	-0.402			
22	44	21	0	0	1.32	-0.878	-0.027	-0.045	0.484	-0.265	0.641	-0.316	0.348	0.14	-0.437	0.697	1.06	1.04	0.544	-0.134	-0.859	-0.509	-0.3974	0.22	-0.3661	0.89	-0.257	0.733	-0.402			
23	44	21	0	0	0.488	-0.0489	-0.027	-0.045	-0.315	-0.799	0.641	-0.116	-0.256	-0.316	0.314	-0.194	-0.027	-0.045	-0.859	-0.509	-0.397	-0.859	-0.509	-0.397	-0.859	-0.509	-0.397	-0.859				
24	44	21	0	0	0.475	1.177	-0.027	-0.045	-0.315	-0.799	0.641	-0.784	-0.119	-0.384	-0.448	-0.548	-0.511	0.132	0.591	0.981	0.794	0.481	-0.297	0.793	-0.463							
25	44	21	0	0	0.821	-0.587	-0.027	-0.045	-0.315	-0.799	0.641	-0.318	-0.117	-0.113	-0.133	-0.214	-0.248	-0.074	-0.084	-0.604	-0.399	-0.398	0.179	0.733	-0.402							
26	44	21	0	0	0.978	-1.42	-0.027	-0.045	-0.315	-1.18	0.134	0.152	1.88	1.86	0.998	-0.436	-0.382	3.03	0.304	-1.02	-0.448	-0.509	0.114	0.274	0.118	-0.334	-0.3061	0.733	-0.402			
27	44	21	0	0	0.931	1.25	-0.027	-0.045	0.484	-0.854	0.641	-0.784	-0.531	-0.494	-0.437	-0.274	-0.223	-0.323	-0.018	-0.048	-0.246	-0.509	-0.612	0.417	-0.818	3.47	-0.394	-0.402				
28	44	21	0	0	0.791	-0.0419	-0.027	-0.045	-0.315	-0.98	3.28	0.31	0.43	1.89	-0.844	-0.42	-0.0718	-0.416	-0.649	-0.773	-0.888	-0.529	-0.531	-0.109	-0.694	-2.39	0.735					
29	44	21	0	0	-0.145	-0.377	-0.027	-0.045	-0.315	-0.899	0.909	2.49	0.487	0.334	-0.419	-0.89	1.28	1.09	-0.113	0.298	-0.806	-0.497	-0.811	-0.773	0.179	-0.28	0.402					
30	44	21	0	0	-0.377	-0.649	-0.027	-0.045	0.484	-0.612	-0.641	0.334	0.139	0.644	0.296	-0.227	-0.414	-0.689	1.01	0.865	0.349	0.277	-0.898	-0.188	-0.676	0.633	0.735					
31	44	21	0	0	1.45	-1.14	-0.027	-0.045	-0.315	-0.899	1.488	1.29	-0.161	-0.244	-0.164	-0.167	-0.165	-0.201	-0.201	1.73	-0.259	-0.259	-0.259	-0.259	-0.259	-0.259	-0.259	-0.259	-0.259			
32	44	21	0	0	0.877	0.724	-0.027	-0.045	-0.315	-0.899	0.641	-0.784	-0.198	-0.0897	0.3	-0.3667	-0.421	1.14	0.048	-0.509	1.29	-0.3658	-0.4046	0.619	0.733	-0.402						
33	44	21	0	0	0.688	-0.479	-0.027	-0.045	-0.315	-0.899	0.641	-0.236	-0.118	-0.0986	-0.0820	-0.0823	-0.237	0.508	-0.0895	-0.509	-0.301	-0.097	-0.302	0.817	0.815	0.154	-0.402					
34	44	21	0	0	0.4056	0.221	-0.027	-0.045	-0.315	-0.764	0.641	-0.784	-0.0894	-0.124	-0.0411	-0.0401	-0.0401	-0.124	-0.124	-0.0411	-0.0416	-0.0416	-0.0416	-0.0416	-0.0416	-0.0416	-0.0416	-0.0416	-0.0416			
35	44	21	0	0	-0.117	0.382	-0.027	-0.045	-0.315	-0.211	1.68	0.62	-0.148	1.25	-0.437	-0.538	0.35	0.147	0.539	-0.839	-0.509	-0.491	-0.288	-0.499	-0.627	0.179	-0.678	3.01				
36	44	21	0	0	0.487	-0.0396	-0.027	-0.045	-0.315	-1.19	0.641	-0.784	-0.388	-0.806	-0.0307	-0.538	-0.0542	0.082	-0.895	-0.509	-0.308	-0.809	-0.305	-0.827	-0.694	0.676	-0.402					
37	44	21	0	0	0.0381	-0.0852	-0.027	-0.045	-0.315	-3.08	0.641	-1.56	-0.158	-0.109	-0.094	-0.105	-0.133	-0.219	1.37	0.724	-0.889	-0.509	-0.3209	-0.02169	-0.021	0.887	0.815	0.058	-0.402			
38	44	21	0	0	0.334	-0.649	-0.027	-0.045	-0.315	-0.899	0.641	-0.236	-0.118	-0.0987	-0.084	-0.084	-0.136	-0.136	-0.084	-0.0893	0.011	-0.889	0.815	0.179	0.735							
39	44	21	0	0	0.442	1.33	-0.027	-0.045	-0.315	-0.439	0.641	-0.784	-0.728	1.2	-0.437	-0.356	-0.507	-0.484	-0.132	1.75	0.876	0.512	0.206	-0.637	-0.131	0.402						
40	44	21	0	0	0.845	-0.107	-0.027	-0.045	-0.315	-0.899	0.641	-0.158	-0.043	-0.152	-0.153	-0.106	-0.157	-0.878	-0.509	1.58	1.23	1.6	1.42	-0.694	-0.38	0.735						
41	44	21	0	0	-0.475	0.724	-0.027	-0.045	-0.315	-0.973	0.641	-0.236	-0.118	-0.0986	-0.0849	-0.0849	-0.136	-0.136	-0.084	-0.089	0.011	-0.1016	-0.0416	0.619	0.733	-0.402						
42	44	21	0	0	1.73	-1.07	-0.027	-0.045	-0.315	-0.446	0.641	-0.236	-0.118	-0.0984	-0.0849	-0.0849	-0.136	-0.136	-0.084	-0.089	0.011	-0.1016	-0.0416	0.735								
43	44	21	0	0	1.18	-0.282	-0.027	-0.045	-0.315	-0.899	0.641	-0.236	-0.118	-0.0986	-0.0849	-0.0849	-0.136	-0.136	-0.084	-0.089	0.011	-0.1016	-0.0416	0.735								
44	44	21	0	0	0.122	0.549	-0.027	-0.045	-0.315	-1.24	0.641	-0.236	-0.118	-0.0984	-0.0849	-0.0849	-0.136	-0.136	-0.084	-0.089	0.011	-0.1016	-0.0416	0.735								
45	44	21	0	0	0.484	-0.0396	-0.027	-0.045	-0.315	-0.643	0.641	-0.156																				

Confidence Interval in Finance: Impact of Outliers



Why is this Important?

In this example, the confidence interval for the expected return is between -50% and -40%, indicating that most outcomes are negative. However, a very rare event pushes the sample mean to 300%, which could give the false impression of high returns. This discrepancy shows that while the sample mean may appear attractive, the confidence interval reveals the underlying risk and variability in the data, emphasizing the need to consider the full range of possible outcomes when making financial decisions.

Why Sample Mean?

Consider a dataset x_1, x_2, \dots, x_n . We consider L2 loss (or squared error loss) function with respect to a constant c as the performance measure P:

$$L(c) = \sum_{i=1}^n (x_i - c)^2.$$

To find the minimizer, differentiate $L(c)$ with respect to c :

$\frac{dL}{dc} = \sum_{i=1}^n 2(x_i - c)(-1) = -2 \sum_{i=1}^n (x_i - c)$. Setting the derivative equal to zero gives:

$$-2 \sum_{i=1}^n (x_i - c) = 0 \implies \sum_{i=1}^n (x_i - c) = 0.$$

Expanding the sum:

$$\sum_{i=1}^n x_i - nc = 0 \implies c = \frac{1}{n} \sum_{i=1}^n x_i.$$

Thus, the minimizing value of c is the sample mean: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Different Prediction

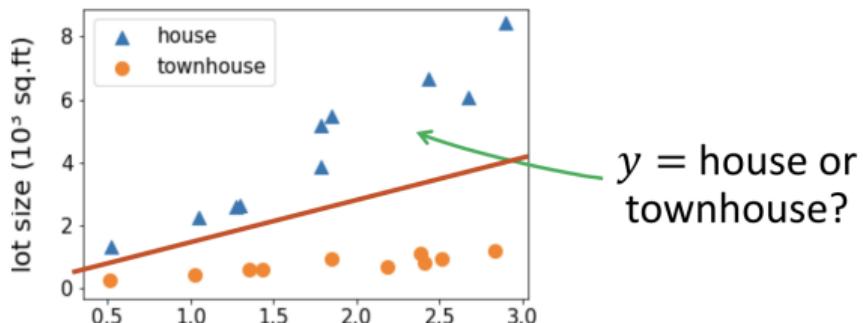
- Point Prediction : retrun $\hat{f}(x)$ since it returns a number.
- Interval Prediction , e.g., Y will be within an interval $[l, u]$ with probability $1 - \alpha$
- distributional prediction , e.g. Y will follow an $N(m, v)$ distribution.

Classification

Classification

- Regression : if $y \in \mathbb{R}$ is a continuous variable
- classification : the label is a discrete variable

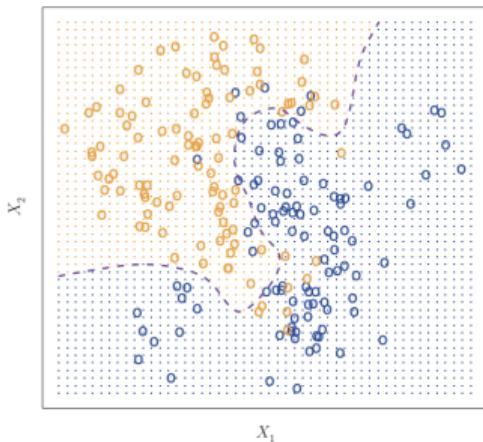
(size, lot size) → house or townhouse ?



Classification as Regression: Bayes Classifier

training error rate: $\frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{y}_i)$

Here the function $I(y_i \neq \hat{y}_i)$ is an indicator variable that equals 1, if $y_i \neq \hat{y}_i$ and 0 otherwise. If $y_i \neq \hat{y}_i$, then the i -th observation was classified incorrectly; otherwise it was not misclassified.



Consider random label: $\mathbb{P}(Y = j | X = x_0)$.
The Bayes classifier returns

$$1 - \max_j \mathbb{P}(Y = j | X = x_0)$$

produces the lowest possible test error rate,
called the *Bayes error rate* is given by

$$\underbrace{1 - \mathbb{E}\left[\max_j \mathbb{P}(Y = j | X)\right]}_{\text{Irreducible}}$$

x and y in Computer Vision

Task. Image Classification

$x = ?, y = ?$

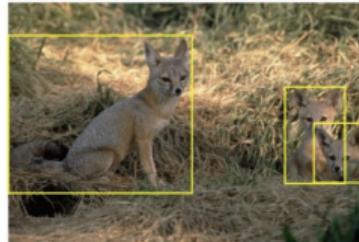
ILSVRC



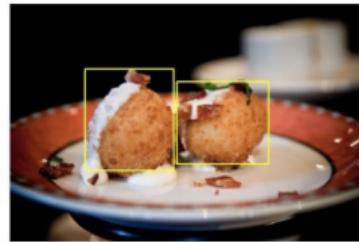
x and y in Computer Vision

Task. Object localization and detection

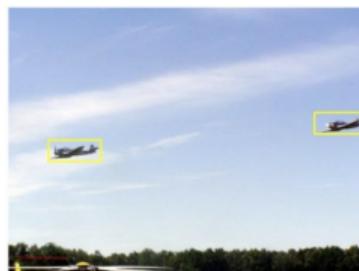
$x = ?, y = ?$



kit fox



croquette



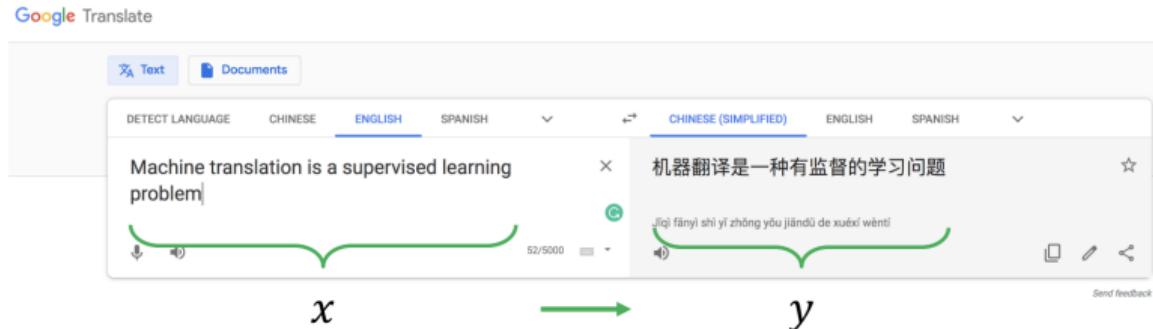
airplane



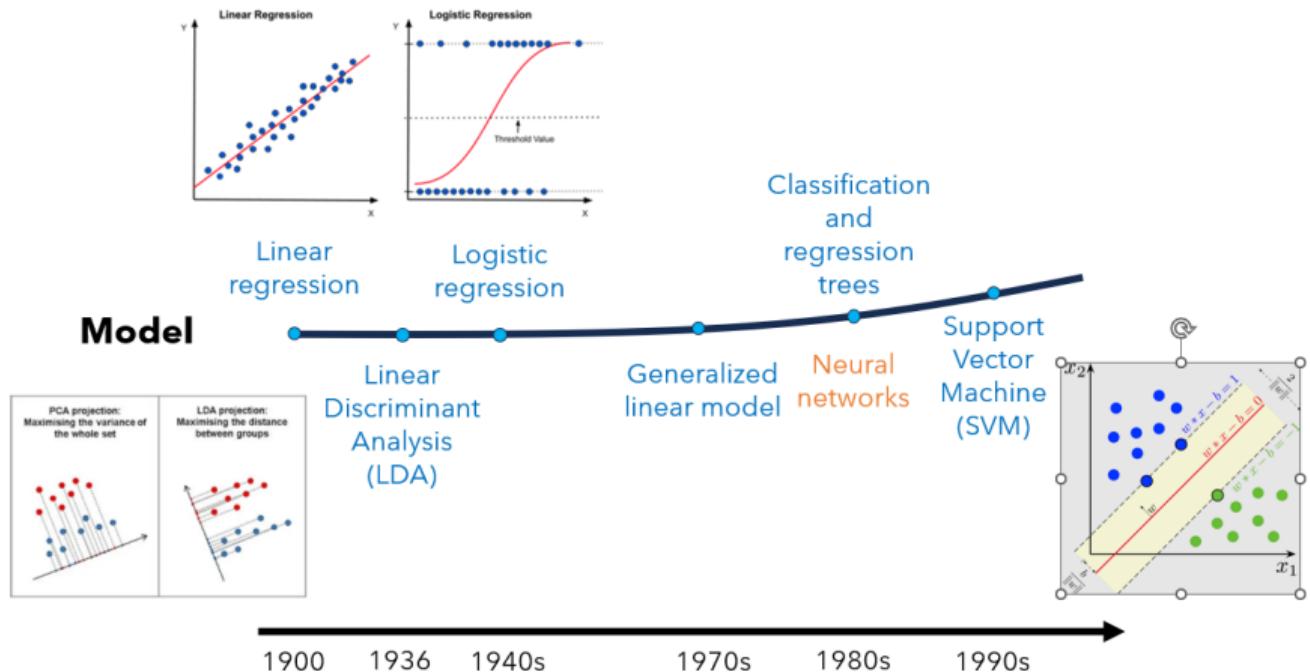
frog

x and y in Natural Language

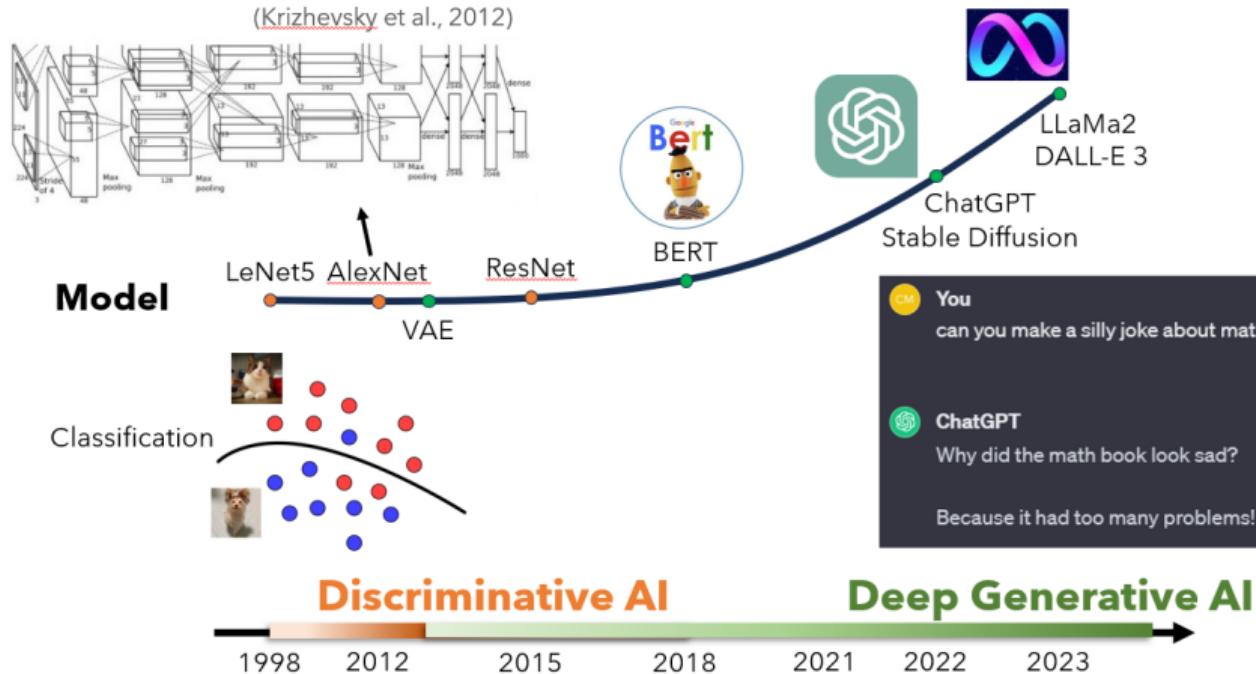
Task. Machine Translation d $x = ?, y = ?$



Early History



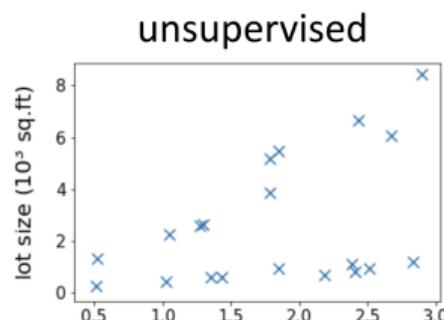
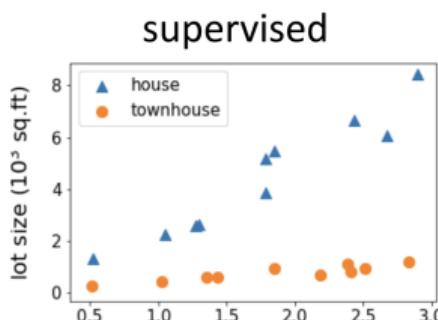
Contemporary Developments



Unsupervised Learning

Unsupervised Learning (Clustering)

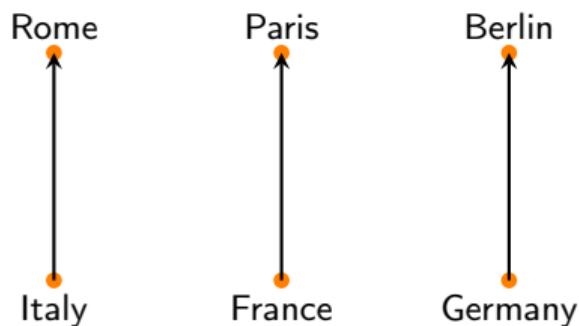
- ❑ Dataset contains **no** labels: $x^{(1)}, x^{(2)}, \dots, x^{(n)}$
- ❑ Goal (**vaguely-posed**): to find interesting structures in the data



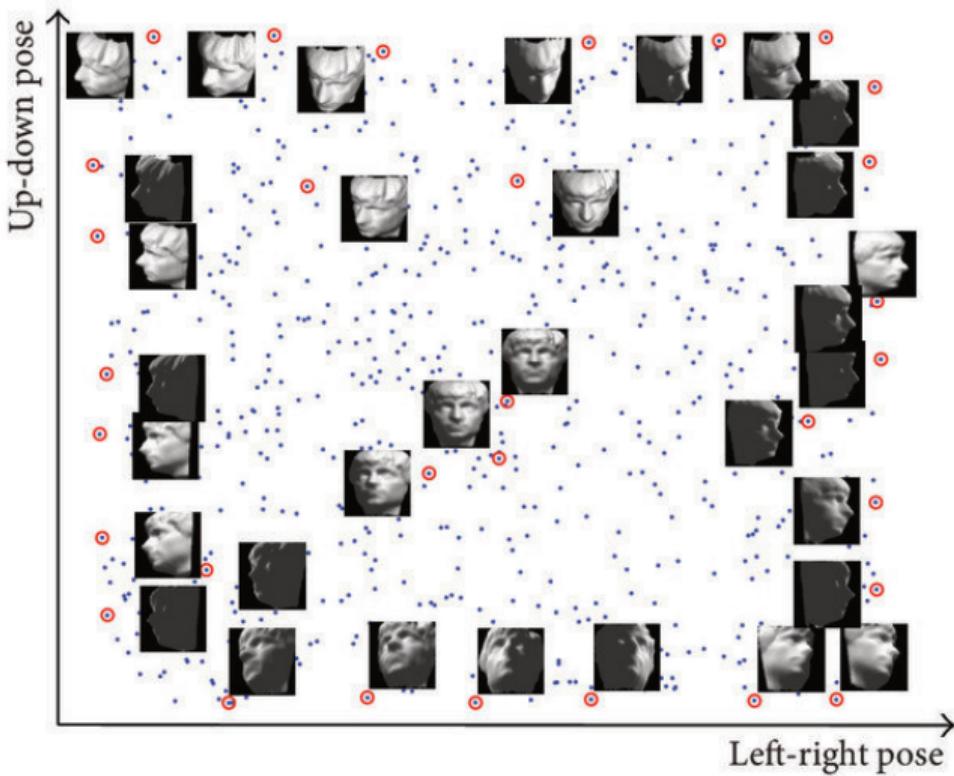
Unsupervised Learning (Feature Extraction)

- Word : Encode as vectors
- Relationship : represent as direction

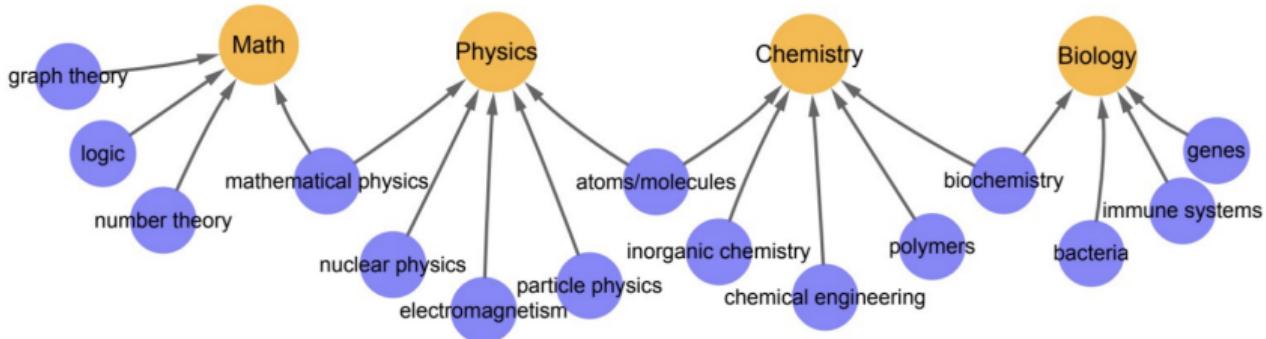
▷ word → encode → vector
▷ relation → encode → direction



Unsupervised Learning (Feature Extraction)

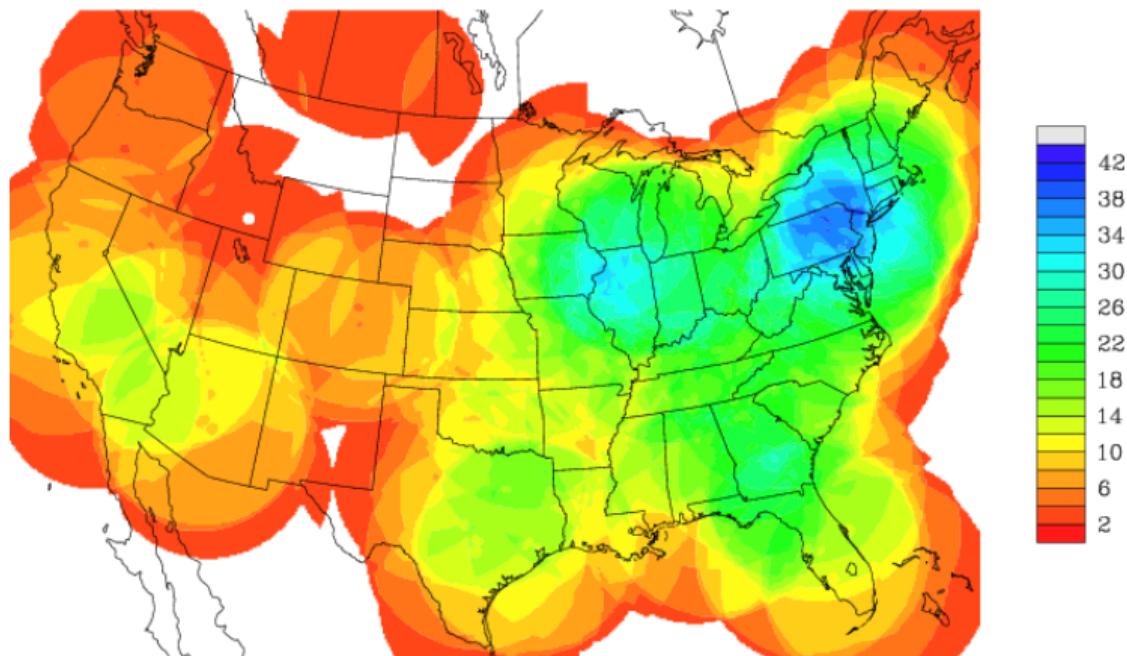


Unsupervised Learning

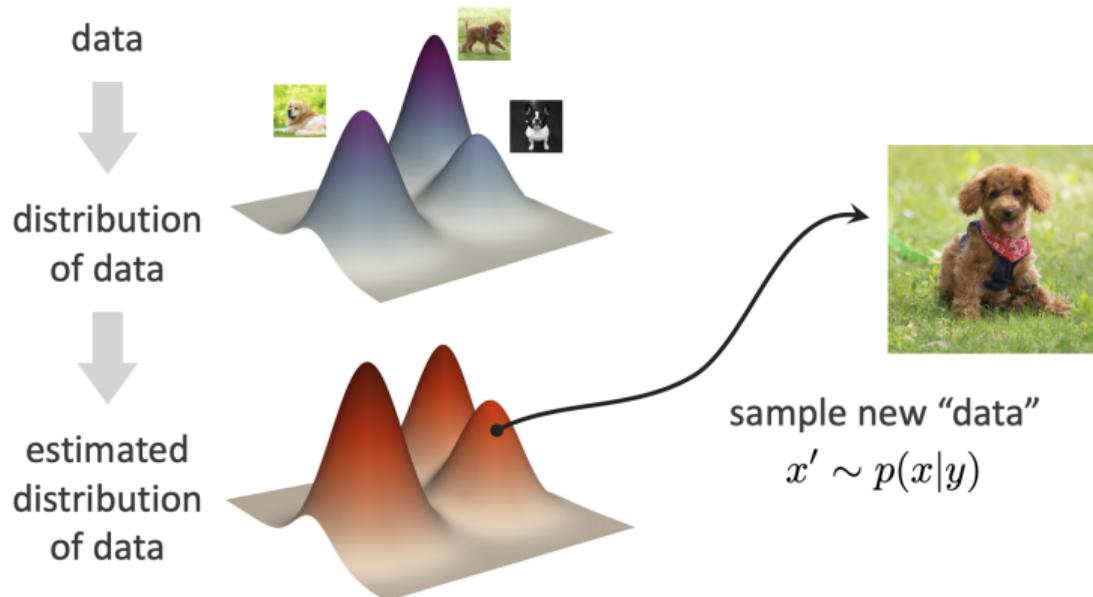


	logic deductive propositional semantics	graph subgraph bipartite vertex	boson massless particle higgs	polyester polypropylene resins epoxy	acids amino biosynthesis peptide
tag	<i>logic</i>	<i>graph theory</i>	<i>particle physics</i>	<i>polymer</i>	<i>biochemistry</i>

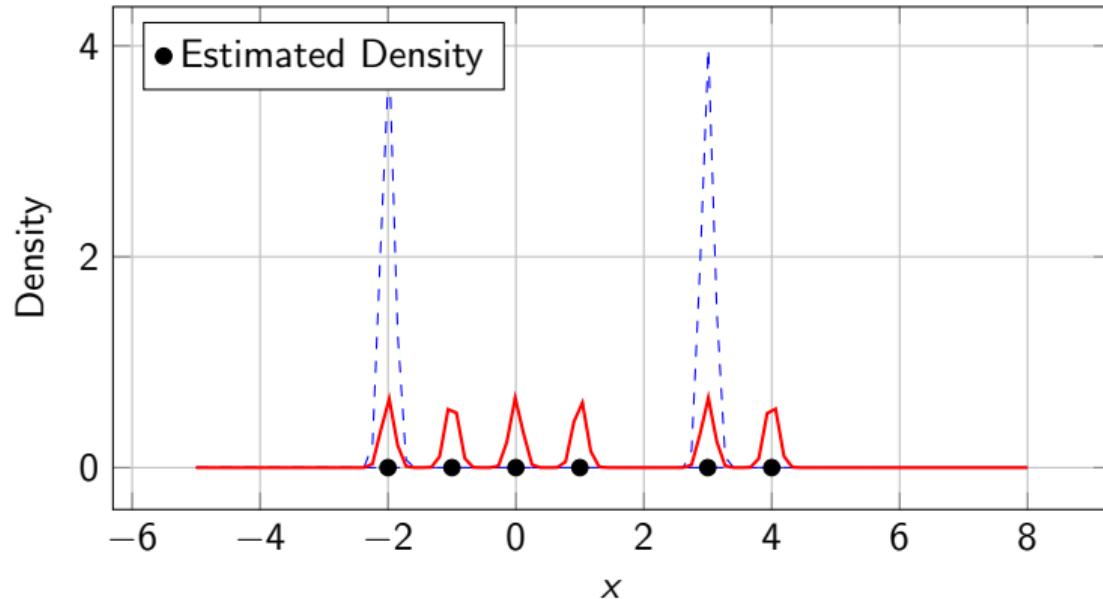
Unsupervised Learning (Density Estimation)



Generative Modeling

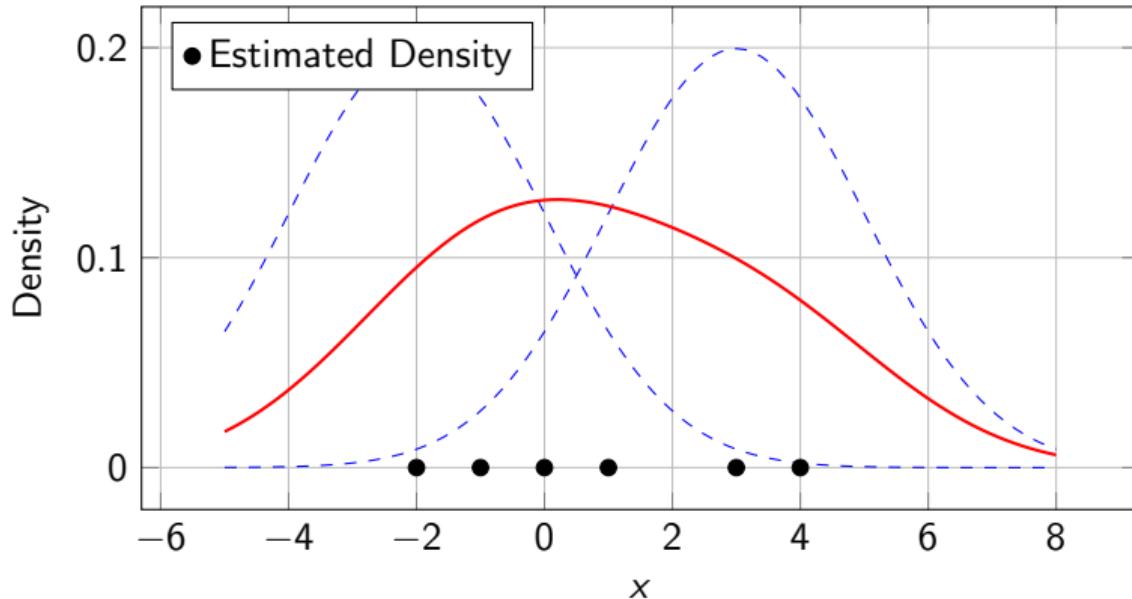


Density Estimation: Bias and Variance Trade-off



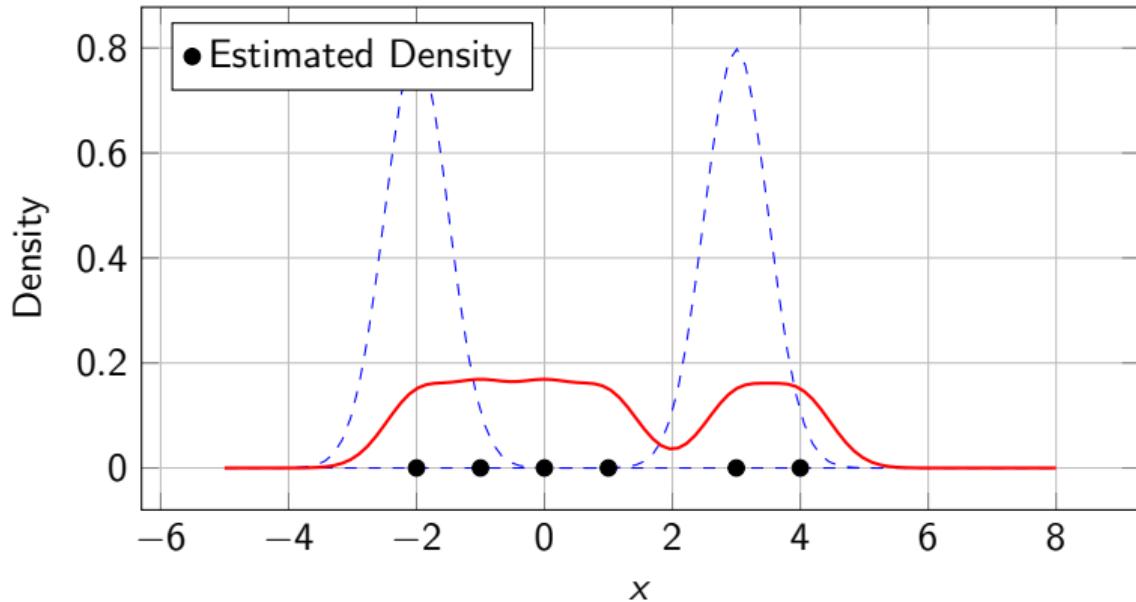
What does this mean in generating images?

Density Estimation: Bias and Variance Trade-off



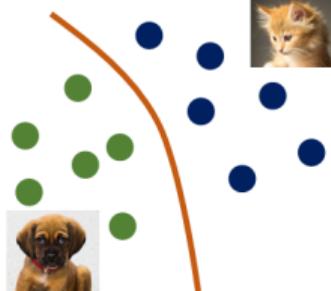
What does this mean in generating images?

Density Estimation: Bias and Variance Trade-off

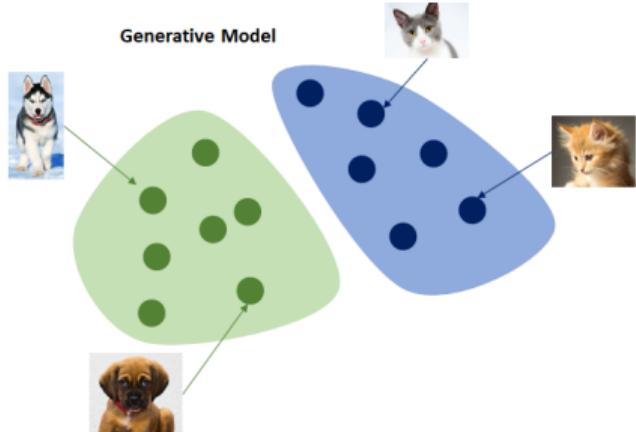


Generative AI

Discriminant Model



Generative Model



$$p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(x|y)p(y)}{p(x)} = p(x|y)\frac{p(y)}{p(x)}$$

Generative AI Case Study: Formulate as $p(x|y)$

- **Text-to-image/video generation**

Prompt: teddy bear teaching a course, with "generative models" written on blackboard

y : text prompt



x : generated visual content

Image generated by Stable Diffusion 3 Medium

Generative AI Case Study: Formulate as $p(x|y)$

- **Text-to-3D structure generation**



Figure credit: Tang, et al. LGM: Large Multi-View Gaussian Model for High-Resolution 3D Content Creation. ECCV 2024

Generative AI Case Study: Formulate as $p(x|y)$

- Class-conditional image generation

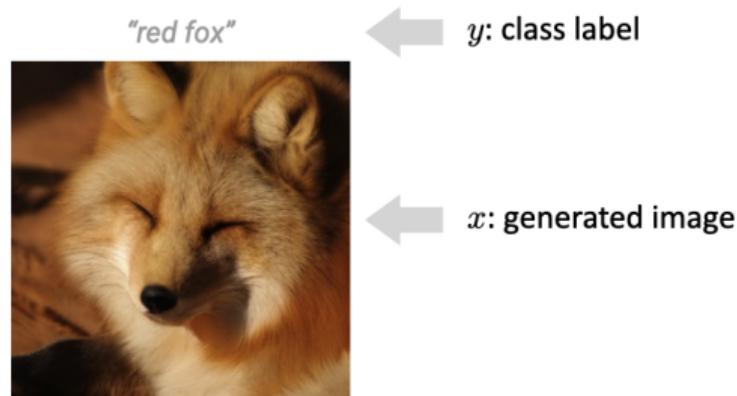


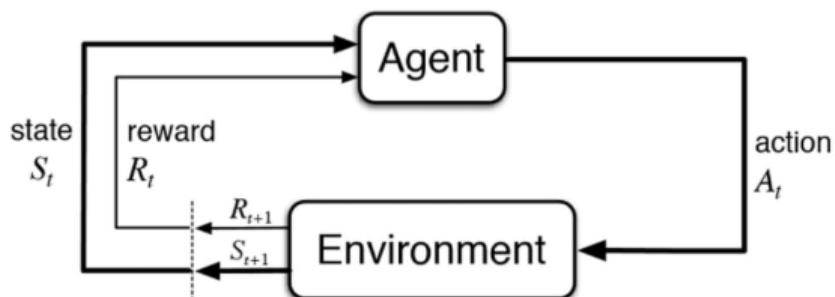
Image generated by: Li, et al. Autoregressive Image Generation without Vector Quantization, 2024

More Examples:

<https://mit-6s978.github.io/schedule.html>

Reinforcement Learning

Learning to make sequential decisions



mathmetical framework called: *markov decision process*

Not included in IEMS 304

Application of RL: Decision Making

What is the agent? What is the action? What is the state? What is the reward?

- AlphaGo
- Robotics