STAT 453: Introduction to Deep Learning and Generative Models

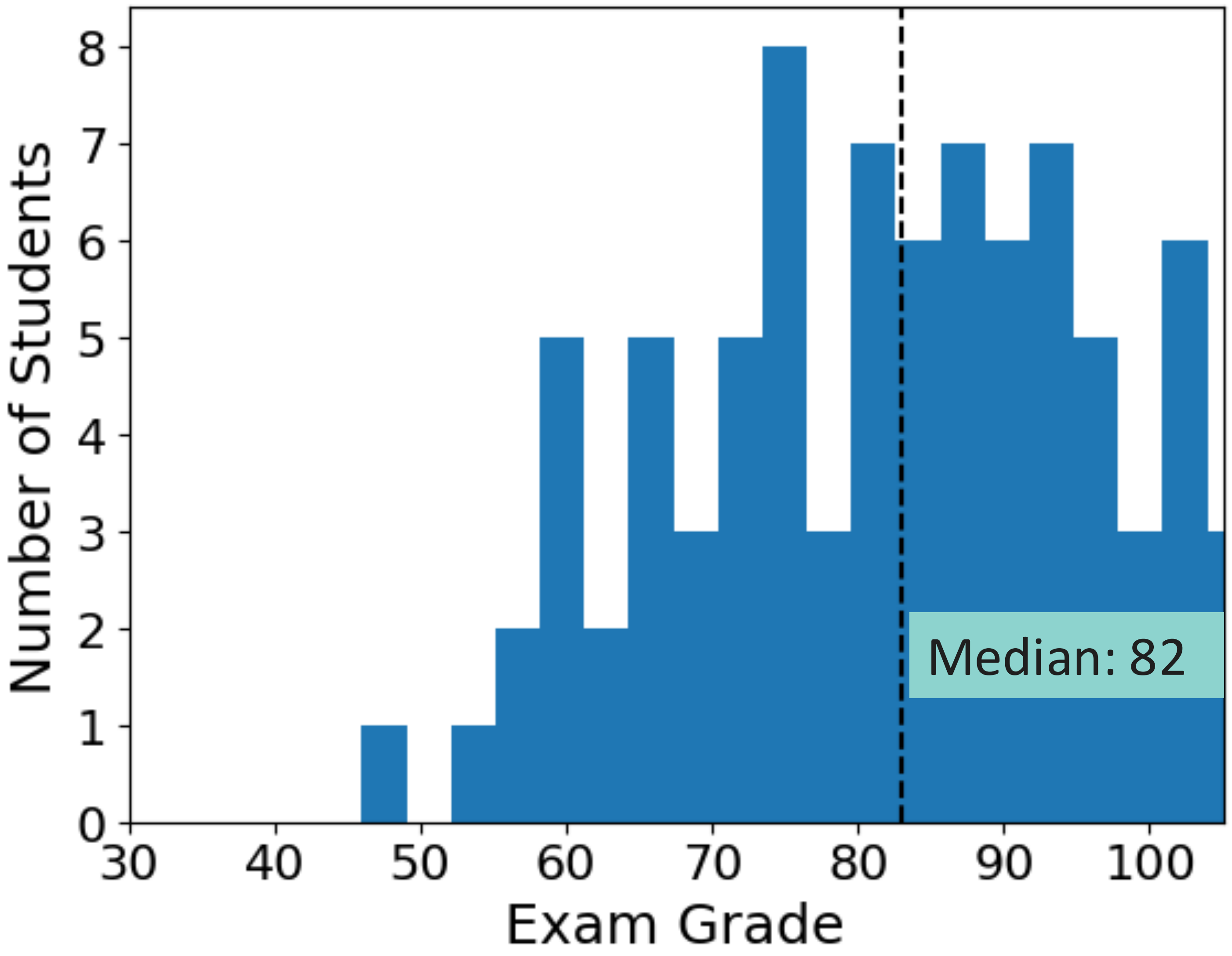
Ben Lengerich

Lecture 15: A Linear Intro to Generative Models

October 27, 2025

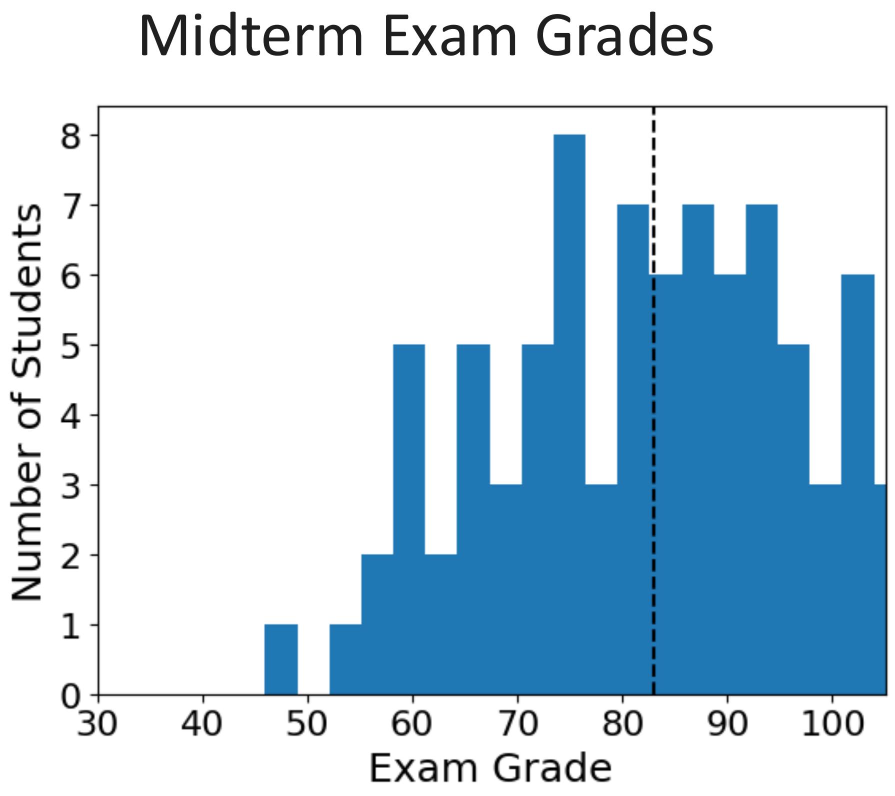
Reading: See course homepage

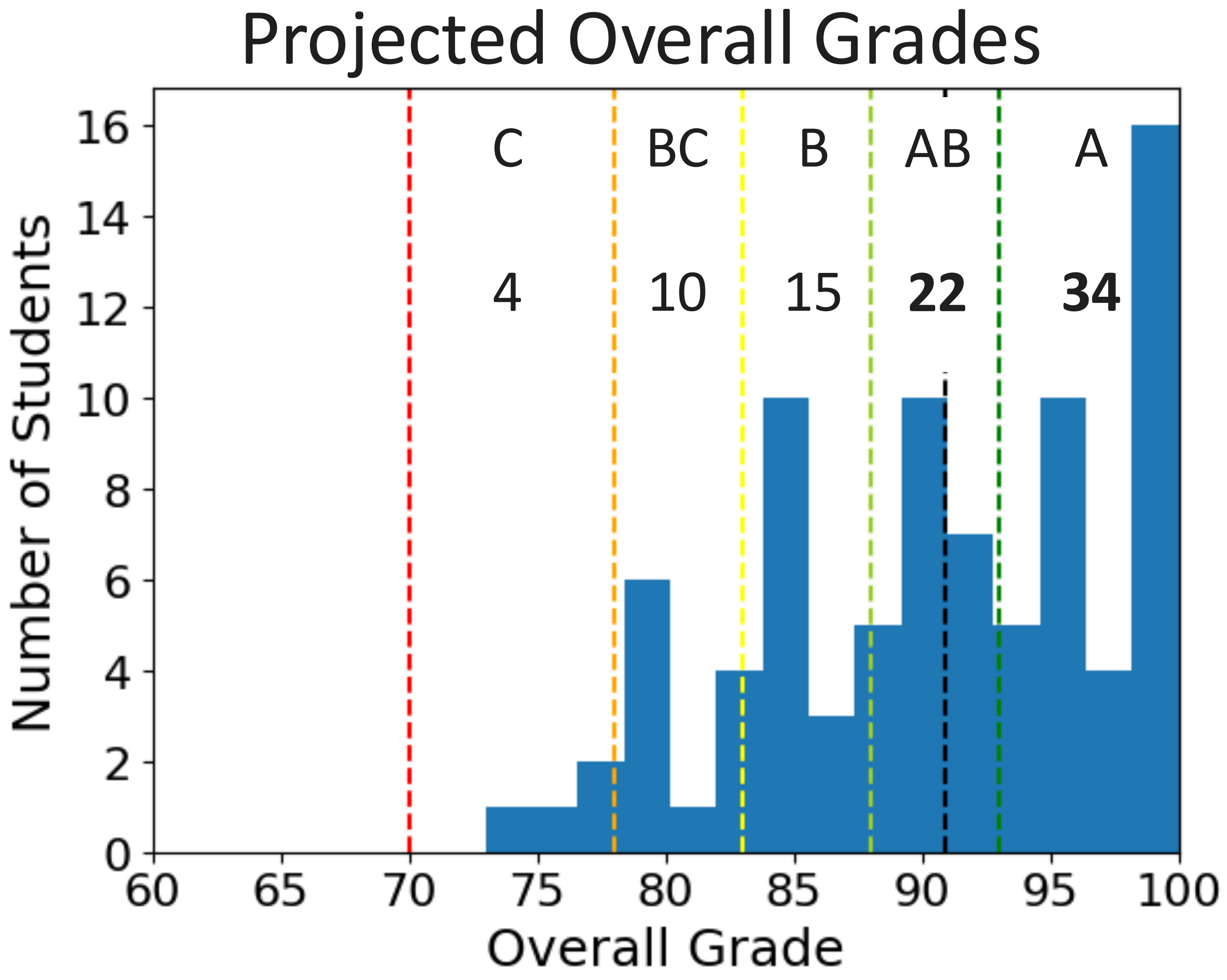
Midterm Exam



Max: 107 (x2)
Median: 82
Mean: 81.5

Midterm Exam in Context





Canvas: put your midterm exam in context

Use Canvas's grades tool to calculate your potential overall grade.

Project Midway Report Project	Nov 7 by 11:59pm	/ 100	
HW4 Assignments	Nov 21 by 11:59pm	/ 100	
HW5 Assignments	Dec 5 by 11:59pm	/ 100	
Project Presentation Project	Dec 10 by 11:59pm	/ 100	
Project Final Report Project	Dec 12 by 11:59pm	/ 300	
Final Exam Exams	Dec 17 by 11:59pm	/ 150	
Lecture Notes (Bonus) Exams	Dec 20 by 11:59pm	25 / 0	
Midterm Exam Adjustment Exams		8 / 0	
Assignments		0%	0.00 / 200.00
Exams		33%	33.00 / 100.00
Project		N/A	0.00 / 0.00
Total		**23.57%**	

Max is 25

Everyone gets 8

Course Schedule / Calendar

Week	Lecture Dates	Topic	Assignments
		Module 1: Introduction and Foundations	
1	9/3	Course Introduction	
2	9/8, 9/10	A Brief History of DL, Statistics / linear algebra / calculus review	HW1
3	9/15, 9/17	Single-layer networks Parameter Optimization and Gradient Descent	
4	9/22, 9/24	Automatic differentiation with PyTorch, Cluster and cloud computing resources	HW 2
		Module 2: Neural Networks	
5	9/29, 10/1	Multinomial logistic regression, Multi-layer perceptrons and backpropagation	
6	10/6, 10/8	Regularization Normalization / Initialization	HW 3
7	10/13, 10/15	Optimization, Learning Rates CNNs	Project Proposal
8	10/20, 10/22	Review, **Midterm Exam**	In-class Exam

Week	Lecture Dates	Topic	Assignments
		Module 3: Intro to Generative Models	
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs	
10	11/3, 11/5	Generative Adversarial Networks, Diffusion Models	Project Midway Report
		Module 4: Large Language Models	
11	11/10, 11/12	Sequence Learning with RNNs Attention, Transformers	HW4
12	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs	
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning	HW5
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research	
15	12/8, 12/10	**Project Presentations**	Project Final Report
16	12/17	**Final Exam**	Final Exam

Your Feedback

- Please fill out our anonymous Google Form

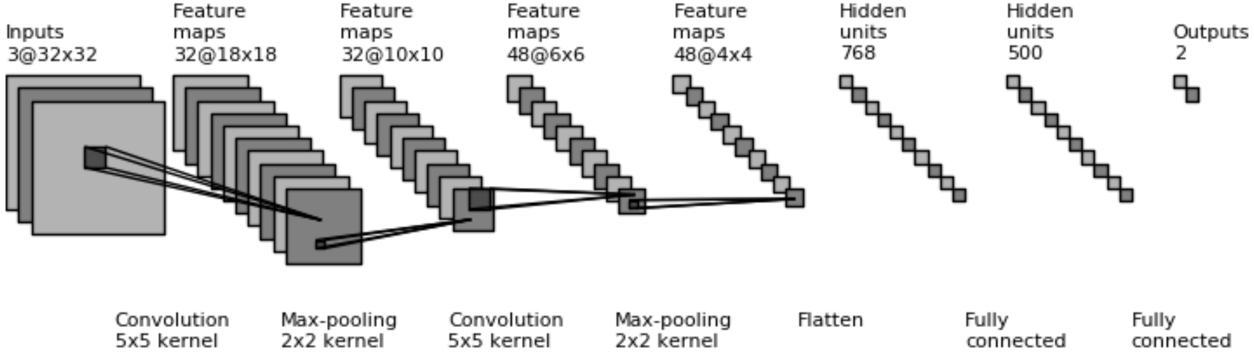
Generative Models

Where we're going: Deep Generative Models

判别模型

Discriminative Model (what we've seen so far)

(predict a label)



Cat or dog?

★(Provide the description of the cat, generate an image of the cat)

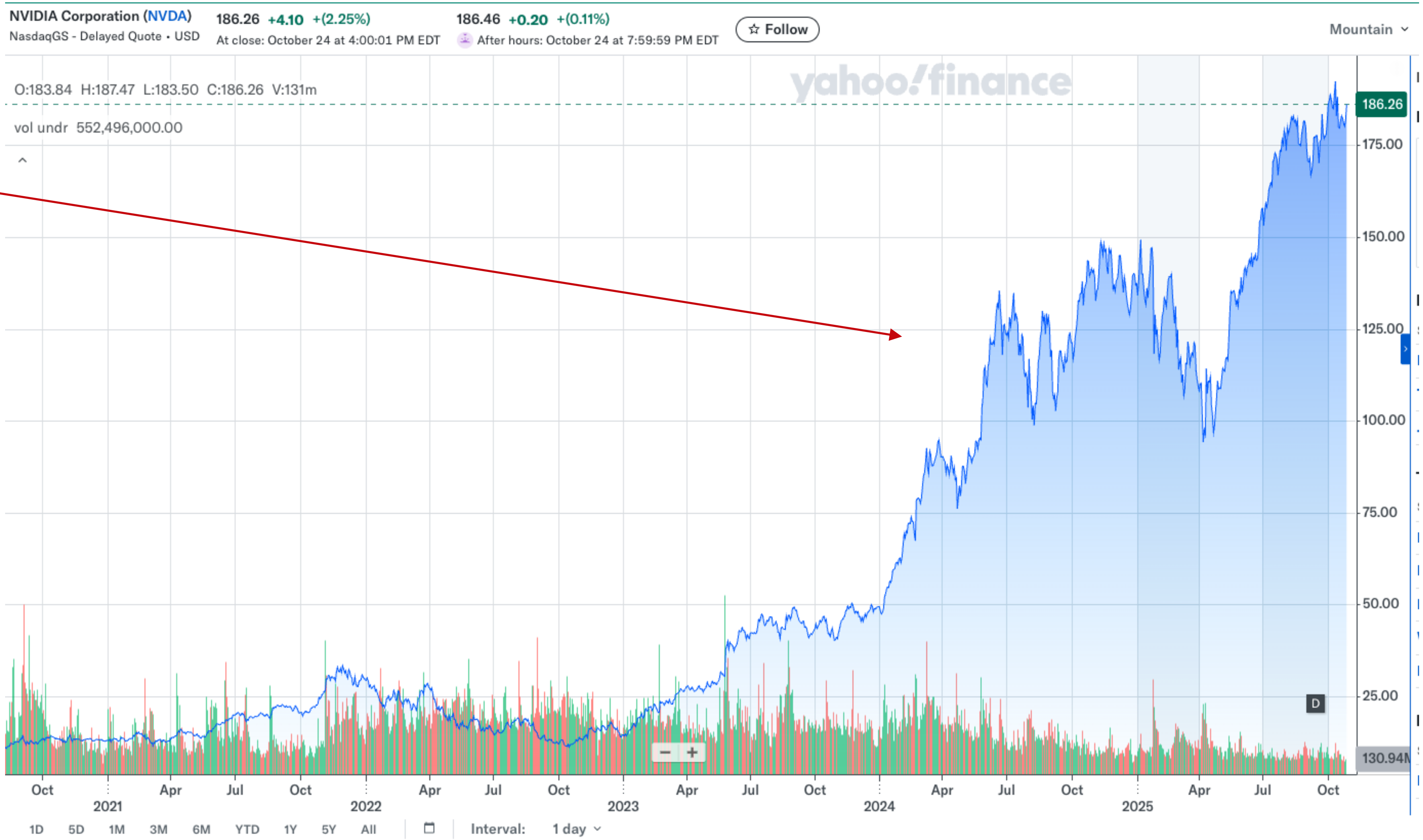
生成模型

Generative Model (what we're going to see)

Gemini Grok deepseek

Where we're going: Deep Generative Models

NVIDIA is now valued at **>$4.5T**



A Linear Intro to Generative Models

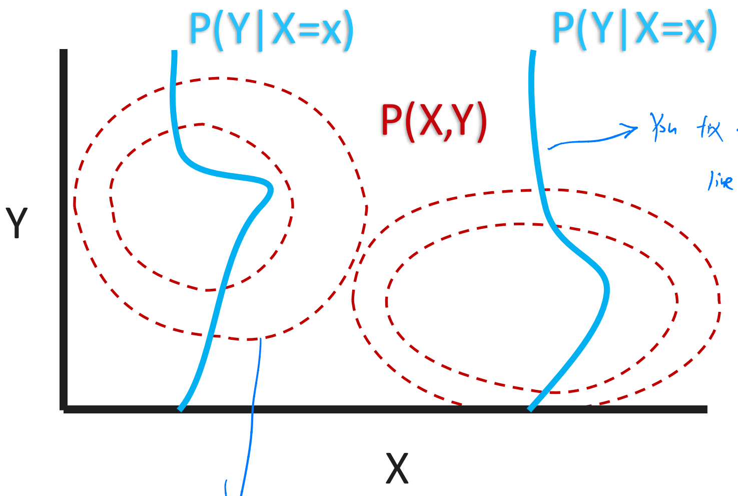
Generative and Discriminative Models

- **Generative**:
  - Models the joint distribution $\mathrm{P}(X, Y)$.

Goal: What kind of X will appear along with Y → Understand the whole picture of the data, and it has the ability to generate new data from this

- **Discriminative**:
  - Models the conditional distribution $\mathrm{P}(Y|X)$.

Goal: Given X, what is the probability of Y
(Find decision lines)



You fix the X, that's why it's a line for discriminative model, only care how Y would distribute based on X

Generative model gives you the contour map

Two paths to P(Y|X)

- **Discriminative:**

Straight line for discriminative model, directly find the relationship between X and Y

☆ (You give as many (X,Y) → learn f(x) to be close to Y as possible)

Observe X, Y → Learn P(Y|X)

- **Generative:**
  - Learn $\mathrm{P}(X|Y), \mathrm{P}(\mathrm{Y})$
  - Calculate $P(X) = \int_Y P(X,Y)dY$

Implicitly learn these things

Observe X, Y

$$\mathrm{P(Y|X)} = \frac{\mathrm{P(X|Y)P(Y)}}{P(X)}$$

Use Bayes theorem to get the P(Y|X) if you want

Two paths to classification

- **Discriminative:**

Observe X, Y $\longrightarrow$ $\hat{Y} = \mathrm{argmax}_{\mathrm{Y}} \mathrm{P}(\mathrm{Y}|\mathrm{X})$

Find the max expression value Y as my predicted $\hat{Y}$

- **Generative:**
  - Learn $\mathrm{P}(X|Y), \mathrm{P}(\mathrm{Y})$
  - ~~Calculate~~ $\cancel{P(X) = \int_Y P(X,Y)dY}$

Observe X, Y $\longrightarrow$ $\hat{Y} = \mathrm{argmax}_{\mathrm{Y}} \mathrm{P}(\mathrm{X}|\mathrm{Y})\mathrm{P}(\mathrm{Y})$

(P(x) is a constant there)

We don't need P(x) here since it is not important when comparing

e.g. $\frac{P(x|Y=1)\ P(Y=1)}{P(x)}$ v.s $\frac{P(x|Y=0)\ P(Y=0)}{P(x)}$

Example Discriminative Model: Logistic Regression

- **Discriminative**:

Observe X, Y $\longrightarrow$ Learn P(Y|X)

- Parameterize:

Linear Combination

parameters / weight $(\theta_0 X_0 + \theta_1 X_1 + \dots)$

probability from 0-1

  - $P(Y = 1|X) = \sigma(\theta^T X)$, where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function.
  - $P(Y = 0|X) = 1 - P(Y = 1|X)$
  - Recall: Why this parameterization?

$$\log \frac{P(Y=1|X)}{P(Y=0|X)} = \log \frac{\sigma(\theta^T X)}{1-\sigma(\theta^T X)}$$

$$= \log \frac{\frac{1}{1+e^{-\theta^T X}}}{1 - \frac{1}{1+e^{-\theta^T X}}} = \log \frac{\frac{1}{1+e^{-\theta^T X}}}{\frac{(1+e^{-\theta^T X})-1}{1+e^{-\theta^T X}}} = \log \frac{\frac{1}{1+e^{-\theta^T X}}}{\frac{e^{-\theta^T X}}{1+e^{-\theta^T X}}}$$

$$= \log \frac{1}{e^{-\theta^T X}} = \log e^{\theta^T X} = \boldsymbol{\theta^T X}$$

linear relationship for log-odds

Example Discriminative Model: Logistic Regression

- **Discriminative**:

  Observe X, Y $\longrightarrow$ Learn P(Y|X)

- Parameterize:
  - $P(Y=1|X) = \sigma(\theta^T X)$, where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function.
  - $P(Y=0|X) = 1 - P(Y=1|X)$

- Estimate $\hat{\theta}$ from observations:
  - $\hat{\theta} = \text{argmax}_\theta \prod_i P(Y_i|X_i;\theta)$

    $= \text{argmax}_\theta \sum_i [Y_i \log \sigma(\theta^T X_i) + (1-Y_i)\log(1 - \sigma(\theta^T X_i))]$

  Train, find the θ

- Calculate $P(Y=1|X) = \sigma(\theta^T X)$ $\rightarrow$ predict

Example Generative Model: Naïve Bayes

Learn $\mathrm{P}(X|Y), \mathrm{P}(\mathrm{Y})$

Observe X, Y

(We don't directly go for it, it is implicit)

$$\mathrm{P(Y|X)} = \frac{\mathrm{P(X|Y)P(Y)}}{P(X)}$$

- Parameterize:

Naive Bayes: Assume if Y is determined, all features X_i is conditionally independent $P(X_1, \dots X_d|Y) = P(X_1|Y)\,P(X_2|Y)\dots$

  - Assume $P(X|Y) = \prod_{j=1}^{d} P(X_j|Y)$,
  - $P(X_j|Y) = N(\mu_{jk}, \sigma_{jk}^2)$ → (Assume in Normal Distribution, we train μ and σ^2)

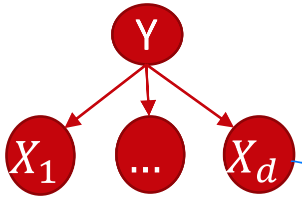
↓

Find μ and σ^2 for each group

Conditional independences of features X | Y

$$P(Y = k) = \frac{\#\, of\ samples\ with\ Y{=}k}{Total\ samples}$$

Frequency of labels



→ No shared interaction between X

Example Generative Model: Naïve Bayes

Learn $\mathrm{P}(X|Y), \mathrm{P}(\mathrm{Y})$

Observe X, Y

$$\mathrm{P(Y|X)} = \frac{\mathrm{P(X|Y)P(Y)}}{P(X)}$$

- Parameterize: Naive Bayes Assumption
  - Assume $P(X|Y) = \prod_{j=1}^{d} P(X_j|Y)$,

$$P(Y=k) = \frac{\#\ of\ samples\ with\ Y=k}{Total\ samples}$$

- Estimate: → Train process
  - $\hat{\mu}, \hat{\sigma} = \mathrm{argmax}_{\mu,\sigma} P(X|Y)$

predict process, directly put it in the formula below

- Calculate $P(Y=1|X) = \dfrac{\prod_{j=1}^{d} P(X_j|Y=1) P(Y=1)}{P(X)}$

Summary

- **Discriminative**:

Observe X, Y $\rightarrow$ Learn $\mathrm{P(Y|X)}$

- **Generative**:
  - Learn $\mathrm{P}(X|Y), \mathrm{P(Y)}$
  - Calculate $P(X) = \int_Y P(X,Y)dY$

Observe X, Y

$$\mathrm{P(Y|X)} = \frac{\mathrm{P(X|Y)P(Y)}}{P(X)}$$

What about MAP / Regularization?

we add a prior here as restriction (MAP)

Prior $\mathrm{P}(\theta)$

Logistic Regression:

Observe X, Y → Learn $\mathrm{P}(\mathrm{Y}|\mathrm{X};\theta)$

- Parameterize:
  - $P(Y=1|X)=\sigma(\theta^T X)$, where $\sigma(z)=\frac{1}{1+e^{-z}}$ is the sigmoid function.
  - $P(Y=0|X)=1-P(Y=1|X)$
- Estimate $\hat{\theta}$ from observations:
  - $\hat{\theta}=\mathrm{argmax}_\theta \prod_i P(Y_i|X_i;\theta) P(\theta)$

    $=\mathrm{argmax}_\theta \sum_i [Y_i \log \sigma(\theta^T X_i)+(1-Y_i)\log(1-\sigma(\theta^T X_i))] - R(\theta)$
- Calculate $P(Y|X)$

We add the prior here, want tradeoff between fitting / model simplicity

MLE: Find the θ to let the probability of the data biggest, but has nothing information on prior P(θ)

Regularizer / penalty term

Discriminative vs Generative Models

- Discriminative models optimize the conditional likelihood:

$$\widehat{\theta_{disc}} = \operatorname{argmax}_{\theta} P(Y|X;\theta)$$

P(Y|X) with θ

- Generative models optimize the joint likelihood:

$$\widehat{\theta_{gen}} = \operatorname{argmax}_{\theta} P(X,Y;\theta)$$

P(X,Y) with θ

Are these the same optimization?

Discriminative vs Generative Models

- Discriminative models optimize the conditional likelihood:

$$\widehat{\theta_{disc}} = \text{argmax}_\theta P(Y|X;\theta) = \text{argmax}_\theta \frac{P(X|Y;\theta)P(Y;\theta)}{P(X;\theta)}$$

- Generative models optimize the joint likelihood:

$$\widehat{\theta_{gen}} = \text{argmax}_\theta P(X,Y;\theta) = \text{argmax}_\theta P(X|Y;\theta)P(Y;\theta)$$

Are these the same optimization?

Same optimization when $P(X;\theta)$ is invariant to θ

☆ (If θ tells nothing of X, then they become the same)

Logistic Regression vs Naïve Bayes

Logistic Regression	Naïve Bayes
Discriminative	Generative
Defines $P(Y\|X;\theta)$	Defines $P(X,Y;\theta)$
Estimates $\widehat{\theta_{lr}} = \text{argmax}_\theta P(Y\|X;\theta)$	Estimates $\widehat{\theta_{nb}} = \text{argmax}_\theta P(X,Y,\theta)$
Lower asymptotic error on classification	Higher asymptotic error on classification
Slower convergence in terms of samples	Faster convergence in terms of samples

As we add more samples, the Logistic Regression tends to do better than Naive Bayes

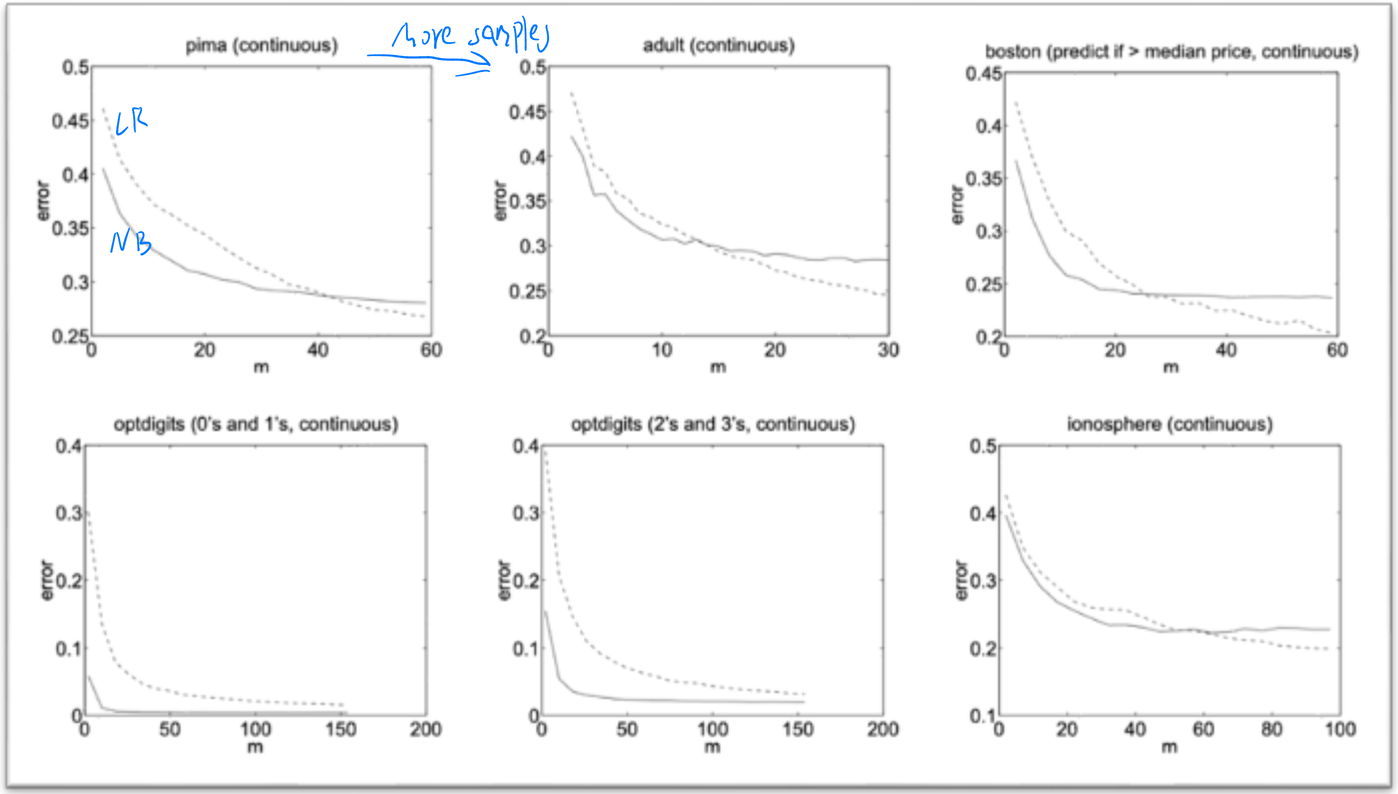
Discriminative vs Generative: A Proposition

- "While discriminative learning has lower asymptotic error, a generative classifier may also approach its (higher) asymptotic error much faster."

Why?

.... LR
—— NB

Ng & Jordan 2001



Discriminative vs Generative: A Proposition

- "While discriminative learning has lower asymptotic error, a generative classifier may also approach its (higher) asymptotic error much faster."
- Underlying assumption of this statement:
  - Generative models of the form $\boldsymbol{P(X, Y, \theta)}$ make **more simplifying assumptions** than do discriminative models of the form $\boldsymbol{P(Y|X, \theta)}$.

    Higher error in long run, but much faster
  - **Not always true** → Only for Naive Bayes that assume simple relationships, not hold for other generative models
  - "So far there is no theoretically correct, general criterion for choosing between the discriminative and the generative approaches to classification of an observation **x** into a class *y*; the choice depends on the relative confidence we have in the correctness of the specification of either $p(y|\mathbf{x})$ or $p(\mathbf{x}, y)$ for the data."

Xue & Tittering 2008

Modern Deep Generative Models (DGMs)

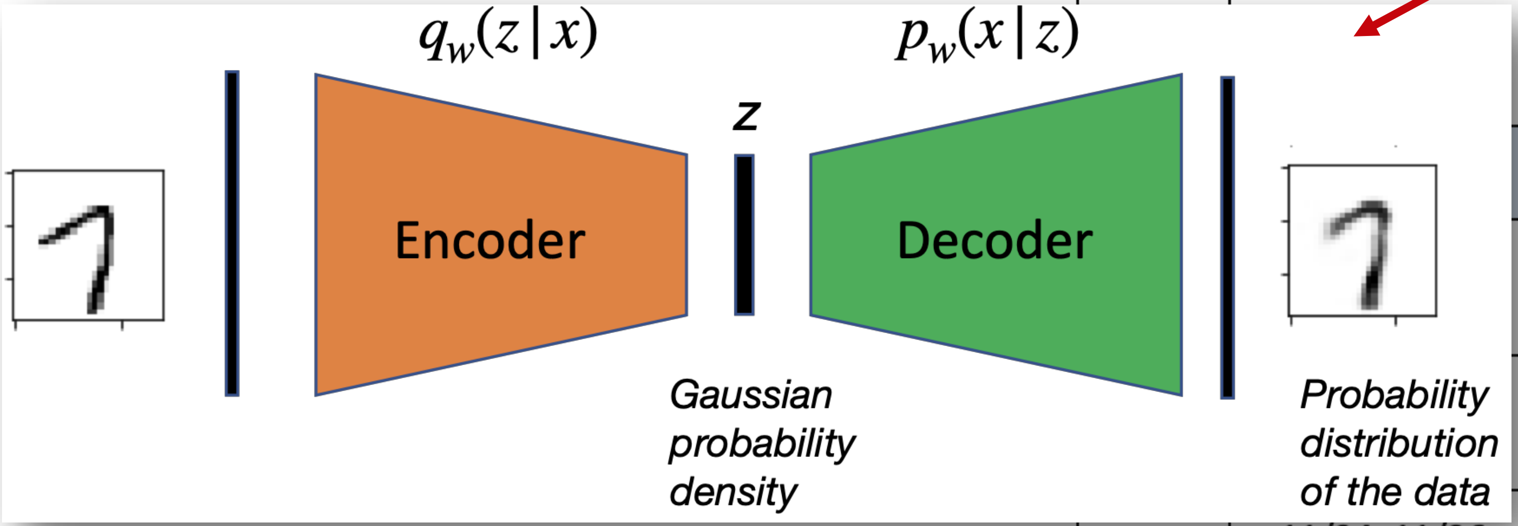
- Goal: Generative models of the form $\boldsymbol{P(X, Y, \theta)}$ without strong simplifying assumptions.
- Hidden structure z that explains high-dim. x

  DGM
- Fundamental challenge: We never observe z
- This makes two core computations difficult:
  - **Marginal likelihood:** $p_\theta(x) = \int p_\theta(x, z)dz$
  - **Posterior inference:** $p_\theta(z \mid x) \propto p_\theta(x \mid z)p(z)$
- Each type of DGM makes a tradeoff

Coming up…

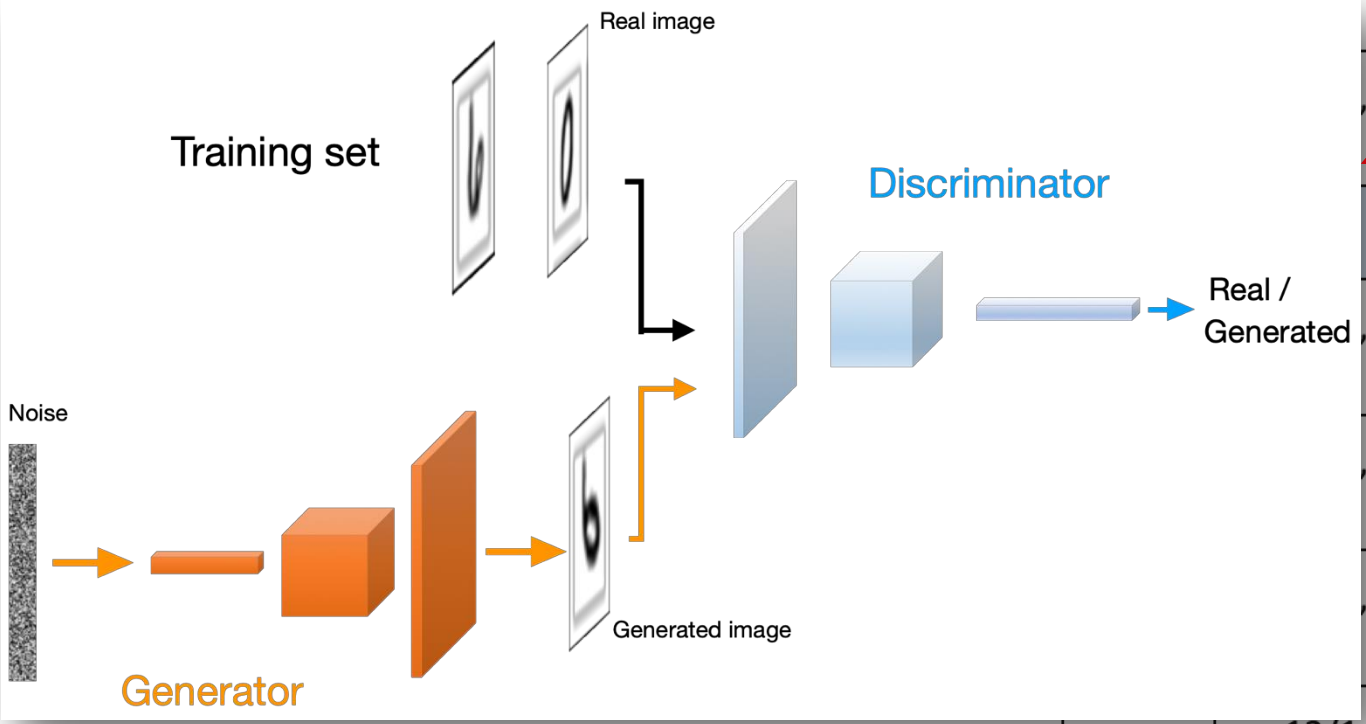
Week	Lecture Dates	Topic	Assignments
Module 3: Intro to Generative Models			
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs	
10	11/3, 11/5	Generative Adversarial Networks, Diffusion Models	Project Midway Report
Module 4: Large Language Models			
11	11/10, 11/12	Sequence Learning with RNNs Attention, Transformers	HW4
12	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs	
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning	HW5
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research	
15	12/8, 12/10	**Project Presentations**	Project Final Report
16	12/17	**Final Exam**	Final Exam

Coming up…



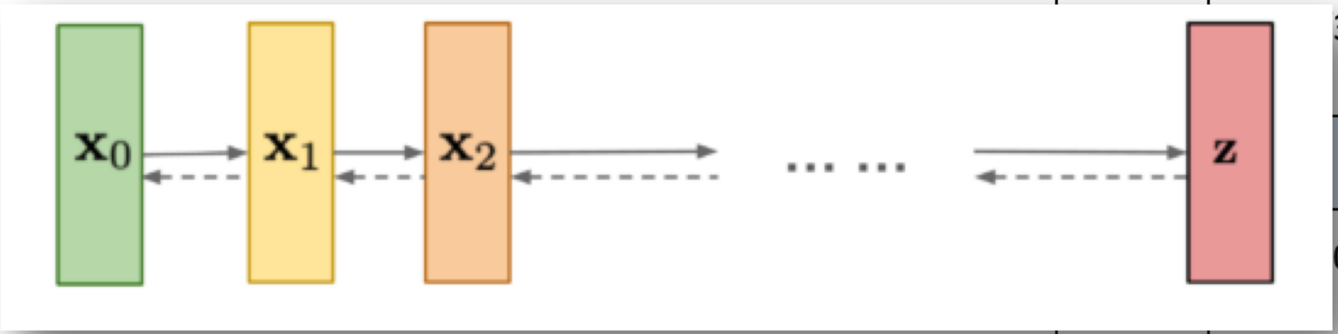
Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
		Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
		Sequence Learning with RNNs Attention, Transformers
		GPT Architectures, Unsupervised Training of LLMs
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

Coming up…



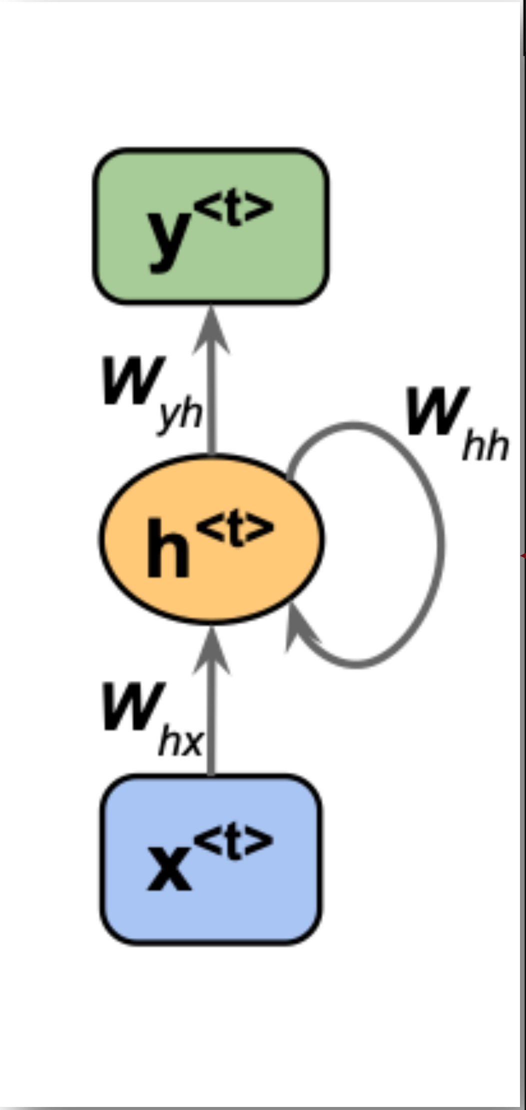
Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
	, 11/5	Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
	, 11/12	Sequence Learning with RNNs Attention, Transformers
	, 11/19	GPT Architectures, Unsupervised Training of LLMs
	, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

Coming up…



Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
	3, 11/5	Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
	0, 11/12	Sequence Learning with RNNs Attention, Transformers
12	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

Coming up…



Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
10	11/3, 11/5	Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
11	11/10, 11/12	Sequence Learning with RNNs Attention, Transformers
12	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., 2017. Attention Is All You Need.

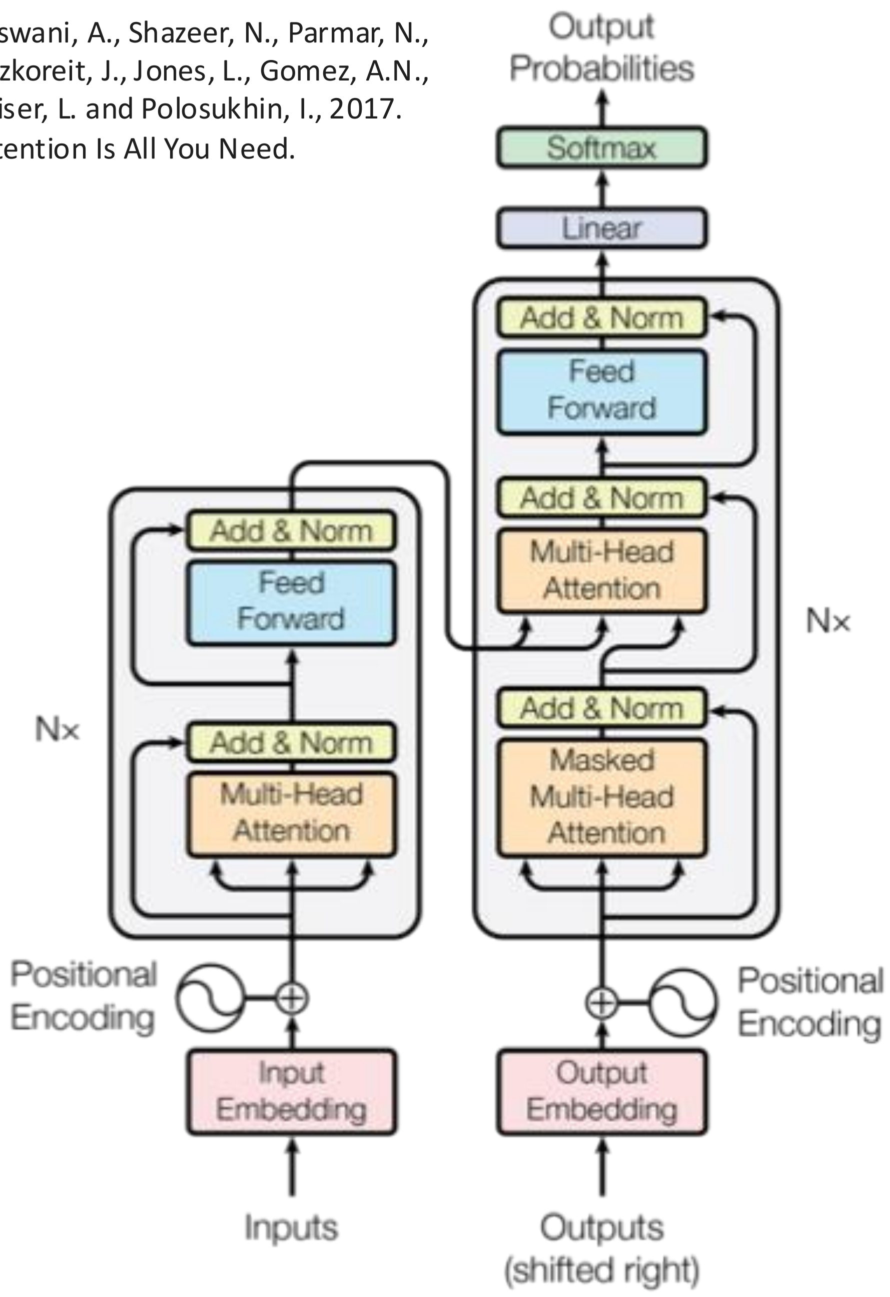
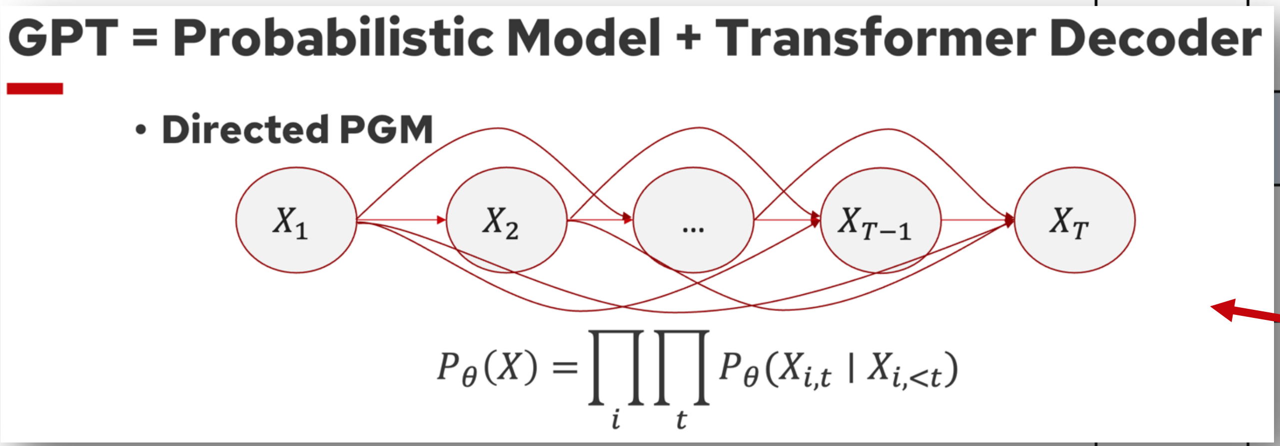


Figure 1: The Transformer - model architecture.

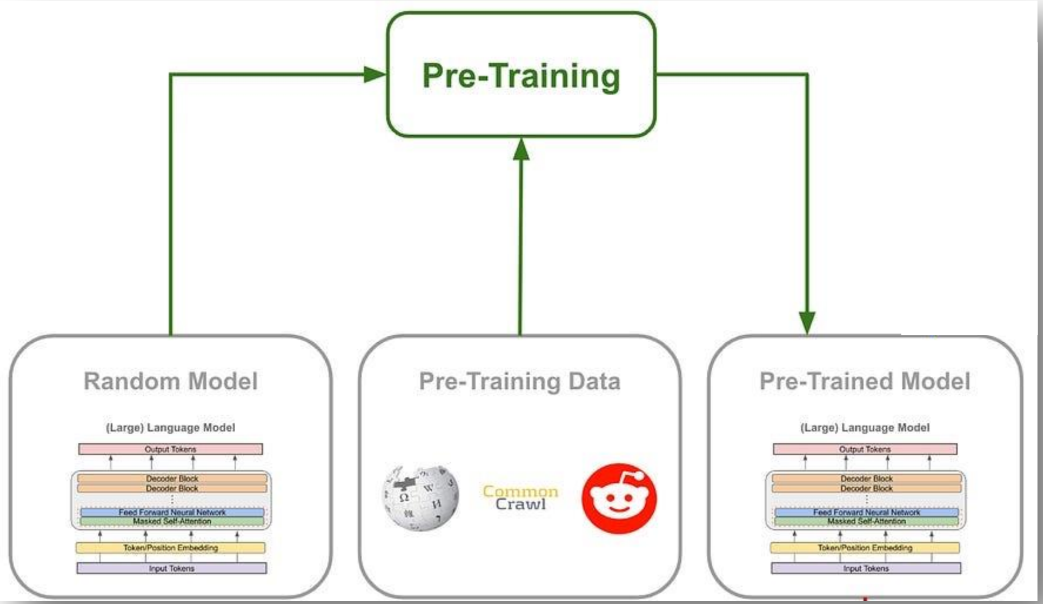
Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
10	11/3, 11/5	Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
11	11/10, 11/12	Sequence Learning with RNNs Attention, Transformers
12	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

Coming up…



Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
	11/3, 11/5	Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
	11/10, 11/12	Sequence Learning with RNNs Attention, Transformers
	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

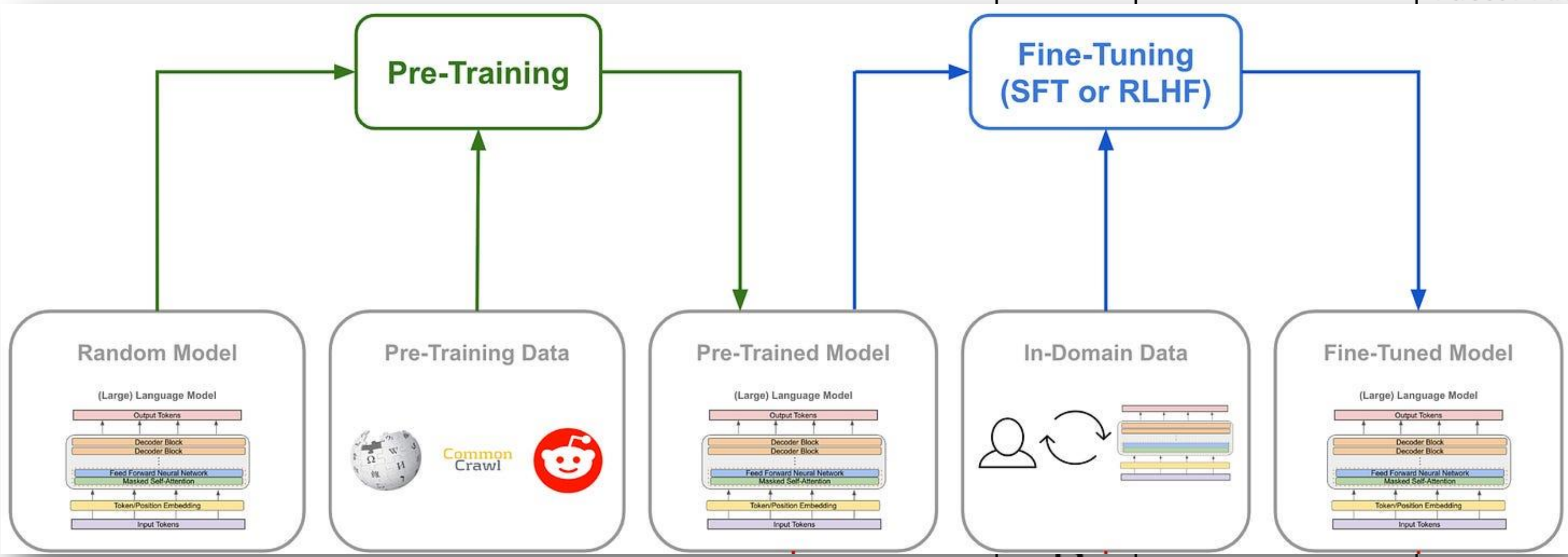
Coming up…



https://cameronrwolfe.substack.com/p/understanding-and-using-supervised

Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
10	11/3, 11/5	Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
11	11/10, 11/12	Sequence Learning with RNNs Attention, Transformers
12	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs
13	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

Coming up…

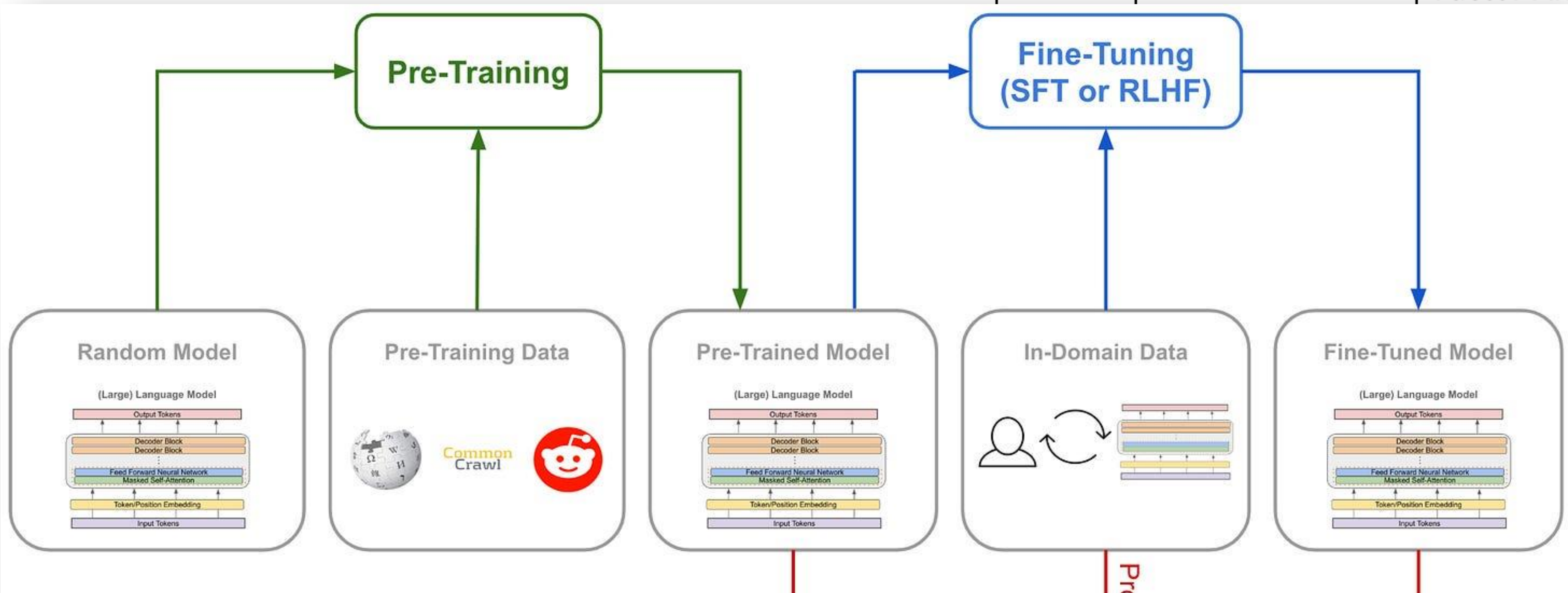


https://cameronrwolfe.substack.com/p/understanding-and-using-supervised

Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
		ve Adversarial Networks, Models
		odule 4: Large Language Models
		Learning with RNNs , Transformers
		itectures, ised Training of LLMs
		ed Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

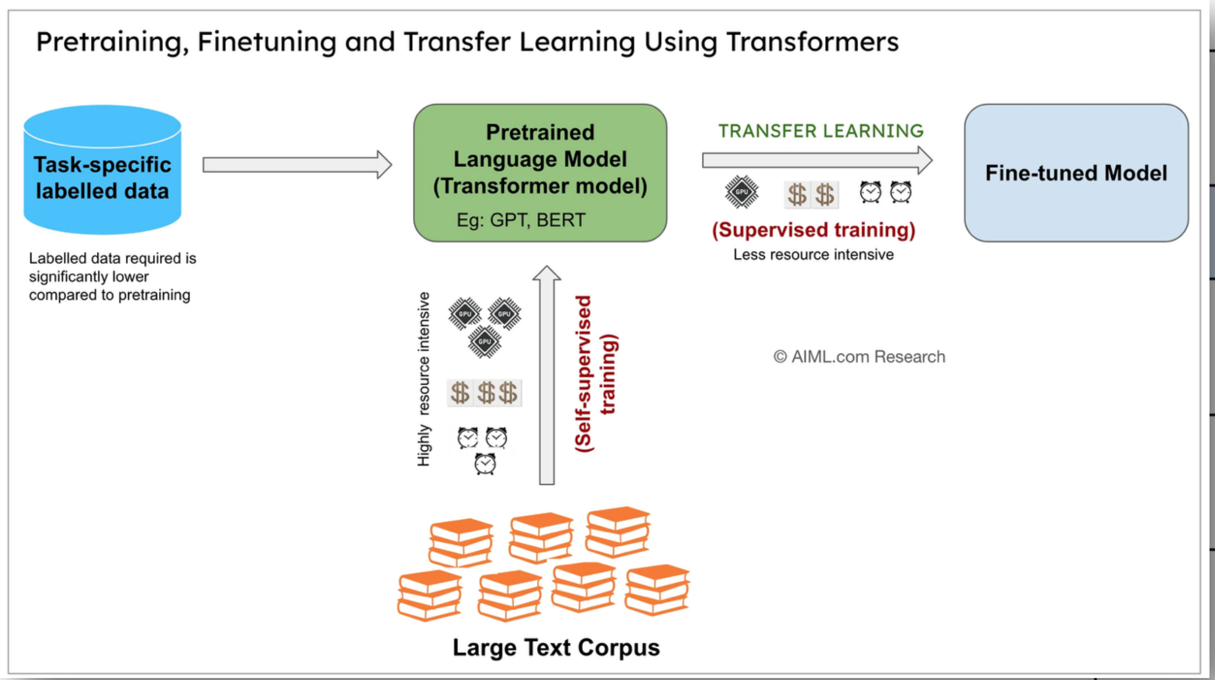
Coming up…

Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
		…ve Adversarial Networks, … Models
		…odule 4: Large Language Models
		… Learning with RNNs, …, Transformers
		…itectures, …vised Training of LLMs
		…ed Fine-tuning of LLMs, …and In-context learning
		…on models, alignment, explainability …ections in LLM research
		…resentations
16	12/17	**Final Exam**



https://cameronrwolfe.substack.com/p/understanding-and-using-supervised

Coming up…



Source: AIML.com Research

Week	Lecture Dates	Topic
		Module 3: Intro to Generative Models
9	10/27, 10/29	A Linear Intro to Generative Models, Factor Analysis, Autoencoders, VAEs
	11/3, 11/5	Generative Adversarial Networks, Diffusion Models
		Module 4: Large Language Models
	11/10, 11/12	Sequence Learning with RNNs Attention, Transformers
	11/17, 11/19	GPT Architectures, Unsupervised Training of LLMs
	11/24, 11/26	Supervised Fine-tuning of LLMs, Prompts and In-context learning
14	12/1, 12/3	Foundation models, alignment, explainability Open directions in LLM research
15	12/8, 12/10	**Project Presentations**
16	12/17	**Final Exam**

Questions?

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON