

# **Introduction to Machine Learning and Neural Networks**

**Deep Learning Bootcamp**

**Feb 14, 2025**

# What is this bootcamp about?

- Quick introduction to Machine Learning and Deep Learning.
- Hands on coding projects related to business application.
- Assume basic background in Python (for/while loop, if else condition, class)

# Why this bootcamp?

- A lot Machine Learning and Deep Learning course available.
  - STA314 and CSC311: Intro ML
  - CSC413: Neural networks and deep learning
  - STA 414/2104: Probabilistic Machine Learning
  - <https://d2l.ai> Dive into Deep Learning
- Too many materials, covers a lot topic not directly related to business application.

# What we will cover?

- Topics depend on your interest. Current plan:
  - Classical NLP – Feature-Based Models & Simple Generative Methods
  - Transformers & Attention Mechanisms (LLMs from Scratch)
  - Large Language Models in Business (LLM Agents & LangChain)
  - Fine-Tuning LLMs for Business Applications
  - Deep Choice Models – Predicting Customer Preferences
  - Deep Reinforcement Learning for Operations & Supply Chain
- Feel free to propose other topics related.

# For This Session

- Overview of Machine Learning
- Multi layer perceptron
- Stochastic Gradient Descent, Backpropagation, chain rule
- MNIST digit recognition

# What is Machine Learning?

- Types of ML:
  - Supervised learning: given input-output pairs  $\{(x_i, y_i)\}_{i=1}^N$ , the goal is to learn the mapping from inputs  $x$  to outputs  $y$
  - Unsupervised learning: Discover patterns or structure in unlabeled data  $\{x_i\}_{i=1}^N$ .
  - Semi-supervised learning: We are given limited of labeled data  $(x_i, y_i)$  and lots of unlabeled  $x_i$ 's.
  - Self-supervised learning: Learn representations by predicting parts of the input from other parts (pertaining of LLM).
  - Reinforcement learning: Learn actions to maximize cumulative rewards in an environment.

# What is Machine Learning?

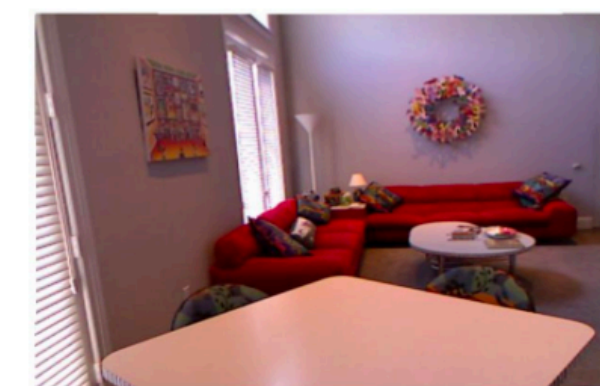
Computer vision: Object detection, semantic segmentation, pose estimation, and almost every other task is done with ML.



Figure 4. More results of Mask R-CNN on COCO test images, using ResNet-101-FPN and running at 5 fps, with 35.7 mask AP (Table 1).



Instance segmentation - [▶ Link](#)



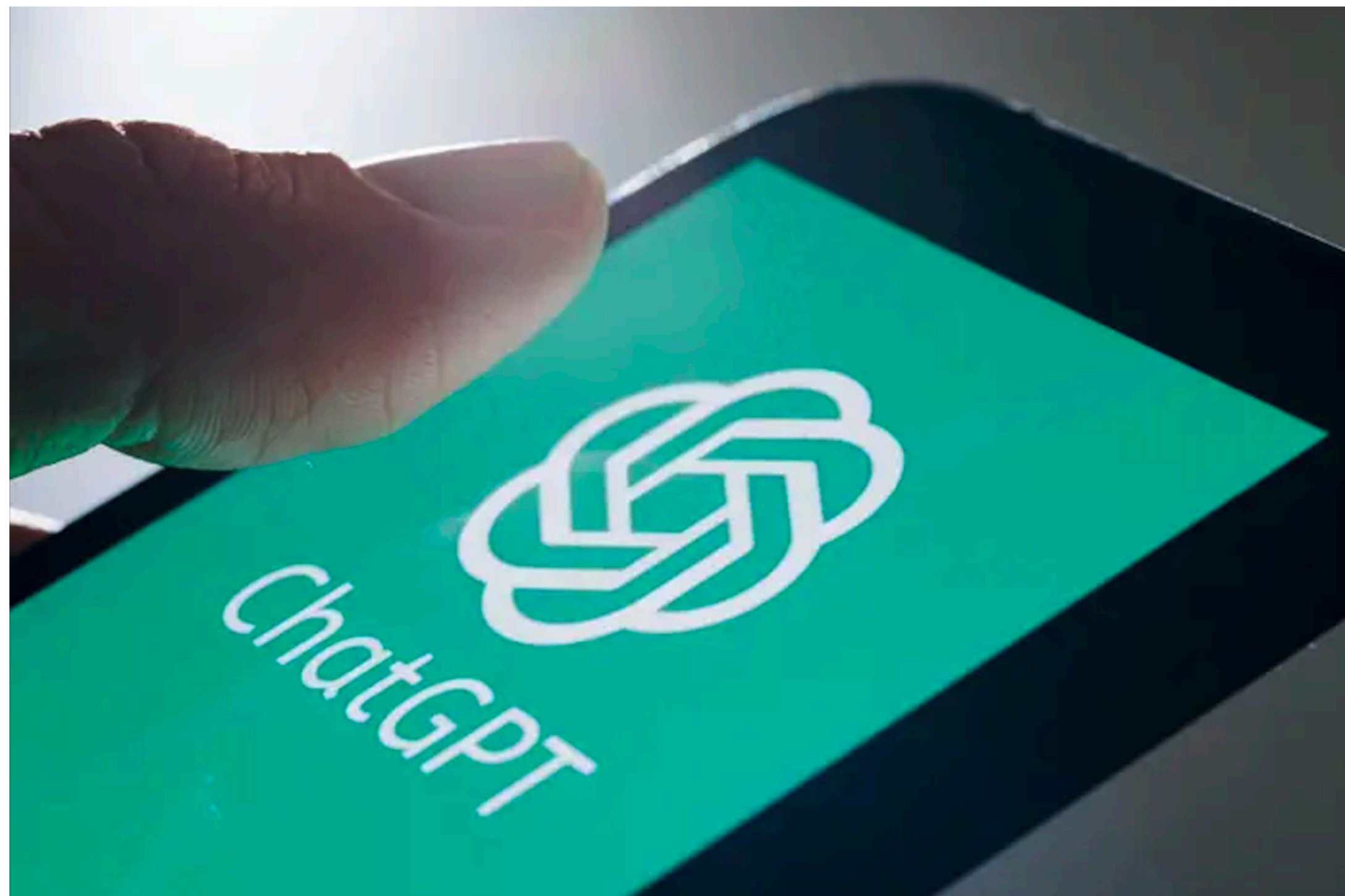
DAQUAR 1553  
**What is there in front of the sofa?**  
Ground truth: table  
IMG+BOW: **table** (0.74)  
2-VIS+BLSTM: **table** (0.88)  
LSTM: **chair** (0.47)



COCOQA 5078  
**How many leftover donuts is the red bicycle holding?**  
Ground truth: three  
IMG+BOW: **two** (0.51)  
2-VIS+BLSTM: **three** (0.27)  
BOW: **one** (0.29)

# What is Machine Learning?

Speech: Speech to text, personal assistants, speaker identification...



# What is Machine Learning?

NLP: Translation, sentiment analysis, topic modeling, spam filtering.

## Real world example:

*The New York Times*

LDA analysis of 1.8M New York Times articles:



# Tools

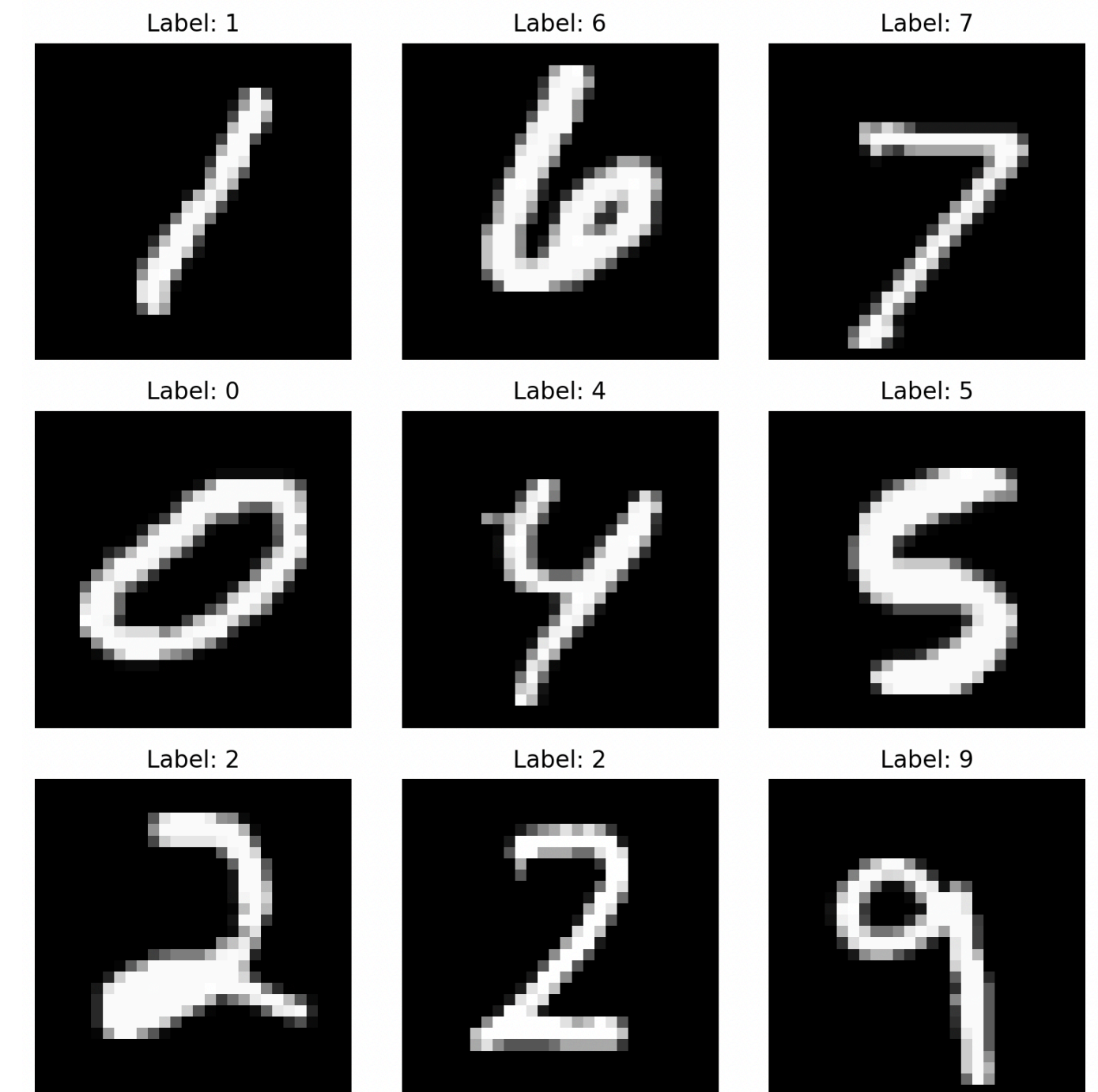
- Python: A programming language widely used for deep learning and AI development.
- PyTorch: A Python package that provides gradient computation and GPU acceleration for building and training neural networks.
- pip: A package manager for Python that allows easy installation and management of packages, including deep learning frameworks like PyTorch and TensorFlow.
- Jupyter Notebook: An interactive web-based environment for writing, executing, and visualizing Python code, commonly used in data science and deep learning.

# Tools

- Colab: A cloud-based Jupyter Notebook environment from Google that provides free GPU and TPU resources for deep learning experiments.
- IDE: An Integrated Development Environment, such as PyCharm or VS Code, that provides tools for writing, debugging, and running Python code efficiently.
- Git: A version control system that tracks code changes, enabling collaboration and efficient management of software projects.
- GitHub: A cloud-based platform that hosts Git repositories, allowing developers to collaborate, manage projects, and contribute to open-source code.

# ML From a Modelling Perspective

- MINIST dataset contains 60000 digit images of size (28 \* 28) with their label.
- $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$
- $\mathbf{x}_i$  is the 784 dim vector and  $y_i \in \{0,1,\dots,9\}$ .



# ML From a Modelling Perspective

## Linear Regression

- $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$
- $\mathbf{x}_i$  is the 784 dim vector and  $y_i \in \{0, 1, \dots, 9\}$ .
- Assume  $y = \mathbf{a}^\top \mathbf{x} + \mathbf{b} + \epsilon$ , with  $\epsilon \sim N(0, \sigma^2)$ .
- This implies  $y \mid \mathbf{x} \sim N(\mathbf{a}^\top \mathbf{x} + \mathbf{b}, \sigma^2)$ .
- This is not a good model and we will see it performs bad.

# ML From a Modelling Perspective

## Linear Regression

- Assume  $y = \mathbf{a}^\top \mathbf{x} + \mathbf{b} + \epsilon$ , with  $\epsilon \sim N(0, \sigma^2)$ .

- This implies  $y \mid \mathbf{x} \sim N(\mathbf{a}^\top \mathbf{x} + \mathbf{b}, \sigma^2)$ .

- $$p(y \mid \mathbf{x}; \mathbf{a}, \mathbf{b}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - (\mathbf{a}^\top \mathbf{x} + \mathbf{b}))^2}{2\sigma^2}\right).$$

- Likelihood of  $\mathcal{D}$ : 
$$L(\mathcal{D}) = \prod_{i=1}^N p(y_i \mid \mathbf{x}_i; \mathbf{a}, \mathbf{b}, \sigma^2).$$

- $$\log L(\mathcal{D}) = \sum_{i=1}^N \log p(y_i \mid \mathbf{x}_i; \mathbf{a}, \mathbf{b}, \sigma^2) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - (\mathbf{a}^\top \mathbf{x}_i + \mathbf{b}))^2$$

# ML From a Modelling Perspective

## Linear Regression

- $\log L(\mathcal{D}) = \sum_{i=1}^N \log p(y_i \mid \mathbf{x}_i; \mathbf{a}, \mathbf{b}, \sigma^2) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - (\mathbf{a}^\top \mathbf{x}_i + \mathbf{b}))^2$
- Want to select parameters  $\mathbf{a}, \mathbf{b}$  such that  $\log L(\mathcal{D})$  is maximized:
  - $a^*, b^* = \arg \min_{a, b} \sum_{i=1}^N (y_i - (\mathbf{a}^\top \mathbf{x}_i + \mathbf{b}))^2$
- Maximize log likelihood is equivalent to minimize the mean square error  $\frac{1}{N} \sum_{i=1}^N (y_i - (\mathbf{a}^\top \mathbf{x}_i + \mathbf{b}))^2$ .
- Demo 1.

# ML From a Modelling Perspective

## Multi-logistic Regression

- $y$  should be treated as class label, or categorical distribution over  $\{0,1,\dots,9\}$  given  $\mathbf{x}$ .
- We need 10 values to describe  $p(y \mid \mathbf{x})$ , basically  $p(y = j \mid \mathbf{x}), j = 0,1,\dots,9$ .
- Assume  $p_{\theta}(y \mid \mathbf{x}) = \text{softmax}(\mathbf{W}\mathbf{x} + \mathbf{b})$ ,  $\theta = (\mathbf{W}, \mathbf{b})$  is the parameter.
  - $\mathbf{W} \in \mathbb{R}^{C \times 784}$ ,  $\mathbf{b} \in \mathbb{R}^C$ ,  $C = 10$  is the number of class.
  - $\text{softmax}(\mathbf{W}\mathbf{x} + \mathbf{b})$  is a vector of dim  $C$ . Elements of  $\mathbf{W}\mathbf{x} + \mathbf{b}$  are called logits.
  - The  $i$ -th element of  $\text{softmax}(\mathbf{o})$  is  $\frac{\exp o^{(i)}}{\sum_j \exp(o^{(j)})}$ .

# ML From a Modelling Perspective

## Multi-logistic Regression

- The  $i$ -th element of  $\text{softmax}(\mathbf{W}\mathbf{x} + \mathbf{b})$  is  $\frac{\exp(W_i\mathbf{x})}{\sum_j \exp(W_j\mathbf{x})}$ .
- For given sample  $(\mathbf{x}_i, y_i)$ , the likelihood is  $p(y = y_i \mid \mathbf{x}_i) = \frac{\exp(W_{y_i}\mathbf{x})}{\sum_j \exp(W_j\mathbf{x})}$ .
- $$L(\mathcal{D}) = \prod_{i=1}^N p(y_i \mid \mathbf{x}_i) = \prod_{i=1}^N \frac{\exp(W_{y_i}\mathbf{x})}{\sum_j \exp(W_j\mathbf{x})}.$$
- Loglikelihood

- $$\log L(\mathcal{D}) = \sum_{i=1}^N \log\left(\frac{\exp(W_{y_i}\mathbf{x})}{\sum_j \exp(W_j\mathbf{x})}\right)$$

# ML From a Modelling Perspective

## Multi-logistic Regression

- "Assume  $p_{\theta}(y \mid \mathbf{x}) = \text{softmax}(\mathbf{W}\mathbf{x} + \mathbf{b})$ ,  $\theta = (\mathbf{W}, \mathbf{b})$  is the parameter."
- Is this model realistic?
- Now assume  $p_{\theta}(y \mid \mathbf{x}) = \text{softmax}(\text{MLP}(\mathbf{x}))$ , where  $\text{MLP}(\mathbf{x}) = g(\mathbf{W}_2 g(\mathbf{W}\mathbf{x}_1 + \mathbf{b}_1) + \mathbf{b}_2)$ ,  $\theta = (\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_1, \mathbf{b}_2)$ .

# Deep Learning

- Deep learning is a subset of machine learning that utilizes deep neural networks, which consist of multiple layers to learn complex patterns from data.
- AlexNet: A major breakthrough in deep learning occurred in 2012 with AlexNet, a deep convolutional neural network (CNN) designed for image classification.
- Model depth significantly improves performance, though it requires substantial computational power. This was made feasible by leveraging graphics processing units (GPUs) for efficient training.

