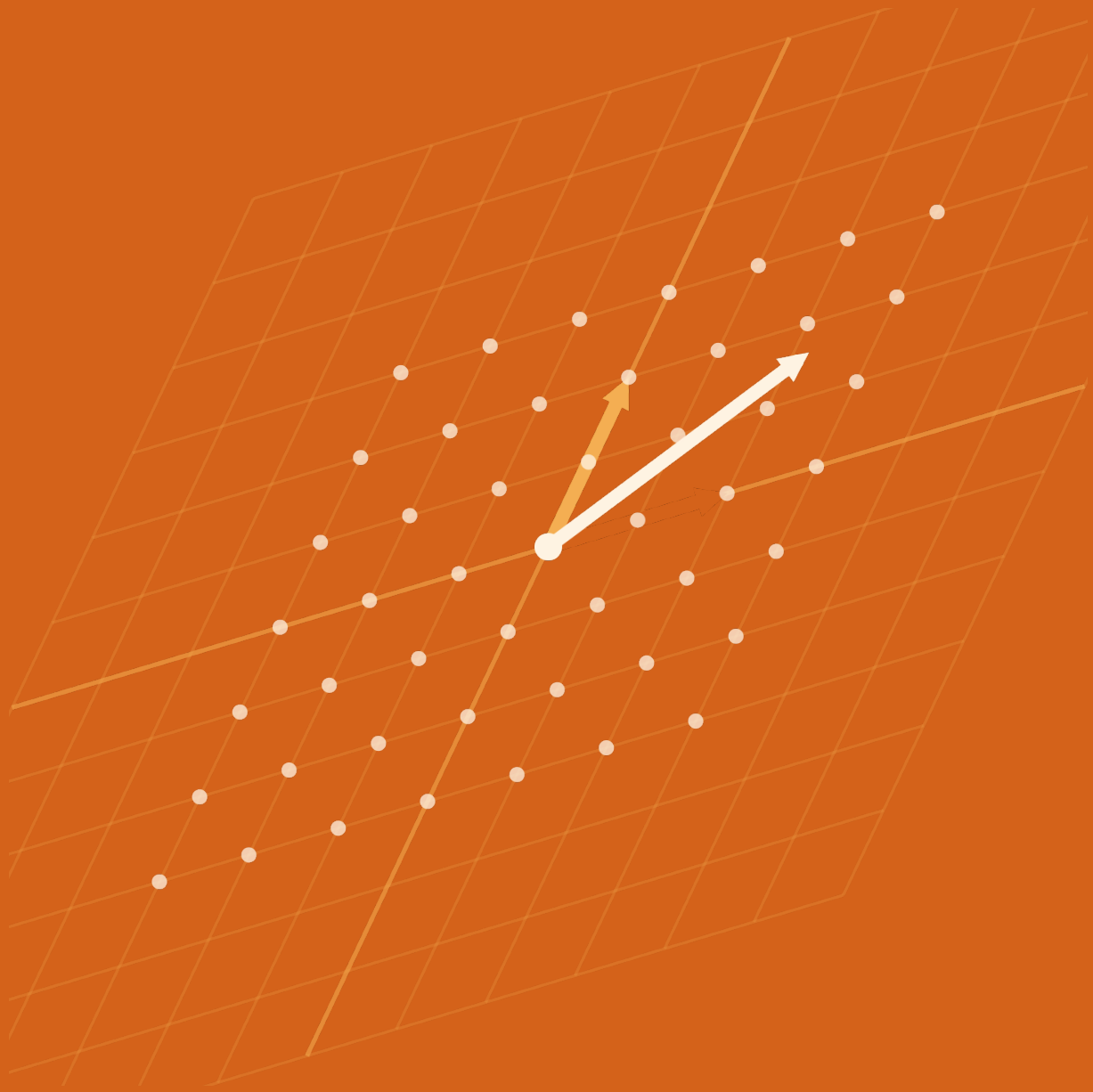
LINEAR ALGEBRA FOR AIML AND 6G

Introduction to Linear Algebra for AIML and 6G

Lecture 0

Prof. Aditya K. Jagannatham
IIT Kanpur



Outline

1. **Evolution of AI and ML**
   From the first artificial neuron to transformers and foundation models

2. **Evolution of 6G**
   Mobile generations from analog voice to the IMT-2030 vision

3. **Linear algebra for AI and ML**
   Vectors, matrices and decompositions at the heart of learning

01

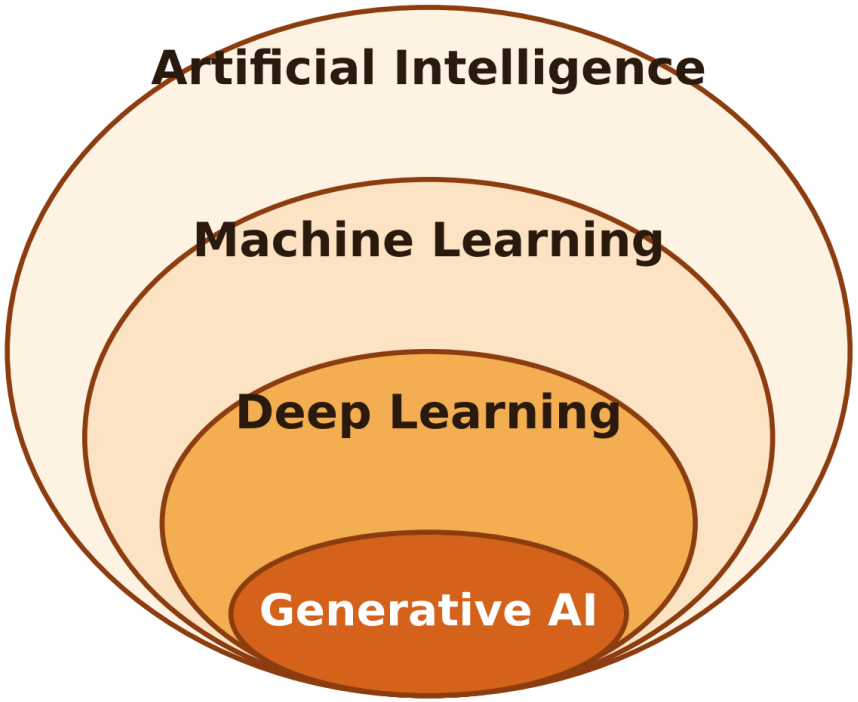
Evolution of AI and ML

From the first artificial neuron to transformers and foundation models



AI, ML and deep learning

- **Artificial intelligence**: machines doing tasks that need human intelligence
- **Machine learning**: learning patterns from data instead of hand-written rules
- **Deep learning**: machine learning with many-layered neural networks
- **Generative AI**: models that create text, images, audio and code



Early foundations: 1943–1969

- **1943:** McCulloch and Pitts propose a mathematical model of a neuron
- **1950:** Turing asks "Can machines think?"
- **1956:** the Dartmouth workshop coins the term **artificial intelligence**
- **1958:** Rosenblatt's **perceptron** learns a linear classifier

$$y = \text{sign}\left(\bar{\mathbf{w}}^T \bar{\mathbf{x}} + b\right)$$

- **1969:** Minsky and Papert show the limits of a single perceptron, leading to the first **AI winter**

Statistical learning: 1970s–2000s

- **Expert systems** encode human knowledge as if–then rules
- **1986: backpropagation** makes multilayer neural networks trainable
- **1989–1998:** LeCun's **convolutional networks** read handwritten digits
- **1995: support vector machines** find the maximum-margin hyperplane
- **2001: random forests** combine many decision trees

Deep learning revolution: 2006–2016

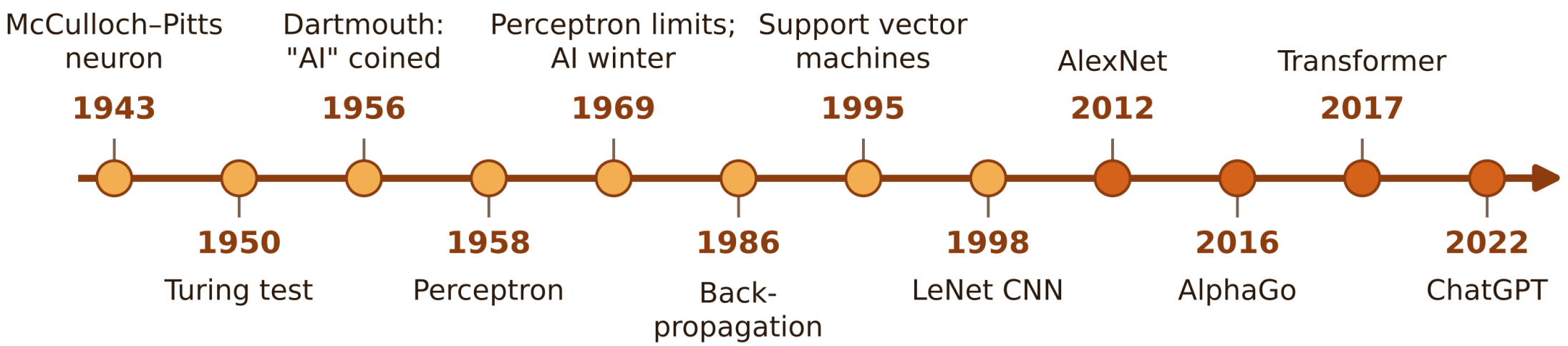
- Three ingredients came together: **big data**, **GPUs** and **better training methods**
- **2009: ImageNet** provides over 14 million labelled images
- **2012: AlexNet** cuts the ImageNet top-5 error from 26% to 15%
- **2015: ResNet** trains networks with 152 layers
- **2016: AlphaGo** defeats Go champion Lee Sedol

Foundation models: 2017 onwards

- **2017:** the **Transformer** replaces recurrence with attention
- **2018–2020:** BERT and GPT-3 (175 billion parameters) are pre-trained on web-scale text
- **2022: ChatGPT** brings large language models to the public
- **Today:** multimodal, reasoning and agentic models built on the same matrix operations

AI and ML timeline

- Each wave was driven by new mathematics, more data and more compute



02

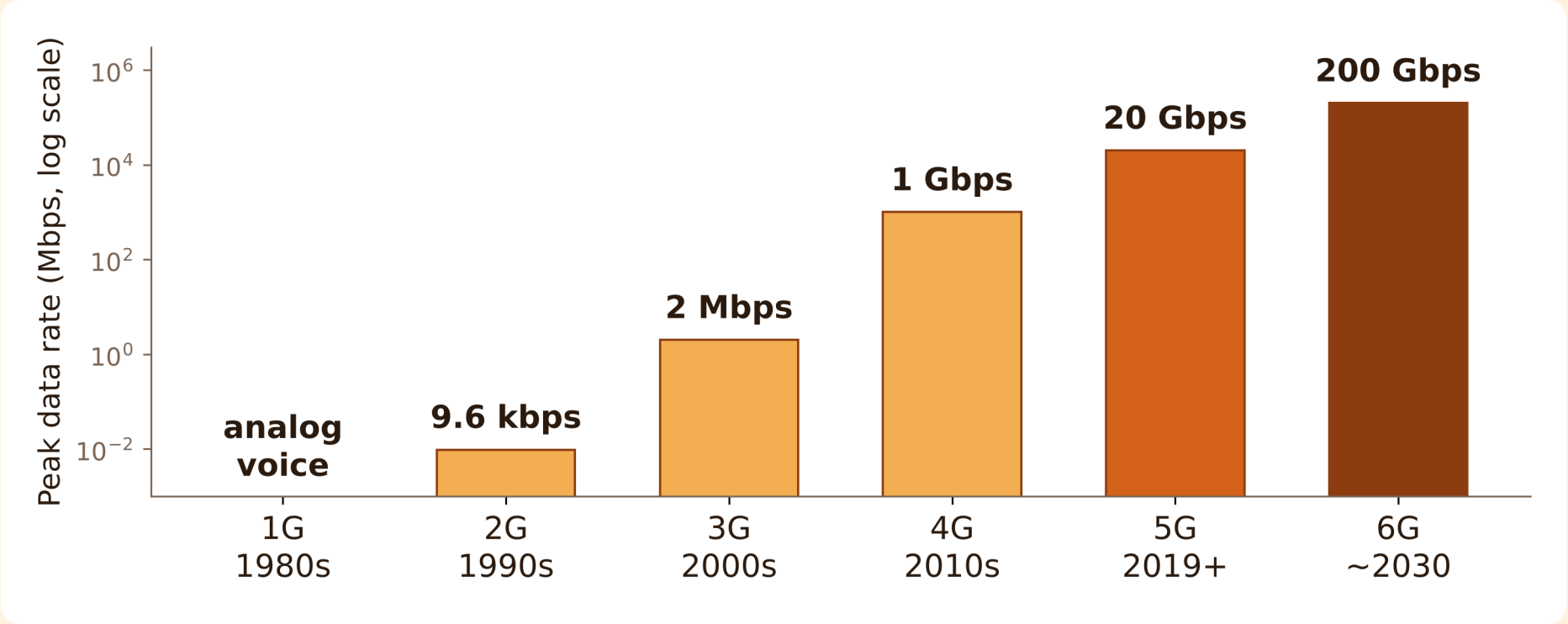
Evolution of 6G

Mobile generations from analog voice to the IMT-2030 vision



Generations of mobile networks

- A new generation arrives roughly every decade; peak rates for 3G–6G are ITU targets



1G to 4G

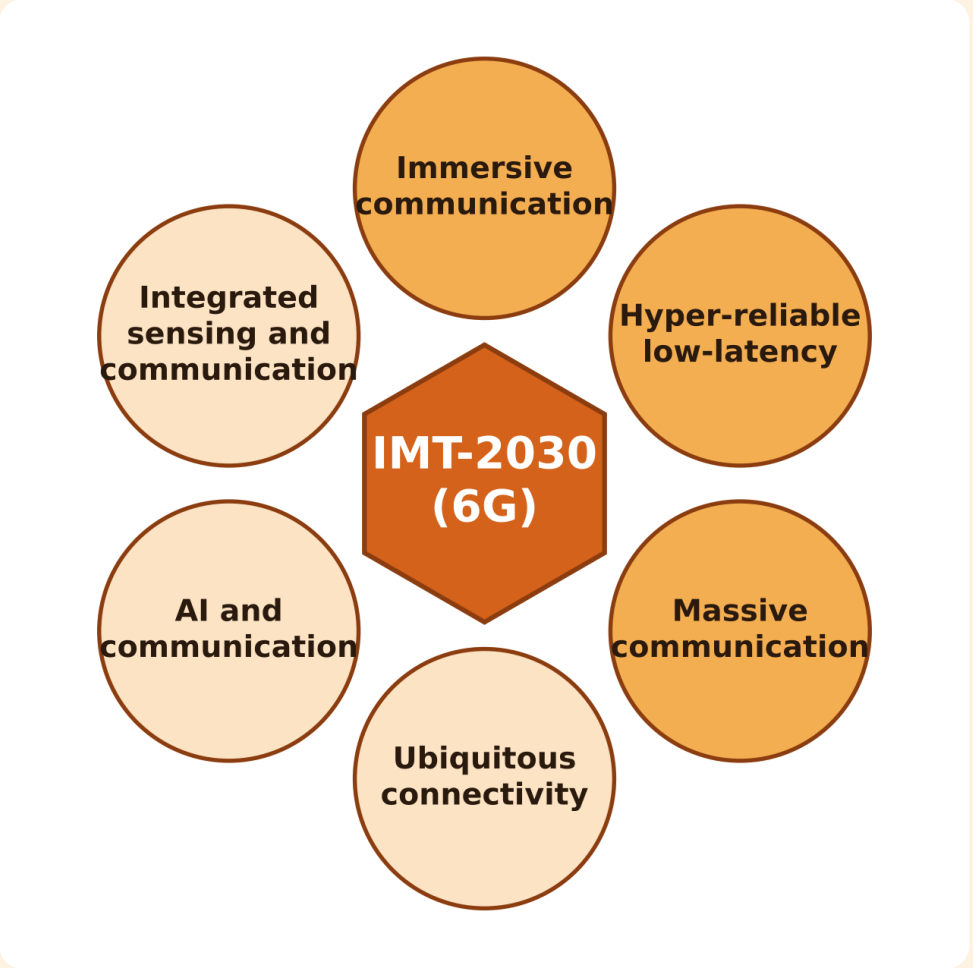
- **1G** (1980s): analog voice using FDMA
- **2G** (1991): digital voice and SMS; GSM uses TDMA
- **3G** (2001): mobile internet using WCDMA
- **4G LTE** (2009): all-IP broadband using **OFDM** and **MIMO**
- OFDM is built on the **DFT matrix**; MIMO is the matrix model $\bar{\mathbf{y}} = \mathbf{H}\bar{\mathbf{x}} + \bar{\mathbf{n}}$

5G: 2019 onwards

- First commercial launches in 2019, based on 3GPP Release 15
- **eMBB**: enhanced mobile broadband, up to 20 Gbps peak
- **URLLC**: ultra-reliable low-latency communication, 1 ms
- **mMTC**: massive machine-type communication, 10^6 devices per km²
- Key technologies: **massive MIMO**, **mmWave**, beamforming and network slicing

Vision for 6G: IMT-2030

- ITU-R framework for IMT-2030, Recommendation M.2160 (2023)
- Extends three 5G scenarios: immersive, hyper-reliable low-latency and massive communication
- Adds three new ones: **ubiquitous connectivity**, **AI and communication**, **integrated sensing and communication**



6G targets and timeline

- **IMT-2030 targets**
  - Peak data rate 50–200 Gbps, latency 0.1–1 ms
  - Connection density 10^6–10^8 devices per km²
  - Positioning accuracy 1–10 cm
- **Standardisation**
  - 3GPP Release 20: 6G studies, 2025–2027
  - 3GPP Release 21: first 6G specifications, 2027–2028
  - First commercial 6G systems expected around 2030

6G enabling technologies

- **Sub-THz and THz** bands for very wide bandwidths
- **Extremely large and cell-free MIMO**: very large channel matrices
- **Reconfigurable intelligent surfaces** that shape the radio channel
- **Integrated sensing and communication** (ISAC)
- **AI-native air interface** and **non-terrestrial networks** (satellites, drones)

03

Linear algebra for AI and ML

Vectors, matrices and decompositions at the heart of learning



Data as vectors and matrices

- A sample with n features is a vector $\bar{\mathbf{x}} \in \mathbb{R}^n$
- A dataset of m samples is a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$
- A grayscale image is a matrix of pixel values
- Words and tokens are mapped to **embedding vectors**
- Similarity of two embeddings uses the angle between them

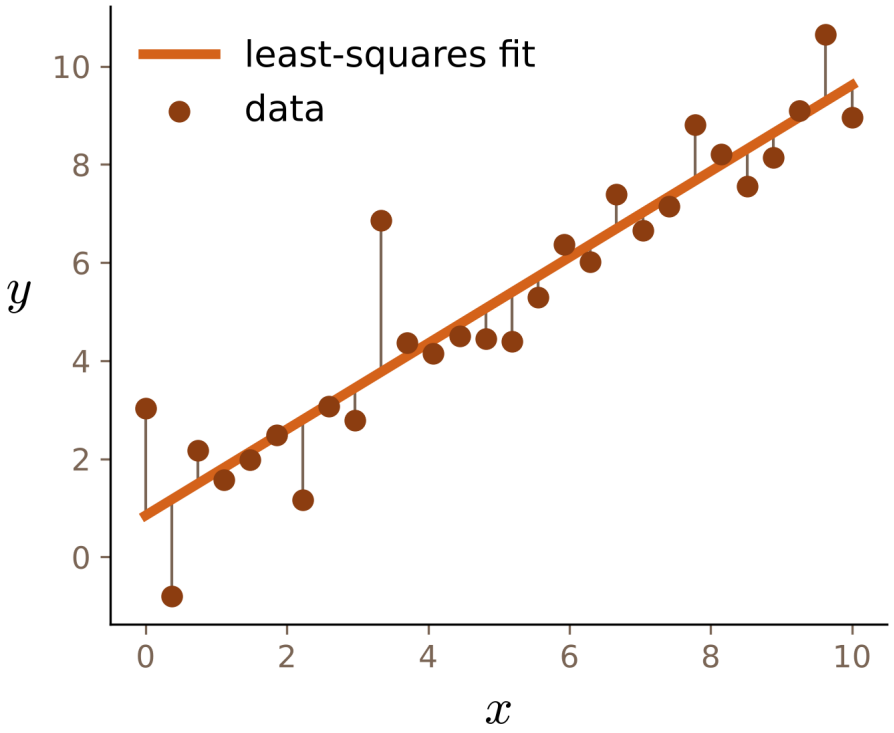
$$\cos\theta = \frac{\bar{\mathbf{u}} \cdot \bar{\mathbf{v}}}{\|\bar{\mathbf{u}}\| \, \|\bar{\mathbf{v}}\|}$$

Linear regression

- Fit a linear model $\bar{\mathbf{y}} \approx \mathbf{X}\bar{\mathbf{w}}$
- Minimise the squared error $\|\bar{\mathbf{y}} - \mathbf{X}\bar{\mathbf{w}}\|^2$
- The **normal equations** give the least-squares solution

$$\bar{\mathbf{w}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\bar{\mathbf{y}}$$

- Uses transpose, inverse, rank and projection

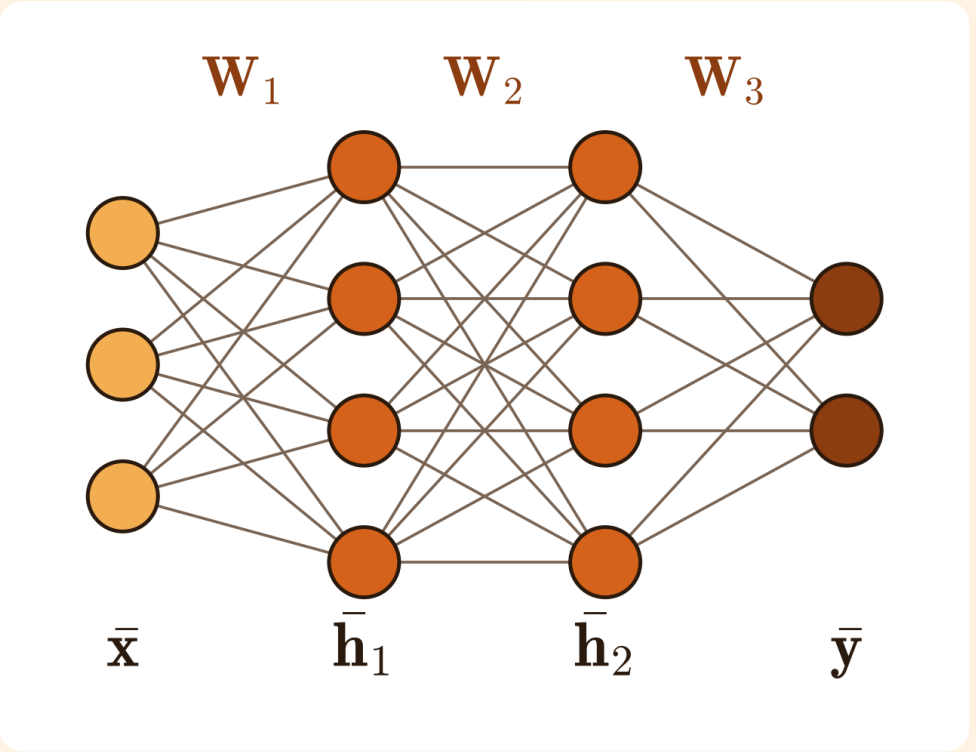


Neural networks

- Each layer is a matrix–vector product followed by a nonlinearity

$$\bar{\mathbf{h}}_1 = \sigma\left(\mathbf{W}_1\bar{\mathbf{x}} + \bar{\mathbf{b}}_1\right)$$

- A deep network composes many such layers
- **Backpropagation** computes gradients as a chain of matrix products
- **GPUs** are built for fast matrix multiplication



Attention in transformers

- Token embeddings form the rows of $\mathbf{X}$

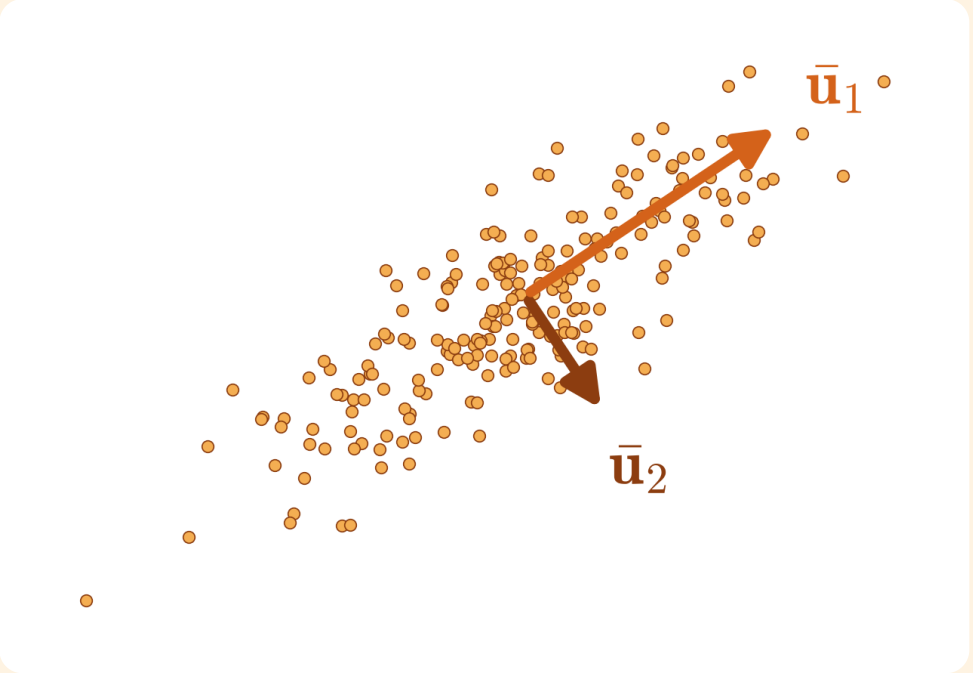
$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \mathbf{K} = \mathbf{X}\mathbf{W}_K, \mathbf{V} = \mathbf{X}\mathbf{W}_V$$

$$\text{Attention}\left(\mathbf{Q}, \mathbf{K}, \mathbf{V}\right) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}$$

- $\mathbf{Q}\mathbf{K}^T$ holds the inner products of queries and keys
- Large language models perform billions of such matrix products

PCA and the SVD

- Covariance of centred data: $\mathbf{C} = 1/m\mathbf{X}^T\mathbf{X}$
- **Eigendecomposition**: $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$
- Principal components are the eigenvectors with the largest eigenvalues
- **SVD**: $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ gives the best low-rank approximation



Linear algebra across this course

- **Vectors, norms, inner products**: embeddings, similarity, beamforming
- **Matrices, linear equations**: linear models, MIMO detection
- **Vector spaces, bases, rank**: features, low-rank models
- **Eigenvalues, SVD**: PCA, recommender systems, MIMO capacity
- **Least squares, pseudo-inverse**: regression, channel estimation
- **Matrix calculus**: training neural networks

Summary

- AI grew from the 1943 neuron model to **deep learning** and **transformers**
- Mobile networks evolve every decade, from analog **1G** to **5G**
- **6G** (IMT-2030) adds AI, sensing and ubiquitous coverage, around 2030
- Data, images and embeddings are **vectors and matrices**
- Regression, neural networks, attention and PCA are **linear algebra** at work



Thank You!

Questions?