UNIVERSITY OF ICELAND

Introduction to Deep Learning

Based on "Dive into Deep Learning" by Zhang et al.

Instructors: Guðmundur Einarsson & Hlynur Davíð Hlynsson

University of Iceland

Teacher Introduction

Guðmundur Einarsson

- Teaches **weeks 1–5**
- Theoretical foundations of deep learning
- gue20@hi.is

Hlynur Davíð Hlynsson

- Teaches the **latter part** of the course
- Applied aspects of deep learning

> Together, you'll get a solid theoretical foundation followed by hands-on, applied experience.

Deep Neural Networks – Course Overview

This course introduces **deep neural networks** and the powerful methods associated with them.

We will cover:

- Design and function of networks for various tasks
- Image, audio, and text analysis
- Solid theoretical foundation
- Practical implementation using PyTorch

The course emphasizes **active participation** through assignments and quizzes, concluding with a larger **final project** where you'll tackle a real-world problem.

Learning Outcomes

Upon completion of the course, you will be able to:

1. **Explain** the fundamental concepts and mechanics of deep neural networks
2. **Train** deep neural networks to solve practical problems
3. **Utilize** the PyTorch library for model implementation
4. **Differentiate** between and apply major types of neural networks:
    - Convolutional Neural Networks (CNNs)
    - Recurrent Neural Networks (RNNs)
    - Generative Models
5. **Design and implement** solutions for a variety of tasks using deep neural networks

Course schedule (Weeks 1-6)

Week	Topic	Assessment
1	Course introduction, supervised learning, multi-layer neural networks	-
2	Training neural networks: Optimization, regularization, backpropagation	Quiz 1
3	Convolutional Neural Networks (CNNs) for images and time series	Assignment 1
4	CNNs (continued) and Transfer Learning	Quiz 2
5	Recurrent Neural Networks (RNNs) and sequence models	Assignment 2
6		Quiz 3

**Subject to change*

Course Schedule (Remaining Weeks)

> From this point on, **Hlynur Davíð Hlynsson** will take over the course and present the remaining material, focusing on applied aspects of deep learning.

The detailed schedule for these weeks will be announced by Hlynur when he takes over.

Course Logistics

📚 Textbook

Dive Into Deep Learning (Zhang et al., 2022)

Available online free of charge at d2l.ai

📅 Schedule

Lectures: Thursdays 10:00–12:20 in Gr-321
Problem Sessions & Quizzes: Tuesdays 8:20–9:50 in Gr-321

Problem sessions provide an opportunity to work on assignments with assistance

🌐 Course Website

Canvas is the main platform for the course

All teaching materials, assignments, announcements, and discussions

⚠️ Check Canvas regularly!

Assessment & Grading

Grade Distribution

10%	25%	15%	50%
Assignments (2% each)	Quizzes (5% each)	Final Project	Final Exam

Passing Requirements

- **Undergraduate:** Total grade ≥ 5.0, Final exam ≥ 5.0
- **Master's:** Total grade ≥ 6.0, Final exam ≥ 6.0

Eligibility for Final Exam

⚠️ By end of week 7, must have:

- Submitted 2 of first 3 assignments
- Taken 2 of first 3 quizzes

Prerequisites

Mathematical Background

- Basic knowledge of **linear algebra** and **calculus**
  - Derivatives and matrix/vector operations
- Familiarity with **data science concepts**
  - Supervised learning, classification, regression

Programming Skills

- Good **programming skills**
  - Python will be used (experience with Java, C, or Matlab is beneficial)

Collaboration Policy

✅ Encouraged

- Discuss course material and assignments
- Study together and share learning resources
- Help each other understand concepts

⚠️ Required

- Write and submit your own implementation
- State if you collaborated with others

❌ Forbidden

- Copy solutions or share your own solutions
- Submit work that is not your own

Contact Information

Instructor

Guðmundur Einarsson

gue20@hi.is

Course Questions

Use Canvas forums for course-related questions

This helps everyone benefit from Q&A

Before the Revolution

The State of AI

- Computer vision relied on hand-crafted features (SIFT, HOG)
- Training limited to CPUs - weeks for modest networks
- Lack of large labeled datasets
- Deep networks suffered from vanishing gradient problem
- Best ImageNet accuracy: ~74% (2011 results)

Sources: ImageNet Large Scale Visual Recognition Challenge 2011 Results (https://image-net.org/challenges/LSVRC/2011/results)

The AlexNet Moment

September 30, 2012

ImageNet Challenge Results:

- AlexNet achieves 15.3% top-5 error rate
- Runner-up: 26.2% (10.9 percentage points behind!)
- First use of GPUs for deep learning at scale

AlexNet's Key Innovations

- 8-layer deep CNN trained on 2 GPUs
- ReLU activation (instead of sigmoid/tanh)
- Dropout for regularization
- Data augmentation techniques
- Local Response Normalization

> "We trained a large, deep convolutional neural network... achieving record-breaking results"
>
> - Krizhevsky, Sutskever & Hinton

Immediate Impact (2013-2016)

Rapid Progress

- 2014: VGGNet, GoogLeNet
- 2015: ResNet (3.57% error)
- 2016: Exceeds human performance
- Industry adoption explodes

Key Breakthroughs

- Deeper networks (100+ layers)
- Skip connections
- Batch normalization
- Transfer learning

Every major tech company established AI research labs

Today's Landscape

Foundation Models

- Transformers revolutionize NLP
- GPT, BERT, and beyond
- Multimodal models

New Frontiers

- Diffusion models for generation
- AI agents and reasoning
- Embodied AI in robotics

> "That moment was pretty symbolic to the world of AI because three fundamental elements of modern AI converged for the first time. The first element was neural networks. The second element was big data, using ImageNet. And the third element was GPU computing."
>
> - Fei-Fei Li, 2024

Sources: The Turing Post - Computer Vision History (https://www.turingpost.com/p/cvhistory6) · IEEE Spectrum - AlexNet Source Code (https://spectrum.ieee.org/alexnet-source-code) · Interview with Fei-Fei Li (https://www.youtube.com/watch?v=JgQ1FJ_wow8&t=811s)

What Can Deep Learning Do?

Traditional Domains

- Computer Vision
- NLP
- Speech Recognition
- Reinforcement Learning

Emerging Applications

- Self-driving cars
- AI agents and autonomous systems
- Embodied intelligence
- Scientific discovery

> Deep learning is transforming industries across the spectrum: healthcare diagnostics, climate science, drug discovery, and creative applications in art and media.

About This Course

Our Mission

Make deep learning approachable by teaching:

- **Concepts** - The theoretical foundations
- **Context** - When and why to use techniques
- **Code** - Practical implementation

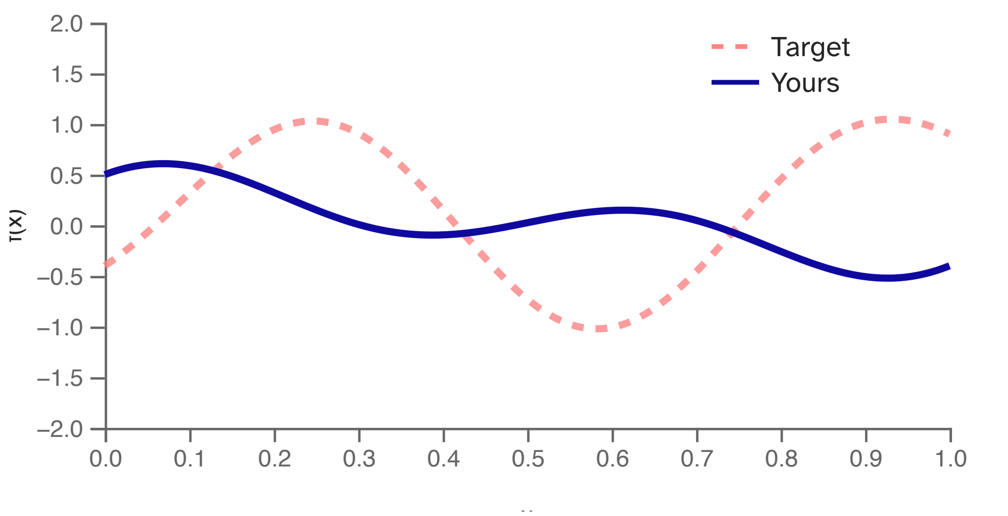
💡 Key Philosophy

Combine rigorous mathematics with runnable code and interactive visualizations

Interactive Demo

$$f(x) = a_1 \sin(b_1 \pi x) + a_2 \cos(b_2 \pi x) + a_3 x^2$$

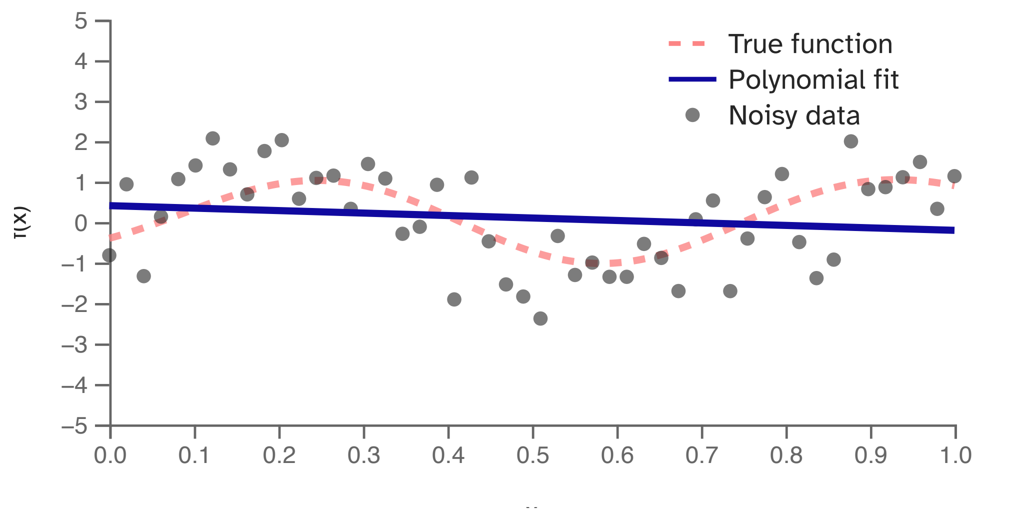
Interactive applet in web based slides



Interactive Demo: Overfitting

Fit a polynomial to **50 noisy data points** — what happens as you raise the degree?

Interactive applet in web based slides



Learning by Doing

Traditional textbooks: exhaustive detail on each topic

Our approach: Just-in-time learning

- Learn concepts when you need them
- Taste success early - train your first model quickly
- One working example per notebook
- Real datasets, practical applications

> 🚀 Start building, then understand deeply

Course Structure

Getting Started

- **Chapter 1: Introduction**
  - The deep learning revolution
  - Key concepts and terminology
  - Course overview and goals
- **Chapter 2: Preliminaries**
  - Data manipulation and preprocessing
  - Linear algebra essentials
  - Calculus and automatic differentiation
  - Probability and statistics

Course Structure

Linear Models

- **Chapter 3: Linear Neural Networks for Regression**
  - Linear regression from scratch
  - Gradient descent and optimization
  - Vectorized implementation
- **Chapter 4: Linear Neural Networks for Classification**
  - Softmax regression
  - Cross-entropy loss
  - Fashion-MNIST dataset

Course Structure

First Deep Networks

- **Chapter 5: Multilayer Perceptrons**
  - Hidden layers and activation functions
  - Backpropagation algorithm
  - Weight initialization and dropout
- **Chapter 6: Builders' Guide**
  - Layers, blocks, and models
  - Parameter management
  - Custom layers and saving/loading

Course Structure

Convolutional Neural Networks

- **Chapter 7: Convolutional Neural Networks**
  - Convolution and cross-correlation
  - Padding and stride
  - Pooling layers
  - LeNet - the first ConvNet
- **Chapter 8: Modern Convolutional Neural Networks**
  - AlexNet - deep learning breakthrough
  - VGG - using blocks
  - ResNet - skip connections
  - DenseNet and EfficientNet

Course Structure

Recurrent Neural Networks

- **Chapter 9: Recurrent Neural Networks**
  - Sequence modeling
  - Text preprocessing
  - Language models
  - Backpropagation through time
- **Chapter 10: Modern Recurrent Neural Networks**
  - LSTM - solving vanishing gradients
  - GRU - simplified gating
  - Bidirectional RNNs
  - Encoder-decoder architecture

Course Structure

Attention and Transformers

- **Chapter 11: Attention Mechanisms & Transformers**
  - Attention pooling
  - Bahdanau attention
  - Self-attention and positional encoding
  - The Transformer architecture
  - Multi-head attention

> 💡 Transformers are the foundation of modern NLP

Course Structure

Optimization and Performance

- **Chapter 12: Optimization Algorithms**
  - SGD and momentum
  - AdaGrad and RMSprop
  - Adam optimizer
  - Learning rate scheduling
- **Chapter 13: Computational Performance**
  - Hardware: CPUs vs GPUs vs TPUs
  - Multiple GPUs and parallelization
  - Training on the cloud

Course Structure

Computer Vision

- **Chapter 14: Computer Vision**
  - Image augmentation techniques
  - Fine-tuning pretrained models
  - Object detection: R-CNN family
  - Single shot detection: YOLO
  - Semantic segmentation

> 🖼️ From image classification to pixel-level understanding

Course Structure

Natural Language Processing

- **Chapter 15: NLP Pretraining**
  - Word embeddings: Word2Vec and GloVe
  - Subword embedding
  - BERT for understanding
  - GPT for generation
- **Chapter 16: NLP Applications**
  - Sentiment analysis
  - Natural language inference
  - Question answering

Course Structure

Advanced Topics I

- **Chapter 17: Reinforcement Learning**
  - Markov Decision Processes
  - Q-Learning
  - Deep Q-Networks (DQN)
  - Policy gradient methods
- **Chapter 18: Gaussian Processes**
  - Bayesian approach to ML
  - Kernel methods
  - Uncertainty quantification

Course Structure

Advanced Topics II

- **Chapter 19: Hyperparameter Optimization**
  - Grid and random search
  - Bayesian optimization
  - Automated ML (AutoML)
- **Chapter 20: Generative Adversarial Networks**
  - The GAN framework
  - Deep Convolutional GANs
  - StyleGAN and applications

Course Structure

Real-World Applications

- **Chapter 21: Recommender Systems**
  - Collaborative filtering
  - Matrix factorization
  - Deep learning for recommendations
  - Real-world considerations

📚 All chapters follow the d2l.ai book structure

🔗 Free online with code examples and exercises

Prerequisites & Tools

What You Need

- Basic linear algebra
- Elementary calculus
- Probability basics
- Python programming

What We'll Use

- PyTorch framework
- Jupyter notebooks
- GitHub for code
- Discussion forum

📚 All materials freely available at d2l.ai

Let's Get Started! 🚀

Your First Tasks

1. Set up your development environment
   - Complete installation guide: d2l.ai/chapter_installation
2. Download the course notebooks
   - Direct download: d2l.ai/d2l-en.zip
   - GitHub repository: github.com/d2l-ai/d2l-en
3. Run your first example
   - Open the first notebook and execute the cells
   - Verify GPU setup (if available)

Optional: Join the course forum at discuss.d2l.ai for community support

Mathematical Notation Reference

Throughout this course, we'll use standard mathematical notation

This is a quick reference - we'll cover details as needed

> 💡 Don't worry about memorizing all of this!
>
> Full coverage in Chapter 2: Preliminaries

Reference: d2l.ai notation guide

Numbers and Arrays I

Notation	Meaning	Example
x	Scalar (single number)	$x = 3.14$
$\mathbf{x}$	Vector (1D array)	$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$
$\mathbf{X}$	Matrix (2D array)	$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
X	Tensor (n-D array)	3D, 4D arrays, etc.
$\mathbf{I}$	Identity matrix	$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Numbers and Arrays II

Notation	Meaning	Example
x_i, $[\mathbf{x}]_i$	i-th element of vector	$x_2 = 2$ in $\mathbf{x} = [1, 2, 3]$
x_{ij}, $[\mathbf{X}]_{ij}$	Element at row i, column j	$x_{12} = 2$ in example above

💡 Bold lowercase = vectors, Bold uppercase = matrices

Set Theory I - Common Sets

Notation	Meaning	Example
$\mathbb{R}$	Real numbers	$3.14, -2.5, \sqrt{2}$
$\mathbb{Z}$	Integers	$\ldots, -2, -1, 0, 1, 2, \ldots$
$\mathbb{Z}^+$	Positive integers	$1, 2, 3, 4, \ldots$
$\mathbb{R}^n$	n-dimensional vectors	$\mathbb{R}^3$ = 3D vectors
$\mathbb{R}^{a \times b}$	$a \times b$ matrices	$\mathbb{R}^{2 \times 3}$ = 2×3 matrices

Example: If $\mathbf{x} \in \mathbb{R}^{784}$, then $\mathbf{x}$ is a 784-dimensional vector

(like a flattened 28×28 image)

Set Theory II - Operations

Notation	Meaning	Example
$\lvert\mathcal{X}\rvert$	Cardinality (size)	$\lvert\{1,2,3\}\rvert = 3$
$\mathcal{A} \cup \mathcal{B}$	Union (A or B)	$\{1,2\} \cup \{2,3\} = \{1,2,3\}$
$\mathcal{A} \cap \mathcal{B}$	Intersection (both)	$\{1,2\} \cap \{2,3\} = \{2\}$
$\mathcal{A} \setminus \mathcal{B}$	Difference (A not B)	$\{1,2,3\} \setminus \{2\} = \{1,3\}$

Functions I

Notation	Meaning	Example
$f(\cdot)$	A function	$f(x) = x^2$
$\log(\cdot)$	Natural logarithm (base e)	$\log(e) = 1$
$\log_2(\cdot)$	Logarithm base 2	$\log_2(8) = 3$
$\exp(\cdot)$	Exponential function	$\exp(x) = e^x$
$\mathbf{1}(\cdot)$	Indicator function	$\mathbf{1}(x > 0) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$

Functions II

Notation	Meaning	Example
$\mathbf{X}^\top$	Transpose	$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^\top = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$
$\mathbf{X}^{-1}$	Matrix inverse	$\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}$
$\odot$	Hadamard product	$[1, 2] \odot [3, 4] = [3, 8]$
$[\cdot, \cdot]$	Concatenation	$[[1, 2], [3, 4]] = [1, 2, 3, 4]$
$\mathbf{1}_{\mathcal{X}}(z)$	Set membership indicator	1 if $z \in \mathcal{X}$, else 0

Operators & Norms

Notation	Meaning	Example
$\|\cdot\|_p$	ℓ_p norm	$\|\mathbf{x}\|_p = \left(\sum_i \lvert x_i \rvert^p\right)^{1/p}$
$\|\cdot\|$	ℓ_2 norm (default)	$\|[3,4]\| = 5$
$\langle \mathbf{x}, \mathbf{y} \rangle$	Inner product	$\langle [1,2],[3,4] \rangle = 11$
$\sum$	Summation	$\sum_{i=1}^{3} i = 6$
$\prod$	Product	$\prod_{i=1}^{3} i = 6$

> 💡 $\stackrel{\text{def}}{=}$ means "is defined as"

Calculus

Notation	Meaning	Example
$\frac{dy}{dx}$	Derivative of y w.r.t. x	$\frac{d}{dx}(x^2) = 2x$
$\frac{\partial y}{\partial x}$	Partial derivative	$\frac{\partial}{\partial x}(x^2 + y^2) = 2x$
$\nabla_{\mathbf{x}} y$	Gradient of y w.r.t. vector $\mathbf{x}$	$\nabla_{\mathbf{x}} f = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]$
$\int_a^b f(x)\,dx$	Definite integral	$\int_0^1 x^2\,dx = \frac{1}{3}$
$\int f(x)\,dx$	Indefinite integral	$\int x^2\,dx = \frac{x^3}{3} + C$

Key for DL: Gradients tell us how to update parameters

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \nabla_{\mathbf{w}} L \quad \text{(gradient descent)}$$

Probability I

Notation	Meaning	Example
X	Random variable	X = outcome of dice roll
P	Probability distribution	P = uniform distribution
$X \sim P$	X follows distribution P	$X \sim \mathcal{N}(0, 1)$
$P(X = x)$	Probability of event	$P(\text{dice} = 6) = \frac{1}{6}$
$P(X \mid Y)$	Conditional probability	$P(\text{rain} \mid \text{clouds})$

Probability II

Notation	**Meaning**	**Example**
$p(\cdot)$	Probability density function	$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
$\mathbb{E}[X]$	Expectation (mean)	$\mathbb{E}[\text{dice}] = 3.5$
$X \perp Y$	Independence	Coin flips are independent
$X \perp Y \mid Z$	Conditional independence	$X \perp Y$ given Z

Example: $X \sim \mathcal{N}(\mu, \sigma^2)$ means X follows a normal distribution

Statistics & Information Theory

Notation	Meaning	Example
σ_X	Standard deviation	Spread of data
$\mathrm{Var}(X)$	Variance = σ_X^2	$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2]$
$\mathrm{Cov}(X,Y)$	Covariance	$\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$
$\rho(X,Y)$	Correlation coefficient	$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$
$H(X)$	Entropy	$-\sum_x p(x) \log p(x)$

> 💡 $D_{\mathrm{KL}}(P\|Q)$ = KL divergence

Welcome to Deep Learning!

> *"The goal is to make deep learning concepts intuitive and engaging."*

Questions?

Next lecture: Introduction to the Basics of Deep Learning

⚠️ This repository may change throughout the course. If you have cloned it, `git pull` regularly to get the newest updates.