

# EE-559 – Deep Learning



You can find here the materials for the [EPFL](#) course [EE-559 “Deep Learning”](#), taught by [François Fleuret](#).

Info sheet: [dlc-info-sheet.pdf](#)

We will use the [PyTorch](#) framework for implementations. You can find below a [Linux virtual machine](#) for the practical sessions.

Thanks to Adam Paszke, Alexandre Nanchen, Xavier Glorot, Matus Telgarsky, and Diederik Kingma, for their help, comments, or remarks.

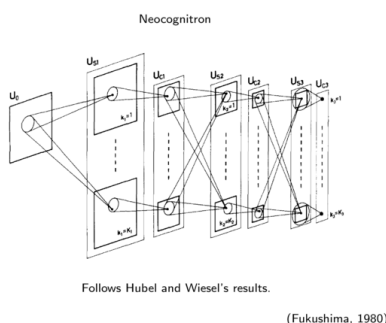
## Course material

You will find here the slides which are full of “animations” and not convenient to use as notes, handouts with two slides per pages, and for some of the lectures videos of voice-over.

## Practical session prologue

Helper python prologue for the practical sessions: [dlc\\_practical\\_prologue.py](#)

## Lecture 1 (Feb 21, 2018) – Introduction and tensors

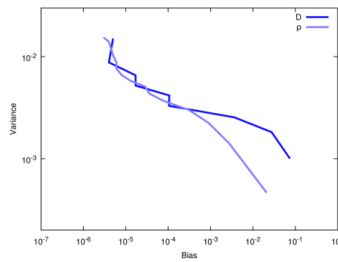


What is deep learning, some history, what are the current applications. `torch.Tensor`, linear regression.

- [slides](#) / [handout](#) / [video](#) (part a)
- [slides](#) / [handout](#) / [video](#) (part b)
- [practical](#)

## Lecture 2 (Feb 28, 2018) – Machine learning fundamentals

From this comes the **bias variance tradeoff**:



Reducing the capacity makes  $f^*$  fit the data less on average, which increases the bias term.

Empirical risk minimization, capacity, bias-variance dilemma, polynomial regression, k-means and PCA.

- [slides](#) / [handout](#)
- [practical](#)

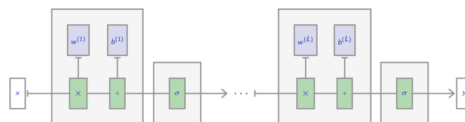
## Lecture 3 (Mar 07, 2018) – Multi-layer perceptrons

We can combine several "layers":

With  $x^{(0)} = x$ ,

$$\forall l = 1, \dots, L, \quad x^{(l)} = \sigma(w^{(l)}x^{(l-1)} + b^{(l)})$$

and  $f(x; w, b) = x^{(L)}$ .

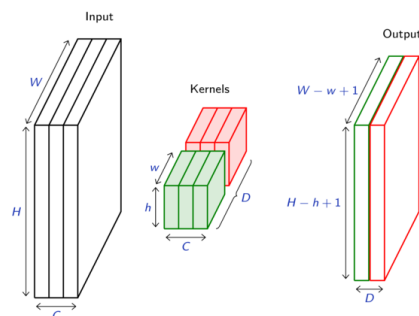


Such a model is a **Multi-Layer Perceptron (MLP)**.

Linear classifiers, perceptron, linear separability and feature extraction, Multi-Layer Perceptron, gradient descent, back-propagation.

- [slides](#) / [handout](#) (part a)
- [slides](#) / [handout](#) (part b)
- [practical](#)

## Lecture 4 (Mar 14, 2018) – Convolutional networks and autograd

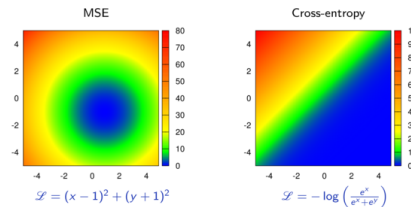


Generalized acyclic graph networks, torch.autograd, batch processing, convolutional layers and pooling, torch.nn.Module.

- [slides](#) / [handout](#) (part a)
- [slides](#) / [handout](#) (part b)
- [practical](#)
- [dlc\\_practical\\_4\\_embryo.py](#)

## Lecture 5 (Mar 21, 2018) – Optimization

Let's consider the loss for a single sample in a two-class problem, with a predictor with two output values. The  $x$  axis here is the activation of the correct output unit, and the  $y$  axis is the activation of the other one.

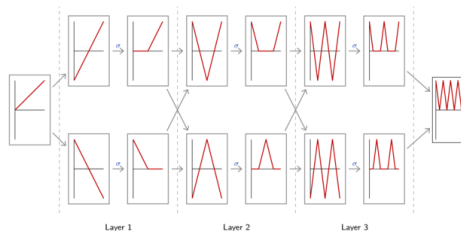


Cross-entropy, L1 and L2 penalty. Weight initialization, Xavier's rule, loss monitoring. `torch.autograd.Function`.

- [slides](#) / [handout](#)
- [practical](#)

## Lecture 6 (Mar 28, 2018) – Going deeper

We can hand-design a network that [almost] reaches the bound:



Theoretical advantages of depth, rectifiers, drop-out, batch normalization, residual networks, advanced weight initialization. GPUs and `torch.cuda`.

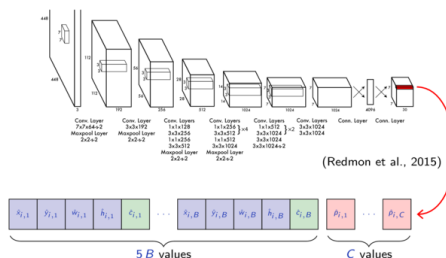
- [slides](#) / [handout](#)
- mini-project

## No lecture (Apr 4, 2018) – Easter holidays

## Lecture 7 (Apr 11, 2018) – Computer vision

The output corresponds to splitting the image into a regular  $S \times S$  grid, with  $S = 7$ , and for each cell, to predict a 30d vector:

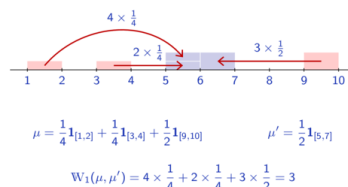
- $B = 2$  bounding boxes coordinates and confidence,
- $C = 20$  class probabilities, corresponding to the classes of Pascal VOC.



- [slides](#) / [handout](#)
- mini-project

## Lecture 10 (May 2, 2018) – Generative Adversarial Networks

An alternative choice is the "earth moving distance", which intuitively is the minimum mass displacement to transform one distribution into the other.



GAN, Wasserstein GAN, Deep Convolutional GAN, Image-to-Image translations, model persistence.

- [slides](#) / [handout](#)
- mini-project

## Lecture 11 (May 9, 2018) – Recurrent networks and NLP

TBD

## Lecture 12 (May 16, 2018) – TBD (guest speaker: Soumith Chintala, Facebook)

TBD

## Lecture 13 (May 23, 2018) – TBD (guest speaker: Andreas Steiner, Google)

TBD

## Lecture 14 (May 30, 2018) – TBD (guest speaker: Andreas Steiner, Google)

TBD

## Virtual machine for the practicals

A Virtual Machine (VM) is a software that simulates a complete computer. The one we provide here includes a Linux operating system and all the tools needed to use PyTorch from a web browser (firefox or chrome).

To use it, first download and install [Oracle's VirtualBox](#) on your machine, then download the image file:

[Deep Learning VM.ova](#) (3.4Gb)