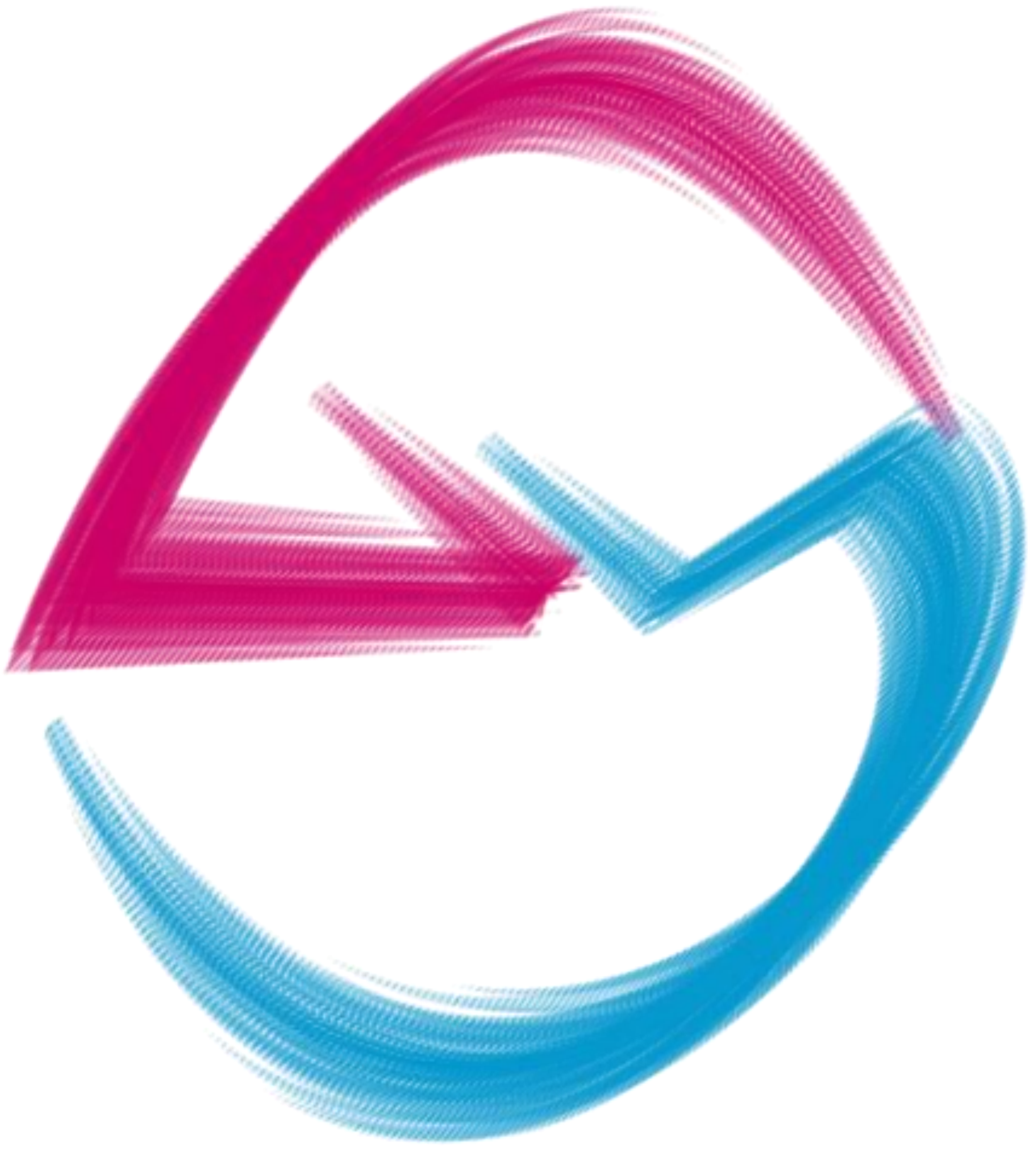
In the name of God

the compassionate the merciful



This curriculum belongs to two courses, so *Deep Learning Engineer* course has two parts. The sections and subsections of each course are contained in the next pages.

Deep learning Engineer

By

Mohammad Ebrahim

Course One

Course 1

Lecture 1, Introduction

Applications

Deep learning is used in different areas today.

What Is Deep Learning?

In general and for supervised approaches, DL is almost a differentiable framework which is able to learn different complicated mappings.

Inputs, outputs, and models of supervised learning

There are different kinds of data.

Boundaries

What are the limitations of deep learning?

A brief history

Before 2012, the number of written papers which were about neural networks was about 150,000.

Lecture 2, Prerequisites

Logic

The necessary and sufficient conditions

Prerequisites

Inner product

Matrices

Cosine similarity

Functions

Norms

Derivative

Gradient

Second derivative

Taylor series

Chain rule

Computation graph

Vector space

Transformation

Rank and range of matrices

Discrete prediction

Continuous prediction

Cost function

Interpretation of the cost with probability

Computation graph for the logistic regression, backward and forward

Formulas for weight updates

Logistic regression and gradient descent

Vectorizing the operations

Limitations of the logistic regression

Lecture 4, Linear Regression

General Concept

Problems with real-valued outputs

Linear Regression Example

A regression example

Properties of lines

What will happen to a line if we change the parameters?

Linear Regression

Linear regression with a single neuron

Higher Dimensions

What will happen if we have many input features?

Cost Function – MSE

Mean squared error is investigated in detail.

Cost Function – MAE

Mean absolute error is investigated in detail.

Solution Finding

How to find the solution?

Drawbacks

When we cannot apply linear regression?

Polynomial Regression

What is the effect of adding polynomial features?

Lecture 5, Shallow Neural Networks

Adding one hidden layer

We improve our model by adding more layers.

Vectorization for Network operations – Forward Pass

We see the calculations of the network for the forward pass. The matrix multiplications are illustrated in detail.

Vectorization across samples – Forward Pass

We see the calculations which are vectorized to perform fast computations across the samples.

Vectorization to Find Loss Function – Backward Pass

The vectorized backprop is explained here using the loss function.

Vectorization to Find Cost Function – Backward Pass

The vectorized backprop is explained here using the cost function.

Convexity

The permutation of nodes is explained here to show the cost function is not convex for multi-layer networks.

Lecture 6, MLP

MLP

We will talk about MLPs and the computations of the forward and backward passes in detail.

Decision Boundaries

We will talk about the different kinds of decision boundaries that can be learned.

Setting the Number of Neurons and Layers

We will talk about setting these hyperparameters.

Multi-Class Classification

Softmax regression is covered in detail.

Multi-label Classification

These are classification tasks which may label each sample with multiple labels from the possible C classes. This can be thought of as predicting properties of a sample that are not mutually exclusive. This is also known as multi-task learning.

MLPs for Regression

MLPs can be utilized for regression tasks as well.

Gradient Checking

This approach is used to check whether our implementation is correct or not.

Lecture 7, Dealing with Overfitting

Model Evaluation

We should be able to investigate whether our model is learning the concept which we want it to learn appropriately or not.

Bias and Variance

What do these concepts mean?

Dataset

How to split our dataset?

Imbalanced Dataset

What does this lead to?

The Distribution of the Dataset

How should be the distributions of data samples?

L2 Regularization

L1 and L2 regularization methods are investigated, and the cost functions that employ them are shown in detail.

Dropout Regularization

The concepts of this method are introduced in detail.

Early stopping

We investigate a kind of model complexity graph.

Data Augmentation

One of the solutions to deal with overfitting was to add more data.

Lecture 8, General Challenges

Normalizing

Numerical stability: related calculations of large and small values should be performed wisely.

Input features which are not normalized will lead to cost functions which are not easy to optimize.

Vanishing Problem

The problem of slow weight updates

Initialization Methods

Having good initial weights can place the neural network close to the optimal solution. This allows the neural network to come to a good solution quicker.

He's method

Glorot's or Xavier's method

Bengio's method

Truncated normal distribution

Symmetry Problem

Bad initialization

Random Restart

Deterministic and nondeterministic algorithms

Gradient Checking

How to check our implementation of the gradient descent is correct?

Lecture 9, Hyperparameters

Hyperparameter

What are they?

Optimization Hyperparameters

Epoch

Iteration

Number of epochs

Mini-batch size

SGD

Batch training

Mini-batch training

Mini-batch gradient descent

Learning rate

The number of parameters and non-convexity

Gradient descent for each direction

Learning rate decay

Some learning rate schedules

Exponential Decay

Manual Decay

Cosine Learning Rate Decay

The challenges of using the learning rate schedules

Model hyperparameters

Number of Hidden Units

Activation functions

Sigmoid and its challenges

Tanh and its challenges

ReLU and its challenges

Sparsity

SmoothReLU

LeakyReLU

PReLU

Softmax

Linear activation

Lecture 10, Optimization Algorithms

Reminder

Minimizer

Second order necessary condition

Second-Order Sufficient Condition (SOSC)

Optimization

Error Surface

How is the error surface of the costs?

Iterative approaches

An example to solve

Newton-Raphson Method

An example which is solved.

Optimization – iterative approaches

Gradient descent is not a very good direction.

Newton method

Wolfe's first condition

Backtracking

Matrix condition number

Ill-conditioned matrices and its effect on our calculations

Momentum and viscosity

Nesterov's approach or NAG

Exponentially weighted average

Course Two

Course 2

Lecture 11, Covariates

Batch Normalization

Speed up

Covariate shift

Solution

Shuffling

i.i.d

Faster convergence

Lecture 12, Hyperparameter Tuning

Hyperparameter Tuning

Is there any standard order to tune the hyperparameters?

Coarse to fine sampling

Different ranges

Good scale

How to look for the hyperparameters?

Lecture 13, Evaluation Criterions

Evaluation

There are different metrics for evaluating the outcomes.

Confusion Matrix

Different situations for different scenarios.

Precision & Recall

Type 1 and 2 errors

Threshold

TPs

FPs

FNs

TNs

Classification and Thresholding

Thresholding

PR-AUC

Area Under the Precision-Recall Curve

Accuracy

Pros and cons

F_1 Score

Single number evaluation

Imbalanced dataset

F_β Score

The importance of precision and recall

ROC Curve

Receiver operating characteristic curve

Further Study

Directions for other methods

Mean Absolute Error

Interpretation

Mean Squared Error

Interpretation

R^2 Score

Interpretation

Lecture 14, CNN

Complexity of the problem

Why image is difficult to learn?

Intuition

How humans classify images?

Structure of Data

Locality

Translation Invariance

Weight sharing

Sparse connections

Problems of MLPs for Images

Many features

Appropriate resolution

Why MLP is bad for images?

From fully connected to locally connected

Why sparse connections are useful?

Number of parameters in conv nets

The challenges of increasing the size of inputs.

Pooling

Calculations

Max pooling

Why it may help?

Reducing the number of parameters

Intuition behind applying max pooling

Max pooling over volume

Average pooling

Global average pooling

Examples

Network

Hierarchical visualization

Hierarchy

Hierarchical features that are extracted

Strided convolutions and pooling are useful for the hierarchy

Translation invariance

Equivariant

Invariant

Feature Extractors

Examples with visualizations

Visualizing CNNs

Investigating the outcomes of a paper for visualizing the CNNs

Training a CNN

Data

Attack

How to fool a CNN?

Hyperparameters of CNNs

CNNs have many hyperparameters. They are investigated in this section.

Transfer learning.

Four scenarios which are possible

The other scenario

Data Augmentation for CNNs

Adding new data by applying different transformations

Why transformed images are helpful?

Warnings

Applying distortions

Applications of CNNs

We briefly introduce the applications in this section.

Classification

Classification with localization

Detection

Neural style transfer

WaveNet

Deep reinforcement learning

Famous Image Datasets

Famous datasets are introduced.

Lecture 15, CNN Architectures

LeNet-5

The model is investigated.

AlexNet

The model is investigated.

VGG

The model is investigated.

NiN

The model is investigated.

ResNet

The model is investigated.

DenseNet

The model is investigated.

GoogLeNet

The model is investigated.

The ImageNet Challenge

The performance over years is investigated.

Image Transformations

We learn different affine transformations that can be applied to images in detail.

Interpolation

Bilinear interpolation is taught in detail.

How to Find the Transformation

Applying different transformation and the calculations

How a transformation should be applied?

A big mistake

Why pooling is not a good idea?

Spatial Transformer Networks

This network is investigated in detail.

Lecture 16, Word Embeddings

Word Representation

Different approaches to represent a word.

Vocabulary

Unique integer

One-hot representation

Tokens

Challenges

Ambiguity

Similar words and different meanings

Different words and similar meanings

Why we demand weight sharing?

Word Embeddings

A key concept in NLP

Making a simple word embedding

Word Embeddings – Similarity

Similarity of words in the new space

Visualizations

Word Embeddings – Analogy

Embeddings can capture relations

Distance measures

Embedding Matrix

The calculations related to an embedding matrix

Word2Vec

Creating embedding matrices with skip gram model

Self-supervised problem

How to make dataset

Different matrices for embeddings

Why the corresponding classification task is difficult?

Different cost functions and their calculations

Negative sampling idea

Negative Sampling

Why not to use the previous method?

Loss function intuition

Loss function calculations in detail

How to pick the words?

Why that distribution?

How is the structure of the mode for this new cost?

Other Methods

There are different challenges for making the embeddings. Other methods are introduced in this section.

Transfer Learning

When and how word embeddings should be used? This is investigated in detail.

Visualizing Word Embeddings

The visualization techniques are introduced.

Self-Attention

This method is investigated in detail.

Debiasing Word Embeddings

The approach is investigated.

Lecture 17, RNN

Sequential Data

Named entity recognition is introduced here as an application.

Examples of Sequence Data

Different applications are introduced to illustrate different types of tasks that we may face when we are dealing with sequence data.

RNN

The architecture and the computations are illustrated, and the main concepts are discussed.

Different Types of RNNs

Different architectures are introduced in detail.

Optimization

Backpropagation through time is explained here.

Language Model

Language modeling in RNNs is explained and visualized in detail.

Sampling with LM

You will learn how to make novel sequences.

Vanishing and Exploding Gradients

The main reason and other side effects are elaborated in this section.

GRU

GRU unit is explained and visualized in detail.

LSTM

LSTM unit is explained and visualized in detail.

BRNN

BRNN unit is explained and visualized in detail.

Deep RNN

Deep RNN unit is explained and visualized in detail.

Seq2Seq

The details and applications of this kind architecture are explained.

RNN Hyperparameters

The results of some papers which are about the hyperparameters of RNNs are explained.

Lecture 18, Autoencoders and GANs

Autoencoders

Different models are explained in detail, and different applications like neural inpainting, image segmentation, and other tasks are explained.

Discriminative and Generative Models

These kinds of models are explained in detail.

GANs

The GAN paper is explained and the theoretical explanations are incorporated. Other studies are also introduced. Different circumstances are elaborated for GANs. Different cost functions are introduced. Other operations like deconvolution and upsampling are elaborated.

Lecture 19, Managing the Challenges Vol. 1

Technical Strategies

In this lecture, we explain what we should do if each part of the network does not work properly. Orthogonalization is explained in detail.

Technical Strategies - Single-Number Evaluation Metric

Single-number evaluation metric can accelerate your ability to select a good classifier.

Technical Strategies - Optimizing and Satisficing Metrics

There are different things that we want to achieve when we train a neural network. These should have a kind of priority.

Technical Strategies – Dataset Distribution

The validation and test sets should come from the same distribution.

Technical Strategies – Size of Data Partitions

The training, validation, and test sets should have appropriate sizes. We should have appropriate strategies to set these.

Technical Strategies – Small Datasets

Deep learning models demand large datasets. What should we do if we do not have large dataset?

Labeled Data

How many labeled data are there for our task? What should we do for different situations?

Technical Strategies – When to change dev/test sets and metrics

Are we noticing exactly to what we demand? Is there something wrong? Are we ignoring something important?

Technical Strategies – All Together

Summary

Technical Strategies – Human-Level Performance

What does that mean, and why is that overly important?

Technical Strategies – Bias Variance

What are our precise solutions for these problems under different circumstances?

Technical Strategies – Improve the Performance

Wrap-up

Technical Strategies – Error Analysis

Analyzing the errors can help us find out the problems of our models, and we can evaluate our assets to see what we can do.

Lecture 20, Managing the Challenges Vol. 2

Technical Strategies – Mismatched Data

What should we do if our training dataset is not like our test set? What are the other circumstances?

End-to-End Deep Learning

The concept is explained, and different circumstances that this technique may be helpful in are explained.

End-to-End Deep Learning - Choosing pipeline components

What are the steps of our task? Should we have a modular system?

End-to-End Deep Learning - Directly learning rich outputs

Nowadays, the outputs are more complicated than before. Is End-to-End learning helpful for complex tasks?

End-to-End Deep Learning - Error analysis by parts

Suppose your system is built using a complex machine learning pipeline, and you would like to improve the system's performance. Which component demands refinement?

End-to-End Deep Learning - Attributing error to one part

Which part is not working properly? How to detect it?

End-to-End Deep Learning - General case of error attribution

The general steps for error attribution.

Error analysis by parts and comparison to human-level performance

Carrying out error analysis on a learning algorithm is like using data science to analyze an ML system's mistakes in order to derive insights about what to do next.

Spotting a flawed ML pipeline

What if each individual component of your ML pipeline is performing at human-level performance or near-human-level performance, but the overall pipeline falls far short of human-level?