Topics:

1. Intro about AI
2. Numbers, Vectors, Matrices, Norms
3. Functions, Continuity, Differentiability
4. Structure model of Single Layer Neural Network

My contacts

- Konstantin Burlachenko
- Phone: +7 925 008 22 53
- Email burlachenkok@gmail.com
- Room: **7092**

What is Artificial Intellegence?

- While media hype is real it is true that both **companies** and **governments** are heavily investing in AI (Artificial Intellegence). Both see AI as an integral part of their competitive strategy.

- There are two ways to look at AI philosophicaly.

- **The first** is what one would normally associate with the AI - the <u>science</u> and <u>engineering</u> of building "intelligent" agents. The inspiration of what constitutes intelligence comes from the types of capabilities that humans possess.

- **The second** views AI as a set of tools. We are simply trying to solve problems in the world, and AI techniques happen to be quite useful for that

What is Artificial Intelligence really if remove all hype?

The philosophical differences do change the way AI researchers approach and talk about their work. As a **big picture** But both views boil down to similar steps:

1. Collecting data
2. Formulate Optimization problem
3. Optimizing a training objective

Let’s setup common vocabulary

AI models is in fact umbrella for all applied math models. We need **math**, but really in all subtle details you will learn math in the University.

Currently we need setup common vocabulary and go through math concepts which we need for purpose of our project

“Fast solvers for feed forward neural net”

Some abbreviations

То есть, другими словами	i.e	От латинского id est
Например	e.g	От латинского “exempli gratia” (“ради примера”).
За и против	pros and cons	От латинского "pro et contra” ("За и против")
И другие	et al	От латинского et alii (И другие)
Смещение	bias	

Mathematical concept - Numbers

Types of numbers:

- **Real number (e.g. 2.3, 3.14, $\sqrt{2}$).**
  The set of such elements is called $\mathbb{R}$

- Natural number (e.g. 0,1,2, 3, 4)
  The set of such elements is called $\mathbb{N}$

- Integer numbers (e.g. -1,0,1)
  The set of such elements is called $\mathbb{Z}$

- Rational numbers (e.g. 3/4, 2/5)
  The set of such elements is called $\mathbb{Q}$

Mathematical concept - Vectors

- Ordered pair is a construction (x,y)
For which two ordered pairs (a,b) and (a′, b′) is equal iff **(if and only if)** a=a′ and b=b′.
Example: $(1,2) = (1,2)\ but\ (1,2) \neq (2,1)$

- Generalization of ordered pair is a tuple (кортеж) or vector (вектор) where $(x_1, x_2, \dots, x_n) = (y_1, y_2, \dots, y_n)$ iff $x_i = y_i$

- The set of tuples $(a_1, \dots, a_n)$ where each a_i if from set A_i is called cartesian product (декартово произведение) of the sets $A_1 x A_2 \dots x A_n$

- If all sets are the same (e.g. like $\mathbb{R}$) it's called $\mathbb{R}^n$

Math concepts - dot product – I/II

The **standard inner product** on $\mathbb{R}^n$ can be written in different ways with the same meaning

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

Math concepts - dot product – II/II

The **standard inner product** on $\mathbb{R}^n$, the set of real n-vectors, is given by:

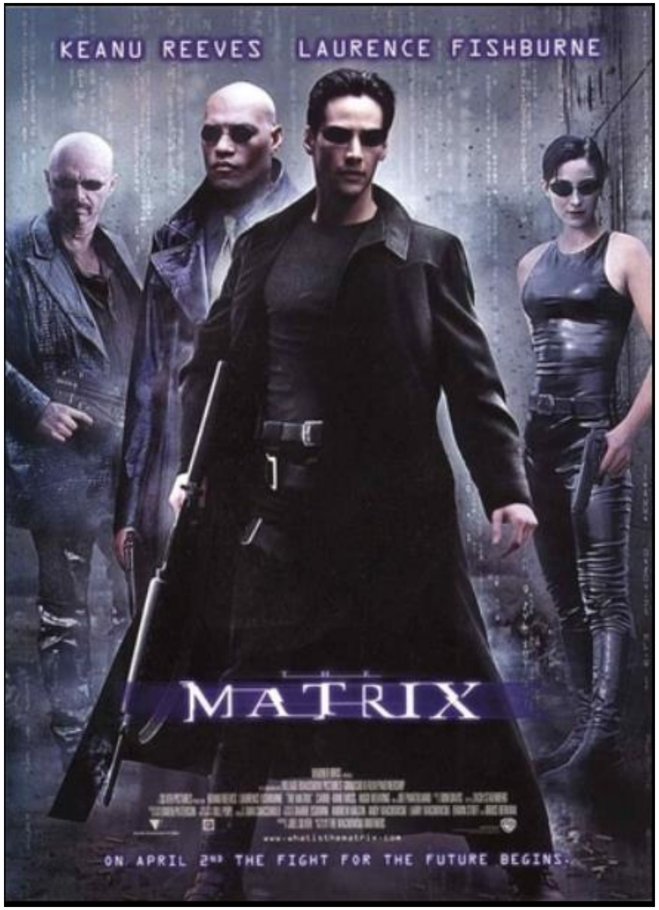
The Euclidean norm, or ℓ2-norm of a vector from $\mathbb{R}^n$ is defined as

$$\|x\|_2 = (x^T x)^{1/2} = (x_1^2 + \cdots + x_n^2)^{1/2}$$

Some meaning of Matrix

https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0_(%D1%84%D0%B8%D0%BB%D1%8C%D0%BC)

We will not go in this way ------->



Math concepts - matrices

By $R^{m\ x\ n}$ we denote a matrix with **m rows** and **n columns**, where the entries of A are real numbers.

By convention the vector from R^{n} is the column vector contains n element in the column.
I.e. this a matrix which **shape** $R^{n\ x\ 1}$

One of operation is transpose – it's means:
- Take rows and write new matrix columns
- Take columns and write new matrix by rows

Math concepts - matrices

Matrix A can be written in different form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \qquad A = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}$$

Math concepts
matrix-vector product

Given a matrix **A** from $R^{m\ x\ n}$ and a vector **x** from R^n their product is a vector y = Ax from R^m.

$$y = Ax = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

$$y = Ax = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1 \end{bmatrix} x_1 + \begin{bmatrix} a_2 \end{bmatrix} x_2 + \ldots + \begin{bmatrix} a_n \end{bmatrix} x_n$$

Couple interpretations of y = Ax

- x is **input** or **action**; y is **output** or **result**
- y = Ax defines a function or transformation
- From algebra point of view for $\boldsymbol{y_m} = \boldsymbol{A}_{[\boldsymbol{m},\boldsymbol{n}]}\boldsymbol{x}_{[\boldsymbol{n},\mathbf{1}]}$ we have $\boldsymbol{y_i} = \sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{a_{ij}} \cdot \boldsymbol{x_j}$

- i-th row of A correspond to i-th output
- j-th column of A correspond to j-th input
- $a_{35} = 0$ means third out $\boldsymbol{y_3}$ does not depend on 5-th in x_5

- If $|a_{32}| \gg |a_{3j}|$ for $j \neq 2$ then it means $\boldsymbol{y_3}$ mainly depends on $\boldsymbol{x_2}$

Math concepts
matrix-matrix product

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}$$

$$C = AB = A \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

Math concepts – outer product

Special form of matrix-matrix multiplication when you multiply $x \in R^{n\,x\,1}$ by matrix $y^T \in R^{1\,x\,m}$ is called **outer product**

$$xy^T \in \mathbb{R}^{m\times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

Math concepts – when outer product usefull

- Matrix which consist on the **same column**.

$$\begin{bmatrix} | & | & & | \\ x & x & \cdots & x \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_1 & x_1 & \cdots & x_1 \\ x_2 & x_2 & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_m & \cdots & x_m \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} = x\mathbf{1}^T$$

- One more view on **matrix multiplication**

$$C = AB = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & b_1^T & — \\ — & b_2^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix} = \sum_{i=1}^{n} a_i b_i^T$$

Properties of matrix multiplication

- Matrix multiplication is associative
  (AB)C = A(BC)

- Matrix multiplication is not distributive
  A(B + C) = AB + AC

- Matrix multiplication is, in general, not commutative, i.e. $AB \neq BA$

Properties of matrix transpose

$(A^T)_{ij} = A_{ji}$

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A + B)^T = A^T + B^T$

Norms for elements from R^n

A function $f: R^n \to R$ with dom(f) = R^n is called a norm if:

1. **f** is nonnegative: **f(x) ≥ 0** for all x ∈ R^n
2. **f** is definite: **f(x)=0** only **if *x* =*0***
3. **f** is homogeneous: f(tx) = |t|f(x) for all x ∈ R^n
4. **f** satisfies the triangle inequality:

   f(x + y) ≤ f(x) + f(y)

Norm can be used to define $distance(x, y) = |x - y|$

Examples of norm

- Sum-absolute-value, or L1-norm

$$\|x\|_1 = |x_1| + \cdots + |x_n|$$

- Chebyshev or L-infinity norm

$$\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$$

- L-p norm, p greater or equal to 1

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

Continuity of function

Continuity of function f : X->Y can be defined in different ways.

We take two points x_1, x_2.
As $|x_2 - x_1|$ decreases it should imply that $|y_2 - y_1|$ decreases.

Mathematically can be setuped as for all $\epsilon > 0$

$$y \in \mathbf{dom}\, f, \quad \|y - x\|_2 \le \delta \implies \|f(y) - f(x)\|_2 \le \epsilon$$

Differentiablity of the function

- $f(x): R^n \to R^m$ is some function
- Let's fix some point x_0 and think about when $f(x) =? f(x_0)$ as variable x is varying
- Typically this is not hold let's slightly reformulate representation
  $f(x) = f(x_0) + R(x)$

Here we have thrown unknown term R(x).
Only what know that once x will have x_0 value then we should have $f(x_0) = f(x_0) + R(x_0)$ => $R(x_0) = 0$

Differentiablity - definiotion

Differentiability means
$f(x) \approx f(x_0) + A(x - x_0)$

In mathematics (and science in general) people try to be precise.

What we can do is measure residual and evaluate
$|f(x) - f(x_0) - A(x - x_0)|_2$

If it is smaller then $|x - x_0|_2$ as **x** come near to $\boldsymbol{x_0}$ then it means that function is differentiable.

Function is differentiable if it is differentiable in all point of it's domain

This matrix **A** sometimes called Df or J is called Jacobian.

What differentiability means

- The property of the function to be differentiable in fact means that very locally it can be approximated very good by affine function
- It’s local property, not global property of the function so from first point of view it’s rather useless.
- In reality some functions by their nature can not be approximated by affine plane. One example $y = |x|$ or $y = \max(0, x)$
- Even if infinitely zoom around point 0 from function domain the form of function will still be not the same as “line”

Chain Rule and Derivatives – next time

- Next time…

- Now let's take a look into definition of Single-Leayer Neural Network

Set of functions

- We defined set of real number, etc. at the beginning
- But we can consider sets of more complicated objects like functions
- If $f_1: X \to Y$ and $f_2: X \to Y$ then we can take this two functions and consider a set $\{f_1, f_2\}$

Function class

- Instead of enumerating functions directly we can do one trick to consider infinite set of functions

- We have a form $F(X,W)=Y$ where X is variable, but W is something which encode selection of function