Exercise 3.1

It is estimated that a human brain consists of approx. 10^{11} nerve cells, each of which has about 10^3 to 10^4 synapses. For this exercise we assume 10^3 synapses per neuron.

Let us further assume that a single synapse could save 4 bits of information. Naively calculated: How much storage capacity does the brain have?

> Note 1
>
> In a human brain the information which neuron is connected to which other neuron is also important.

> Note 2
>
> One bit (b) is a value 0 or 1, so it has two possible values.
>
> One byte (B) is equal to eight bits: 1 B = 8 b.
>
> One kilobyte (kB) is 1024 bytes: 1 kB = 1024 B.
>
> One megabyte (MB) is 1024 kilobytes: 1 MB = 1024 kB.
>
> One gigabyte (GB) is 1024 megabytes: 1 GB = 1014 MB.
>
> One terabyte (TB) is 1024 gigabytes: 1 TB = 1014 GB.
>
> One petabyte (PB) is 1024 terabytes: 1 PB = 1014 TB.
>
> One exabyte (EB) is 1024 petabytes: 1 EB = 1014 PB.
>
> One zettabyte (ZB) is 1024 exabytes: 1 ZB = 1014 EB.
>
> One yottabyte (YB) is 1024 zettabytes: 1 YB = 1014 ZB.

Exercise 3.2

Given a perceptron with three input features (x_1, x_2, x_3) and corresponding weights (w_1, w_2, w_3) and a bias term w_0. Calculate the sum of outputs for the following perceptrons:

- perceptron 1: $(x_1, x_2, x_3) = (0.1, 0.2, 0.3), (w_1, w_2, w_3) = (0.5, -0.1, 0.3), w_0 = -0.2$
- perceptron 2: $(x_1, x_2, x_3) = (-0.3, 0.3, -0.2), (w_1, w_2, w_3) = (0.1, 0.1, -0.2), w_0 = -0.4$
- perceptron 3: $(x_1, x_2, x_3) = (0.1, 0.0, 0.2), (w_1, w_2, w_3) = (-0.1, 0.3, -0.2), w_0 = 0.3$
- perceptron 4: $(x_1, x_2, x_3) = (0.4, 0.3, -0.4), (w_1, w_2, w_3) = (-0.2, 0.3, 0.2), w_0 = 0.1$
- perceptron 5: $(x_1, x_2, x_3) = (0.2, 0.2, -0.1), (w_1, w_2, w_3) = (0.1, -0.1, -0.1), w_0 = -0.05$

Exercise 3.3

ReLU is the most used activation function. It is efficient in the learning process of neural network. But it has also a negative side: a neuron can become "dead", i.e. its weights and biases are not updated in the training phase.

Activation functions Leaky ReLU, Parametric ReLU, ELU (Exponential Linear Units), SELU (Scaled Exponential Linear Units) and GELU (Gaussian Error Linear Units) tries to fix the "dead" neuron problem, although they all have their own issues.

The GELU is an activation function that uses the standard normal distribution (mean is zero and standard deviation is one) to make non-linear transformation. The Cumulative Distribution Function (CDF) of the standard normal random variable $X \sim \mathcal{N}(0, 1)$ is defined as the Φ ("uppercase phi") function

$$\Phi(x) = F_X(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

We don't calculate the values of Φ because it is impossible to make the integration with normal calculus. Instead we typically use a table of pre-calculated values (- found in many books etc.) .

The GELU is defined as follows:

$$GELU(x) = xP(X \leq x) = x\Phi(x)$$

The GELU can be approximated with a following clause:

$$GELU(x) \approx 0.5x(1 + tanh(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3)))$$

Make an implementation of GELU that gives the approximated value described above. The function GELU takes an numpy array of values as input and return the approxmated GELU values for every input value.

What is the mean of returned values when the input is the following numpy array: $[0.2, -0.1, 1.2, 0.7, -0.3, 0.1, -0.2]$? Give the answer rounded to three decimals.

Note: GELU is used in GPT-3, GPT-4, BERT and most other Transformers.

Exercise 3.4

Let there be neurons $N = \{n_1, n_2, n_3, n_4, n_5, n_6, n_7\}$ and inputs $X = \{X_1, X_2, X_3\}$.

Let's define connections as follows:

$V_1 = \{(1, n_1), (1, n_2), (1, n_3), (1, n_4),$
$(X_1, n_1), (X_1, n_2), (X_1, n_3), (X_1, n_4),$
$(X_2, n_1), (X_2, n_2), (X_2, n_3), (X_2, n_4),$
$(X_3, n_1), (X_3, n_2), (X_3, n_3), (X_3, n_4)\}$

$V_2 = \{(1, n_5), (1, n_6), (n_1, n_5), (n_1, n_6), (n_2, n_5),$
$(n_2, n_6), (n_3, n_5), (n_3, n_6), (n_4, n_5), (n_4, n_6)\}$

$V_3 = \{(1, n_7), (n_5, n_7), (n_6, n_7)\}$

How many parameters does the neural network defined above have?