Test on machine learning fundamentals

1 History of artificial intelligence

1.1 The different approaches of AI

1.1.a What is the difference between the symbolic and connectionist approaches?

1.1.b What has Frank Rosenblatt invented?

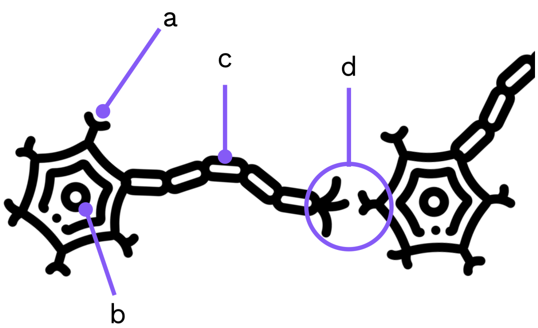
1.1.c True of false : The first neural network ran on a personal computer. Explain your answer.

1.1.d In what year and at what conference was the term ”artificial intelligence” coined?

1.1.e What is deep learning? When is the deep learning revolution considered to have started? Dœs deep learning belong to the connectionist or symbolic approach?

1.2 The artificial neuron

Here is a schema of a biological neuron.



$\boxed{\textbf{1.2.a}}$ Name each part and their function regarding the transmission of the eletric signal.

a : ...

b : ...

c : ...

d : ...

Warren S. McCulloch and Walter H. Pitts Jr formalized the behavior of a neuron for the first time in 1943:

$$\sigma(X) = \begin{cases} 1 \text{ if } \sum_{k=1}^{n} x_k \geq \theta \text{ and i=0} \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

$\boxed{\textbf{1.2.b}}$ Explain the different terms of this equation.

$\sigma(X)$: ...

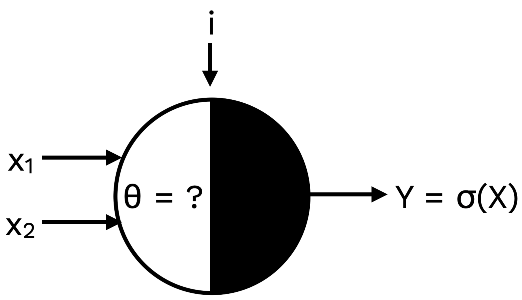
$\sum_{k=1}^{n} x_k$: ...

θ : ...

i : ...

$\boxed{\textbf{1.2.c}}$ In which part of the biological neuron dœs this signal processing take place?

Let's now consider a neuron with two input connexions x_1 and x_2. To simplify, we consider that their values can either be 0 or 1.



$\boxed{\textbf{1.2.d}}$ If $\theta = 1$ and $i = 0$, what are the values of $\sigma(X)$ when:

$x_1 = x_2 = 0$?

$x_1 = 0$, $x_2 = 1$?

$x_1 = 1$, $x_2 = 0$?

$x_1 = x_2 = 1$?

What logical function dœs this remind you of?

1.2.e If $\theta = 2$ and $i = 0$, what are the values of $\sigma(X)$ when:

$x_1 = x_2 = 0$?

$x_1 = 0, x_2 = 1$?

$x_1 = 1, x_2 = 0$?

$x_1 = x_2 = 1$?

What logical function dœs this remind you of?

1.6 Draw a schema that implement the logical function NOT with artificial neurons.

1.7 **BONUS** - Draw a schema that implement the logical function XOR with artificial neurons.

2 Optimization of an artificial neural network

2.1 With an explicit function

We remind you the formula of the gradient descent:

$$\theta^{t+1} = \theta^t - \eta \nabla J(\theta_t) \tag{2}$$

We consider the following loss function: $J(\theta^t) = \frac{1}{4}\left(x^2 - 1\right)^2$

2.1.a What is the gradient of $J(\theta^t)$? Keep in mind that $\frac{d}{dx}[f(g(x))] = f'(g(x)) \cdot g'(x)$

We consider the random initialization of $\theta^{t=0} = 2$ and we choose $\eta = \frac{1}{6}$.

2.1.b Perform two updates of the weight $\theta^{t=0}$. Did the model converge to a local minimum?

2.1.d Perform an update with the initialization $\theta^{t=0} = 0$ and the same learning rate as before. Did the model reach a local minimum ?

2.2 With an unknown loss function

⚠️ It is important to understand that, in practice, we do not know the loss function in advance. We have to estimate it from an estimation of the model's error compared to training data points. In the next questions, we will optimize a linear model by computing a loss on a few data points.

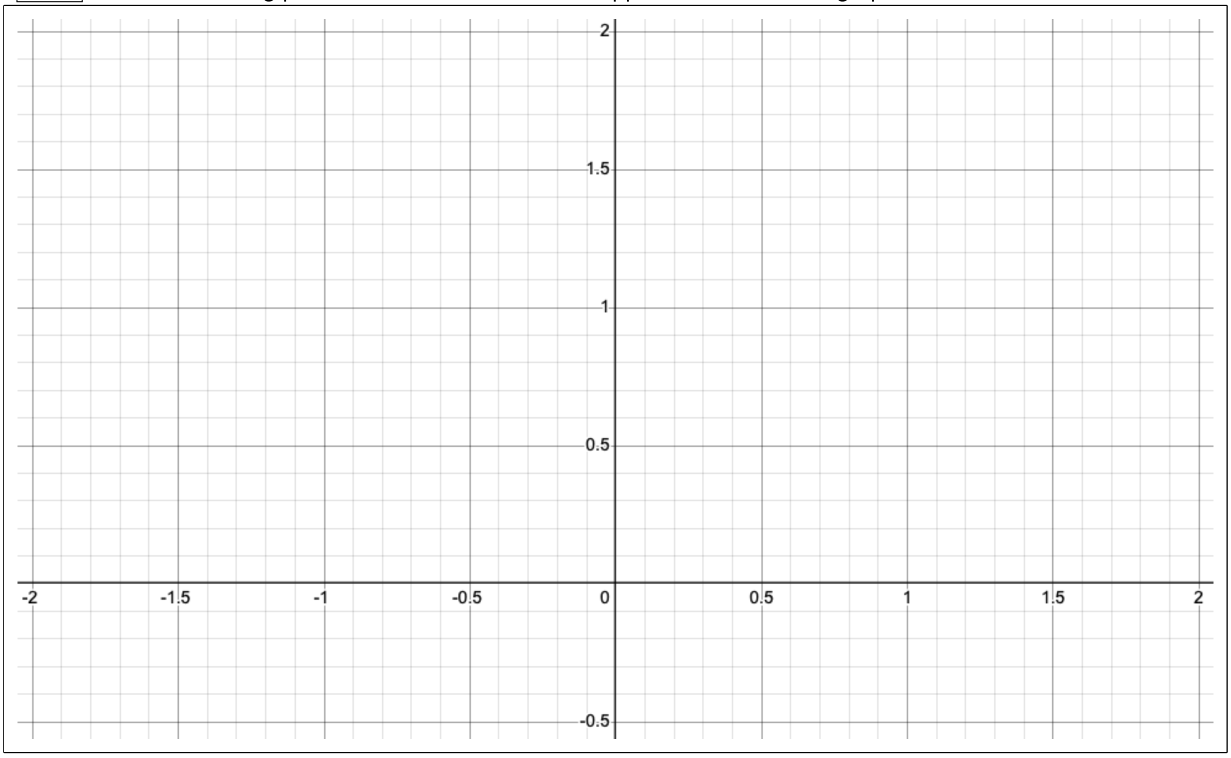
Suppose a linear regression model with parameters θ_0 and θ_1, where the model's prediction is given by $\hat{y} = \theta_0 + \theta_1 x$. At the current optimization step, we obtained $\theta_0^t = 0.5$ and $\theta_1^t = -0.2$.

Our (very small) training data consists of the following pairs (x_i, y_i):

$$(-1, 0.5), (0, 1), (2, 2)$$

For example, these data points could represent temperature variations according to plant growth.

2.2.a Draw the training points and the current model approximation on this graph.



We consider a mean squared error (MSE) as a loss function. We remind you that the MSE is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

where N is the number of training examples.

2.2.b Compute the loss of the model for the given training set.

2.2.c Perform one iteration of the gradient descent with learning rate $\eta = 0.3$. Note that with two parameters, each parameters is updated as follow:

$$\theta_0^{t+1} = \theta_0^t - \eta \frac{\partial J(\theta_0, \theta_1))}{\partial \theta_0} \tag{4}$$

and

$$\theta_1^{t+1} = \theta_1^t - \eta \frac{\partial J(\theta_0, \theta_1))}{\partial \theta_1} \tag{5}$$

2.2.d Draw the new model on the graph in question 2.2. Is the model a better approximation of our training set?

3 The development cycle of machine learning

3.1 Data collection

3.1.a Explain what an image classification task is.

3.1.b Give a few real-world applications of image classification.

3.2 Training

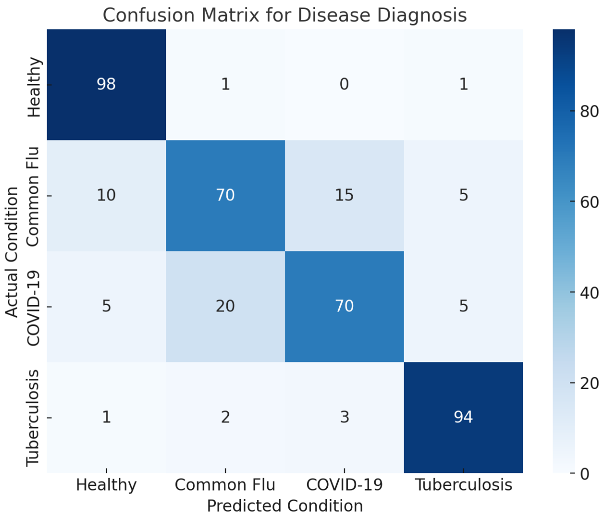
3.2.a Explain what is an epoch. Suppose $N = 100$ images in a training set and $batch\_size = 10$. After one epoch, how many batches were used to train the artificial neural network?

3.2.b True of False? The validation set is usually constructed from the training set. Explain your answer.

3.3 Testing

3.3.a Why should we use a test set?

The figure below represents a confusion matrix of a classification problem of chest X-ray scans in order to classify different lung diseases.



3.3.b What are the two most commonly confounded classes?

3.3.c Is it more dangerous to have high values in the first row or high values in the first column?

The table below depicts training and test accuracies of several machine learning models for the same problem above (4 classes of lung disease).

Name of the model	**Training accuracy**	textbfTest accuracy
Multi-layer perceptron	0.9	0.8
kNN	0.9	0.25
SVM	0.25	0.27
Naives Bayes	0.60	0.50

Table 1. Caption

3.3.d Which models did not learn at all (underfitted)?

3.3.f Which models overfitted?

3.4 Deployment

3.4.a What is the difference between a noisy prediction and a biased prediction?

3.4.b Imagine an example of an unethical bias in a realist deployment of a machine learning model.