Pre-Class Discovery: Deep Learning

Complete before the lecture. Bring your answers.

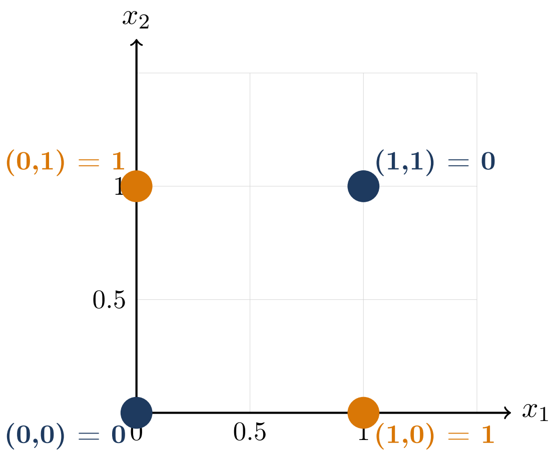
Data Science with Python – BSc Course

Instructions: Work through all six tasks before the lecture. There are no formulas here — just your intuition. Write directly on this sheet or bring a separate page with your answers. Total time: approximately 45 minutes.

Task 1: XOR Problem by Hand

10 min

Four points are plotted on the grid below. The colour tells you their label: **blue = 0**, **orange = 1**.



(a) Try to draw *one straight line* on the grid above that puts both orange points (label 1) on one side and both blue points (label 0) on the other side. Can you do it?

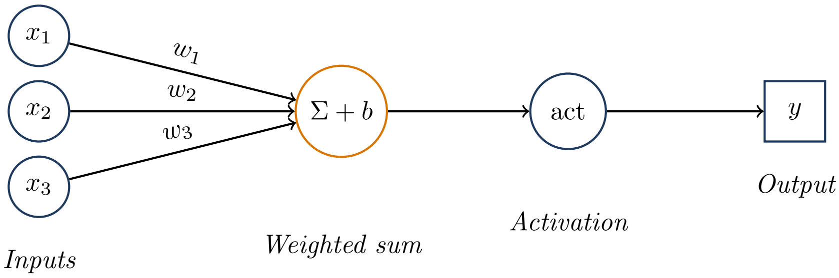
(b) Now describe in words: what shape of boundary *would* separate the orange from the blue points? Could you use two lines? A curve?

(c) A simple perceptron (a single "neuron") can only draw one straight line. What does your answer to (a) tell you about the limits of a single-neuron classifier? What might you need to add?

Task 2: From Neuron to Network

8 min

Below is a diagram of a single artificial neuron. Each input x_i is multiplied by a weight w_i, the products are summed together with a bias b, and the result is passed through an *activation* function to produce an output.



(a) Label each of the four parts above with its name: *input*, *weight*, *sum-and-bias*, *activation*. Add a short one-line description of what each part does.

(b) Now consider a small *two-layer* network: 3 input features, a hidden layer with 4 neurons, and 1 output neuron.

- How many weights are there between the input and hidden layer?
- How many weights are there between the hidden layer and the output?
- How many bias terms are there in total?

(c) The total count of weights and biases is the number of *parameters* the network has to learn. If you made the hidden layer twice as wide (8 neurons instead of 4), roughly how many more parameters would the network have?

Task 3: The Learning Loop

5 min

A child is learning to throw a ball into a bucket.

- **Attempt 1:** Throws too hard. Ball flies past the bucket.
- **Attempt 2:** Throws more gently. Ball falls short.
- **Attempt 3:** Throws a bit harder. Ball lands closer.
- **Attempt 4:** Small adjustment. Ball lands in the bucket.

(a) In this story, what plays the role of the "error"? What information does the child use to know which direction to adjust?

(b) What plays the role of the "adjustment"? Why does the child make *small* adjustments instead of wild ones?

(c) A neural network learns in the same way: it makes a prediction, measures the error, and adjusts its weights. Which of the child's behaviours above would you expect to also see in neural network training? Which would you *not* want (for example, making huge jumps that overshoot)?

Task 4: Sign of Trouble **7 min**

Below are two training logs from two different neural networks. Both were trained for 100 epochs on the same task.

	Training loss	**Validation loss**
Network A	0.020 (stable)	0.050 (stable)
Network B	0.001 (decreasing)	0.100 $\rightarrow$ 0.300 (increasing!)

(a) Which network is *overfitting*? How can you tell?

(b) For Network B, the training loss keeps going down but the validation loss rises. Describe in your own words what the network is probably doing during the later epochs.

(c) If you had to stop one network early to save the best model, when would you stop? Describe a simple rule ("stop when . . . ").

Task 5: Droplets of Noise

8 min

Imagine a classroom of 30 students preparing for a series of exams. Before each exam, the teacher flips a coin for every student and tells a random 30% of the class: "You must stay completely silent during practice today — you can listen but not contribute." The silenced students change from day to day.

(a) After many practice sessions, what happens to the class as a whole? Will the average student be:

- more independent (can solve problems without always relying on the same two loud classmates), or
- less capable (because they practice less)?

(b) Why might random silencing prevent the class from becoming *over-reliant* on just a few dominant students?

(c) Now translate this story to a neural network: what plays the role of the "students"? The "silencing"? The "exam"? This is the intuition behind a very common technique for preventing overfitting — what do you think it is called?

Task 6: Three Questions

5 min

After working through Tasks 1–5, you have seen several ideas: the limits of a single neuron, the building blocks of a network, the learning loop, overfitting signals, and the idea of adding noise to improve generalization.

Write **three questions** you want answered during the lecture. These can be about anything — how gradients are computed, how many layers are enough, why some activations work better than others, how deep learning compares to classical machine learning, or connections to finance and real applications.

Question 1:

Question 2:

Question 3: