

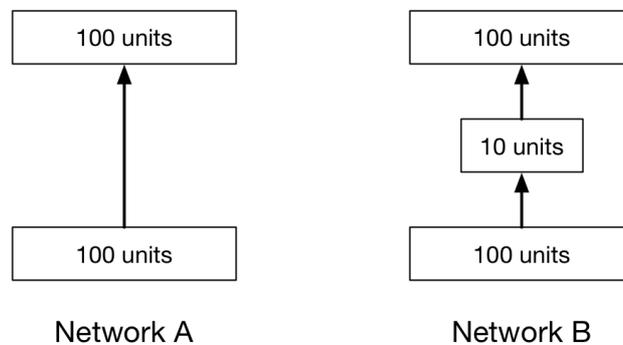
CLASS Exercises: WORD CLASSIFICATION

Exercise 1

We have a multilayer perceptron for classification with a single hidden layer with the hard threshold activation function. The output layer uses the softmax activation function with cross-entropy loss. What will happen when training with gradient descent?

Exercise 2

We have two alternative networks with multilayer perceptrons. Consider that all of the layers use linear activation functions.



- 2a. Mention and argue one advantage of Network A over Network B.
- 2b. Mention and argue one advantage of Network B over Network A.

Exercise 3

Let's imagine that you are training a neural network for a classification task. In this process, you get a much lower training loss than the validation loss. Mention what problem are you facing and how can you solve it.

Exercise 4

The task to perform is to classify movie review text as either positive or negative sentiment, and either action, comedy or romance movie genre. To perform these two related classification tasks, we use a neural network that shares the first layer, but branches into two separate layers to compute the two classifications. The loss is a weighted sum of the two cross-entropy losses.

4a-Given the previous description, rewrite the equations by completing the information marked with ? on the following equations:

$$\begin{aligned}h &= \text{RELU}(W_0 X + b_0), h \in \mathbb{R}^{10}, W_0 \in \mathbb{R}^{? \times ?}, \\ \hat{y}_1 &= \text{softmax}(W_1 ? + b_1), \hat{y}_1 \in \mathbb{R}^?, W_1 \in \mathbb{R}^{? \times 10} \\ \hat{y}_2 &= \text{softmax}(W_2 ? + b_2), \hat{y}_2 \in \mathbb{R}^?, W_2 \in \mathbb{R}^{? \times 10}\end{aligned}$$

$$J = \alpha CE(y_1, ?) + \beta(y_2, ?)$$

4b-When training the model, we see that the model is underfitting? What does it mean? Provide solutions for this.