

[Return to Assessments page](#)

Name: \_\_\_\_\_

Score: \_\_\_\_\_ / \_\_\_\_\_

## Worksheet 09 - RNN challenge

### Part 1: RNN challenge

In this challenge, we to train a classifier for sequences of genetic code.

Each sequence is represented by a string of letters ['A', 'C', 'G', 'T'] and belongs to one of five categories/classes labelled [0,...,4].

For training purposes, you will find 400 labelled sequences, each of length 400 characters (sequences: `data_x`, labels: `data_y`).

To validate your model, you have a further 100 labelled sequences (`val_x`, `val_y`) with 1200 characters each.

Finally, you have 250 unlabelled sequences (`test_x`, 2000 characters) which need to be classified.

**Hint:** Training recurrent networks is very expensive! Do not start working on this challenge to late or you will not manage to finish in time.

#### Attachments

rnn-challenge-data.npz

Your task is to train an RNN-based classifier and make a prediction for the missing labels of the test set (`test_x` in the attached archive). Store your prediction as a one-dimensional **numpy.ndarray**, save this array as **prediction.npy**, and upload this file to the KVV.

You will receive points according to the achieved accuracy according to the following table:

accuracy	points
Loading [MathJax]/extensions/MathZoom.js	10

accuracy	points
$\geq 95\%$	5
$\geq 90\%$	2

Click "Browse" to locate your file and then click "Upload" to upload your file.

File:

*Answer Point Value: 10.0 points*

## Part 2

Please state the names of all the students you worked with on this assignment:

*Answer Point Value: 0.0 points*

*Model Short Answer: -----*