Tests & Quizzes

Worksheet 08 - RNN Challenge

Return to Assessment List

Part 1 of 1 - 10.0 Points

Question 1 of 1

10.0 Points

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 unlabeled sequences (**test_x**, 2000 characters) which need to be classified.

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

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
$\geq 95\%$	10
$\geq 90\%$	7
$\geq 85\%$	5

rnn-challenge-data.npz 3051 KB

prediction.npy (2.13 KB)

- Whiteboard-Startseite
- Klausurübersicht
- Räume
- Site Browser
- Lehr- und Studiumsplanung
- Vorlesungsverzeichnis
- Accessibility Information

- Powered by Sakai
- Copyright 2003-2020 The Apereo Foundation. All rights reserved. Portions of Sakai are copyrighted by other parties as described in the Acknowledgments screen.