Practical Assignment for Knowledge-Based Control Systems (SC42050)

Introduction

This assignment based on MATLAB and Python is a compulsory part of the course Knowledge-Based Control Systems (SC42050). It will be graded and the mark counts for 20% in the final grade of the course (the exam grade is 60% of the final grade, and the literature assignment grade is 20%). The assignment is carried out in groups of two[1] students, and should take around 25 hours per person to solve, depending on your experience with MATLAB, Simulink, and Python. You can start signing up on **Thursday 21 February 2019**. The assignment must be worked out in the form of a short written report (in English, one report per group), to be delivered along with the corresponding MATLAB/Python software by **Thursday 11 April 2019, 12:00 (noon) at the latest** via a BrightSpace assignment. Do not forget to include your names and student numbers on the title page of the report. The expected length of the report is ca. 10 pages (excluding title page and references, but including figures, tables, code snippets, etc.) with a maximum of 15 pages. Note that it is strictly forbidden to take over results from other students or make your results available to others.

Use MATLAB version 6.5 or higher and mention the version you used. Please submit your report in a PDF format and your software as a ZIP file with the file names `GROUPID.pdf/zip`, where GROUPID stands for your group number.

Please post your questions on the BrightSpace discussion forum. For organizational issues contact the course assistant Vasos Arnaoutis (**V.Arnaoutis@student.tudelft.nl**). On Thursday 28 March 2019 from 15:45 to 17:30, there will be a question hours session for the Practical Assignment in IO-PC hall 1 (ENTER). You can bring your own computer here with you, or use the computers in the computer room.

The assignment consists of three problems. The first problem concerns data-driven black-box modeling using a feedforward neural network. The second problem is based on reinforcement learning. The third problem is about model-based control.

You can get a total of 100 points (corresponding to a grade of 10) for this assignment. 10 points are for the quality of the report, the remaining 90 as indicated below.

Matlab programming

Strive for a compact and elegant MATLAB code, use functions where suitable, avoid loops (`for`, `while`, etc.) and also `if-then` constructs at places where you can easily use vector and matrix operations. Search for "vectorization" in Matlab help for helpful tips on the proper MATLAB programming style.

If you are unfamiliar with programming in Matlab, here are some pointers that should help you to quickly learn the basics. To access the Matlab documentation, type `doc` at the command line. A good place to start is the "Getting Started" node of the Matlab documentation. Focus especially on "Matrices and Arrays" and "Programming". A minimal knowledge of "Graphics" is required in order to present your results in a graphical form. For a more in-depth introduction, see "Mathematics" > "Matrices and Linear Algebra", and under this node: "Matrices in Matlab" and "Solving Linear Systems of Equations".[2]

[1] If you absolutely cannot find a partner, you may work alone. Note that groups of three or more students are not allowed.

[2] These pointers assume the documentation structure in Matlab 7.3. While the the structure may vary in other versions, you should still be able to easily find these topics.

Problem 1. Bicycle rental prediction in TensorFlow (20 Points)

In this assignment you will be asked to predict the number of bicycles that will be rented out, based on weather and seasonal data. This will be done by training a neural network using `TensorFlow`, a computational library. A Python script is provided, to which you will make changes to improve the predictions. You have several options of running the code. Either you run it in an online environment or you install it on your personal computer.

Running TensorFlow in an Online Environment

`Google Colaboratory` has all the required components pre-installed, you will need a Google account. Create a new "Python 3 notebook" and copy and paste the provided Python script. The only additional thing you need to do, is to upload the "hour.csv" file, see Figure 1.

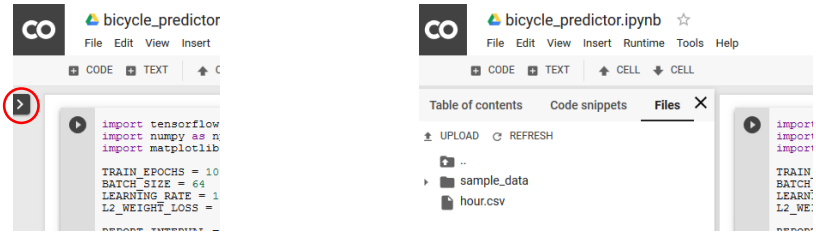

Figure 1: Click on the arrow circled in red in the left image and upload "hour.csv" in the "Files" tab.

Installing TensorFlow on Your Computer

The assignment requires `Python 3`, `TensorFlow`, and `matplotlib` to be installed. See the installation instructions on the respective websites, we provide some tips below. If you encounter any problems, Google is your friend, ask fellow students for help (this is encouraged for this step, and this step only), or post a question on the Brightspace discussion forum.

Intallation tips

Windows If you are using Windows, you might first have to install Python 3 first. Tensorflow requires the 64-bit version of Python (`Windows x86-64`). Unfortunately, the standard download page links to the 32-bit version. So download it from the Windows specific page. Using the option "Add Python 3.X to PATH" is recommended to be able to call `pip3` and `python` from any directory. Afterwards, open the Windows `Command Prompt` and use `pip3 install --upgrade tensorflow` to install the CPU version of TensorFlow and `pip3 install --upgrade matplotlib` to install matplotlib.

Mac
1. Download Xcode from the app store.
2. Install Homebrew by executing the following command in the terminal: `ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"`
3. Install Python 3: `brew install python3`
4. Use the steps here to install the *Python 3 version* of TensorFlow **with the following exception:** For **step *3***, use the command `python3 -m venv directoryOfYourChoice` instead. This will prevent issues with matplotlib.
5. install matplotlib with `pip3 install --upgrade matplotlib`

Linux Python 3 should already be installed. Use the instructions here to install the Python 3 CPU version of TensorFlow. Then use `pip3 install --upgrade matplotlib` to install matplotlib.

If you installed TensorFlow in a virtual environment, make sure it is still activated
`source YourPreviouslyChosenDirectory\bin\activate`
when running the exercise files.

Testing the script

The assignment comes with a Python script: `bicycle_predictor.py` and a dataset: `hour.csv`. There is also a `Readme.txt` file which describes the dataset.

Test if your TensorFlow installation was successful by running the script. This can be done by extracting the exercise files `bicycle_predictor.py` and `hour.csv` to a directory and using the following terminal command[3] form that directory:

`python bicycle_predictor.py`

or by clicking the run button in your online environment.

You should get a list of training and validation errors per episode in the terminal as well as a plot resembling Figure 2. If you installed TensorFlow you need to close the plot to return to the terminal, on Colaboratory you might need to restart the runtime to run the complete script again after you made some changes.

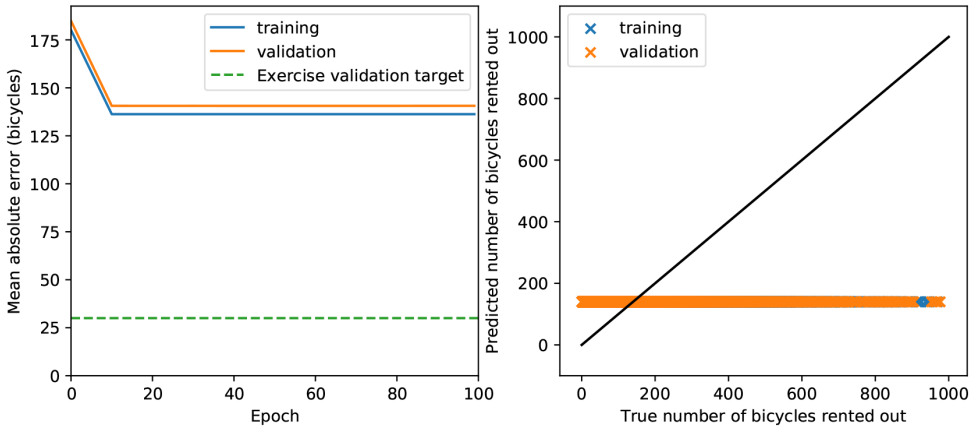

Figure 2: Expected output of `bicycle_predictor.py`

Improving the model

When the script is working, it is time to improve it. Open the `bicycle_predictor.py` file in your favorite editor and read through it to get a basic understanding of how it works.

Task 1.1 "Look ma, without any features!"

- On line `75` and above of the python script you can see that the loss is defined as the mean absolute error between the network predictions and the true number of bicycles (The regularization term is set to zero for now on line `8`). On lines `87 - 97` you can additionally see that the prediction of the network currently does not depend on any input, as the inputs are multiplied by zero. As a result, when the network is trained it learns to output a single value (the bias). To what property of the dataset should the output of the network converge according to theory?
- Change line `75` so that the mean *squared* error is used as the training criterion instead (and keep it like that for the rest of the assignment). To what property of the dataset should the output of the

[3]Depending on your installation, the command `python` might default to Python 2, then use `py -3 bicycle_predictor.py` (Windows) or `python3 bicycle_predictor.py` (Linux & Mac)

network converge now?

Task 1.2. Linear regression

Change the `create_neural_network` function starting on line `84` such that the prediction of the number of bicycles is an affine function of the input features (by removing the multiplication with zero). Note down the validation performance (mean absolute error in bicycles) that you get. How would that performance change when adding additional linear layers or changing the number of units in the currently used layer? Explain your answer.

Task 1.3. Nonlinear regression

Change the function further to use a ReLU (Rectified Linear Unit) activation function for the hidden layer. Run the script and use the image it generates to discuss whether you think the model might be under-fitting or over-fitting. What does your conclusion mean for the number of training iterations and the model complexity? Should you increase or decrease them to get better validation performance?

Task 1.4 Going deep

Change the `create_neural_network` function further so that you get a neural network with two hidden ReLU layers. How does this change the performance?

You could also make the activation function in the final layer ReLU instead of linear. This would prevent the model from predicting that a negative number of bicycles is rented out. However, under some conditions this might prevent the model from learning anything. Why is that? And what would be an alternative to the ReLU activation that prevents this phenomenon?

Task 1.5. Hyper-parameters

At the start of the Python file several hyper-parameters are defined. Change these hyper-parameters and the neural network architecture such that the mean absolute validation error is around 30 bikes. Report and *motivate* the changes you make.

Optional Bonus question

Even though the neural network might predict the number of rented out bicycles fairly accurately on most days, it still makes large errors on other days. Use the `largest_data_point_errors` function of the `BicycleRentalPredictor` class to see for which dates your network makes the largest mistakes. Can you use the knowledge that the rental data is from Washington DC to explain one of these mistakes?

Problem 2. Reinforcement Learning (35 Points)

Most of the theory needed to answer the questions in this assignment can be found in the book "Reinforcement Learning: an Introduction" by Sutton and Barto (S&B), Chapters 1, 2, 3, 4 and 6. An online version of this book can be found here: `http://incompleteideas.net/book/the-book.html`[4]. You can also look at the lecture slides.

The goal of this exercise is to have a robot soccer player swing up a ball with its arm, even though the torque it can apply to its shoulder joint is not enough to do this in one go. This is called the "underactuated pendulum swing-up" problem. By answering the theoretical questions and implementing their solutions you will construct a temporal difference reinforcement learning solution to this problem using the tabular SARSA(0) algorithm.

The Matlab code for this exercise contains five main files:

`assignment.m`	Main function. Run it to learn and test your controller.
`assignment_verify.m`	Verification harness. Run it after implementing each question to verify your code.
`swingup.m`	Implements the SARSA learning loop described in S&B (Section 6.4, Figure 6.9), and a testing loop. The file is partly incomplete and your task is to complete it (search for TODO).
`swingup_initial_state.m`	Sets the initial state of the arm as a slightly perturbed bottom position.
`body_straight.m`	Simulates the dynamics of the body.

In this problem, you will go through questions and implementation tasks which eventually will lead the robot to perform an arm swing-up.

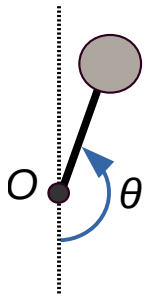

Figure 3: The angle θ is equal to 0 when arm is in the bottom position and is equal to π in the upright position.

Task 2.1. Understanding the code

Read `assignment.m` and `swingup.m`. Note how the `swingup` function can be used it three settings: learning, testing and verification. Compare the structure of the learning part to Figure 6.9 from the textbook.

a) How many simulation steps are executed in a trial?

Now run `assignment_verify.m`. This will report any basic errors in your code.

b) What does it report?

c) Find the source of the error. Why is this value not correct? Think about what it means in terms of the learning algorithm.

[4]The second edition does not allow for easy links to specific chapters, so the links in the remainder of the exercise are to the first edition.

Task 2.2. Setting the learning parameters

Look at the `get_parameters` function in `swingup.m` and set the random action rate to 0.1, and the learning rate to 0.25.

a) Learning is faster with higher learning rates. Why would we want to keep it low anyway?

b) Set the position discretization such that there is exactly one state for every $\pi/15$ rad.

c) Assuming that the velocity stays in the interval $[-5\pi, 5\pi]$ rad s^{-1}, set the velocity discretization such that there is exactly one state for every $\pi/3$ rad s^{-1}.

Set the action discretization to 5 actions. Set the amount of trials to 2000.

Run `assignment_verify` to make sure that you didn't make any obvious mistakes.

Task 2.3. Initialization

The initial values in your Q table can be very important for the exploration behavior, and there are therefore many ways of initializing them (see S&B, Section 2.7). This is done in the `init_Q` function.

a) Pick a method and give a short argumentation for your choice.

b) Implement your choice. The Q table should be of size $N \times M \times O$, where N is the number of position states, M is the number of velocity states, and O is the number of actions.

Run `assignment_verify` to find obvious mistakes.

Task 2.4. Discretization

In Task 3.2, you determined the amount of position and velocity states that your Q table can hold, and the amount of actions the agent can choose from. The state discretization is done in the `discretize_state` function.

a) Implement the position discretization. The input may be outside the interval $[0, 2\pi]$, so be sure to wrap the state around (hint: use the `mod` function). The resulting state must be in the range $[1, \texttt{par.pos\_states}]$. This means that π (the "up" direction) will be in the middle of the range. See the pendulum model shown in Figure 3.

b) Implement the velocity discretization. Even though we assume that the values will not exceed the range $[-5\pi, 5\pi]$, they must be clipped to that range to avoid errors. The resulting state must be in the range $[1, \texttt{par.vel\_states}]$. This means that zero velocity will be in the middle of the range.

c) What would happen if we clip the velocity range too soon, say at $[-2\pi, 2\pi]$?

Now you need to specify how the actions are turned into torque values, in the `take_action` function.

d) The allowable torque is in the range $[-\texttt{par.maxtorque}, \texttt{par.maxtorque}]$. Distribute the actions uniformly over this range. This means that zero torque will be in the middle of the range.

Run `assignment_verify`, and look at the plots of continuous vs. discretized position. Are they what you would expect?

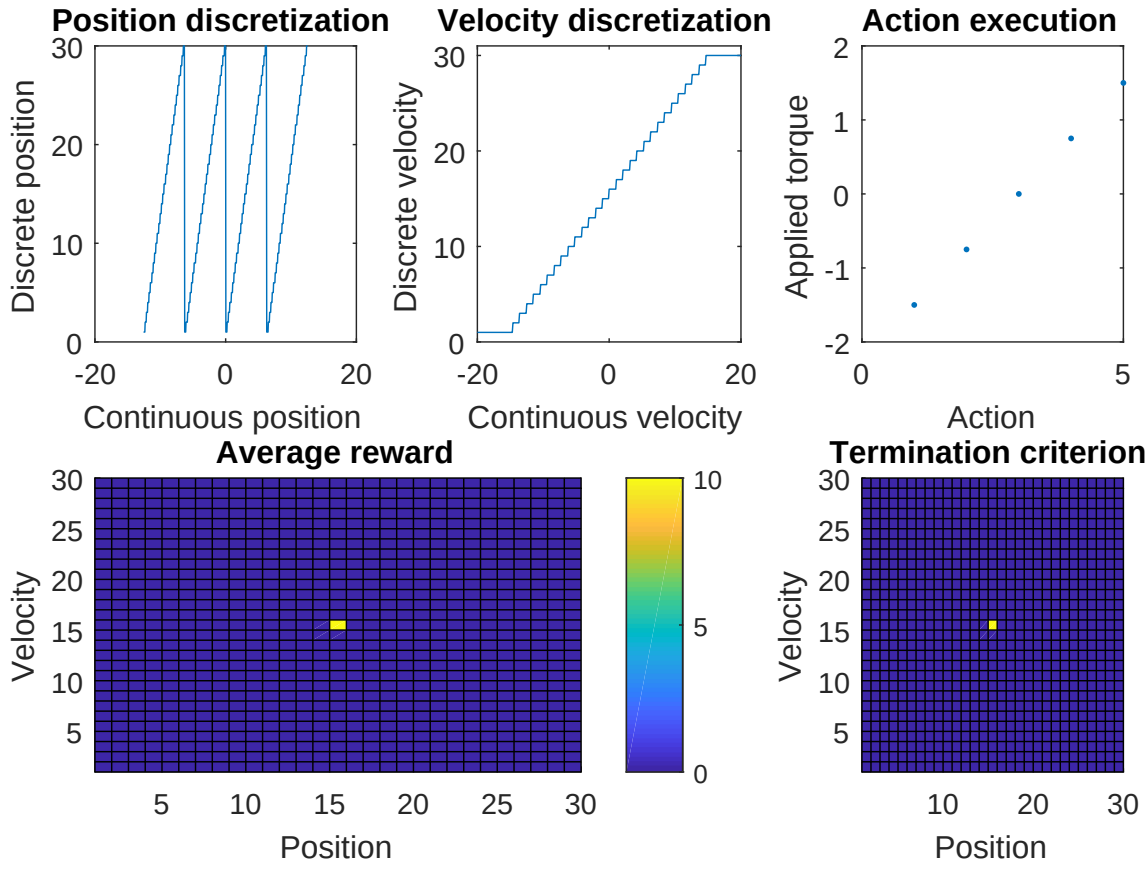

Figure 4: Output of `assignment_verify` after completing Tasks 3.1-3.6.

Task 2.5. Reward and termination

Now you should determine the reward function, which is implemented in `observe_reward`.

a) What is the simplest reward function that you can devise, given that we want the system to balance the pendulum at the top?

b) Implement `observe_reward`.

Run `assignment_verify`, and verify in the lower left plot that you have indeed implemented the reward function you wanted.

You also need to specify when a trial is finished. While we could learn to continually balance the pendulum, in this exercise we will only learn to swing up into a balanced state. The trial can therefore be ended when that goal state is reached.

c) Implement `is_terminal`.

Run `assignment_verify`, and verify that your termination criterion is correct.

Task 2.6. The policy and learning update

It is time to implement the action selection algorithm in `execute_policy`. See S&B, Sections 2.2 and 6.4.

a) Implement the greedy action selection algorithm.

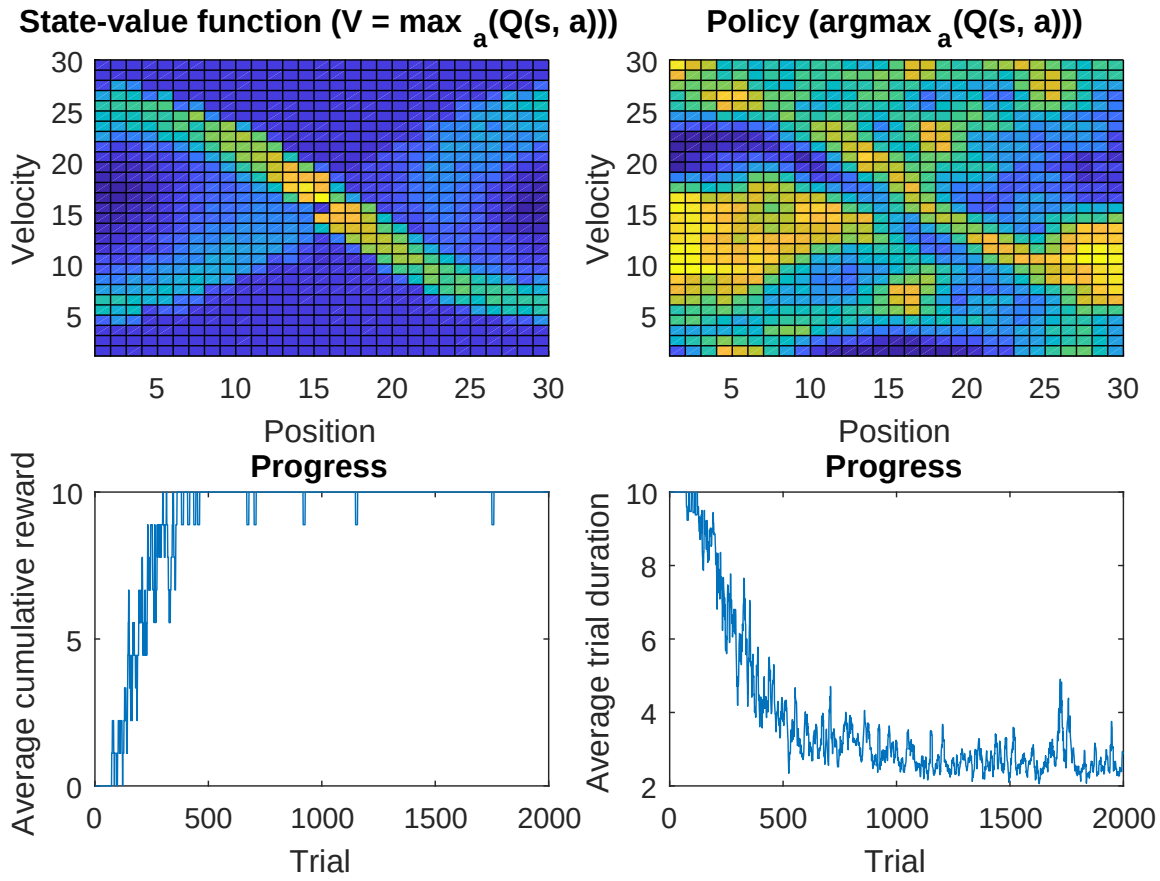

Figure 5: Output of a successful run of `assignment.m`

b) Modify the chosen action according to the ε-greedy policy. Hint: use the `rand` and `randi` functions.

c) Finally, implement the SARSA update rule in `update_Q`.

Run `assignment_verify` a final time to check for errors. The result should be similar to Figure 4.

Task 2.7. Make it work

Finally, use Figure 6.9 from S&B and complete all the code of the learning section in `swingup` (initializations of outer and inner loops, calculation of torque, learning and termination). Basically you need to call all functions prepared in Tasks 3.3-3.6 in a right order. Also make sure that initial state is always slightly perturbed, i.e., that `swingup_initial_state.m` is used for initialization.

It is time to see how your learning algorithm behaves! Run `assignment.m` and check the progress. A successful run looks somewhat like Figure 5.

a) How many simulations steps on average does a swing-up take (after learning has finished)? Will it be wise to reduce the number of steps per trial during learning?

b) Large parts of the policy in the upper-right graph are quite noisy. What reasons can you name?

c) Test your code with greedy and ε-greedy policies. Which method allows the algorithm to converge faster and which method results in a higher cumulative reward (on average)? Explain the reason.

d) Try several values of discount rate, spanned across the $[0, 1]$ interval. What discount rate allows the algorithm to converge faster? Explain the reason.

Problem 3. Model-base Control (35 Points + Bonus)

Your task is to design a model-based controller for a simulated 2 link robot arm that is tracking an ellipse.

Task 3.1. Warming Up

1. Run `controller_0.m`, `controller_1.m`, `controller_2.m`.
2. `controller_0` is a simple tracking PD controller on the joint level, try 3 other gain settings to improve the performance.
3. `controller_1` and `controller_2` use a model-based control approach (with the perfect analytical model). Note that the PD gains are a lot lower. There are subtle differences in how the model is used. Which control structures discussed in the lectures do they correspond to? Switch off the feedback (PD) in both controllers. What happens? Set the initial position to the desired initial position for both controllers. What happens? Do the effects of switching off the feedback and setting the initial position correspond to the properties of the controllers discussed in the lecture?

Task 3.2. Design your own Controller

The goal is to replace the analytical model in the feedforward part by a data-driven model (GP, neural network, fuzzy system, basis functions, etc.) or a qualitative one (naïve physics, knowledge-based, etc.). That is, you cannot make use of the physical equations and values of the analytical model. With feedback gains of `Kp = [500; 500]; Kd = [50; 50];` your model needs to get a lower RMSE than the pure PD controller as defined in `controller_0`; and all that for a range of the rotational velocity `tp.w` between 70 and 80, also see `controller_yours.m` and `controller_yours_evaluate.m`

For this evaluation only the feedforward model `ff_yours.m` can be modified (its input parameters are the current joint position and velocity, as well as the desired joint position, velocity and acceleration, *not* the current joint acceleration), the rest of the code (besides loading the model, variables, etc. and passing them to `ff_yours`) should remain functionally unchanged. For collecting data, training the model, etc. you can modify more things.

You can use any toolboxes you like, however, `controller_yours_evaluate.m` needs to be directly run-able on a standard TUD installation (`https://weblogin.tudelft.nl`) after unzipping.

Task 3.3. Bonus Points

You can get full points for the assignment without this task. With this task you can get bonus points to make up for points you missed, the maximum grade is still a 10. You can get up to 10 points for this task plus an additional bonus if your group is among the top 10 `RMSE x` scores (lowest RMSE of all groups gets 10 points, second lowest 9 points, etc. until tenth place 1 point).

The task is to make your controller robust to additional variations in the initial joint positions (we will evaluate a secret test set with deviations of up to ± 30 deg per joint compared to the desired initial position).