Homework #3

HTTP Request Arrival Estimation (4 points)

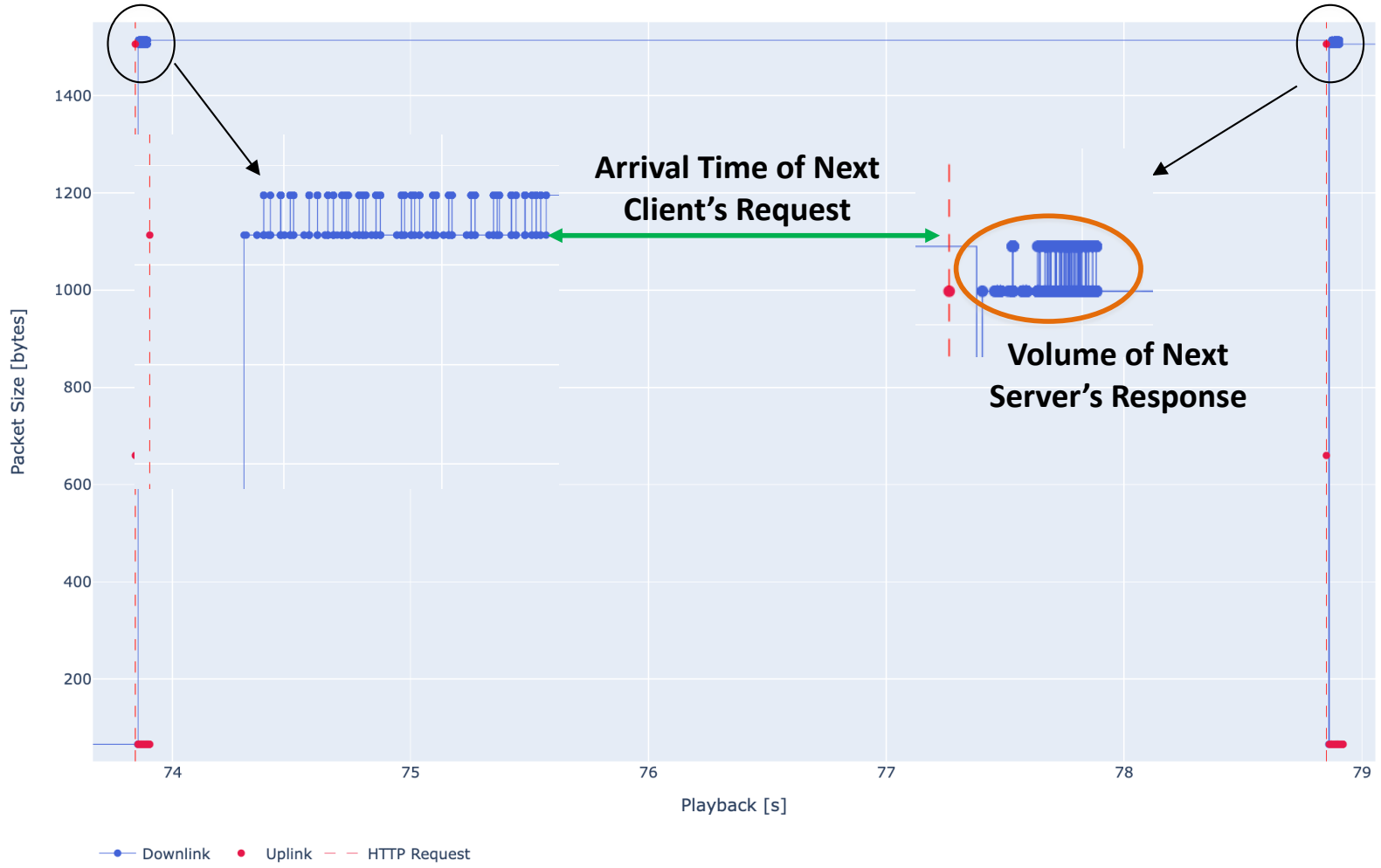
Deadline: 14/04/2022

Assignments

- You are requested to perform and evaluate the performance of the following **regression** tasks:
  - **Predict when the next UL Request is sent by the Video Client**
  - **Predict how large is the response of the Server to the next UL Request**

- For the tasks at hand, you are provided with **20 traffic captures** (.csv files) of ~180 [s] long Youtube Video Sessions

Assignments



Regression vs Classification

- Classification is the task of predicting a discrete class label
- **Regression is the task of predicting a continuous quantity** (e.g., arrival time, burst volume, etc.)
- For this homework, you can use the regression version of the Random Forest Classifier:
  - `from sklearn.ensemble import RandomForestRegressor`
- The training process works as well as what seen during class
- **For the test phase, use Root Mean Square Error as performance metric**

Assignments (1 points)

1. For each video session, extract features from the corresponding traffic trace:
   a) Parse DNS responses and find Server(s) IP
   b) Keep only the dominant traffic flow
   c) Extract features:
      a) Size of last observed HTTP Request
      b) Inter Request-Response Time
      c) Server Burst Download Time
      d) Server Burst Volume
      e) # of Packets in Server Burst
      f) Playback Time
2. Join sessions' data in a single dataset (vertical concatenation of single session dataset)

What to expect as dataset output

	Request_Size	Inter_RR_Time	DL_Time	DL_Vol	DL_Size	PB_Time
0	660	0.009949	0.015693	512272	340	31.785389
1	660	0.009757	0.039912	1016740	674	33.786951
2	660	0.011463	0.279520	509764	339	36.788042
3	660	0.010261	0.122553	1027372	682	38.789221
4	660	0.003953	0.392465	1029212	682	41.793379

Assignments (3 points)

3. For each video session, extract groundtruth from the corresponding traffic trace to:
   a) (**2 Points**) Estimate the arrival time of next HTTP Request:
      - **HINT**: The arrival time of the next HTTP Request is the time between the last DL packet and the next “large” UL packet…
   b) (**1 Point**) Predict the size of next burst from the server
      - **HINT**: this groundtruth comes from free from "Server Burst Volume" feature…

What to expect as groundtruth output

	Next_Request_Time	Next_Response_Vol
0	1.975919	1016740.0
1	2.951421	509764.0
2	1.710192	1027372.0
3	2.871343	1029212.0
4	1.608527	510772.0

My prediction performance:

a) Estimate the arrival time of next HTTP Request:
   - **RMSE = 4.85 [s]**

b) Predict the size of next burst from the server
   - **RMSE = 389.73 [KB]**

Hands On!

You can use

Homework_Skeleton.ipyn

for this homework!