UNIVERSITE DE GENEVE
Faculté des sciences

Section de Mathématiques

Statistical Machine Learning I - WS 2021
Project

You can choose one of the following two projects. The final product should be a report detailing what you did as well as the code. You are asked to train a model (in the first case it's a regression in the second a classification task) that performs well on the given task and explain what you did.

In both tasks you should make clear what role regularization plays in your solution. I.e. showcase overfitting in the regime of not enough regularization and how the choice of hyperparameters influences the end-result.

Project 1. On moodle you will find a regression dataset of artificially generated data $[y, X]$. Your goal is to find the underlying function and isolate the significant features. To do this, fit a LASSO first to extract the linear terms (if any) followed by an artificial neural network or random forest on the residuals to see whether you can improve/find a nonlinear component.

Project 2. You will train a classifier on the Fashion MNIST dataset. It is publically available for you to download. In case you use tensorflow, you can just import it using

```
from tensorflow.keras.datasets import fashion_mnist
(X_train, Y_train), (X_test, Y_test) = fashion_mnist.load_data()
```

You can choose your approach of training. Suggested methods are either to use Boosting with decision trees or Neural Networks or even a combination of both. In either case you are allowed to use whatever framework you want a long as you can explain what they do. Please put a focus on visualizing the role of regularization in the final presentation.