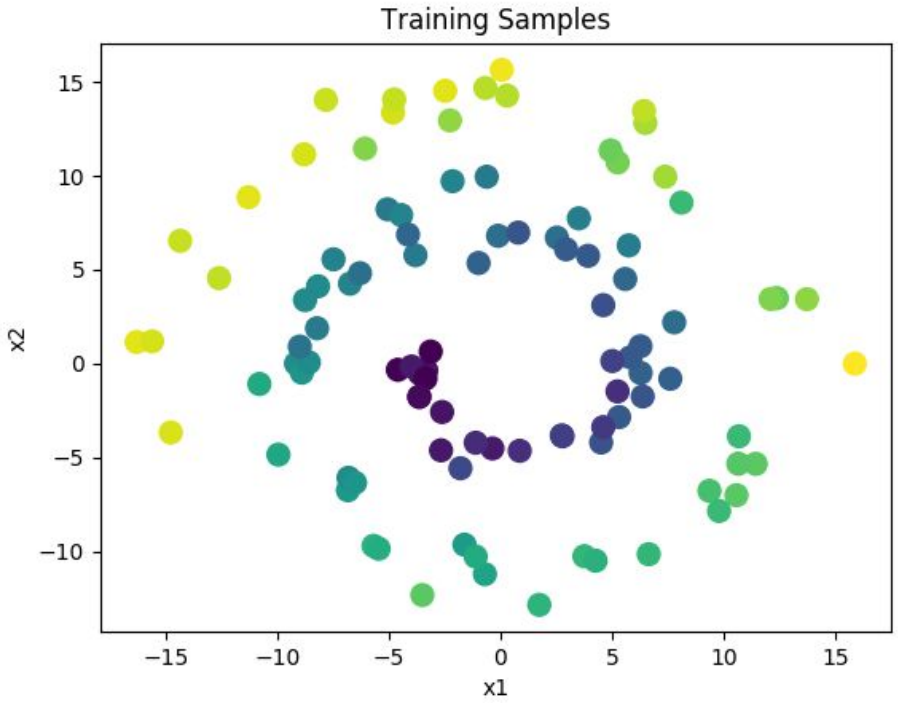
Assignment 3

Machine Learning
708.064 19S

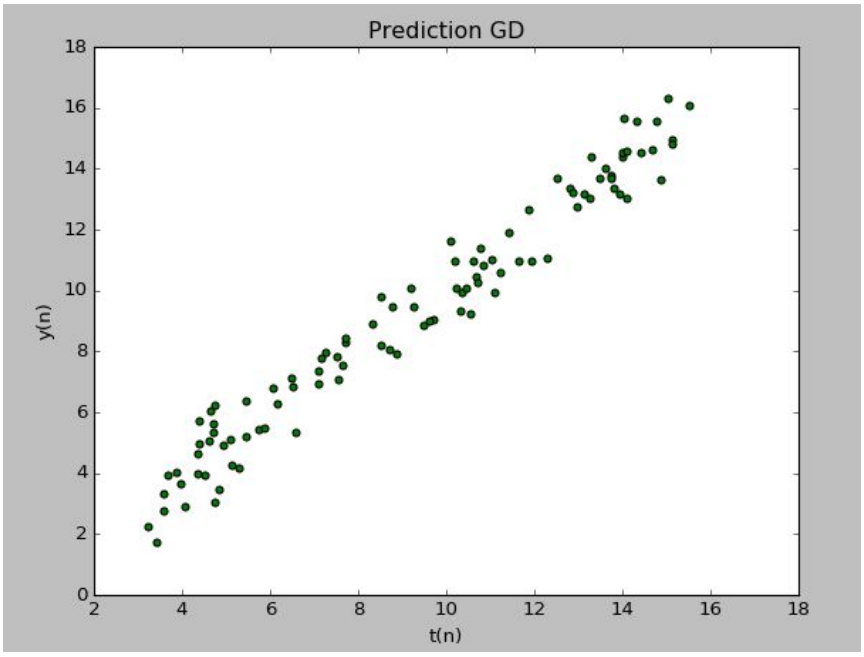
Last Name	First Name	Matr. Number
Konanur Ramanna	Keerthi Datta	01647641
Gajanovic	Stefan	01431869

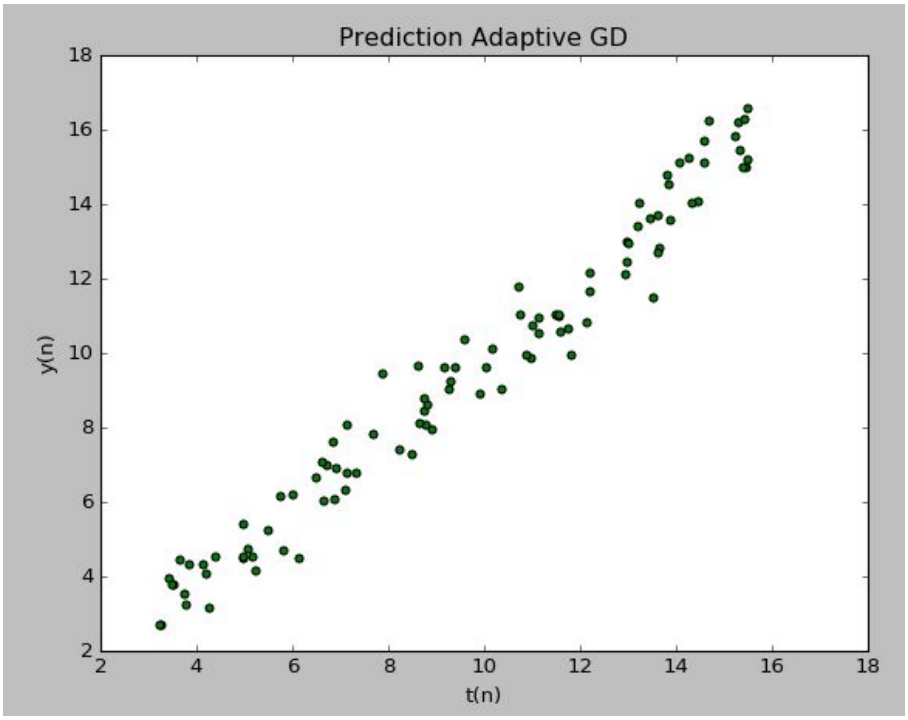
1 Support Vector Regression: the primal problem (10 points)

1. Sampled data from Swiss-roll distribution:

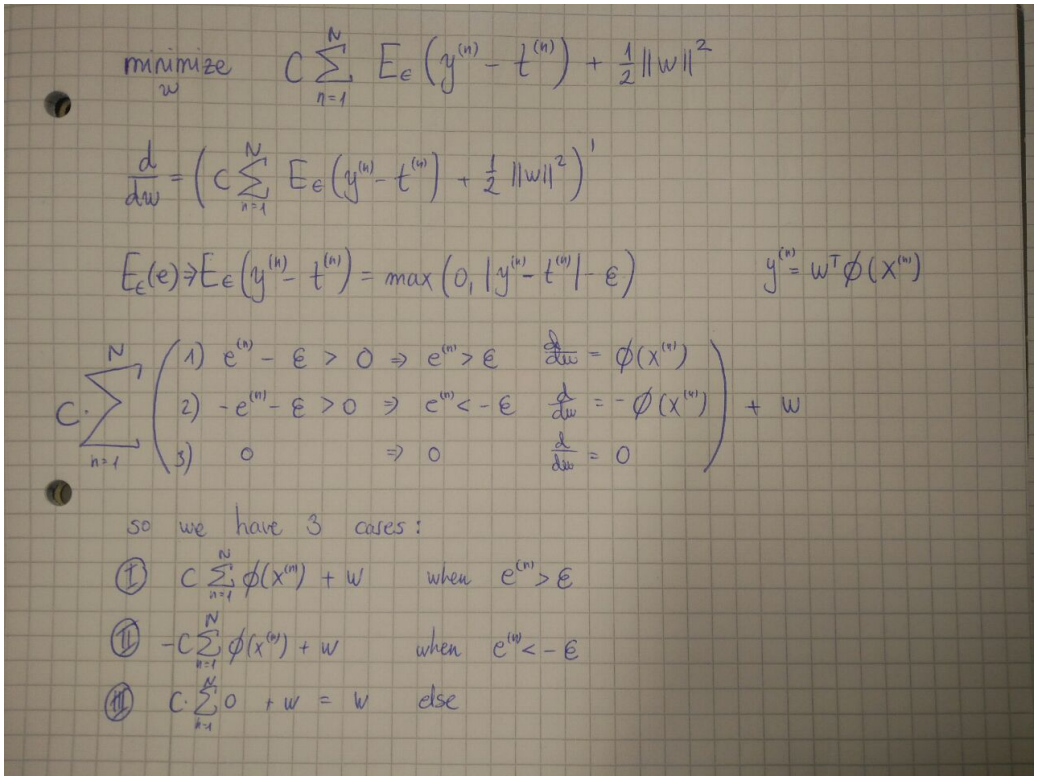


Scatter plot on freshly generated samples, that were not used during training: Predictions were made based on learnt parameters from Gradient descent optimization and adaptive Gradient descent algorithm. Plots are below:

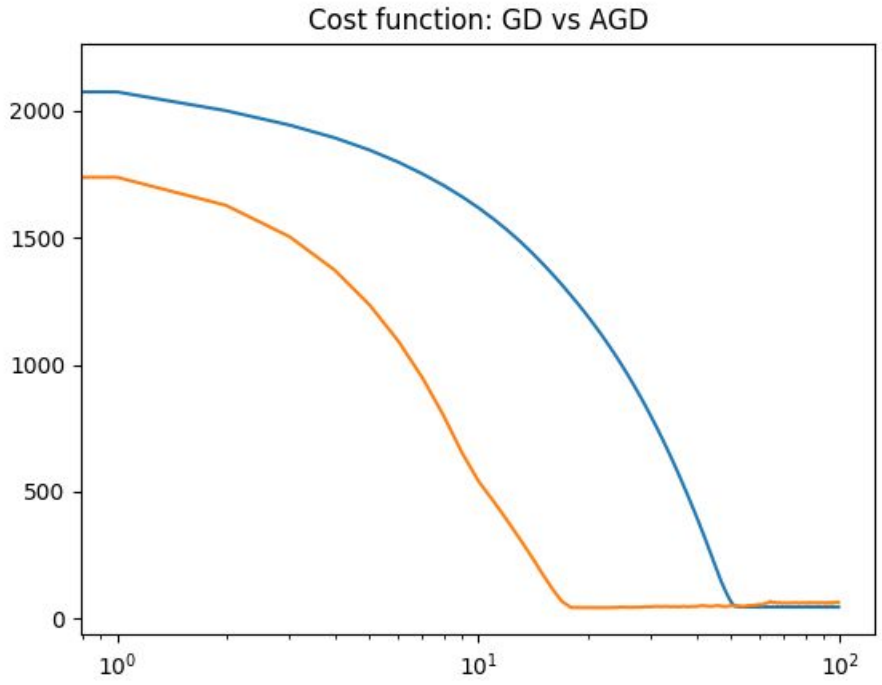




Derivatives used in GD:

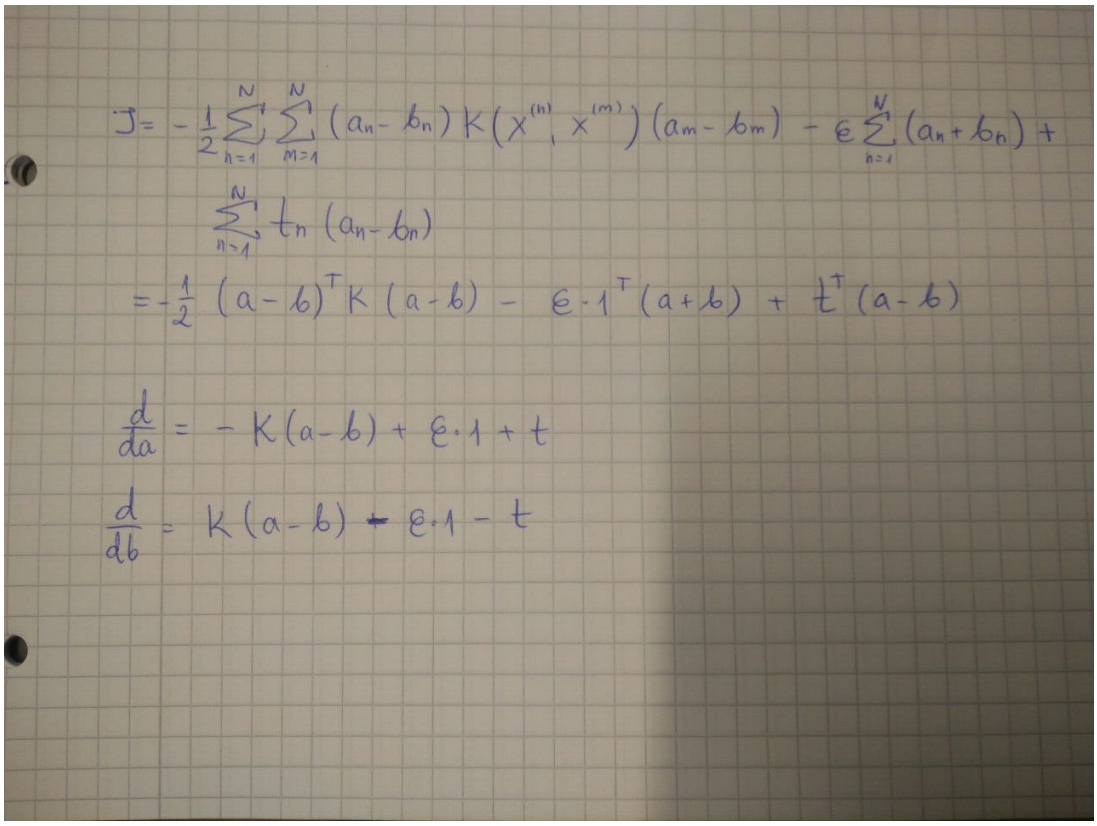


2. When using the Adaptive Gradient Descent (AGD) the prediction was very similar, but the cost function goes much faster in the convergence. Step size: **0.0001**:



(GD: blue; AGD: orange)

2. Support Vector Regression : the dual problem (10 points)



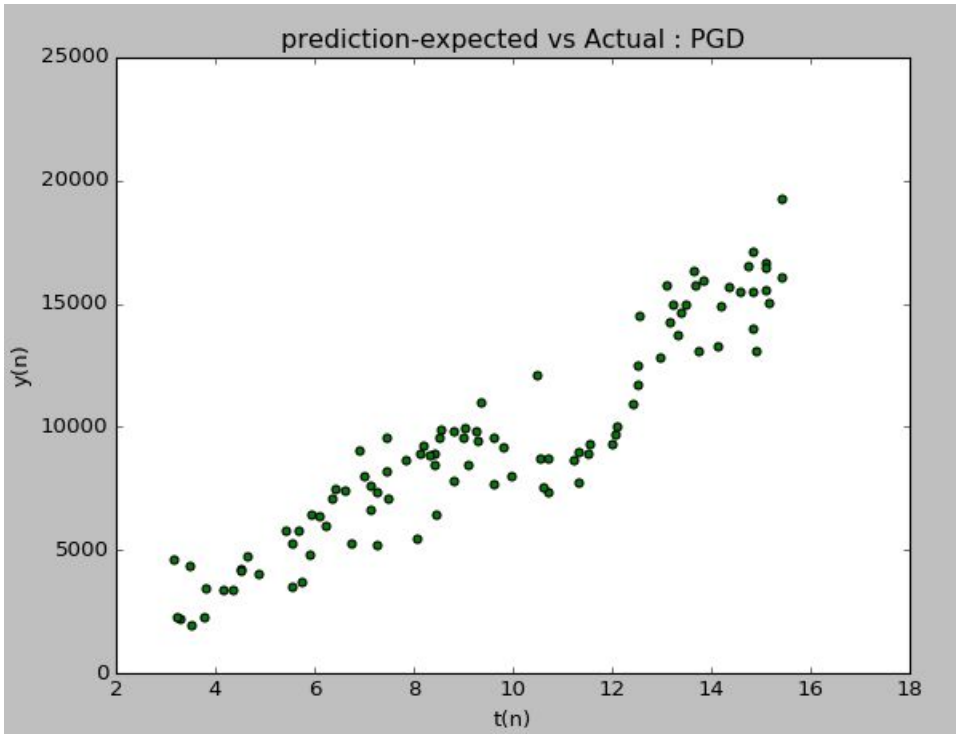
$$J = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(a_n - b_n)\,K(x^{(n)}, x^{(m)})\,(a_m - b_m) - \varepsilon\sum_{n=1}^{N}(a_n + b_n) + \sum_{n=1}^{N} t_n (a_n - b_n)$$

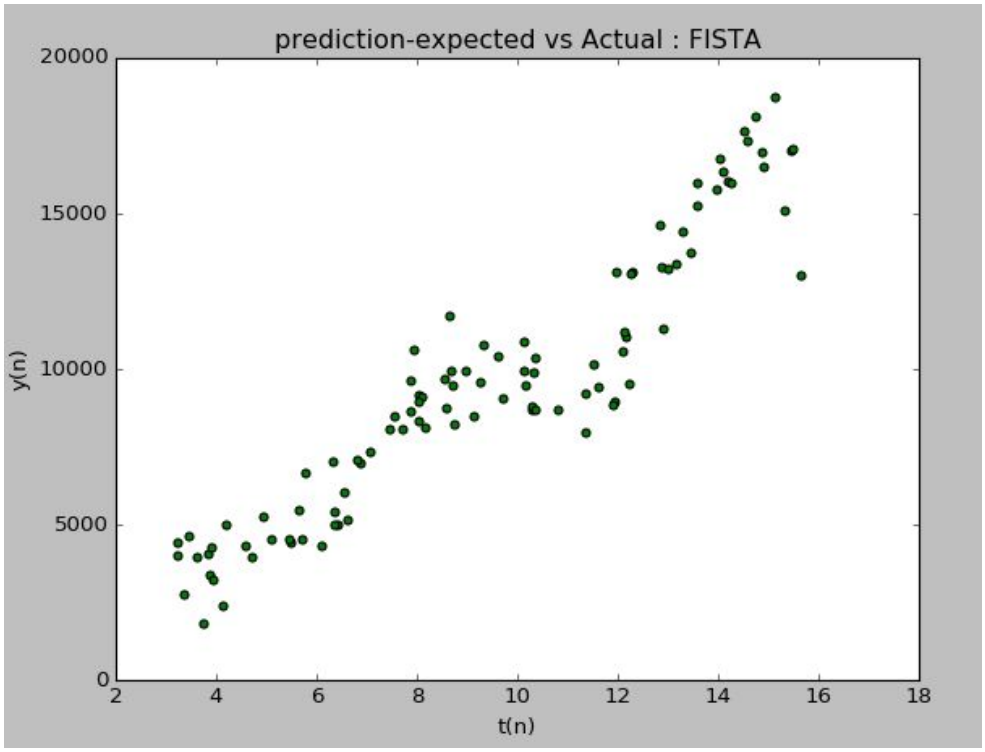
$$= -\frac{1}{2}(a-b)^T K (a-b) - \varepsilon \cdot 1^T (a+b) + t^T (a-b)$$

$$\frac{d}{da} = -K(a-b) + \varepsilon \cdot 1 + t$$

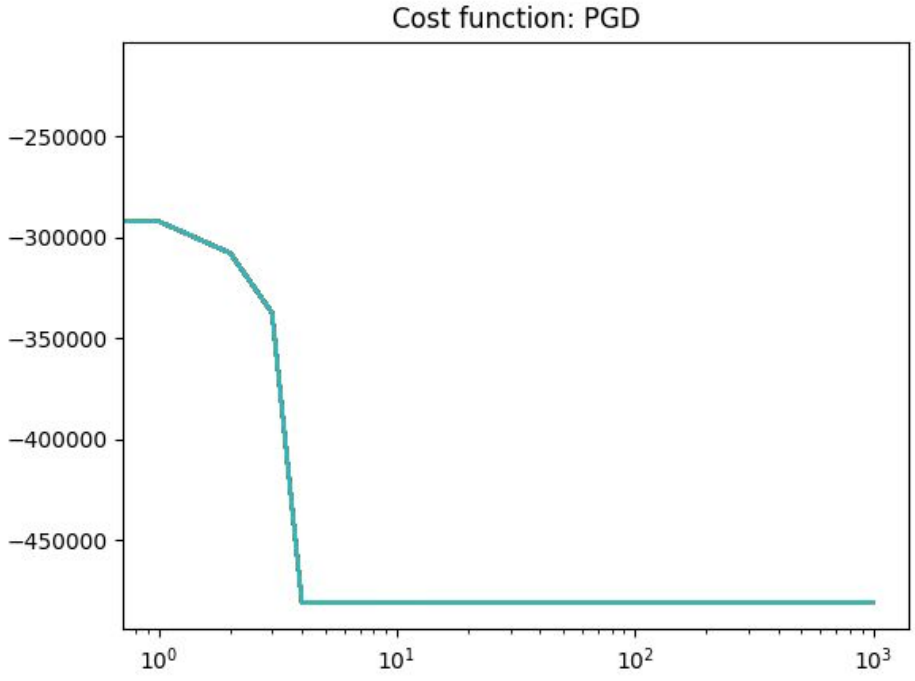
$$\frac{d}{db} = K(a-b) + \varepsilon \cdot 1 - t$$

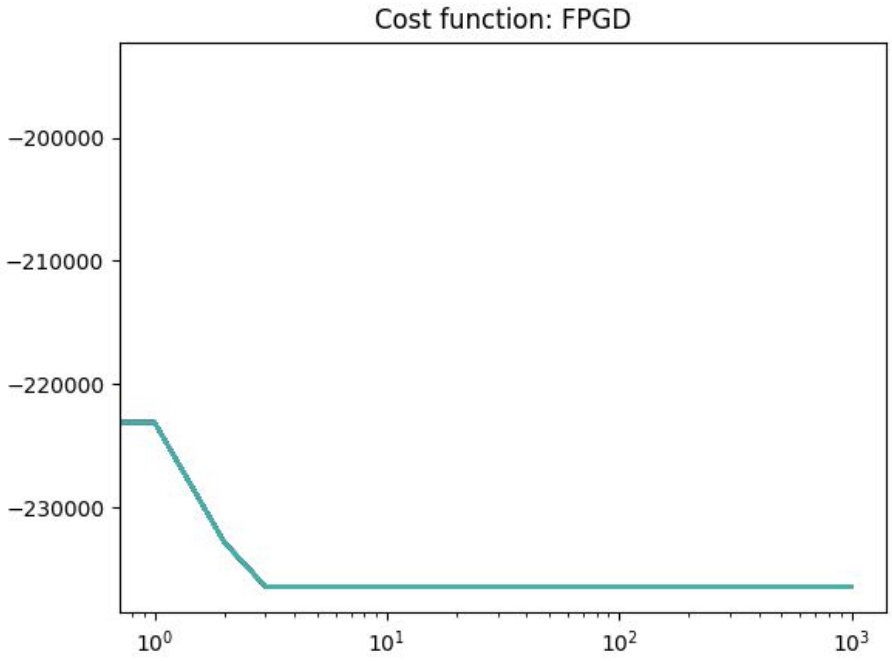
Similar to what we performed in Task1, predictions were made for test data based on the parameters a, b learnt from projected Gradient descent algorithm and FISTA. Plots are shown below for the target vs actual output.





2. Investigations (5 points)


Cost function: PGD
−250000
−300000
−350000
−400000
−450000
10^0
10^1
10^2
10^3


Cost function: FPGD
−200000
−210000
−220000
−230000
10^0
10^1
10^2
10^3