There are multiple parameters in the project which can be tuned.

* Gamma : gamma is the parameter describing the step of the gradient descent and of the stochastic gradient descent. At each iteration, the new weight will be calculated by adding to the current weights the derivative of the weights times gamma. It mustn’t be too high or we will go in the direction of the good weights but miss them and go further. And at each iteration, we will further ourselves from the weights. If it is too small, we will eventually reach the best weights, but it will take a lot of time. Thus, we searched iteratively for the smallest gamma for which we got a reasonable computing time.
  + Least square GD and Least square SGD : gamma=0.000003.
  + Logistic regression and regularized logistic regression : 0.00003
* Max iterations: this parameter describes how many steps will we take before stopping the algorithm. The higher, the more finely tuned the weights will get, and the smaller, the shorter the algorithme will be.
  + Least square GD and Least square SGD : max iterations=1000.
  + Logistic regression and regularized logistic regression : max iterations=800
* Lambda: This parameter is used to avoid overfitting. For this, we did a cross-validation (with 4 folds) and check the test error. We tried it for 30 different lambdas and we saw that there was no lambdas for which the test error was extremely different from the train error. We can then conclude that we don’t have overfitting. This is also because we use linear models. So lambda=0
* Which function we will use to estimate the weights. We decided to use the least square method. We chose this before the least square GD and SGD since it yielded a better mse (0.3423 against 0.3616). We chose Least square over logistic regression, since when we applied the weights to the input data, classified it, and compared it to the original y, we saw that Least square was more effective than logistic regression (64695 misclassifications against 68921).