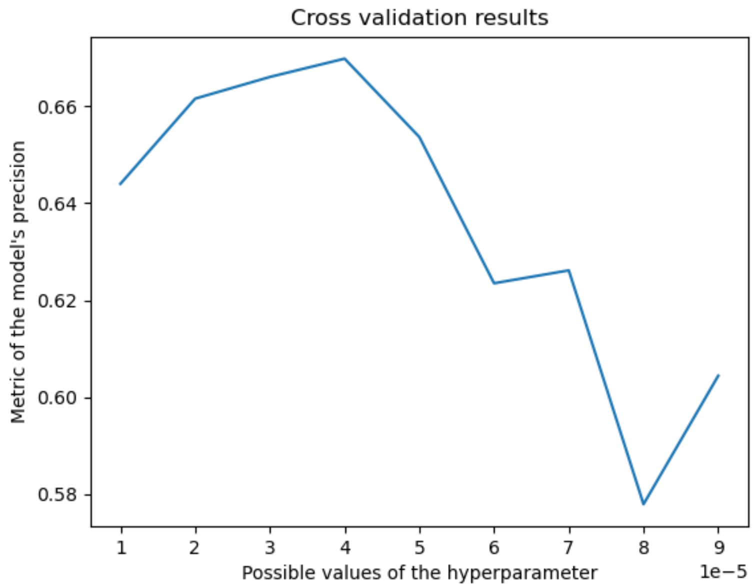
LOGISTIC REGRESSION

We perform gradient descent on the loss function of randomly generated weights to find the weights that generate a minimum of loss. To do this, we have multiple parameters: max_iters which is the maximum number of iterations during the gradient descent, and the learning rate which is the rate of descent. We found that 100 was a good value for max_iters, as a value too big would lead to a sub-optimal runtime and not a notable improvement in results. For the learning rate, we initially tested with a huge range of different values like for example the powers of 10 from -5 to 5 to see which scale of values it should take. We then progressively narrowed the possible values to a little range:

[10^{-5}*i for i in range(1,10)]. Afterwards, we use k-fold cross validation (k=4) on these possible values to find the best value.

For logistic regression, the best value generates the biggest value of the metric which is the macro F1 score. We can see on the following graph that lr=$4 * 10^{-5}$ is the most optimal value.



<u>Score without cross-validation(lr=1e-5,max_iters=100):</u> accuracy=78.80710659898476, macro F1 score=0.6548834675847545

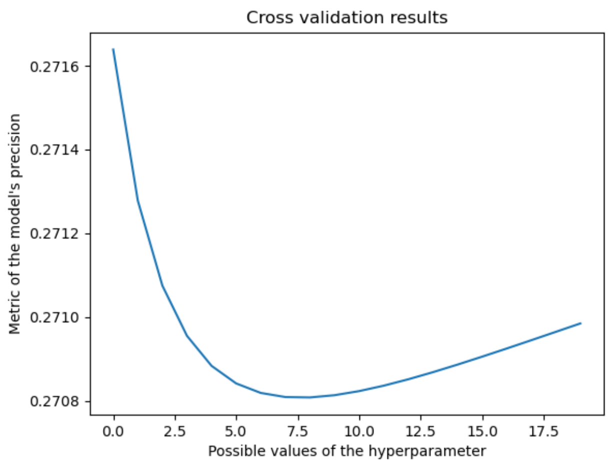
<u>Score with cross validation(lr=4e-5,max_iters=100)</u>:accuracy=79.69543147208121,macro F1 score=0.7168428180741989

<u>Conclusion:</u> Using cross validation, we get a significant improvement in the macro F1 score of about 0.06 which represents around 8% in relative value. Concerning the accuracy we get an improvement of about 0.89 which represents around 1% in relative value. These improvements don’t guarantee a notable difference between the two methods because these results are also due to a certain degree of randomness, but chances are that this follows the general trend and that cross validation is more precise.

LINEAR AND RIDGE REGRESSION

Linear and Ridge regression are very similar methods, where we obtain a closed form for the weights of our linear model, setting the gradient of the loss function to 0 and solving the equation. The only difference is a slight difference in the equation where the lambda parameter of ridge regression intervenes. Linear regression is basically ridge regression with lambda=0. To find the optimal value of lambda, we at first tested with a very large range and narrowed it down to the interval of integers between 1 and 20. Afterwards, we once again use k-fold cross validation (k=4) to obtain the best value from this range.

For linear/ridge regression, the best value achieves the smallest value of the metric: the mean squared error. We can see on the following graph that lambda=8 is the best value.



<u>Score without cross validation(lambda=1):</u> We obtain a final loss of 0.45681751676468

<u>Score with cross validation(lambda=8):</u> We obtain a final loss of 0.4542817400007059 (with an obtained metric of 0.2708087214529691 during the validation step for lambda=8).

<u>Score without ridge regression(linear regression):</u> We obtain a final loss of 0.45761932660137167.

<u>Conclusion:</u> We can observe that default ridge regression improves the loss result by approximately 0.1% (0.0008 absolute difference) whereas cross-validated improves the loss result by approximately 0.7% (0.003 absolute difference) compared to regular linear regression.

FINAL RESULTS

	Linear Regression	Ridge Regression	Logistic Regression
CROSS VALID.	/	≈0.4542 MSE	≈ 79.7% acc, ≈0.72 F1 score
NO CROSS VALID.	≈0.4576 MSE	≈0.4568 MSE	≈78.8% acc ≈0.65 F1 score