Assignment 1

Rajeev Kumar
Department of Computer Science
Indian Institute of Technology Kanpur
rajeevks21@iitk.ac.in

Divyansh Chhabria
Department of Computer Science
Indian Institute of Technology Kanpur
divyanshc21@iitk.ac.in

Sandeep Nitharwal
Department of Computer Science
Indian Institute of Technology Kanpur
nsandeep21@iitk.ac.in

Praneat Data
Department of Mathematics
Indian Institute of Technology Kanpur
praneat21@iitk.ac.in

Vrinda Sharma
Department of Mathematics
Indian Institute of Technology Kanpur
vrindas21@iitk.ac.in

1 Breaking Simple XORRO PUF into a linear model

Let the *config* bits be $\mathbf{a} \stackrel{\text{def}}{=} [a_0, a_1, a_2, ..., a_{R-1}]$.
Let δ^i_{00}, δ^i_{01}, δ^i_{10}, and δ^i_{11} be the time that the i^{th} XOR gate takes before giving its output when the input to the gate is, respectively 00, 01, 10, and 11 where the right bit in each input is a_i.
Let t^j_i be the time taken by the signal to reach the i^{th} XOR gate when the oscillating input to the zeroth XOR gate is j.
Let us define t_i as

$$t_i \stackrel{\text{def}}{=} t^0_i + t^1_i$$

By above definitions,

$$t^0_1 = \delta^0_{00}(1 - a_0) + \delta^0_{01}a_0$$
$$t^1_1 = \delta^0_{10}(1 - a_0) + \delta^0_{11}a_0$$
$$t_1 = t^0_1 + t^1_1 = \delta^0_{00}(1 - a_0) + \delta^0_{01}a_0 + \delta^0_{10}(1 - a_0) + \delta^0_{11}a_0$$
$$t_1 = (\delta^0_{01} + \delta^0_{11} - \delta^0_{00} - \delta^0_{10})a_0 + \delta^0_{00} + \delta^0_{10}$$

Similarly,

$$t^0_2 = \delta^1_{00}(1 - a_0)(1 - a_1) + \delta^1_{01}(1 - a_0)a_1 + \delta^1_{10}a_0(1 - a_1) + \delta^1_{11}a_0a_1 + t^0_1$$
$$t^0_2 = \delta^1_{00}(1 - a_0)(1 - a_1) + \delta^1_{01}(1 - a_0)a_1 + \delta^1_{10}a_0(1 - a_1) + \delta^1_{11}a_0a_1 + t^1_1$$
$$t_2 = t^0_2 + t^1_2 = \delta^1_{00}(1 - a_0)(1 - a_1) + \delta^1_{01}(1 - a_0)a_1 + \delta^1_{10}a_0(1 - a_1) + \delta^1_{11}a_0a_1 + t_1$$
$$t_2 = (\delta^1_{01} + \delta^1_{11} - \delta^1_{00} - \delta^1_{10})a_1 + \delta^1_{00} + \delta^1_{10} + t_1$$

Similarly,

$$t_3 = (\delta^2_{01} + \delta^2_{11} - \delta^2_{00} - \delta^2_{10})a_2 + \delta^2_{00} + \delta^2_{10} + t_2$$
$$\vdots$$
$$t_i = (\delta^{i-1}_{01} + \delta^{i-1}_{11} - \delta^{i-1}_{00} - \delta^{i-1}_{10})a_{i-1} + \delta^{i-1}_{00} + \delta^{i-1}_{10} + t_{i-1}$$

We wish to write t_R exclusively in terms of the *config* bits and delays of individual XOR gates. This is done using a simple linear recursion, the outcome of which is as following:

$$\begin{aligned} t_R = &(\delta_{01}^{R-1} + \delta_{11}^{R-1} - \delta_{00}^{R-1} - \delta_{10}^{R-1})a_{R-1} + (\delta_{00}^{R-1} + \delta_{10}^{R-1}) \\ &+ (\delta_{01}^{R-2} + \delta_{11}^{R-2} - \delta_{00}^{R-2} - \delta_{10}^{R-2})a_{R-2} + (\delta_{00}^{R-2} + \delta_{10}^{R-2}) \\ &\vdots \\ &+ (\delta_{01}^{i} + \delta_{11}^{i} - \delta_{00}^{i} - \delta_{10}^{i})a_i + (\delta_{00}^{i} + \delta_{10}^{i}) \\ &\vdots \\ &+ (\delta_{01}^{0} + \delta_{11}^{0} - \delta_{00}^{0} - \delta_{10}^{0})a_0 + (\delta_{00}^{0} + \delta_{10}^{0}) \end{aligned}$$

Now we represent t_R in terms of matrices as follows:

$$t_R = \begin{bmatrix} \delta_{01}^{R-1} + \delta_{11}^{R-1} - \delta_{00}^{R-1} - \delta_{10}^{R-1} \\ \delta_{01}^{R-2} + \delta_{11}^{R-2} - \delta_{00}^{R-2} - \delta_{10}^{R-2} \\ \vdots \\ \delta_{01}^{i} + \delta_{11}^{i} - \delta_{00}^{i} - \delta_{10}^{i} \\ \vdots \\ \delta_{01}^{0} + \delta_{11}^{0} - \delta_{00}^{0} - \delta_{10}^{0} \end{bmatrix}^T \begin{bmatrix} a_{R-1} \\ a_{R-2} \\ \vdots \\ a_i \\ \vdots \\ a_0 \end{bmatrix} + \sum_{i=0}^{R-1} (\delta_{00}^{i} + \delta_{10}^{i})$$

$$t_R = \mathbf{w}_0^T \mathbf{x} + b_0$$

where

$$\mathbf{w}_0 = \begin{bmatrix} \delta_{01}^{R-1} + \delta_{11}^{R-1} - \delta_{00}^{R-1} - \delta_{10}^{R-1} \\ \delta_{01}^{R-2} + \delta_{11}^{R-2} - \delta_{00}^{R-2} - \delta_{10}^{R-2} \\ \vdots \\ \delta_{01}^{i} + \delta_{11}^{i} - \delta_{00}^{i} - \delta_{10}^{i} \\ \vdots \\ \delta_{01}^{0} + \delta_{11}^{0} - \delta_{00}^{0} - \delta_{10}^{0} \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} a_{R-1} \\ a_{R-2} \\ \vdots \\ a_i \\ \vdots \\ a_0 \end{bmatrix} \qquad b_0 = \sum_{i=0}^{R-1} (\delta_{00}^{i} + \delta_{10}^{i})$$

We are asked to find a linear model to compare the frequencies of two XORROs. Since frequency is the inverse of time period, it is enough to compare the time periods of the two XORROs. It is evident from the above expression that the time period of a XORRO can be determined by a linear model.

Let us define T_U and T_L be the time periods of the upper and lower XORROs respectively. Similarly, define $\mathbf{w}_U$ and b_U, and $\mathbf{w}_L$ and b_L be the parameters of the upper and lower XORROs respectively. Let us define Δ be the difference between their time periods.

$$\Delta = T_L - T_U = (\mathbf{w}_L^T - \mathbf{w}_R^T)\mathbf{x} + (b_L - b_R)$$

$$\Delta = \mathbf{w}^T \mathbf{x} + b$$

Hence, the response to the challenge is given by the following expression:

$$\frac{1 + sign(\mathbf{w}^T \phi(\mathbf{c}) + b)}{2}$$

where

$$\mathbf{w} = \mathbf{w}_L - \mathbf{w}_R \qquad \phi(\mathbf{c}) = \mathbf{x} \qquad b = (b_L - b_R).$$

2 Extending the previous linear model to crack Advanced XORRO PUF

In section 1, we derived

$$t_R = \mathbf{w}_0^T \mathbf{x} + b = \tilde{\mathbf{w}}^T \tilde{\mathbf{x}}$$

$$\tilde{\mathbf{w}} = \begin{bmatrix} \mathbf{w}_0 \\ \\ b \end{bmatrix} \qquad \tilde{\mathbf{x}} = \begin{bmatrix} a_{R-1} \\ \\ a_{R-2} \\ \vdots \\ a_i \\ \vdots \\ a_0 \\ \\ 1 \end{bmatrix}$$

Let T_p and T_q be the time periods of XORROs selected by **p** and **q** selection bits respectively.

$$T_p = (1-p_0)(1-p_1)(1-p_2)(1-p_3)\tilde{\mathbf{w}}_0\tilde{\mathbf{x}} + (1-p_0)(1-p_1)(1-p_2)(p_3)\tilde{\mathbf{w}}_1\tilde{\mathbf{x}} + \ldots\ldots + (p_0p_1p_2p_3)\tilde{\mathbf{w}}_{15}\tilde{\mathbf{x}}$$

$$T_p = \begin{bmatrix} \tilde{\mathbf{w}}_0^T & \tilde{\mathbf{w}}_1^T & \ldots\ldots & \tilde{\mathbf{w}}_{15}^T \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}(1-p_0)(1-p_1)(1-p_2)(1-p_3) \\ \tilde{\mathbf{x}}(1-p_0)(1-p_1)(1-p_2)(p_3) \\ \vdots \\ \tilde{\mathbf{x}}(p_0)(p_1)(p_2)(p_3) \end{bmatrix}$$

$$T_p = \tilde{\mathbf{W}}_p \tilde{\mathbf{X}}$$

Similarly,

$$T_q = \begin{bmatrix} \tilde{\mathbf{w}}_0^T & \tilde{\mathbf{w}}_1^T & \ldots\ldots & \tilde{\mathbf{w}}_{15} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}(1-q_0)(1-q_1)(1-q_2)(1-q_3) \\ \tilde{\mathbf{x}}(1-q_0)(1-q_1)(1-q_2)(q_3) \\ \vdots \\ \tilde{\mathbf{x}}(q_0)(q_1)(q_2)(q_3) \end{bmatrix}$$

$$T_q = \tilde{\mathbf{W}}_q \tilde{\mathbf{X}}$$

Since, we are interested in the differences of time period of the XORROS,

$$T_q - T_p = (\tilde{\mathbf{W}}_q^T - \tilde{\mathbf{W}}_p^T)\tilde{\mathbf{X}}$$

From above, we see that the Advanced XORRO PUF can be cracked by a linear model whose input features are described by the vector **X**. Hence, only a single linear model on the features described by **X** can prove Melbo wrong!

3 Outcomes with different models while tuning hyperparameters

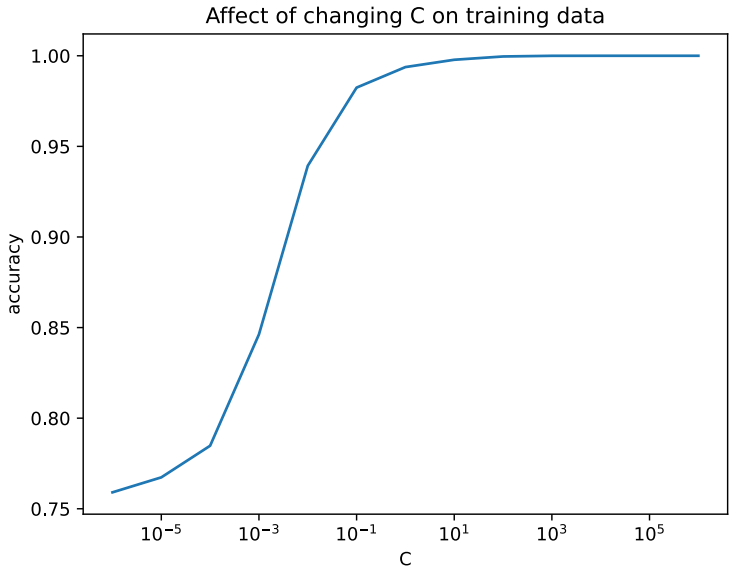
3.1 Changing loss hyperparameter in LinearSVC:

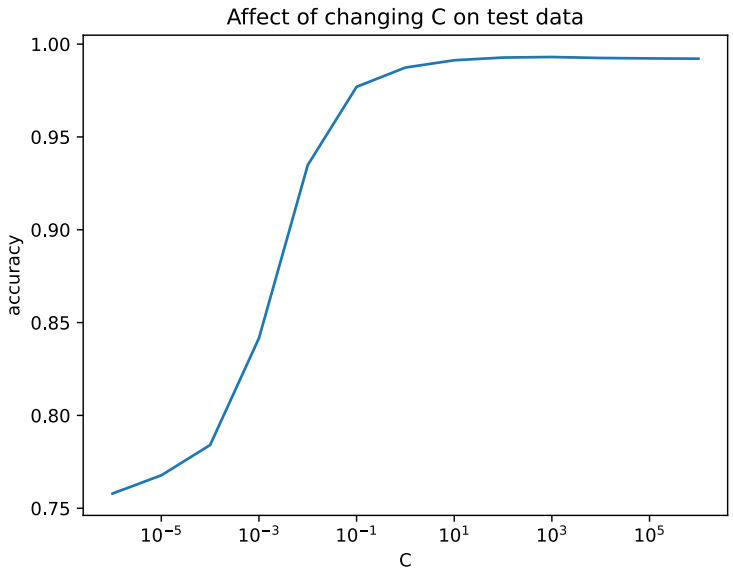
Accuracies on changing the loss function in the LinearSVC model, with default values of hyperparameters and max_iter = 10,000 are as shown:

Loss	Training Data	Test Data
Hinge	99.31	98.69
Squared Hinge	99.72	99.08

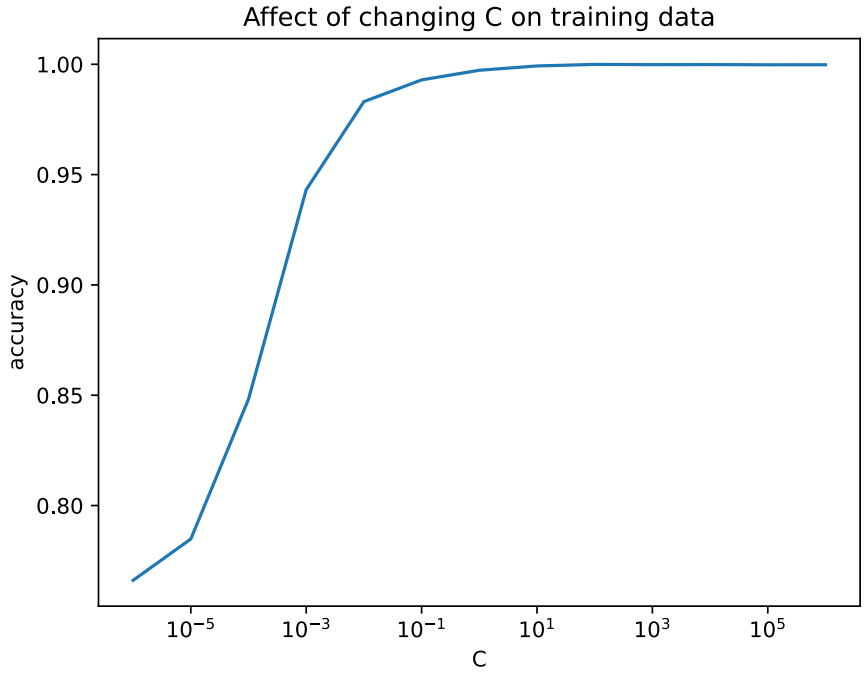
3.2 Tuning C hyperparameter in LinearSVC and LogisticRegression:

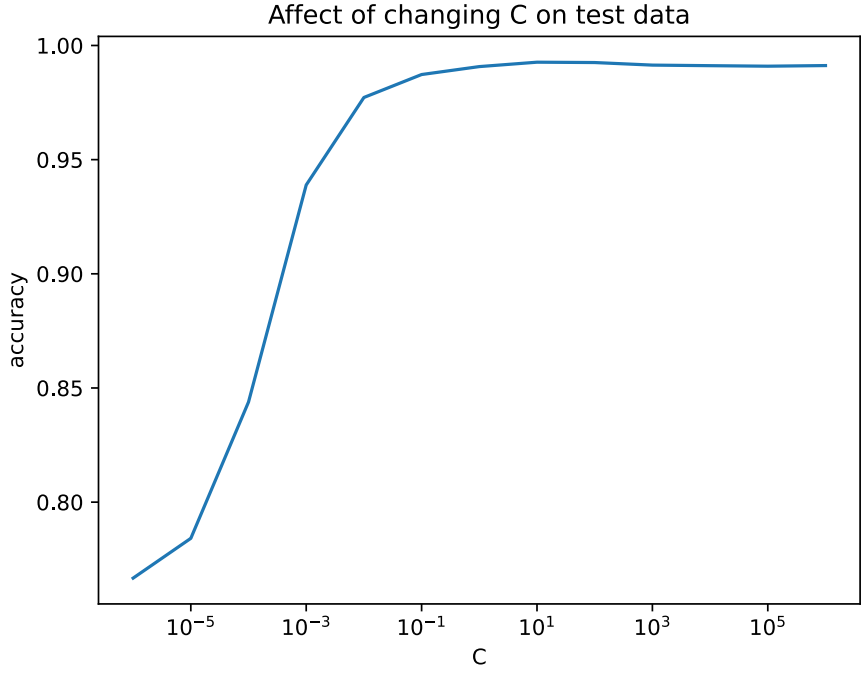
3.2.1 Logistic Regression





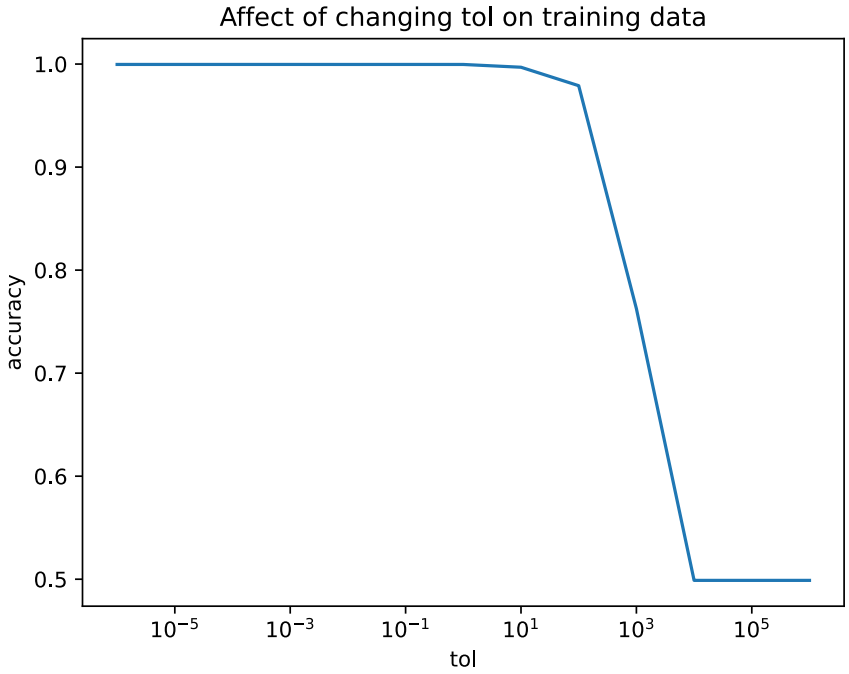
3.2.2 LinearSVC

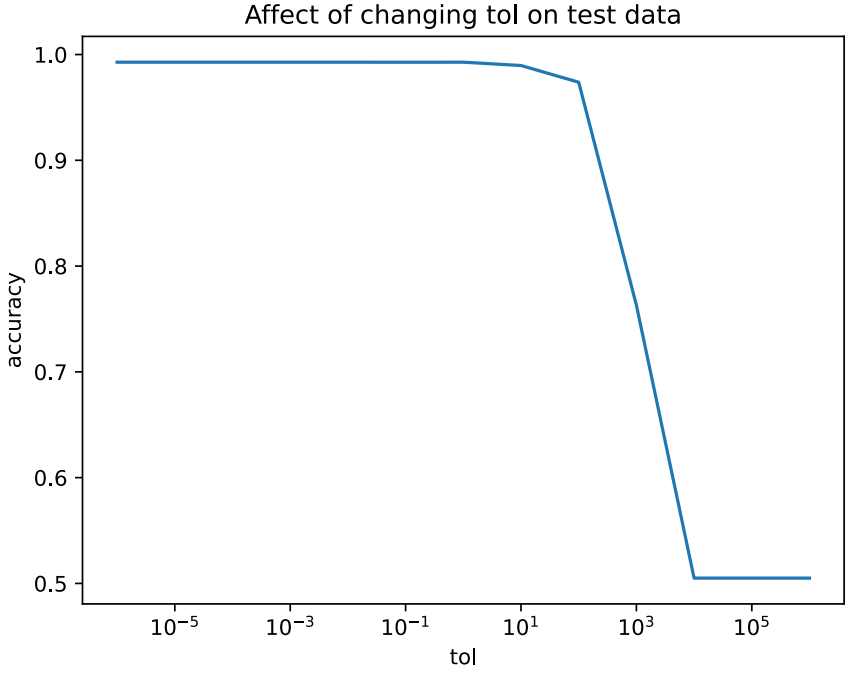




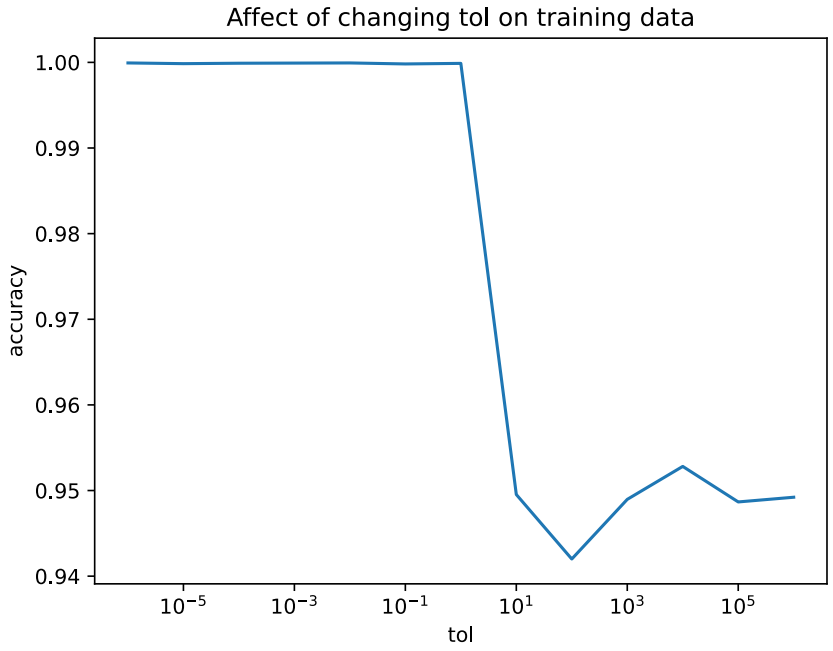
3.3 Tuning tol hyperparameter in LinearSVC and LogisticRegression:

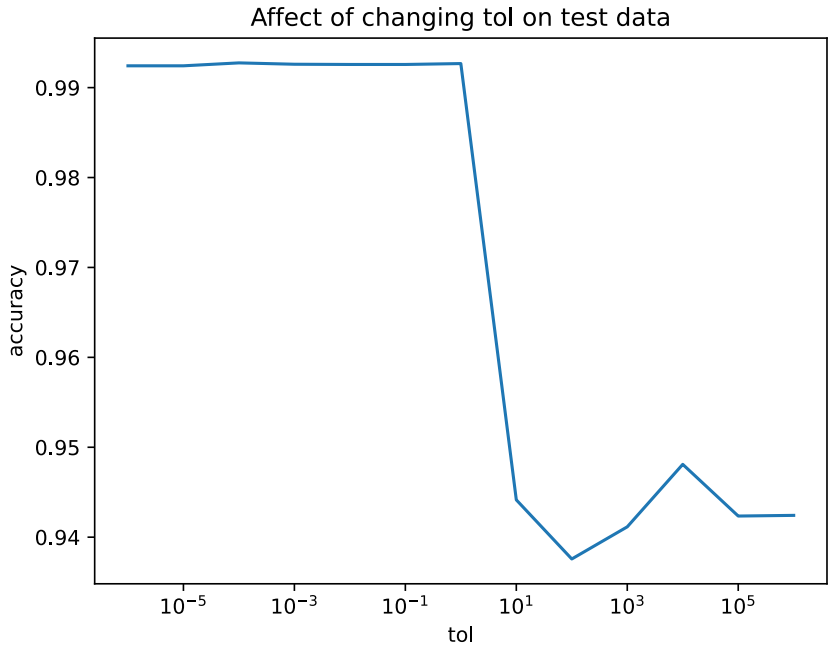
3.3.1 Logistic Regression





3.3.2 LinearSVC





3.4 Changing penalty hyperparameter in LinearSVC and LogisticRegression:

Model	Training Data	Test Data
LinearSVC	L1: 99.82 L2: 99.72	L1: 99.62 L2: 99.07
Logistic Regression	L1: 99.16 L2: 99.38	L1: 98.75 L2: 98.75