There are 2 ways of representing dataset:

First:

In this we represent each column as a feature($f1$, $f2$, ..., fd) and each row as datapoint,

So we can say each row is x_i^T that means transponse of x_i, where x_i = [SL, SW, Pl, PW] for iris, and since we represent it as column vector and to convert it into row vector we take transponse of it.

Y is a column vector of dimension $n*1$, where each y or y_i is the class label or result of x_i^T. Example for iris for any x_i, y_i is one of the { Setossa, versicolor, virginica}.
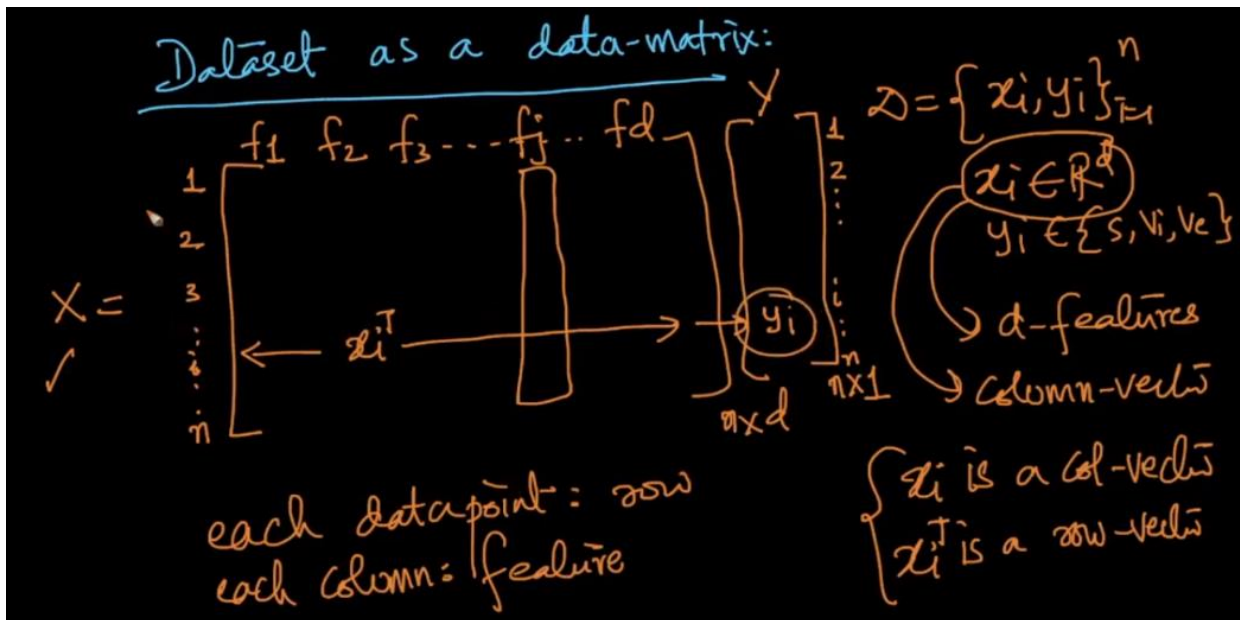
pto

Second: In this we specify each row as feature and each column as datapoint, that means we arrange each x_i in n columns (since x_i is column vector), for example for iris we have rows as SL,SW,Pl,PW and there will be 150 columns. It's found in most of the research papers.

$$X = \begin{array}{c} \\ f_1 \\ f_2 \\ f_3 \\ f_j \\ \vdots \\ f_d \end{array} \begin{array}{cccccc} 1 & 2 & 3 & 4\ldots i\ldots & n \\ & & & \uparrow \\ & & & x_i \\ & & & \downarrow \end{array}$$

$d \times n$

column: data-point
row : feature Variable

$f_1 = PL$
$f_2 = PW$
$f_3 = SL$
$f_4 = SW$

$$X = \begin{array}{c} \\ f_1 \\ f_2 \\ f_3 \\ f_j \\ \vdots \\ f_d \end{array} \begin{array}{cccccc} 1 & 2 & 3 & 4\ldots i\ldots & n \\ & & & \uparrow \\ & & & x_i \\ & & & \downarrow \end{array}$$

$d \times n$

column: data-point
row : feature Variable

$f_1 = PL$
$f_2 = PW$
$f_3 = SL$
$f_4 = SW$