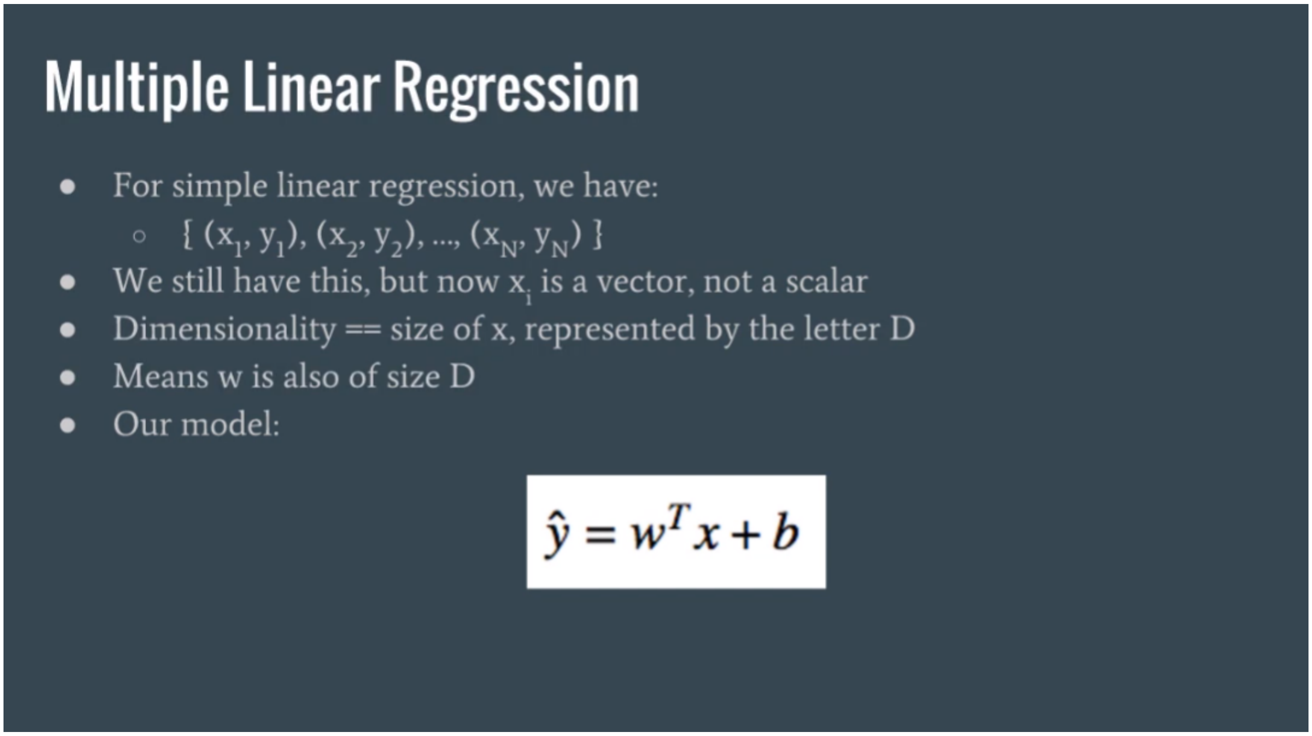
Setting up the problem

- 

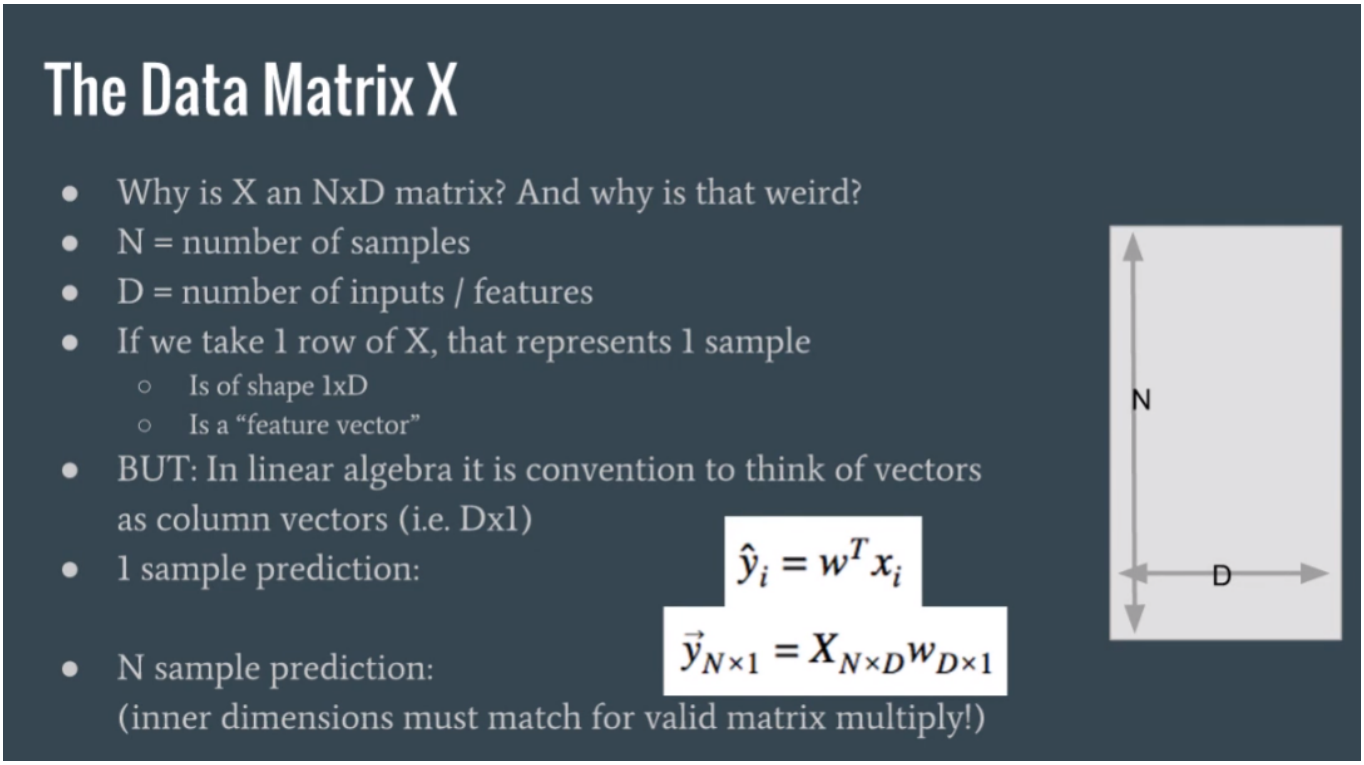
- Before our problem was working with some data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ for n data points, now our data set is $(\vec{x_1}, y_1), (\vec{x_2}, y_2), \dots, (\vec{x_n}, y_n)$ where x_i is now a vector not a scalar.
- Our model before was in the form of $y = ax + b$ and we were solving for 2 parameters a and b to find the most "optimal values", this is learning part of the process.
- Now our model is $\hat{y} = w_1x_1 + w_2x_2, \dots, w_Dx_D + b$ where D to the Dimensionality of x

Bias term

- To simplify future calculations we can actually just absorb b into w be appending an extra term x_0 which is always equal to 1 therefore $w_0x_0 = b$
- Therefore our equation is now

$$\hat{y} = w_0x_0 + w_1x_1 + w_2x_2, \dots, w_Dx_D + b, where\ x_0 = 1$$

The data matrix X



- We can think the sum of all the x_i terms as a matrix instead of row vectors where each row of a matrix X represents 1 sample

$$X = \left\{\begin{array}{cccc} x_{1,1} & x_{1,2} & \dots & x_{1,D} \\ x_{2,1} & x_{2,2} & \dots & x_{2,D} \\ \dots & \dots & \dots & \dots \\ x_{N,1} & x_{N,2} & \dots & x_{N,D} \end{array}\right\}_{N\times D}$$

- Our matrix W needs to be transposed so we can do matrix multiplication, as the inner dimensions must match, therefore W must be a column vector of size $D \times 1$
- So for a single sample prediction $\hat{y}_i = w^T x_i$, we have
- $$y_1 = w^T x_{1=} \left\{\begin{array}{cccc} w_1 & w_2 & \dots & w_D \end{array}\right\}_{1\times D} \times$$

\Bigg{\begin{array}{rcl}
x{i,1}\
x{i,2}\
.. \
x{i,D}\
\end{array}\Bigg}{D\times 1} = w{1}x{i,1}+w{2}x{i,2}+...+w{D}x{i,D}

![[Pasted image 20230519172206.png]] - Instead of calculating each y_{i} on its own we can just do all the calculations by multiplyir