

## Setting up the problem

### Multiple Linear Regression

- For simple linear regression, we have:
  - $\{ (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N) \}$
- We still have this, but now  $x_i$  is a vector, not a scalar
- Dimensionality == size of  $x$ , represented by the letter  $D$
- Means  $w$  is also of size  $D$
- Our model:

$$\hat{y} = w^T x + b$$

- 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  $w$  by appending an extra term  $x_0$  which is always equal to 1 therefore  $w_0x_0 = b$
- Therefore our equation is now

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

## The data matrix $X$

# The Data Matrix X

- Why is X an Nx D matrix? And why is that weird?
- N = number of samples
- D = number of inputs / features
- If we take 1 row of X, that represents 1 sample
  - Is of shape 1xD
  - Is a “feature vector”
- BUT: In linear algebra it is convention to think of vectors as column vectors (i.e. Dx1)
- 1 sample prediction:
 
$$\hat{y}_i = w^T x_i$$
- N sample prediction:
 
$$\vec{y}_{N \times 1} = X_{N \times D} w_{D \times 1}$$
 (inner dimensions must match for valid matrix multiply!)



- 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

$$X = \begin{Bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,D} \\ x_{2,1} & x_{2,2} & \dots & x_{2,D} \\ \dots & \dots & \dots & \dots \\ x_{N,1} & x_{N,2} & \dots & x_{N,D} \end{Bmatrix}_{N \times D}$$

- 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

- $$y_1 = w^T x_1 = \begin{Bmatrix} w_1 & w_2 & \dots & w_D \end{Bmatrix}_{1 \times D} \times$$

$\begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \dots \\ x_{i,D} \end{bmatrix}_{D \times 1}$

$= w_{\{1\}} x_{\{i,1\}} + w_{\{2\}} x_{\{i,2\}} + \dots + w_{\{D\}} x_{\{i,D\}}$

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