

Please hand in this paper to the instructor before the end of the lecture.

Name: _____

Student number: _____ Date: _____

Q. You are given the following data:

	Features			Labels
1	$x_1^{(1)}$	$\dots$	$x_d^{(1)}$	$y^{(1)}$
2	$x_1^{(2)}$	$\dots$	$x_d^{(2)}$	$y^{(2)}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
N	$x_1^{(N)}$	$\dots$	$x_d^{(N)}$	$y^{(N)}$

You are asked to fit a linear model to it. Complete the following tasks.

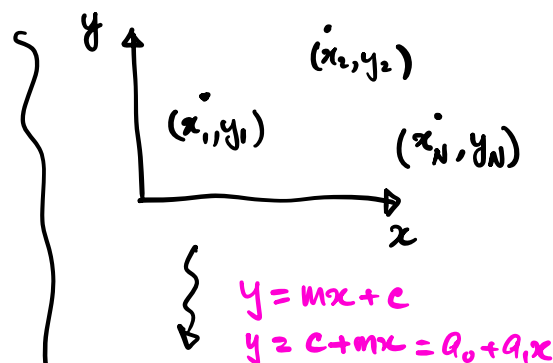
- Express the model mathematically.
- How many parameters this model will have?
- Write down the MSE loss expression for your setup.

$$y = \underline{a_0} + \underline{a_1}x_1 + \underline{a_2}x_2 + \dots + \underline{a_d}x_d$$

$$A = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_d^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_d^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(N)} & x_2^{(N)} & \dots & x_d^{(N)} \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix}$$

$$\theta = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{bmatrix}$$

 $x_1^{(1)}$ $x_1^{(N)}$

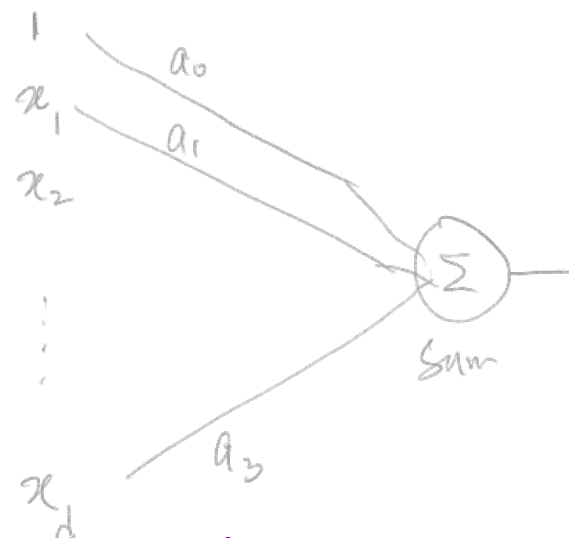
Features

Labels

 x_1 y_1 x_2 y_2 $\vdots$ $\vdots$ x_N y_N

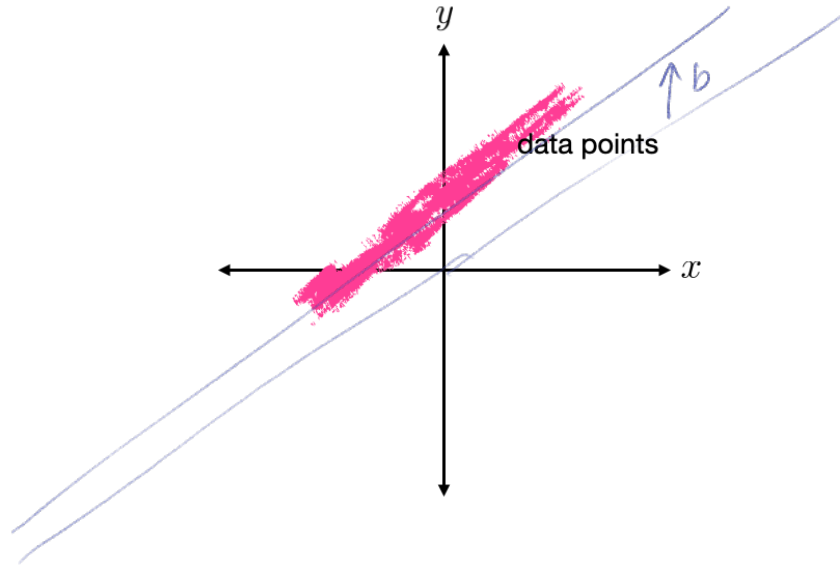
$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$\theta = \begin{bmatrix} m \\ c \end{bmatrix}$$



$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^N (y^{(i)} - a_0 - a_1 x_1^{(i)} - a_2 x_2^{(i)} - \dots - a_d x_d^{(i)})^2 / N \\ &= \sum_{i=1}^N (y^{(i)} - \hat{y}^{(i)})^2 / N \end{aligned}$$

Q. Consider the following setup that shows a collection of data points. Here x -coordinate represents inputs and y -coordinates represents their respective output.



Since both x and y are continuous, we have a *regression* problem at our hand. We are asked to fit the following, single-parameter model to this data:

$$y = mx,$$

where m is the lone model parameter.

Devise a scheme to fit this model to this data? Do you think this model has enough “model complexity” to fit this data well? Can you spot a problem? If there is a problem, can you suggest a fix.