

w06a — Model fitting: solution**Part 1: the tabulated data**

1. **The model.** $\hat{y} = \mathbf{w}^\top \mathbf{x} + b = \sum_{j=1}^d w_j x_j + b$.
2. **Parameters.** d weights plus one bias = $\boxed{d + 1}$. The common slip is answering d and forgetting the bias.
3. **MSE loss.**

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N (y^{(i)} - \hat{y}^{(i)})^2 = \frac{1}{N} \sum_{i=1}^N (y^{(i)} - \mathbf{w}^\top \mathbf{x}^{(i)} - b)^2.$$

Either form works, with or without the $1/N$ or a $1/2$; the minimiser is unchanged.

Part 2: the figure, and fitting $y = mx$

Fitting scheme. One parameter, so minimise $\mathcal{L}(m) = \sum_i (y_i - mx_i)^2$. Setting $d\mathcal{L}/dm = 0$:

$$m = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

A closed form — no iteration needed. Gradient descent also works if correctly derived.

Is the model complex enough? What is the problem?

This is the point of the question. $y = mx$ is a line **forced through the origin**. The data cloud in the figure clearly does *not* pass through the origin — it crosses the y -axis above zero. So the single-parameter model is **underfitting**: no choice of m can both match the slope and pass through the data, and the fitted line will be dragged to a compromise slope that is wrong.

The fix. Add an intercept:

$$y = mx + c$$

Two parameters, still linear least squares, still closed-form.

The specific problem is the *missing intercept*. Answers of the form “it needs more complexity” or “use a polynomial” are weaker: the data is manifestly linear, so a higher-order model would be the wrong remedy and risks overfitting.