Training Error - average error between targets and predictions, or the average cost of our outputs

Generalisation Error - how well will we do on future data?

- We don't know what future data will be, as well as what labels will they have
- However, we know the **possible range of x and y**
    - E.g. all possible pixel values for an image, or all possible classes
- $$E_{gen} = \int error(f_D(\mathbf{x}), y)p(y, \mathbf{x})d\mathbf{x}$$
    - We integrate over all possible values of **x, y**
        - We **integrate** when our x and y are continuous, and when they are categorical, we **average** out
            - Integrate to capture over every possible value of continuous variables
    - **p(y, x) -** a joint distribution, the probability of finding a specific pair of **x** and **y**
    - Same error as before
- Generalisation errors are higher than training errors
- We can't precisely compute generalisation error as we don't know **p(y, x)**
- We can estimate **p(y, x)**, which is basically what neural networks do