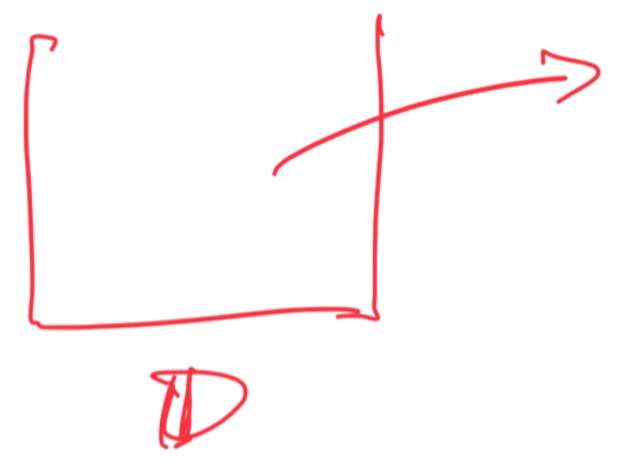


$J(\theta) \rightarrow$ cost / training loss / error

$\theta \rightarrow$ SGD / Newton / MLE

\downarrow
how good is this θ

Test loss



draw new example, $(x, y) \sim D \rightarrow$ test distribution

evaluate:

$$L(\theta) = \mathbb{E}_{(x, y) \sim D} [(y - h_{\theta}(x))^2]$$

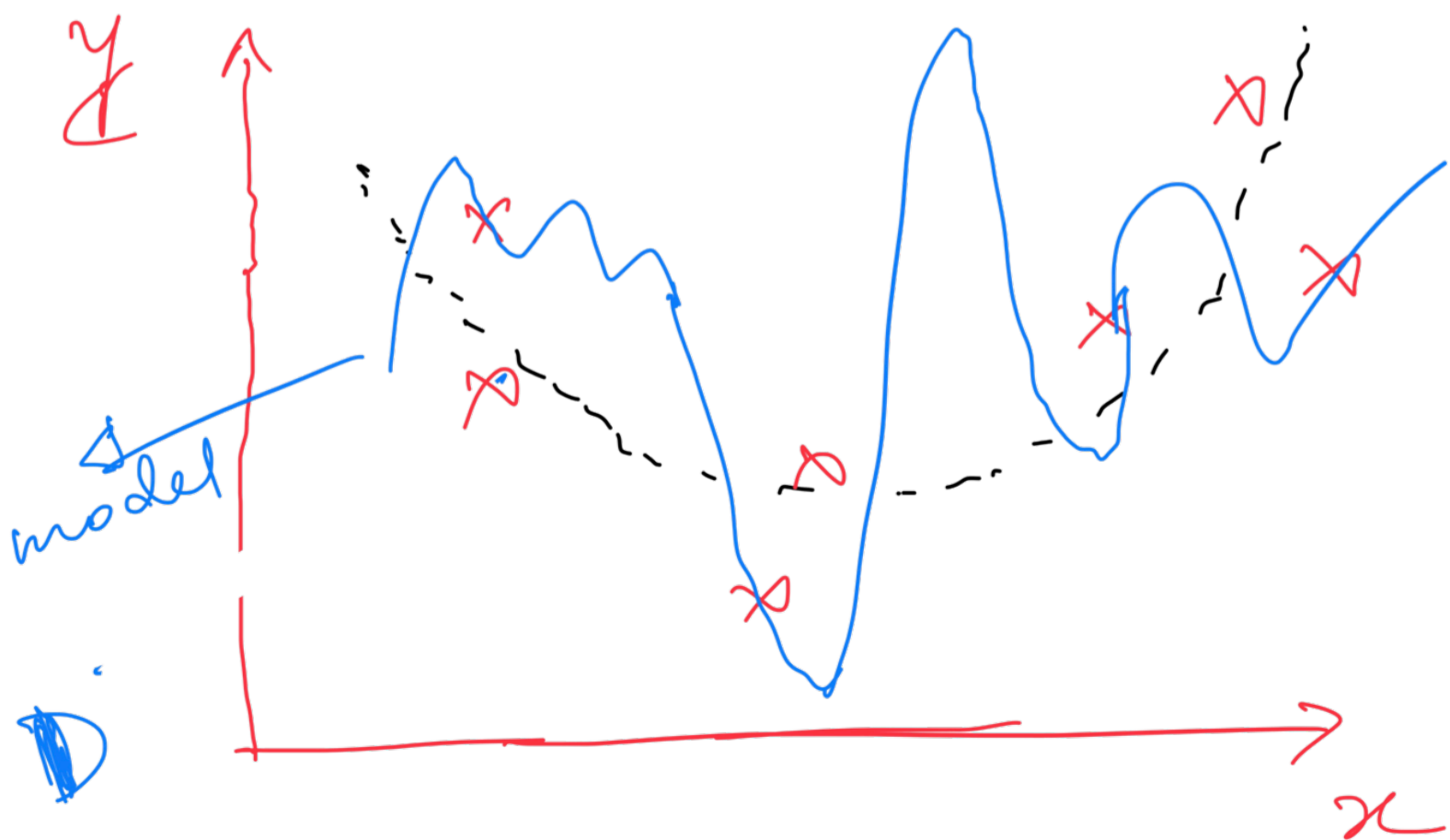
$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}) \dots \dots$ } avg

① $J(\theta)$, $\frac{L(\theta) - J(\theta)}{\quad} \Rightarrow$ Generalization gap
 \downarrow low $\quad \downarrow$ low

② $L(\theta)$ is too big

\rightarrow overfitting : $J(\theta)$ is small, $L(\theta)$ big
 \rightarrow underfitting : $J(\theta)$ is big

Next class $L(\theta) = (\text{bias})^2 + \text{variance} + \xi$



\rightarrow overfitting

$J(\theta) \approx 0$

- linear
- quad.
- 5th deg. poly
- NN