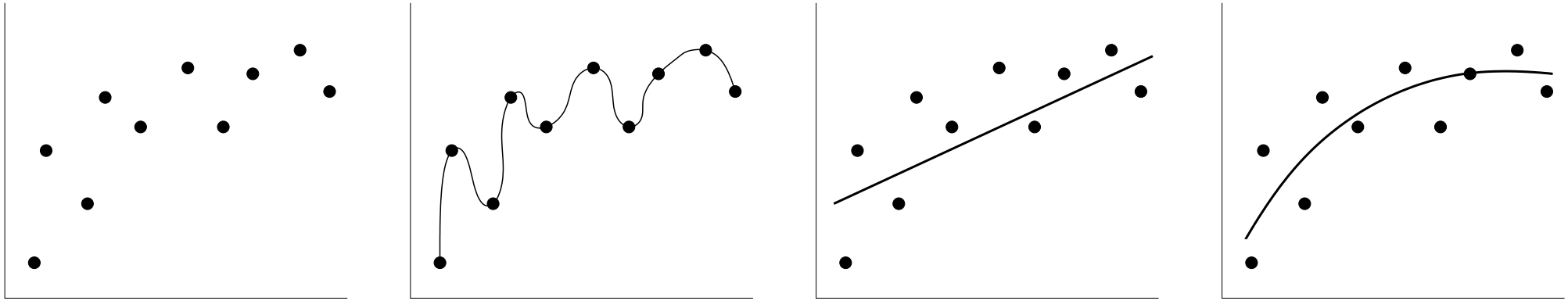
Generalization and overfitting



Learning as Bayesian inference

$$p(M|D) = \frac{p(D|M)\ p(M)}{p(D)}$$

$$-\log p(M|D) = -\log p(D|M) - \log p(M) + \log p(D)$$

$$\text{objective function} = \text{“error”} + \text{“complexity”}$$

Weight decay: Complexity function $\frac{1}{2}w_{ij}^2$ (penalize large weights)

$$\triangle w_{ij}^{[t]} = -\varepsilon \frac{\partial E}{\partial w_{ij}} + \alpha\left(\triangle w_{ij}^{[t-1]}\right) - \lambda w_{ij}$$