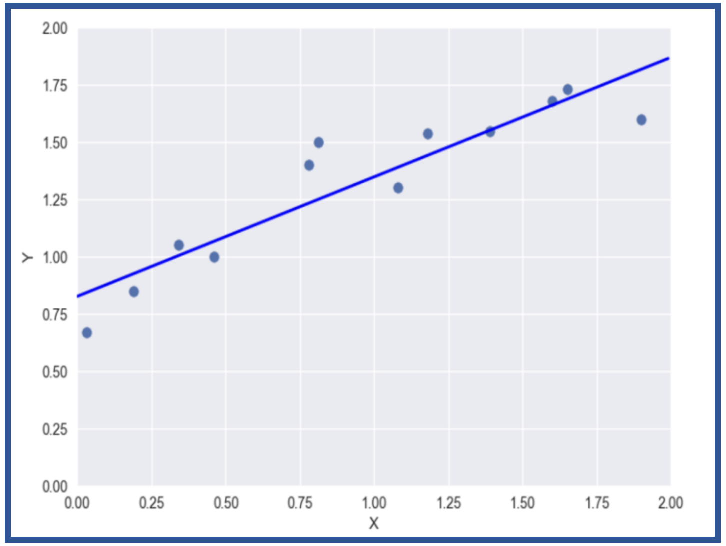
AIL721: Deep Learning

Instructor: James Arambam

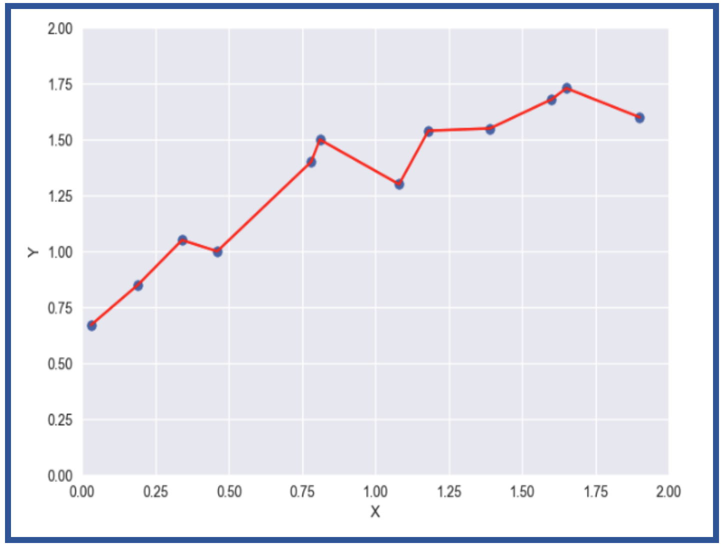
ScAI | Yardi School of Artificial Intelligence
Indian Institute of Technology Delhi

Model Performance

ScAI



Model 1



Model 2

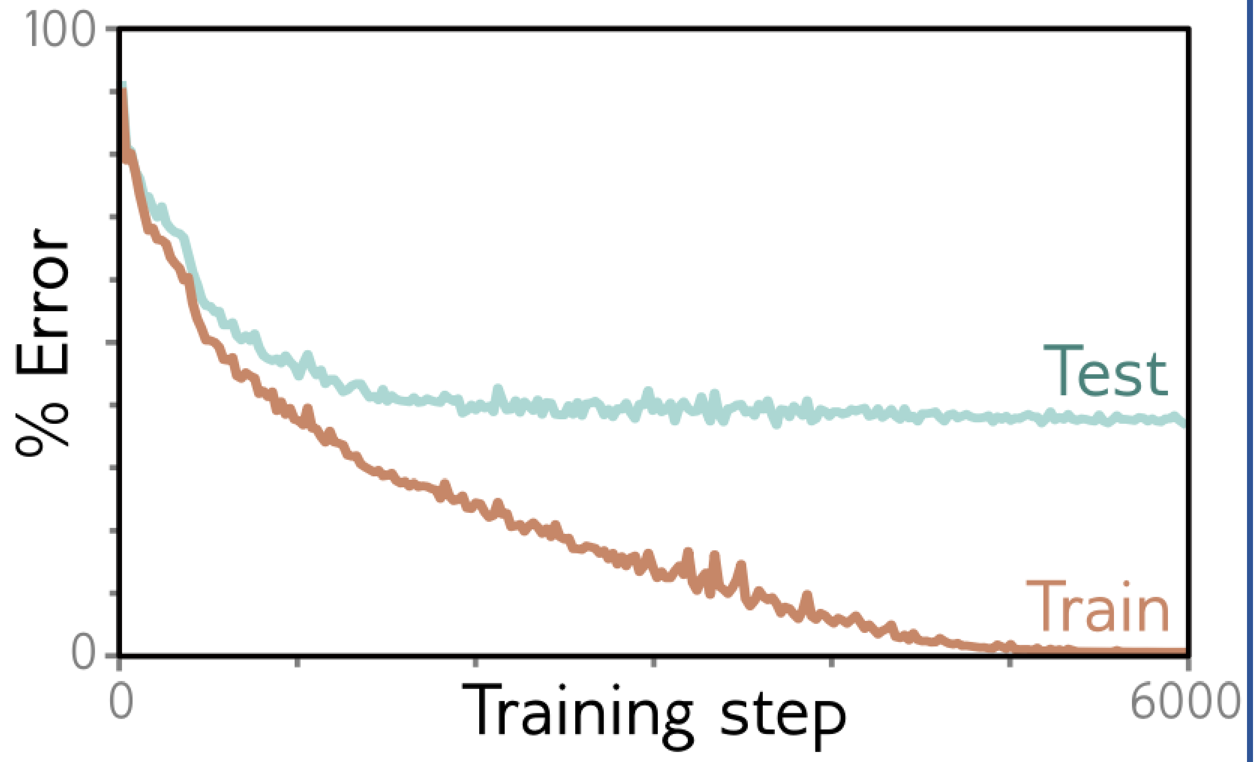
Test Data

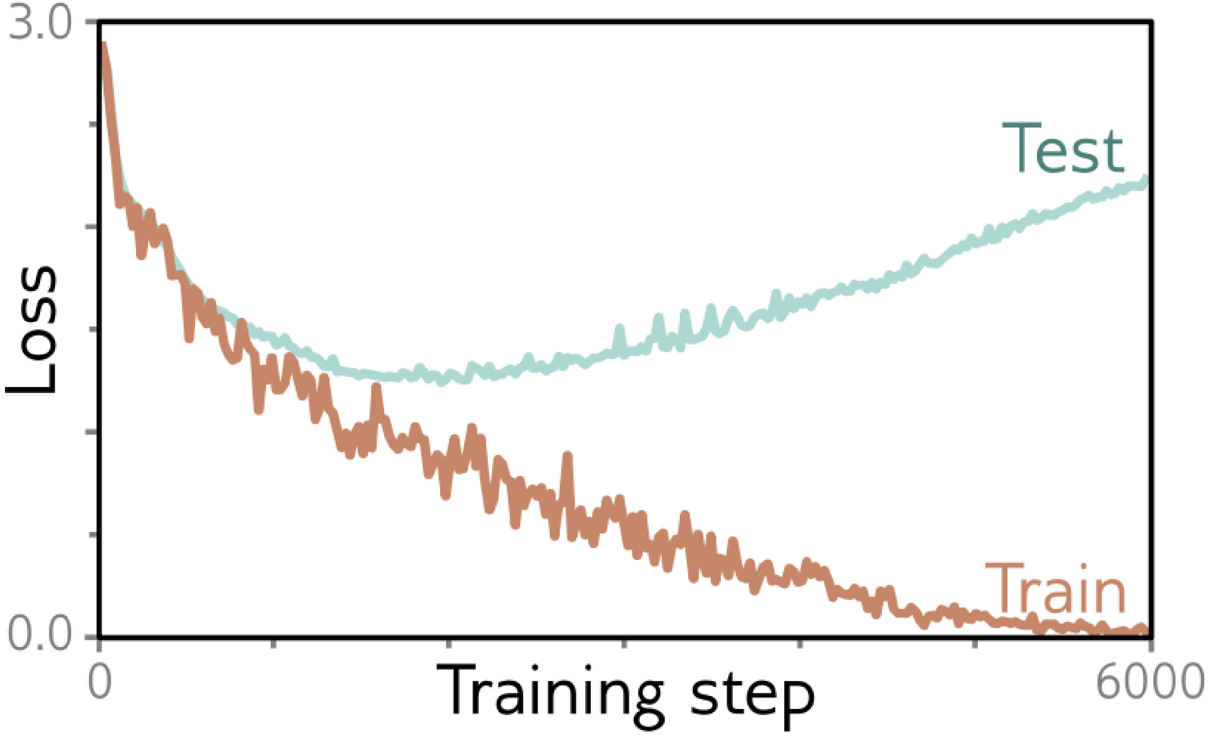
Which model performance is better?

Test Loss?

Training Loss?

Model Performance



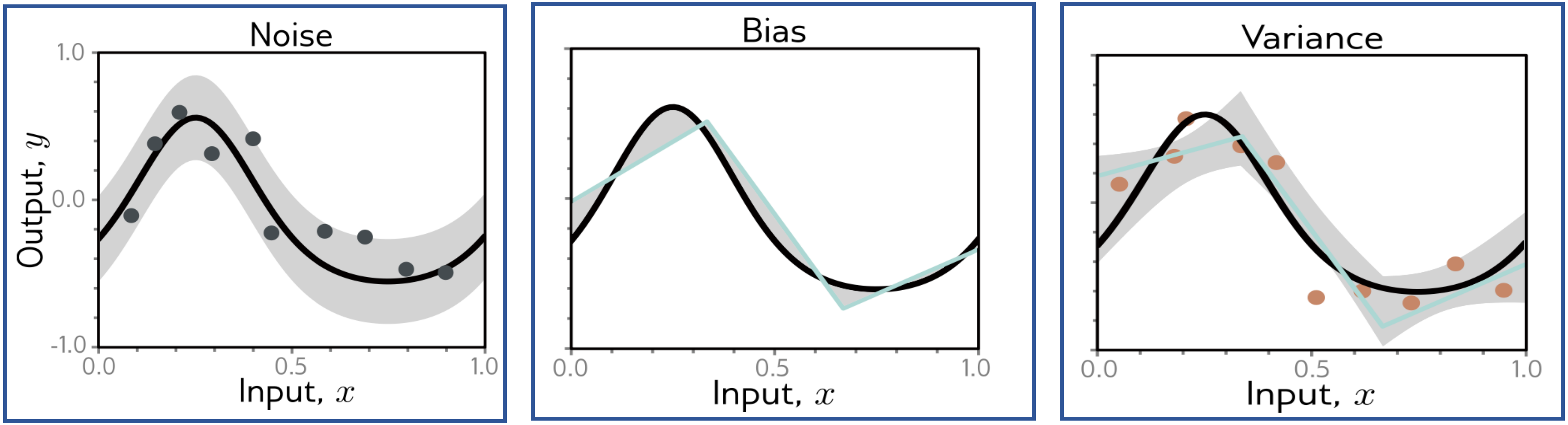


Model failed to generalize to unseen data

[Image from Prince 2023 UDL Book]

Model Performance

❑ Source of Error



[Image from Prince 2023 UDL Book]

Model Performance

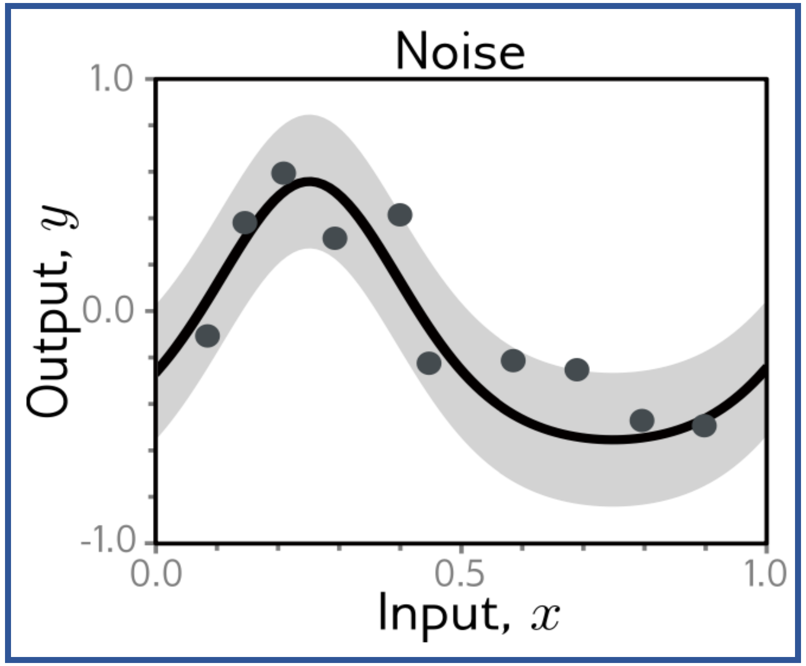
❑ Mathematical formulation of Test Error

[Image from Prince 2023 UDL Book]

Model Performance

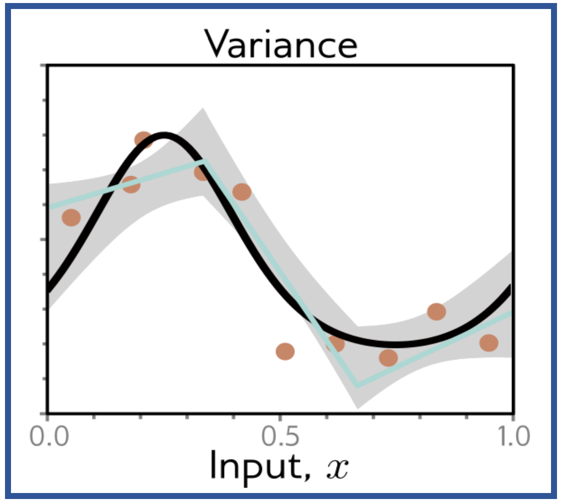
❑ Reducing Noise

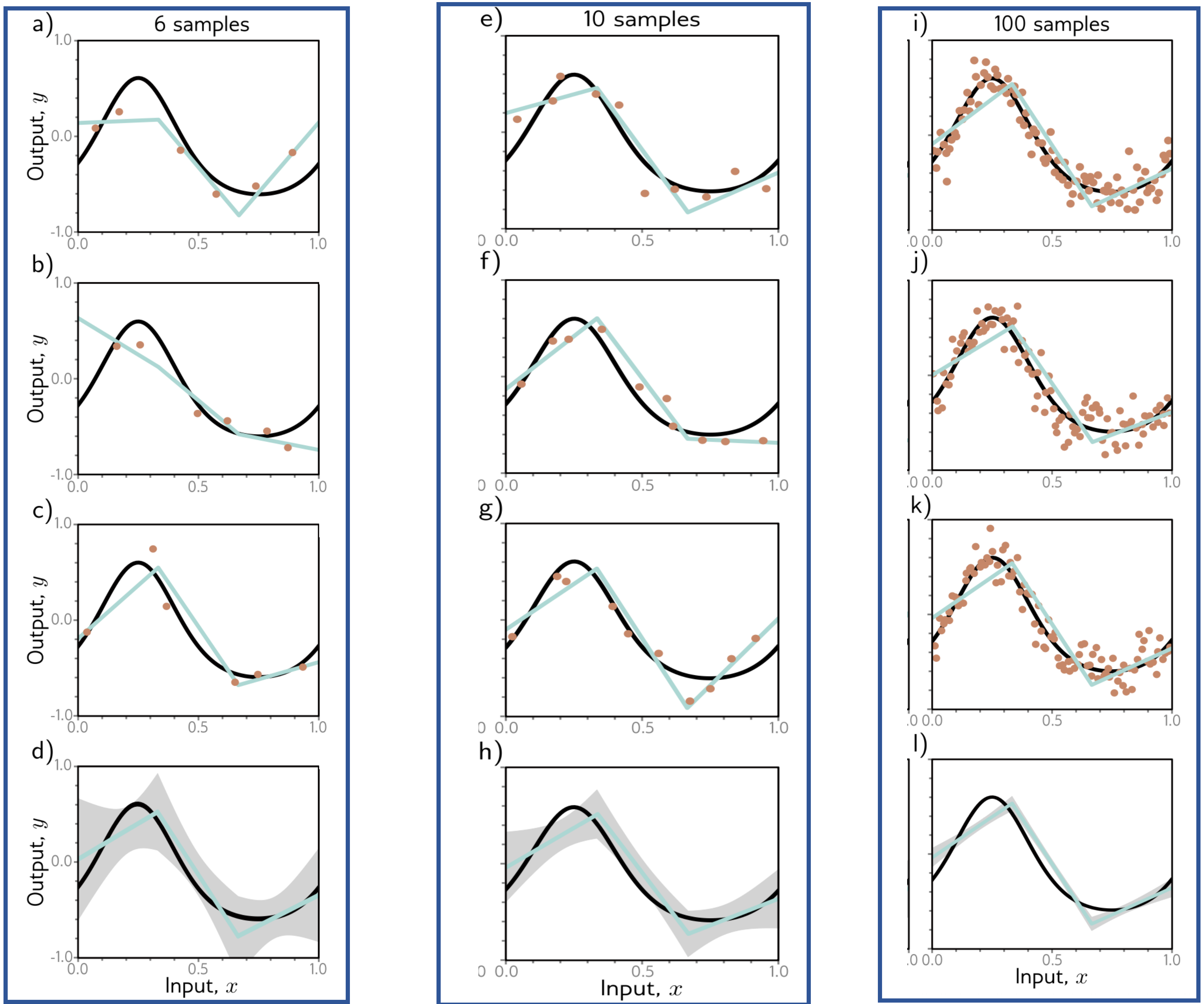
- **Nothing** much we can do there.
- **Fundamental limit** on expected model performance.



[Image from Prince 2023 UDL Book]

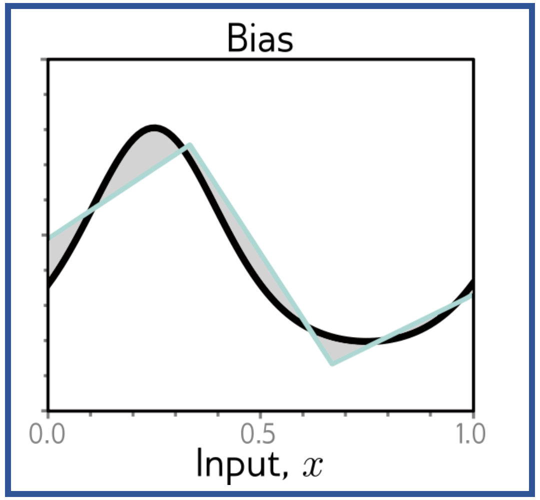
❑ Reducing Variance

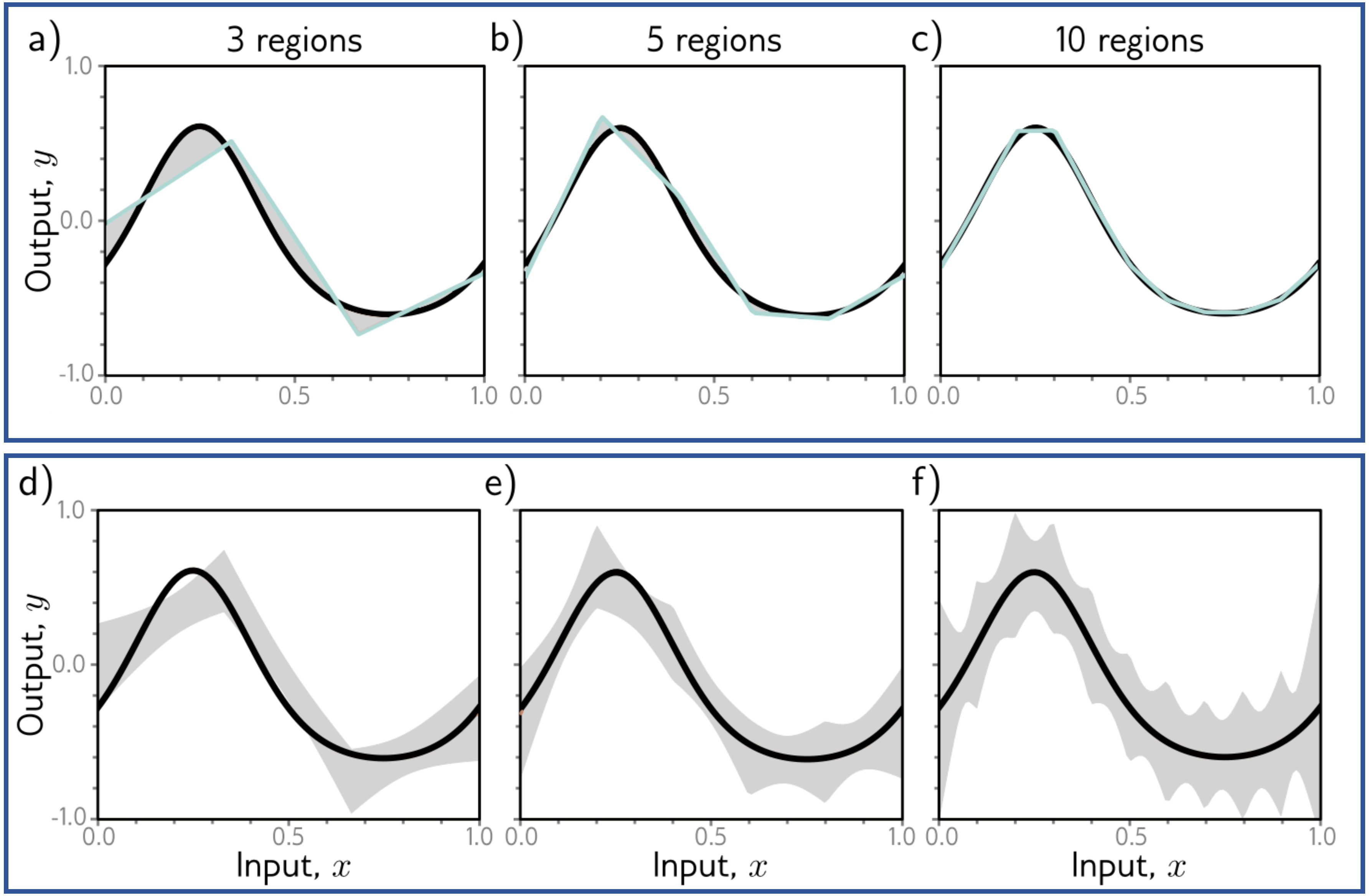




Model Performance

❑ Reducing Bias





Bias-Variance Trade-off

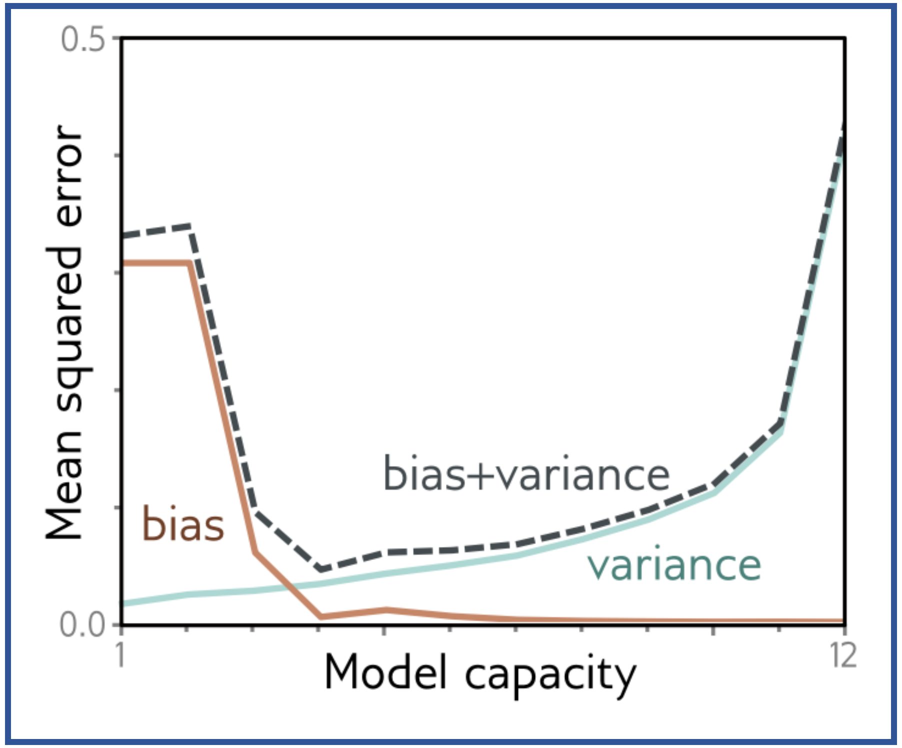
[Image from Prince 2023 UDL Book]

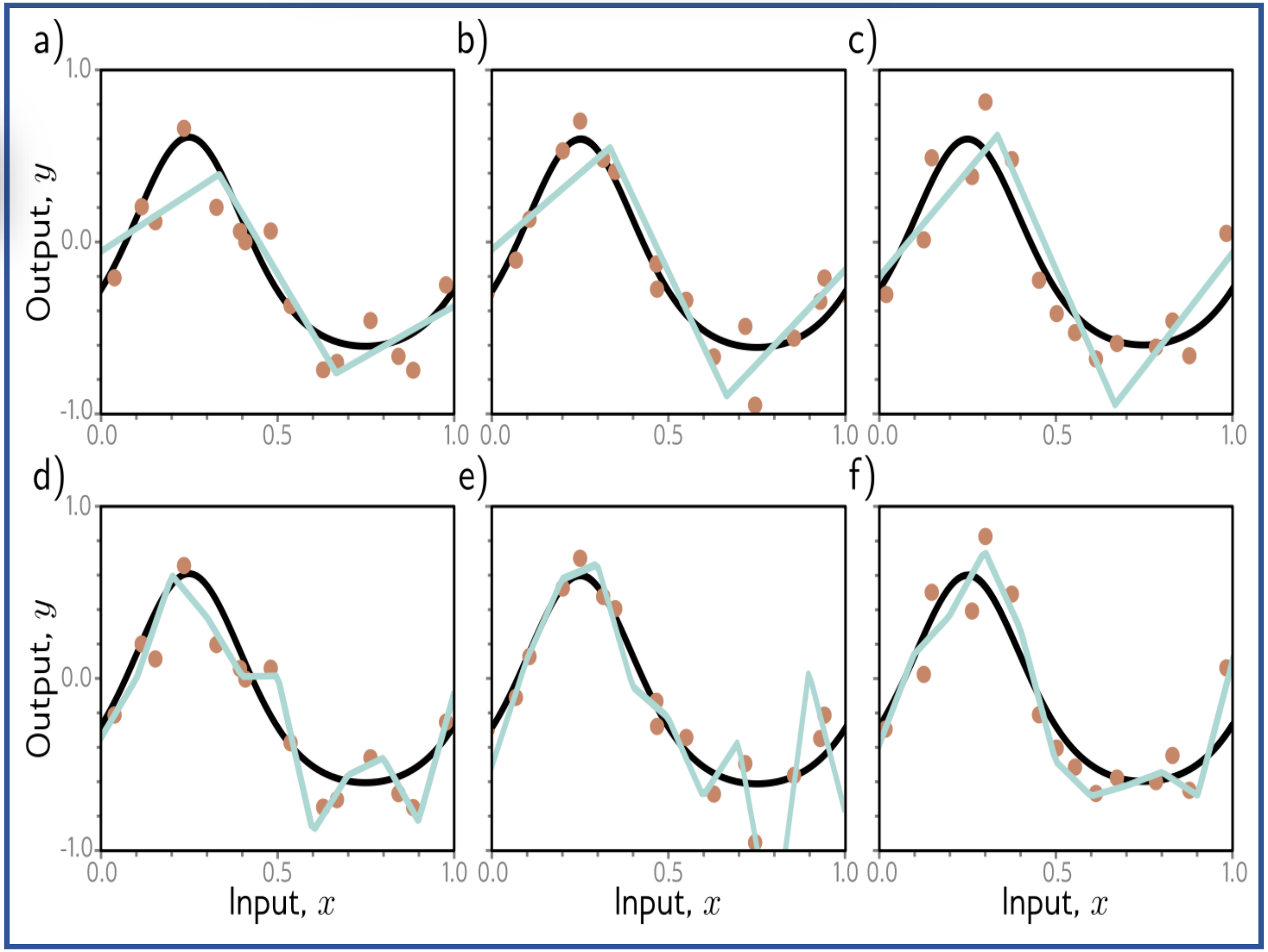
Model Performance

Overfitting

❑ Bias-Variance Trade-Off

What's this phenomenon called?



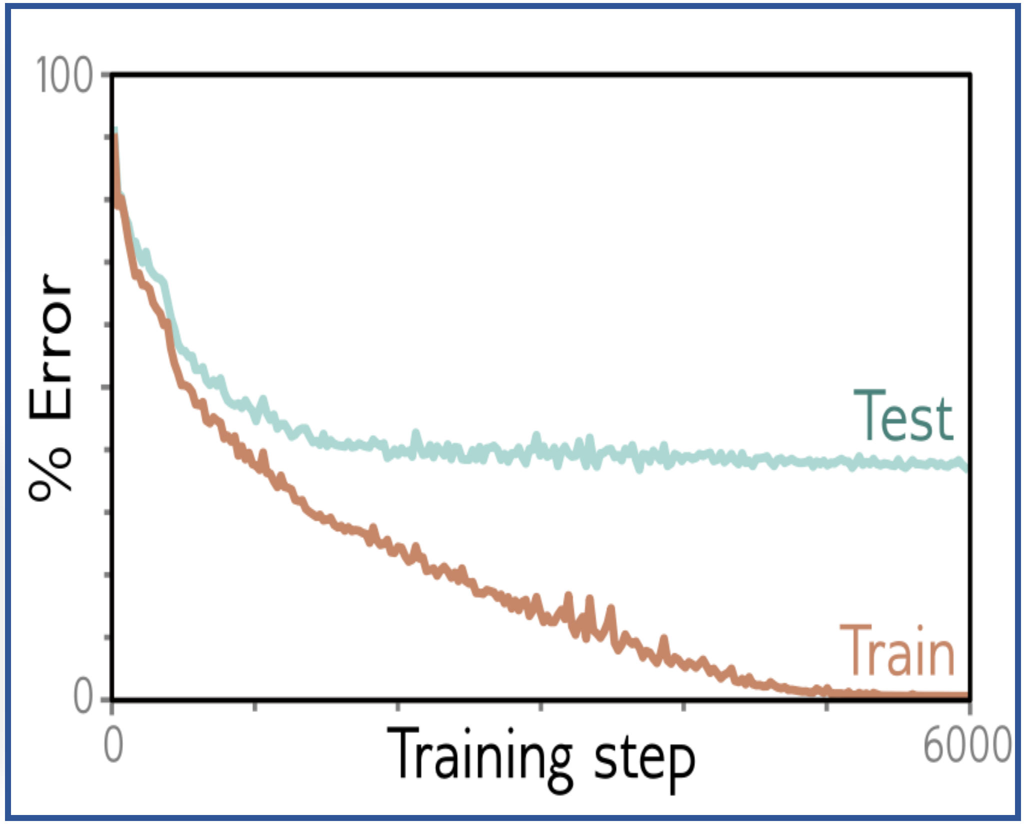


[Image from Prince 2023 UDL Book]

Model Performance

What is the main challenge?

Reduce gap between training and test performance.



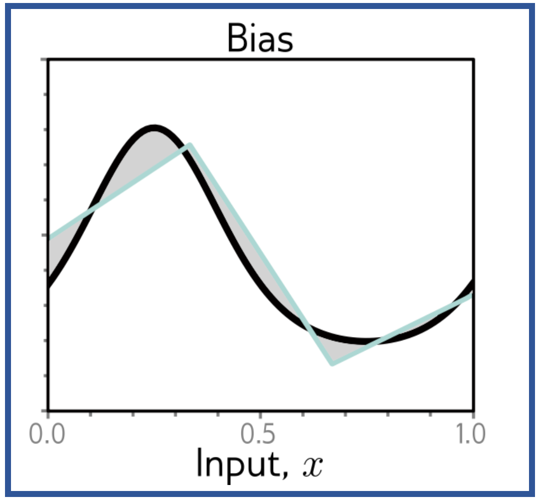
[Image from Prince 2023 UDL Book]

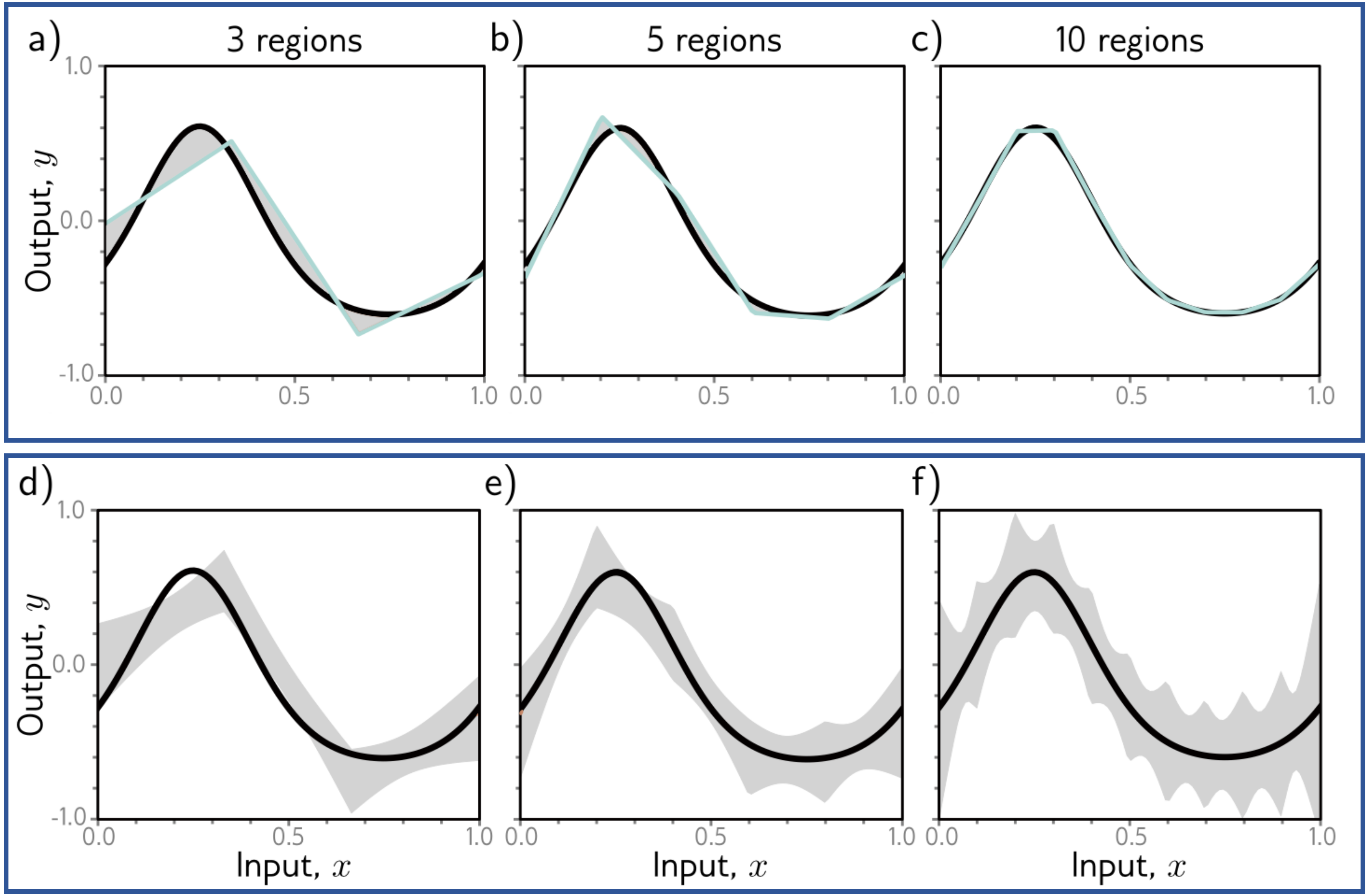
Class Announcement

- Assignment – 1 will be released today in Piazza.
  - **Deadline:** 7th Feb 11:59 pm.
  - Briefly discuss today if there is time.

Model Performance

❑ Reducing Bias





Bias-Variance Trade-off

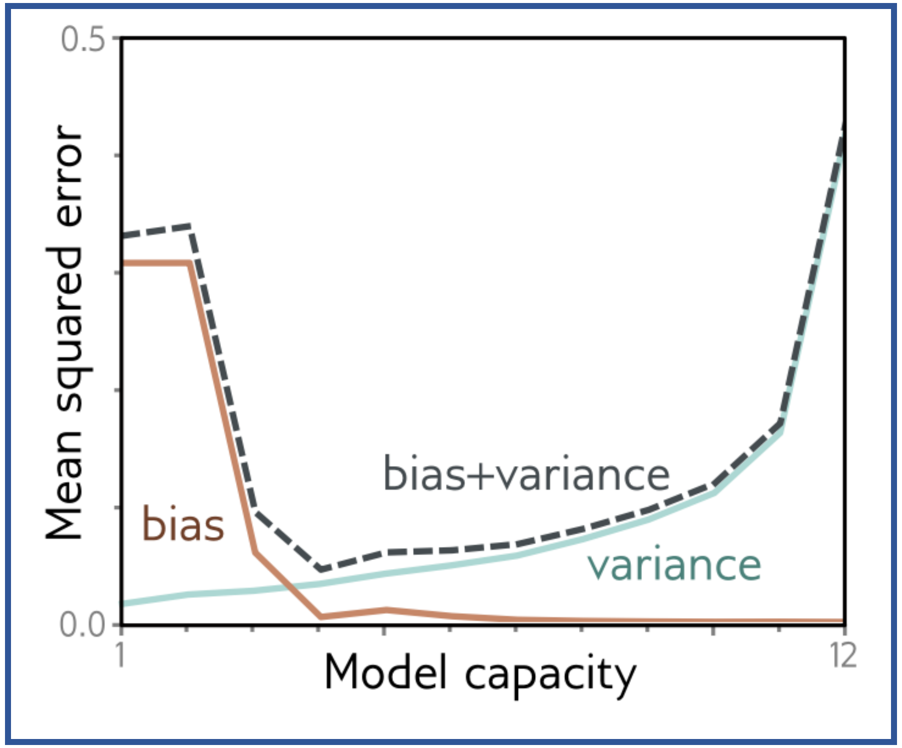
[Image from Prince 2023 UDL Book]

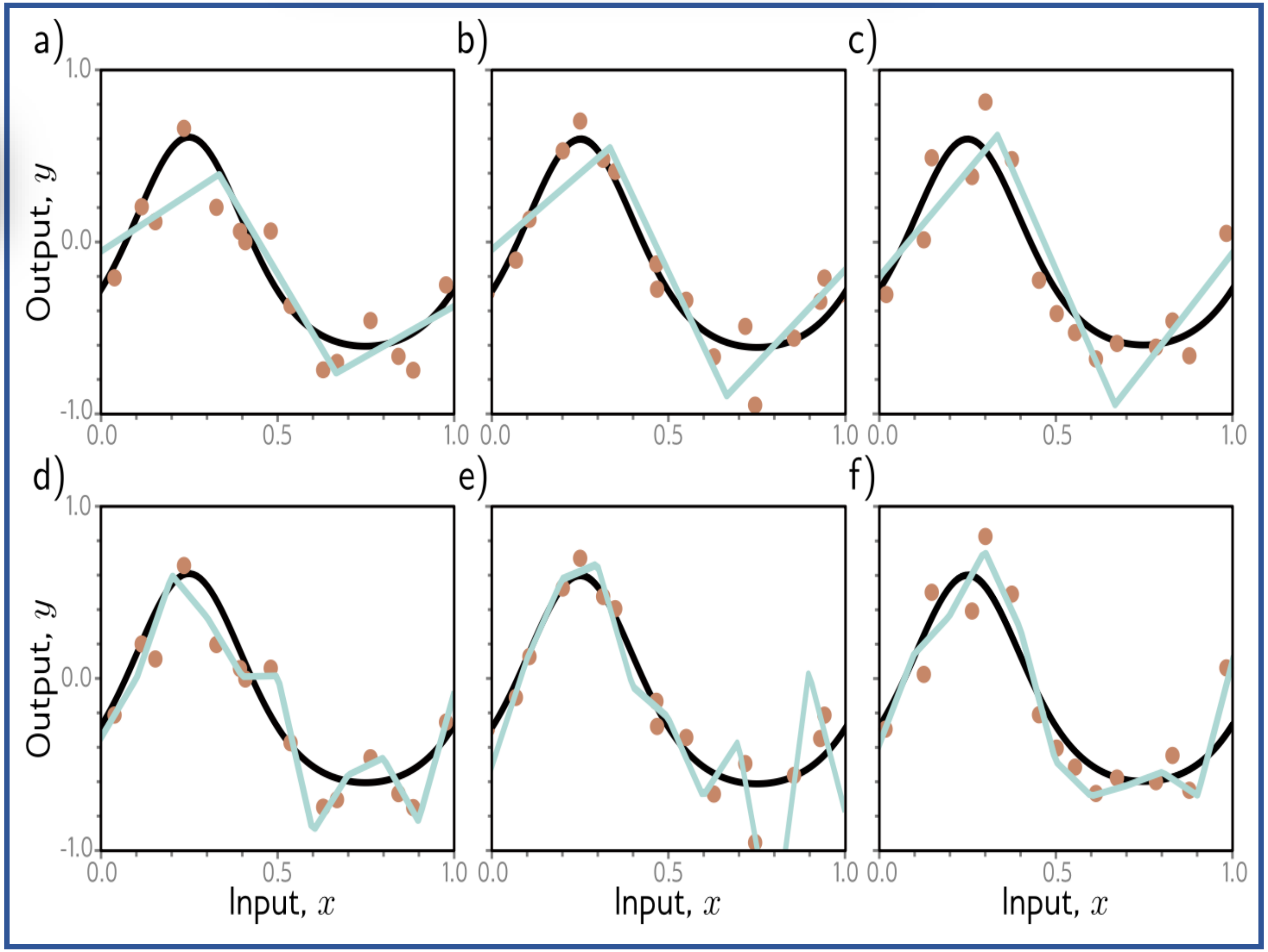
Model Performance

Overfitting

❑ Bias-Variance Trade-Off

What's this phenomenon called?



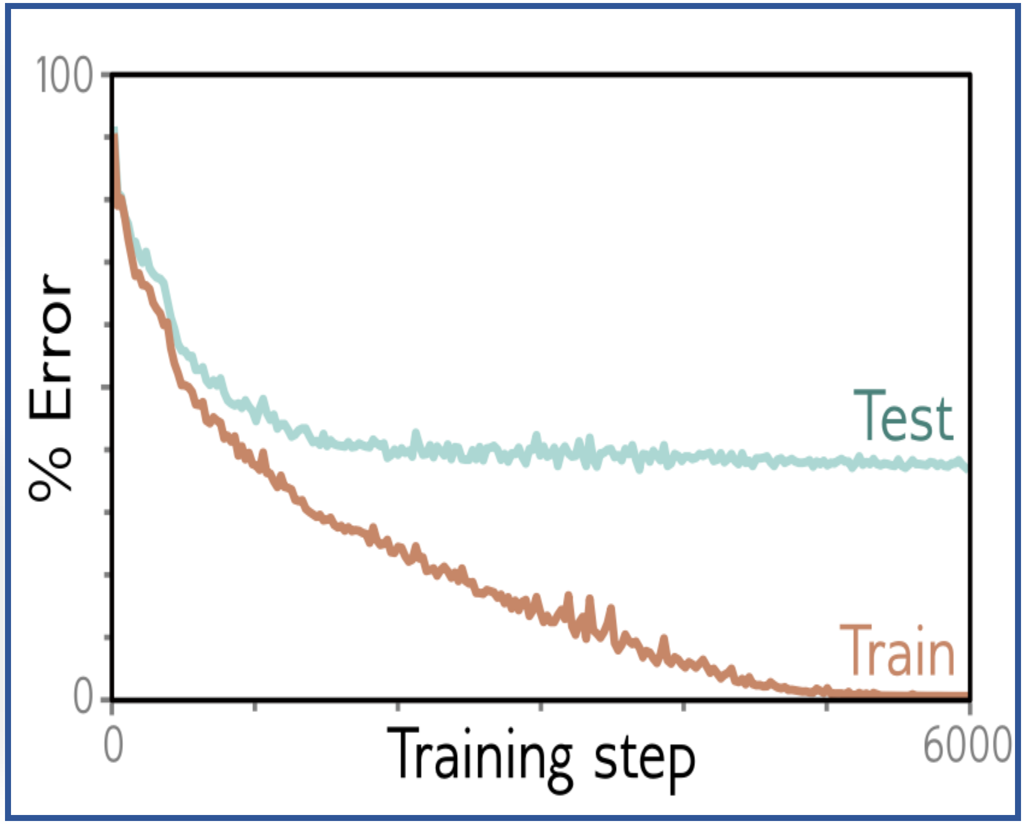


[Image from Prince 2023 UDL Book]

Model Performance

What is the main challenge?

Reduce gap between training and test performance.



[Image from Prince 2023 UDL Book]

Regularization

Regularization

- **Family of methods** that reduce the **generalization gap** between training and test performance.

- Involves adding **explicit terms** to the loss function.

Regularization

❑ Explicit Regularization Technique

$$
\begin{aligned}
\hat{\phi} &= \underset{\phi}{\operatorname{argmin}}\Big[L[\phi]\Big] \\
&= \underset{\phi}{\operatorname{argmin}}\left[\sum_{i=1}^{I} \ell_i[\mathbf{x}_i, \mathbf{y}_i]\right]
\end{aligned}
$$

- Additional regularization term:

Why & How?

$$
\hat{\phi} = \underset{\phi}{\operatorname{argmin}}\left[\sum_{i=1}^{I} \ell_i[\mathbf{x}_i, \mathbf{y}_i] + \lambda \cdot \mathrm{g}[\phi]\right]
$$

Constrained Optimization

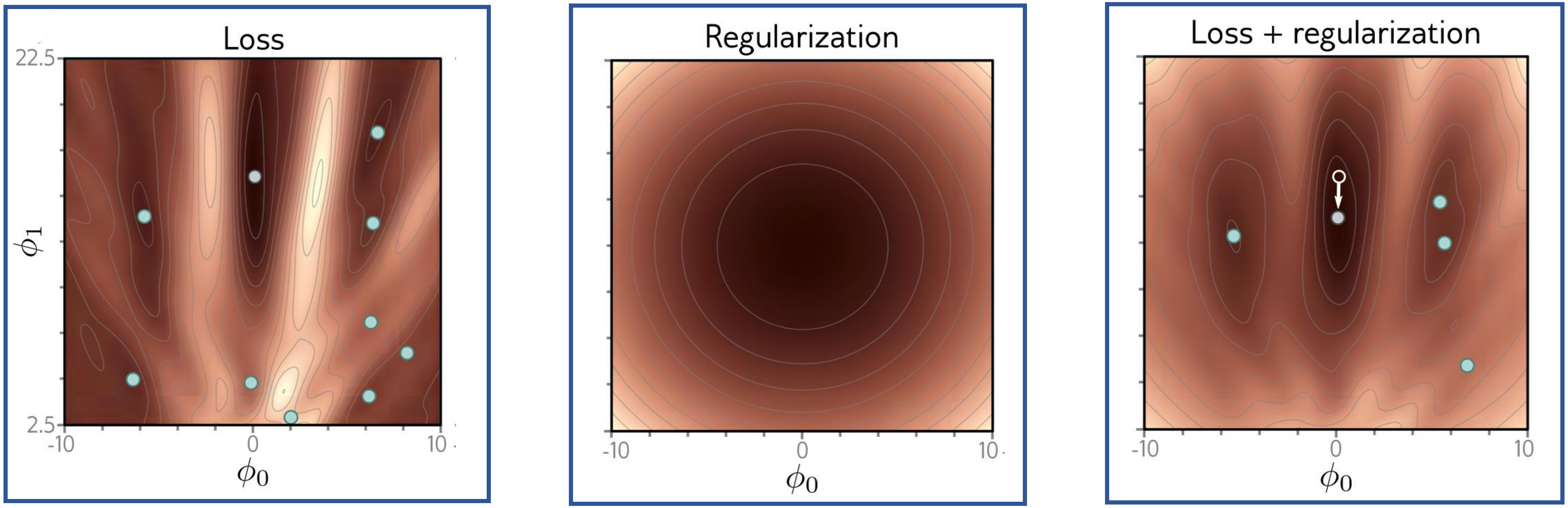
Intuition?

Takes a high value when parameter is not desired.

Regularization

❑ Explicit Regularization

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \left[\sum_{i=1}^{I} \ell_i[\mathbf{x}_i, \mathbf{y}_i] + \lambda \cdot \mathrm{g}[\phi] \right]$$



[Image from Prince 2023 UDL Book]

Regularization

❑ Explicit Regularization

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}}{\operatorname{argmin}} \left[\sum_{i=1}^{I} \ell_i[\mathbf{x}_i, \mathbf{y}_i] + \lambda \cdot \mathrm{g}[\boldsymbol{\phi}] \right]$$

- Probabilistic interpretation:

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}}{\operatorname{argmax}} \left[\prod_{i=1}^{I} Pr(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\phi}) \right]$$

Maximum Likelihood Criterion

- What if we have some knowledge about the parameters?

$$\mathit{prior}\ Pr(\boldsymbol{\phi})$$

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}}{\operatorname{argmax}} \left[\prod_{i=1}^{I} Pr(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\phi}) Pr(\boldsymbol{\phi}) \right]$$

Maximum a posteriori (MAP) Criterion

Negative Log-Likelihood (NLL)

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \left[-\log \left(\prod_{i} Pr(y_i|x_i, \phi) \cdot Pr(\phi) \right) \right]$$

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \left[\sum_{i} -\log Pr(y_i|x_i, \phi) - \log Pr(\phi) \right]$$

$$\lambda \cdot g[\phi] = -\log Pr(\phi)$$

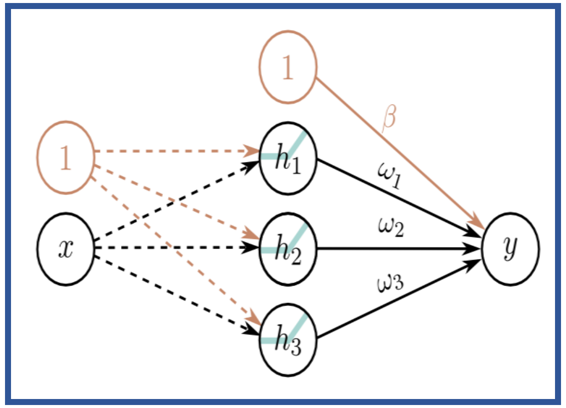
[Image from Prince 2023 UDL Book]

Regularization

❑ L2 Regularization

L2 Norm

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \left[\sum_{i=1}^{I} \ell_i[\mathbf{x}_i, \mathbf{y}_i] + \lambda \sum_{j} \phi_j^2 \right]$$



$$y = \sum_i w_i \cdot h_i + \beta$$

- Applied only to weights, not for biases.

What happens to the weights?

- Encourage smaller weights.
- Leads to smoother functions.

[Image from Prince 2023 UDL Book]

Regularization

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \left[\sum_{i=1}^{I} \ell_i[\mathbf{x}_i, \mathbf{y}_i] + \lambda \sum_j \phi_j^2 \right]$$

ScAI

❑ L2 Regularization

How avoiding overfitting help us?

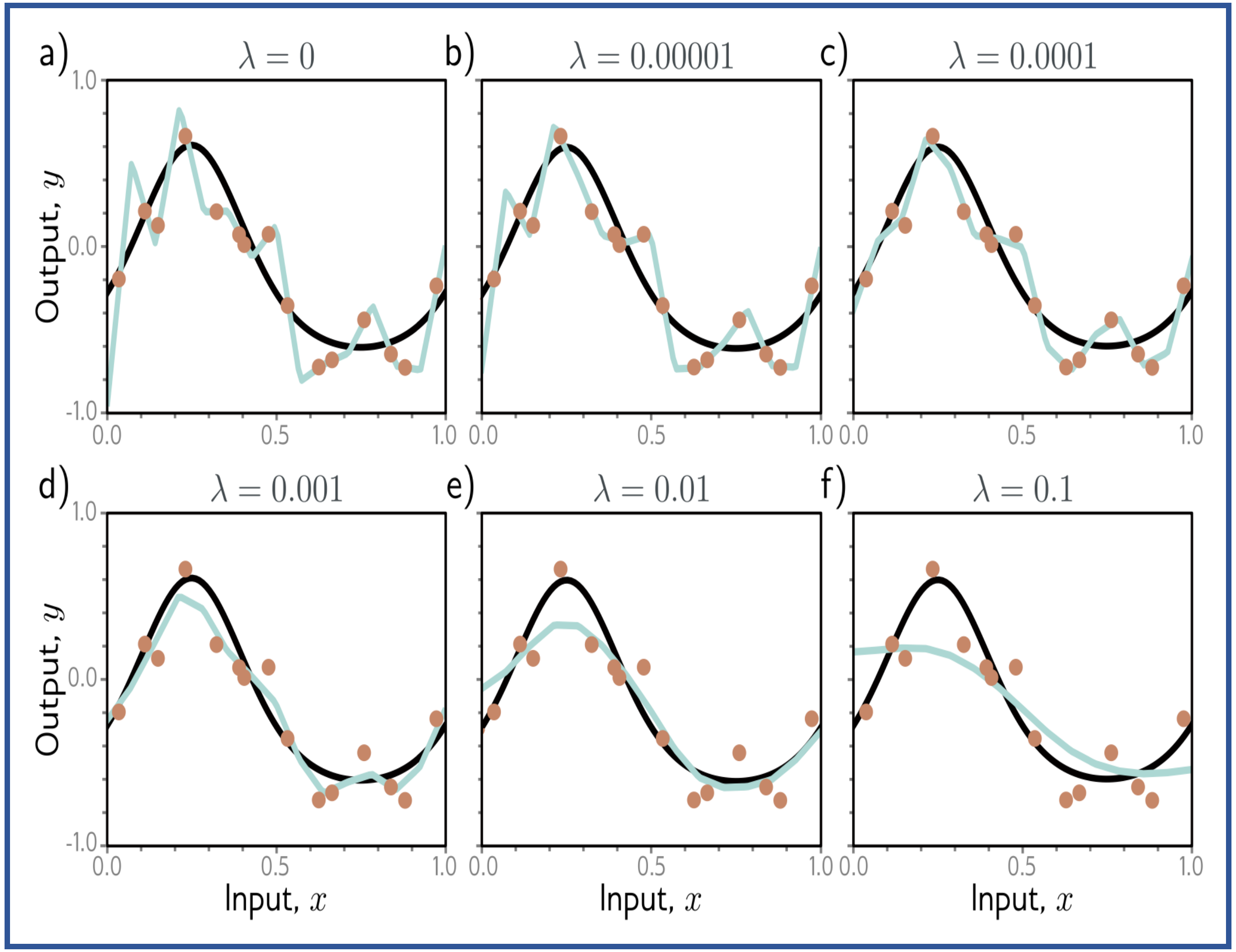
- Improves test error.
  - Noise / Bias / Variance?

Bias Error -> Overfitting -> Variance Error

  - Missing data

Complex vs Smooth function

How smooth fn help us?

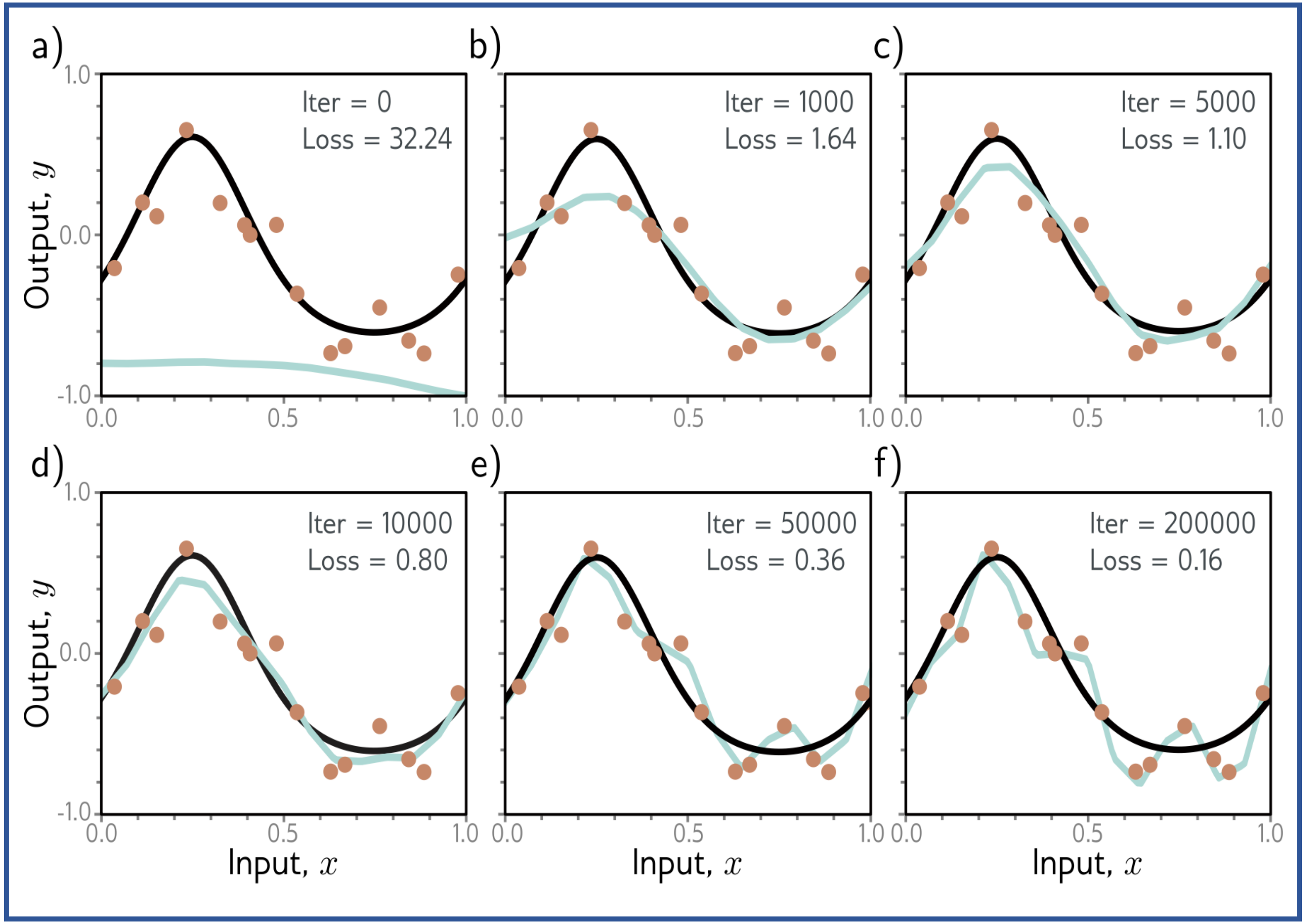


Smoother functions help alleviate overfitting

[Image from Prince 2023 UDL Book]

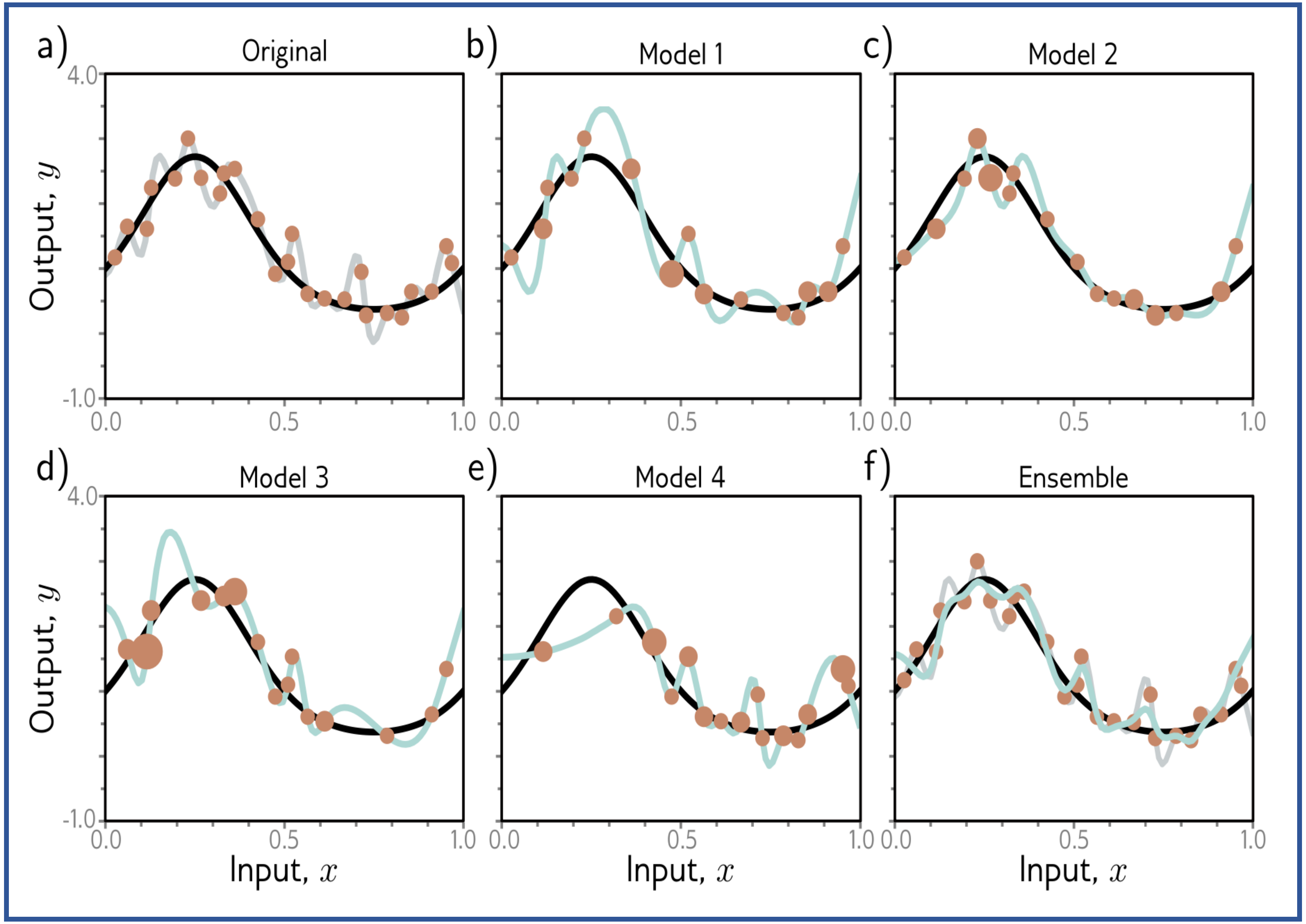
Regularization

❑ Implicit Regularization Technique: Early Stopping



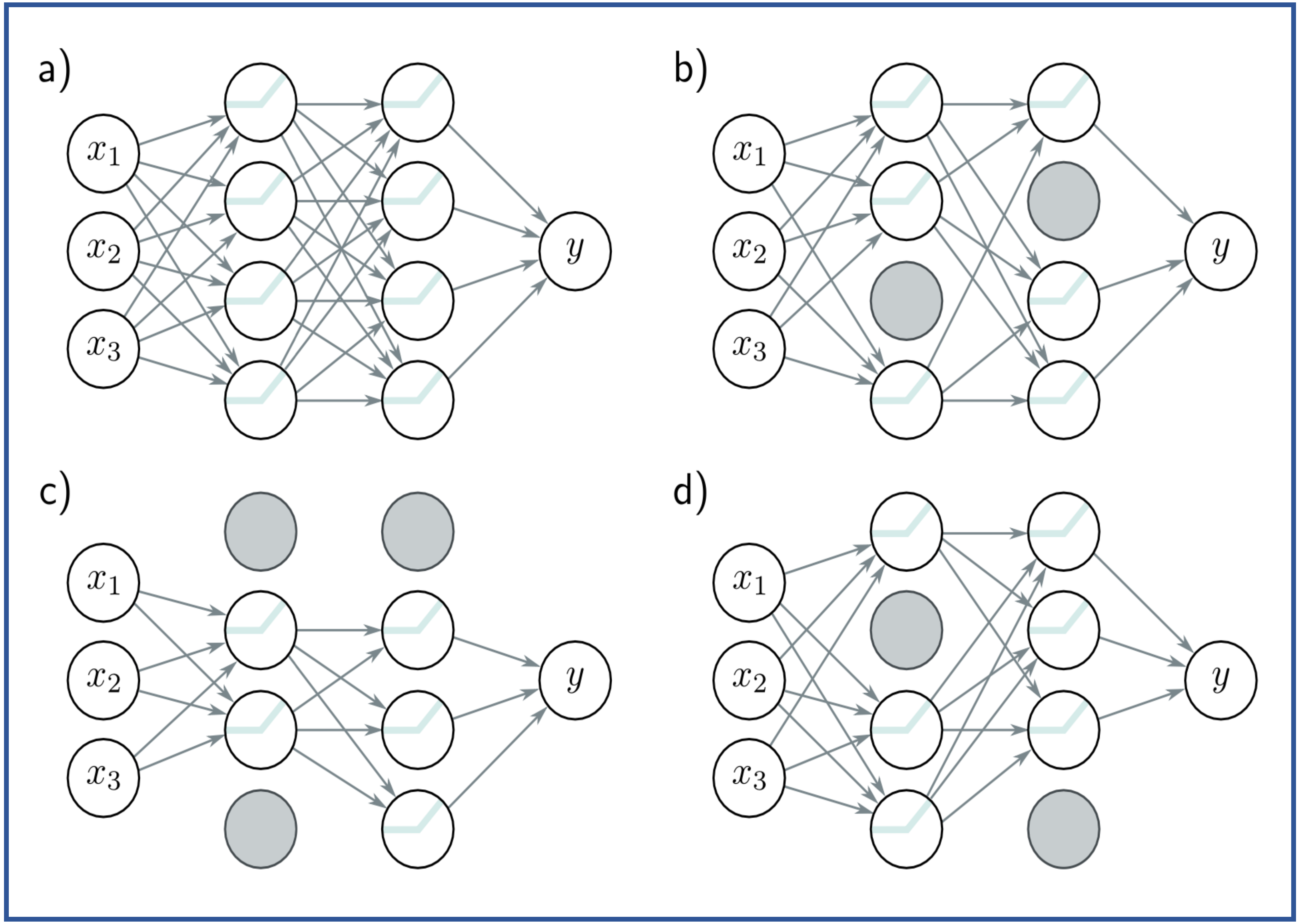
Regularization

❑ Implicit Regularization Technique: Ensembling



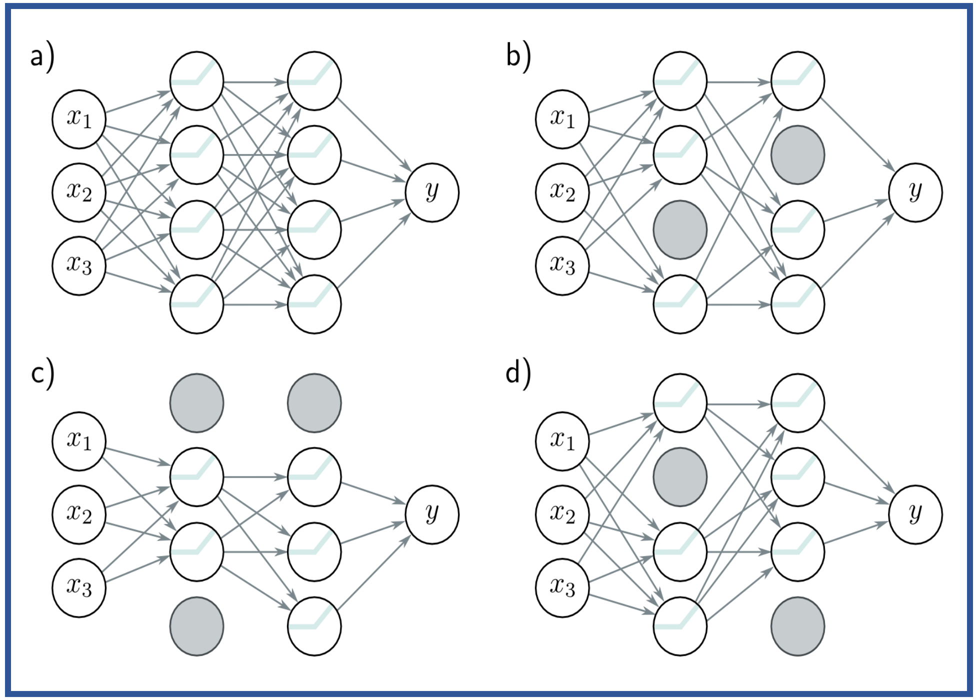
Regularization

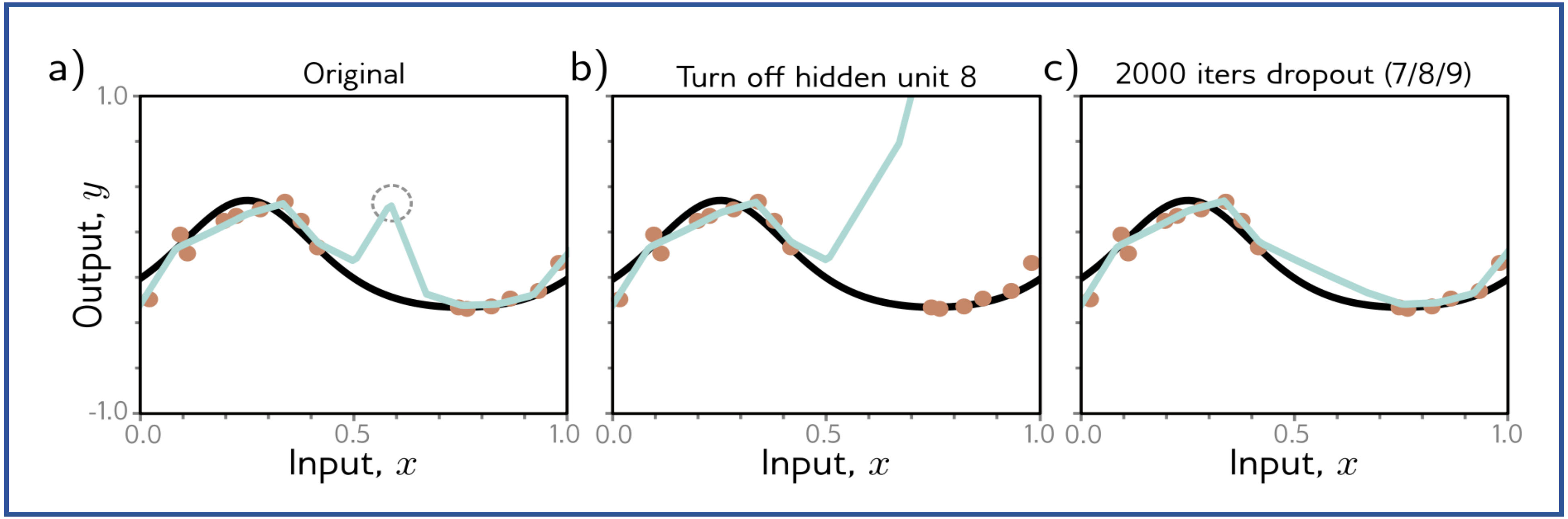
- **Implicit Regularization Technique: Dropout**



Regularization

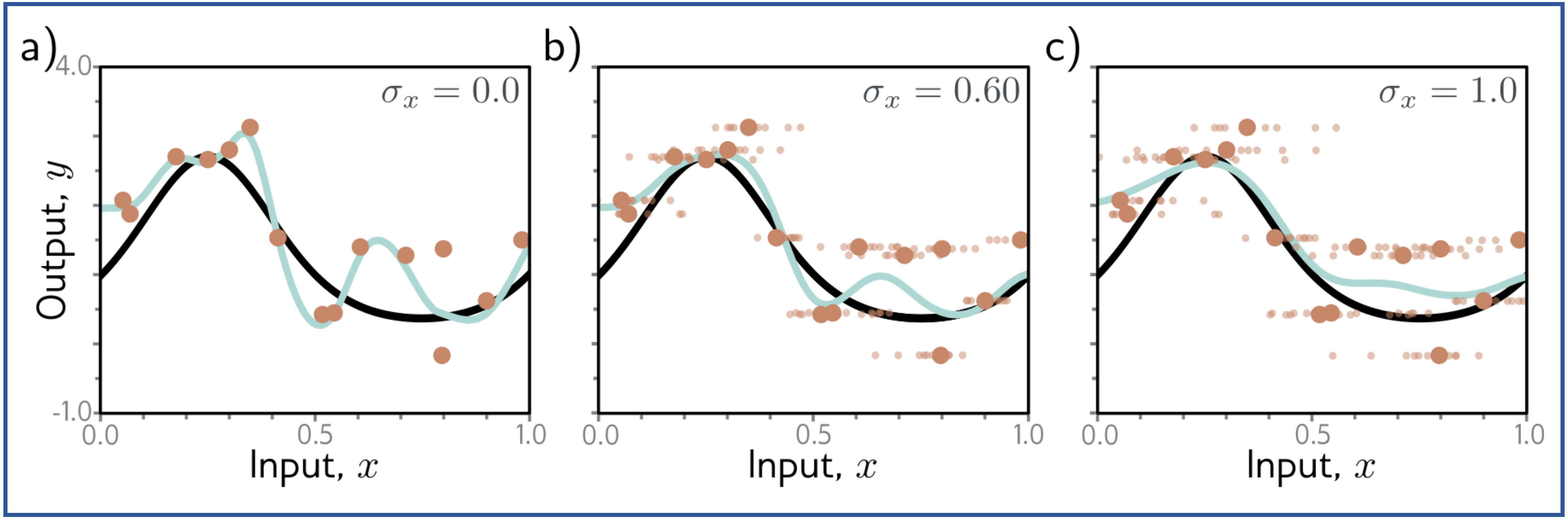
❑ Implicit Regularization Technique: Dropout





Regularization

❑ Implicit Regularization Technique: Applying Noise



ScAI

END