Bias-Variance Trade-off

Y=f(X) + e

$$Err(x) = E\left[\left(Y - \hat{f}(x)\right)^2\right]$$

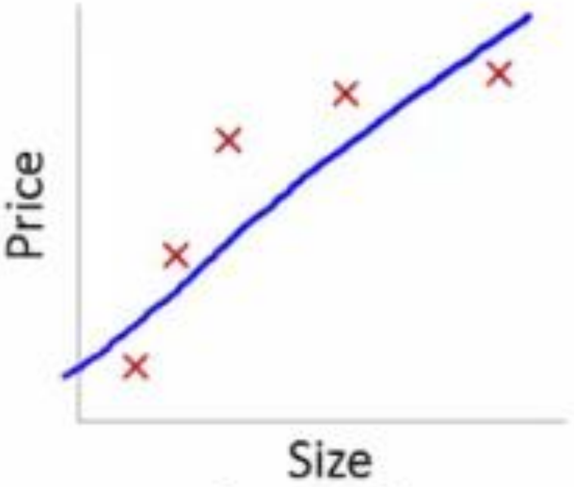
$$\begin{aligned}
\mathbb{E}[(y-\hat{f}(x))^2] &= \mathbb{E}[(f(x)+\epsilon-\hat{f}(x))^2] && (1)\\
&= \mathbb{E}[(f(x)-\hat{f}(x))^2] + \mathbb{E}[\epsilon^2] + 2\mathbb{E}[(f(x)-\hat{f}(x))\epsilon] \\
&= \mathbb{E}[(f(x)-\hat{f}(x))^2] + \underbrace{\mathbb{E}[\epsilon^2]}_{=\sigma_\epsilon^2} + 2\mathbb{E}[(f(x)-\hat{f}(x))]\underbrace{\mathbb{E}[\epsilon]}_{=0} && (2)\\
&= \mathbb{E}[(f(x)-\hat{f}(x))^2] + \sigma_\epsilon^2 && (3)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[(f(x)-\hat{f}(x))^2] &= \mathbb{E}\left[\Big((f(x)-\mathbb{E}[\hat{f}(x)])-(\hat{f}(x)-\mathbb{E}[\hat{f}(x)])\Big)^2\right] && (4)\\
&= \mathbb{E}\left[\Big(\mathbb{E}[\hat{f}(x)]-f(x)\Big)^2\right]+\mathbb{E}\left[\Big(\hat{f}(x)-\mathbb{E}[\hat{f}(x)]\Big)^2\right] \\
&\quad -2\mathbb{E}\left[\Big(f(x)-\mathbb{E}[\hat{f}(x)]\Big)\Big(\hat{f}(x)-\mathbb{E}[\hat{f}(x)]\Big)\right] && (5)\\
&= (\underbrace{\mathbb{E}[\hat{f}(x)]-f(x)}_{=\mathrm{bias}[\hat{f}(x)]})^2+\underbrace{\mathbb{E}\left[\Big(\hat{f}(x)-\mathbb{E}[\hat{f}(x)]\Big)^2\right]}_{=\mathrm{var}(\hat{f}(x))} \\
&\quad -2\Big(f(x)-\mathbb{E}[\hat{f}(x)]\Big)\,\mathbb{E}\left[\Big(\hat{f}(x)-\mathbb{E}[\hat{f}(x)]\Big)\right] && (6)\\
&= \mathrm{bias}[\hat{f}(x)]^2+\mathrm{var}(\hat{f}(x)) \\
&\quad -2\Big(f(x)-\mathbb{E}[\hat{f}(x)]\Big)\Big(\mathbb{E}[\hat{f}(x)]-\mathbb{E}[\hat{f}(x)]\Big) && (7)\\
&= \mathrm{bias}[\hat{f}(x)]^2+\mathrm{var}(\hat{f}(x)) && (8)
\end{aligned}
$$

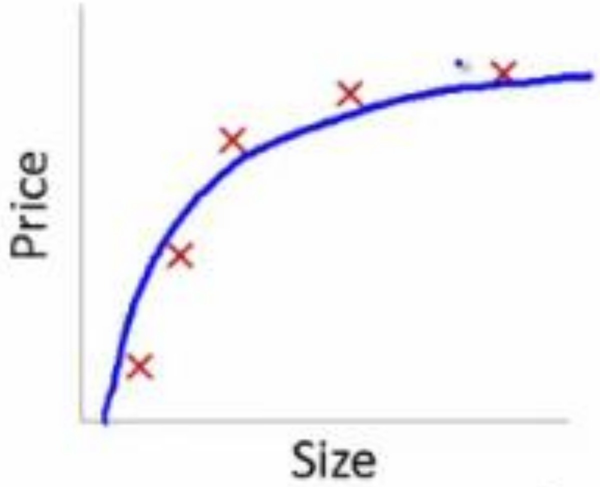
Generalization Error

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

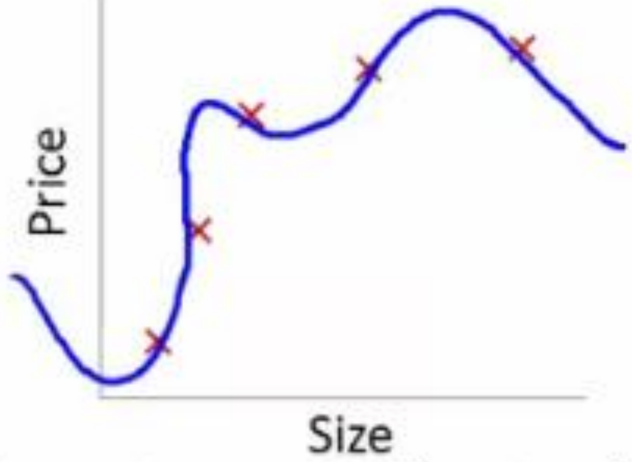
$$Err(x) = \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{Irreducible\ Error}$$



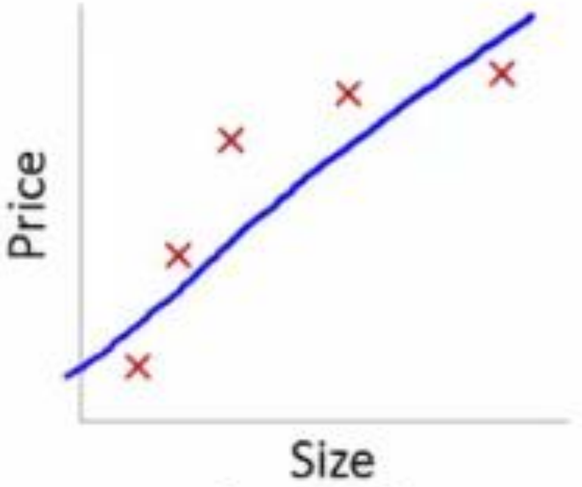
$$\theta_0 + \theta_1 x$$



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

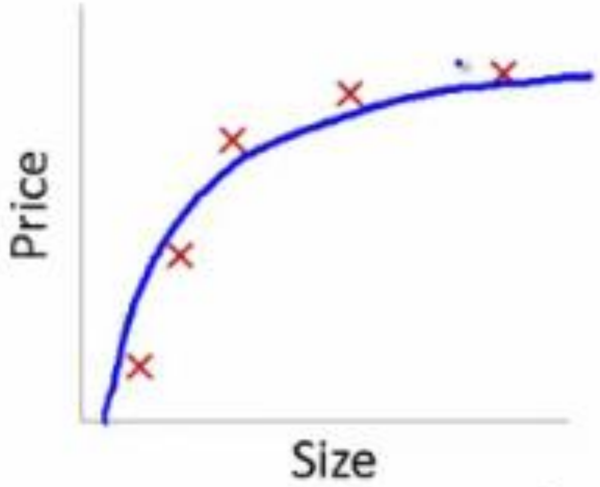


$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$



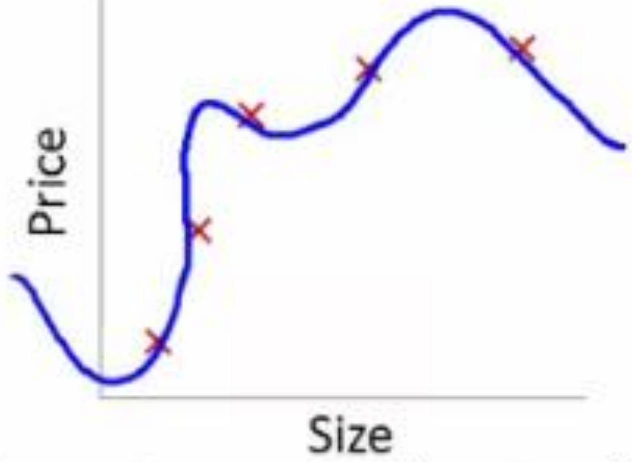
$$\theta_0 + \theta_1 x$$

High bias
(underfit)



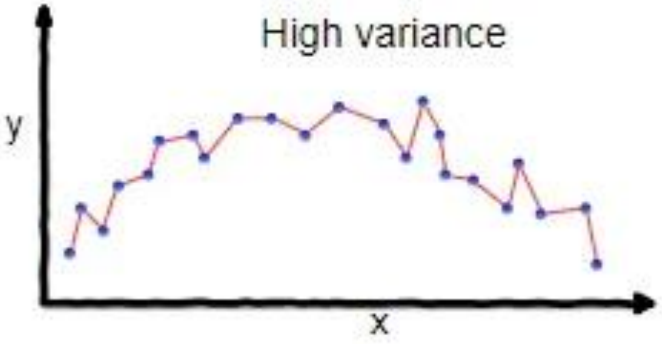
$$\theta_0 + \theta_1 x + \theta_2 x^2$$

"Just right"

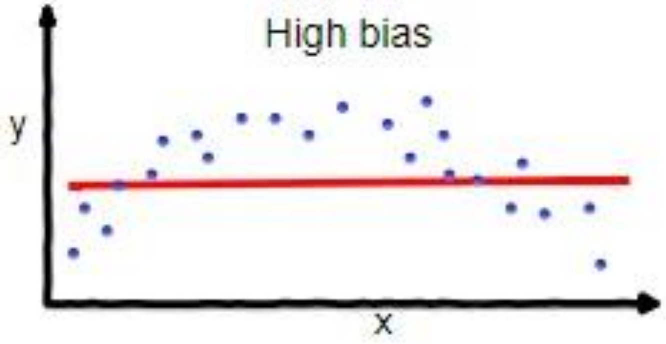


$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

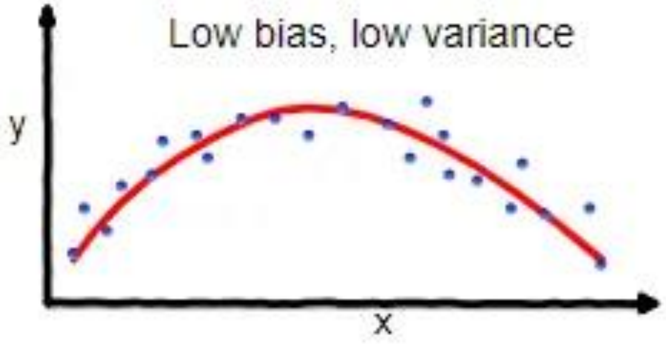
High variance
(overfit)



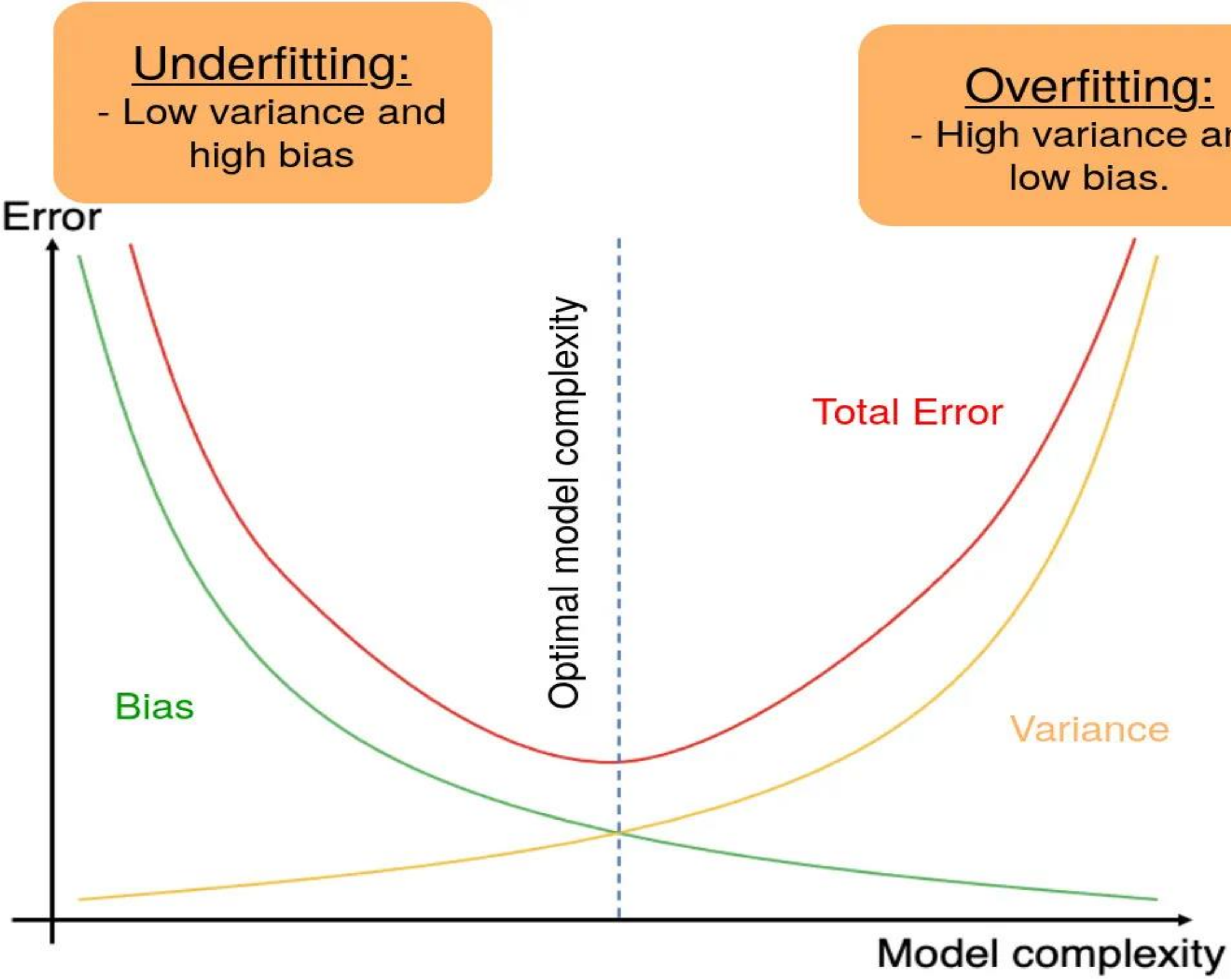
overfitting

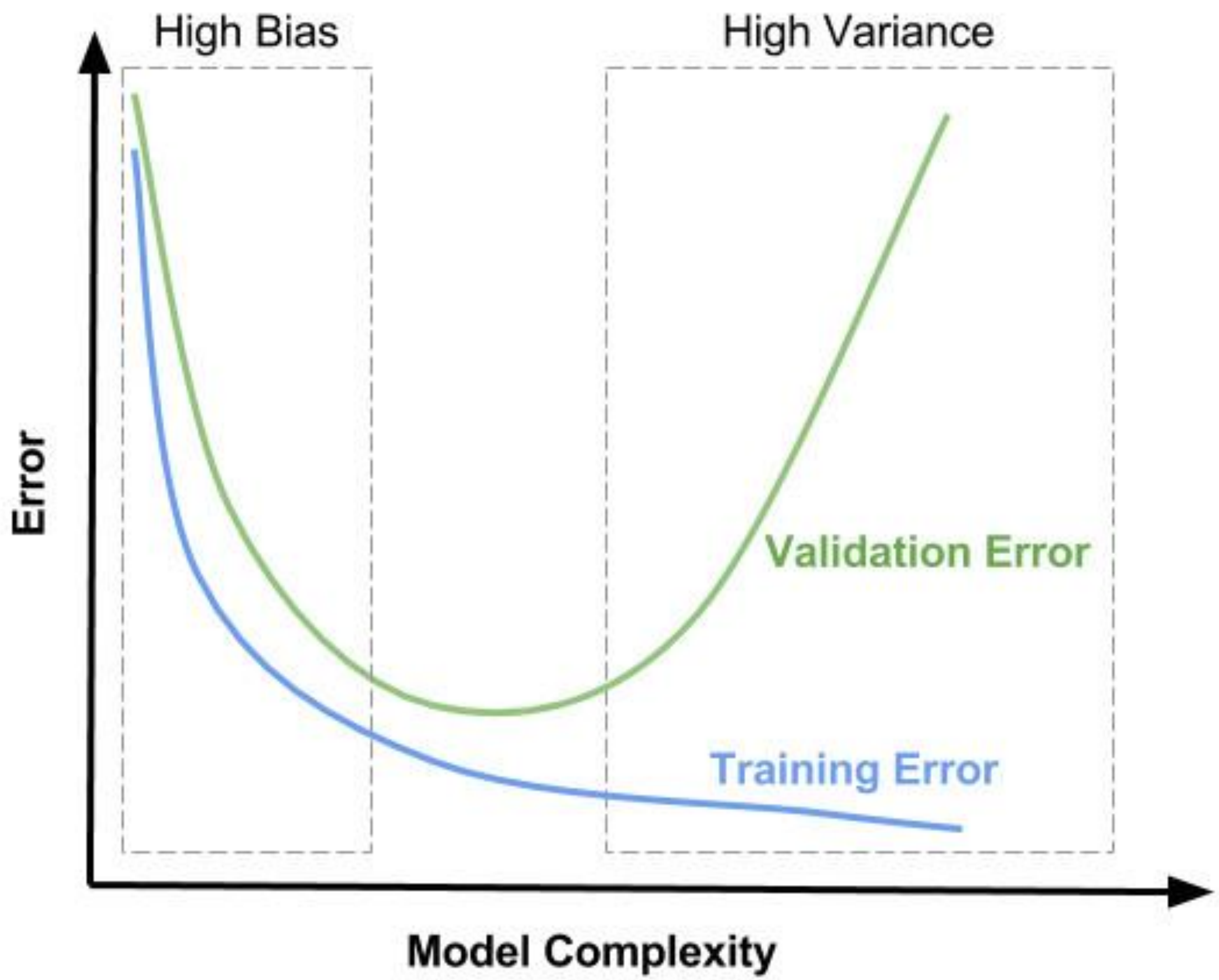


underfitting



Good balance

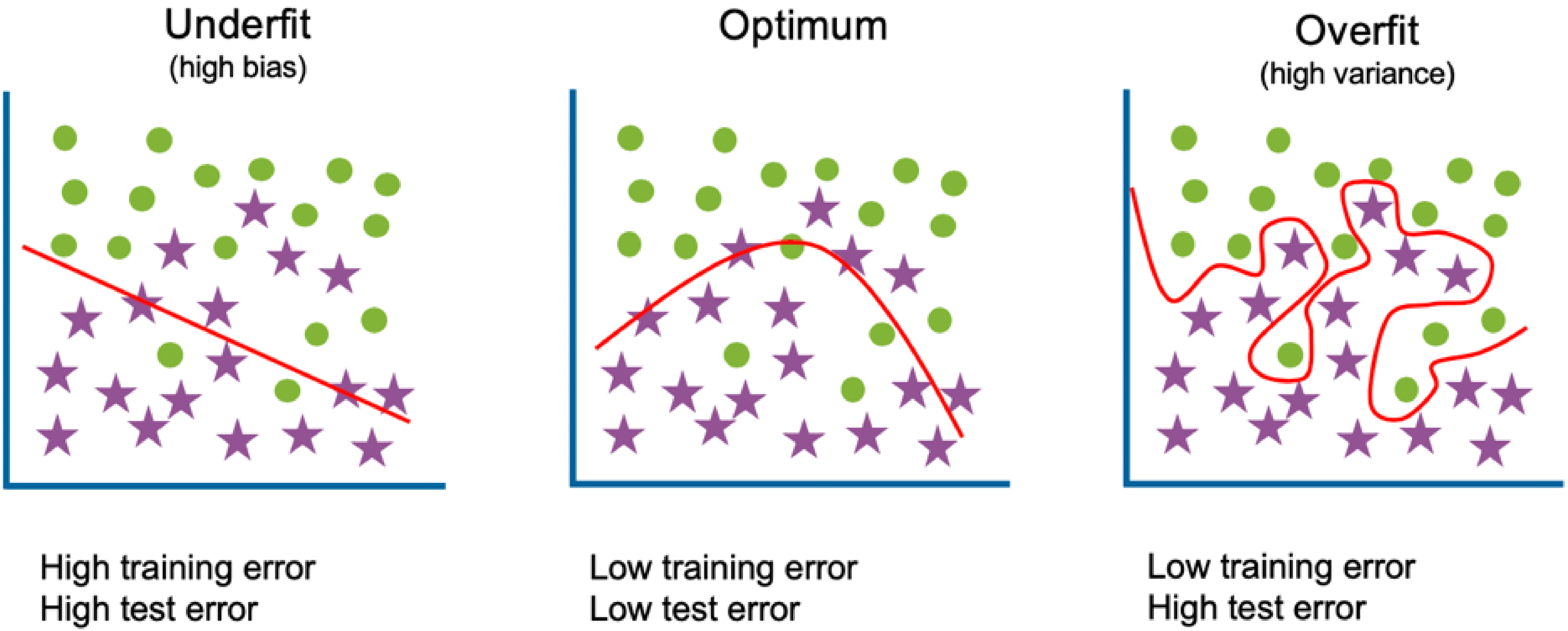

Underfitting:
- Low variance and high bias
Overfitting:
- High variance and low bias.
Error
Optimal model complexity
Total Error
Bias
Variance
Model complexity

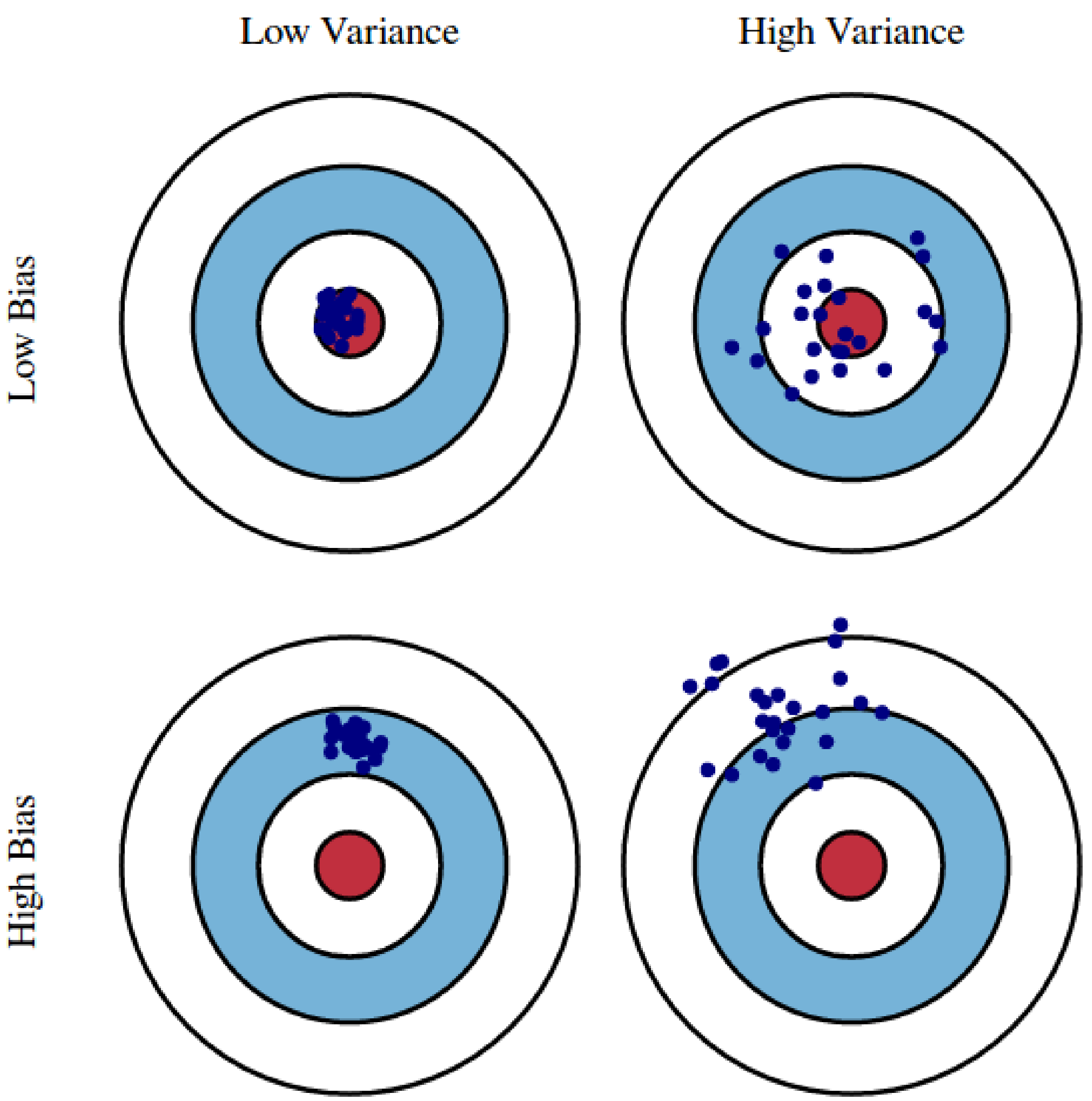

High Bias
High Variance
Error
Validation Error
Training Error
Model Complexity

Why do overfitting and underfitting occur?

Only model complexity?

Any good recipe for machine learning to avoid underfitting and overfitting?


Underfit
(high bias)
Optimum
Overfit
(high variance)
High training error
High test error
Low training error
Low test error
Low training error
High test error


Low Variance
High Variance
Low Bias
High Bias