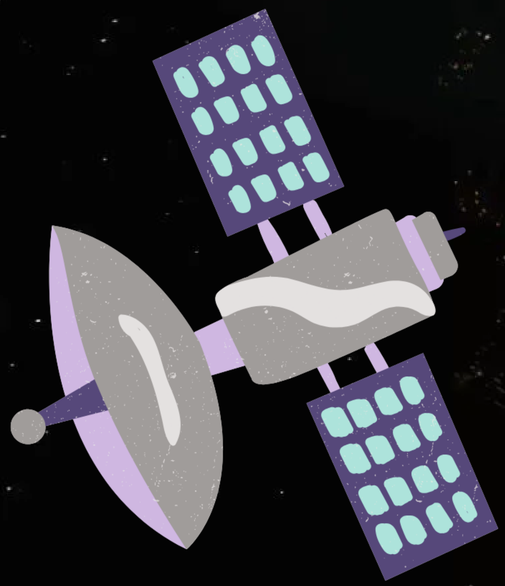
Over-fitting Vs Under-fitting



TUNSA

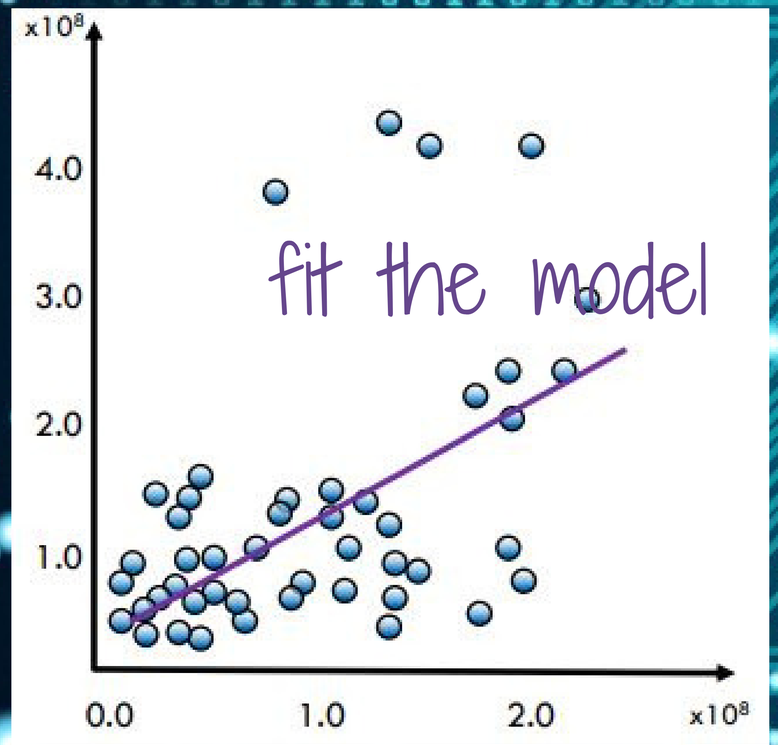
Tunisian Space Association

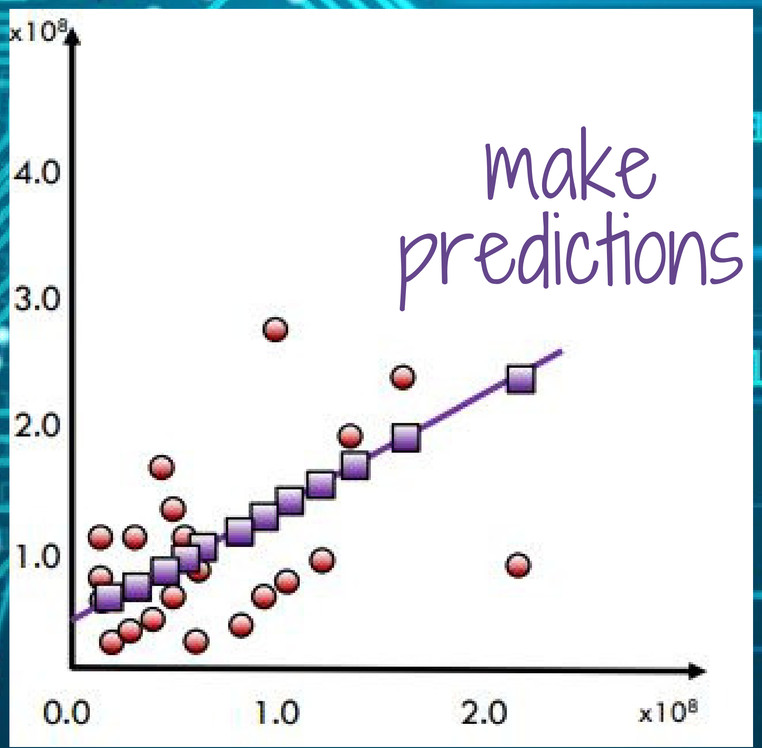
الجمعية التونسية للفضاء

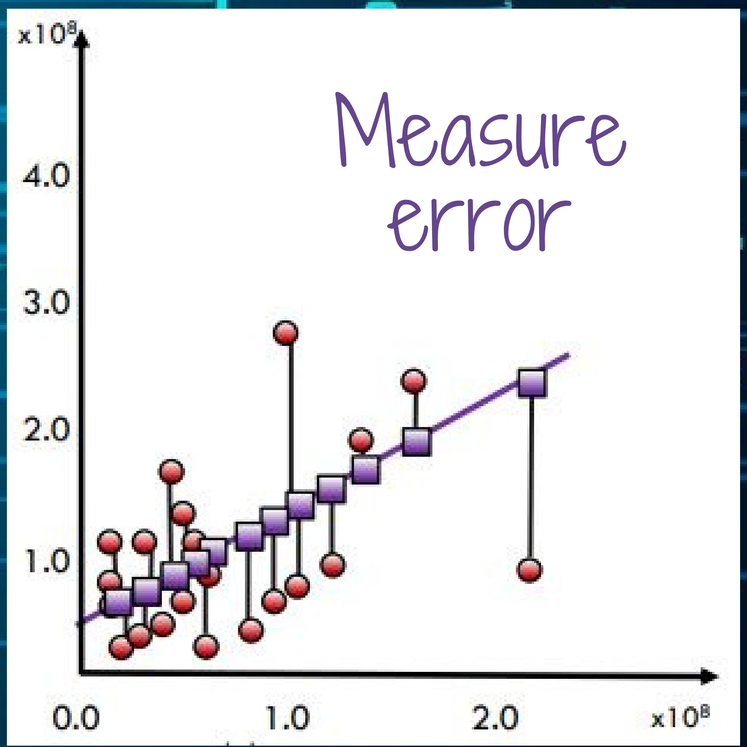

REMINDER

Training Data

Test Data

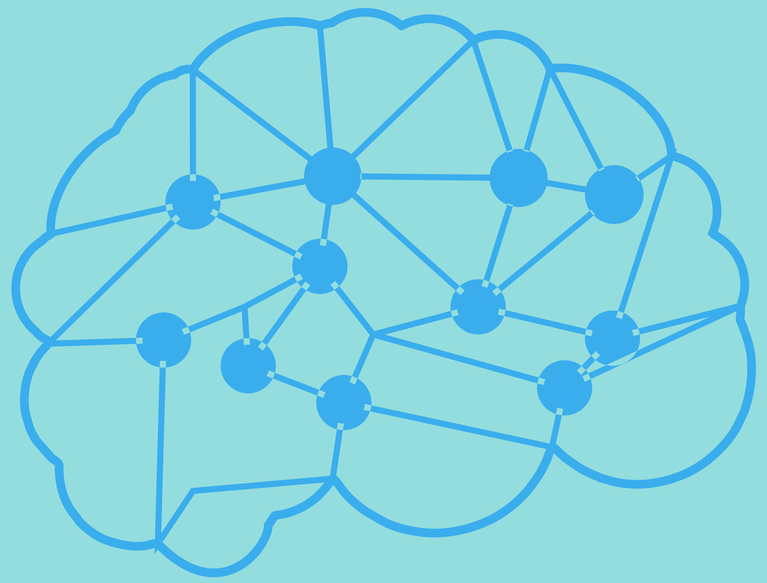




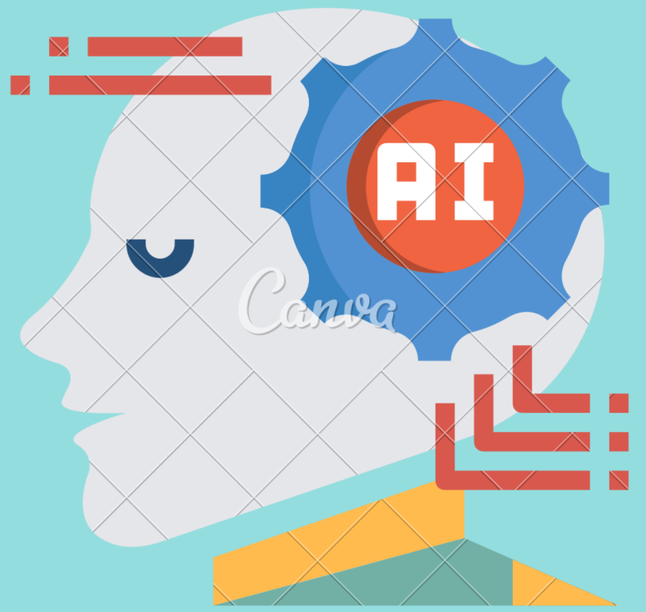


How Well Does the Model Generalize?

Bias-Variance trade-off



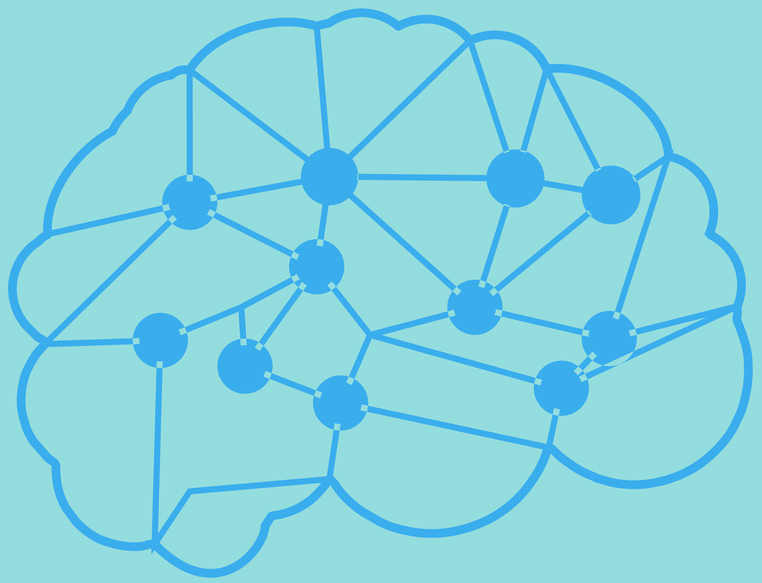
In supervised learning, the prediction error e is composed of the **bias**, the **variance**, and the **irreducible** part.
Bias refers to simplifying assumptions made to learn target function easily.
Variance refers to the sensitivity of the model to changes in the training data.
the goal of parameterization is to achieve a low bias and low variance trade-off.

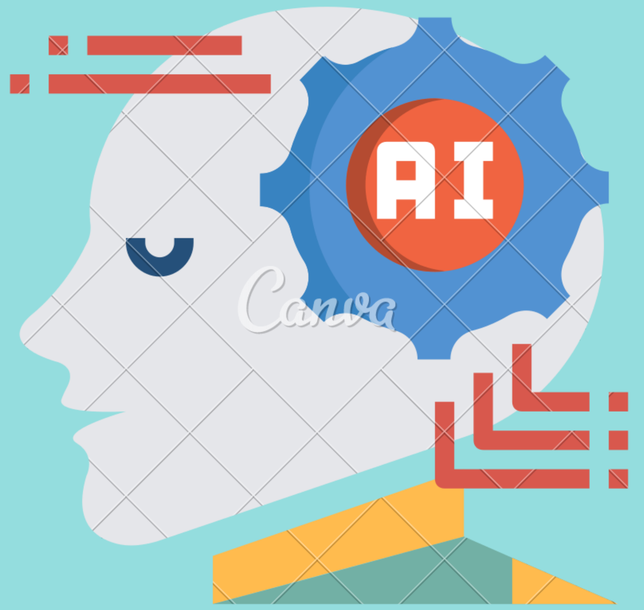


what's Bias?

Error between average model prediction and ground truth.
The Bias of the estimated function tells us the capacity of the underlying model to predict the value.

$$Bias = E(Y) - Y$$



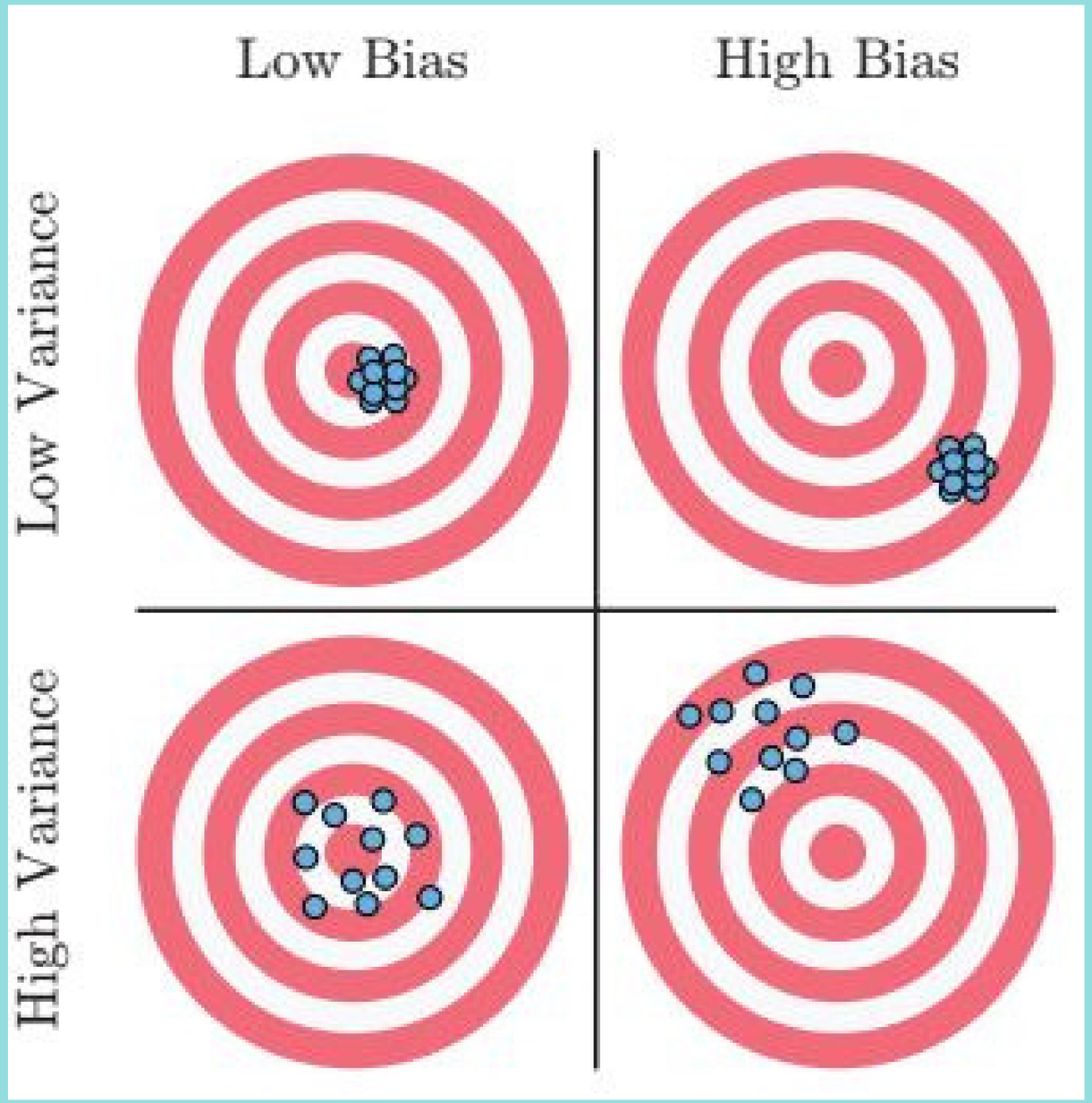


High Bias:

OVERLY-SIMPLIFIED MODEL

High Variance:

OVERLY-COMPLEX MODEL

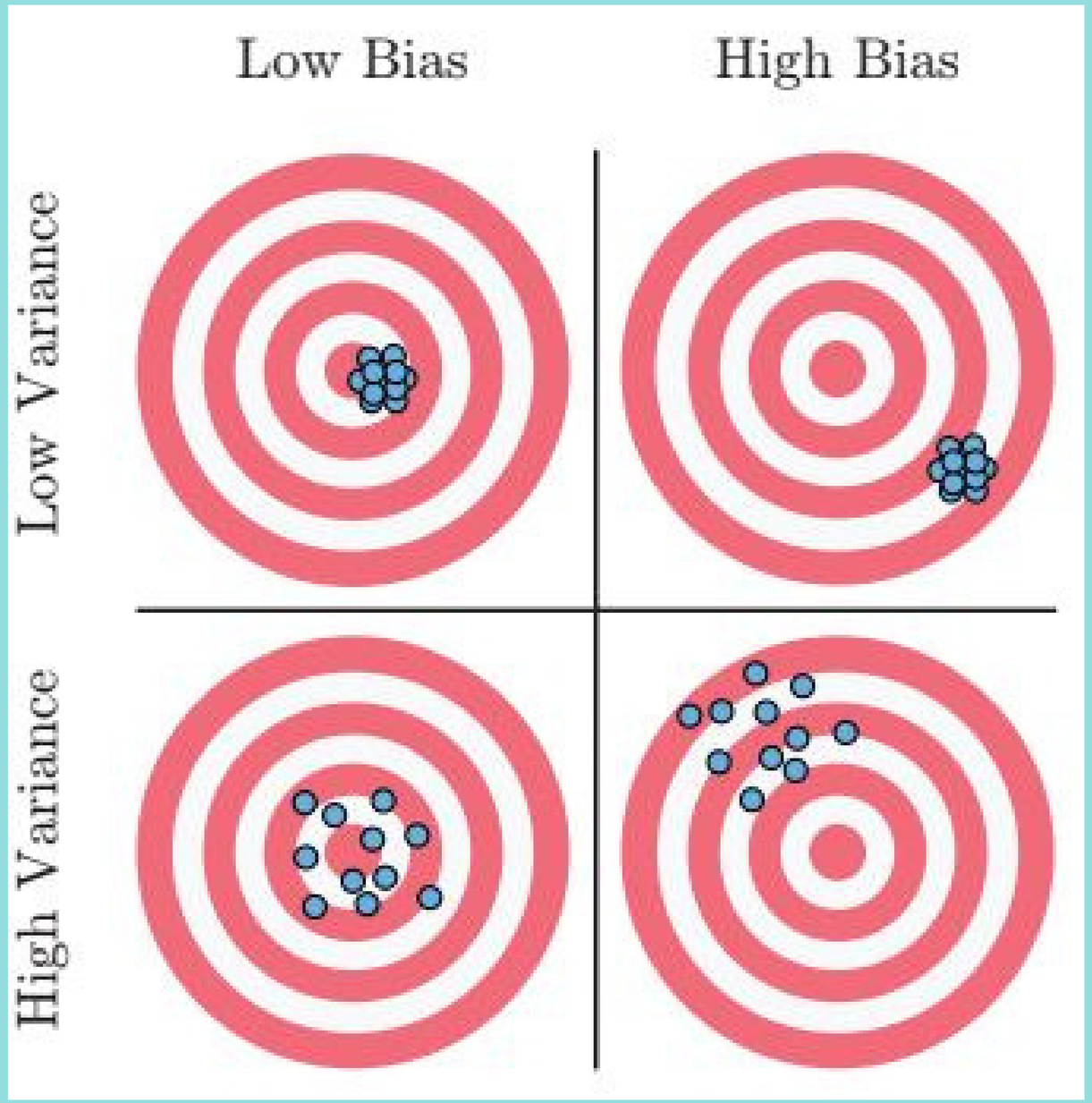


High Bias:

OVERLY-SIMPLIFIED MODEL

High Variance:

OVERLY-COMPLEX MODEL

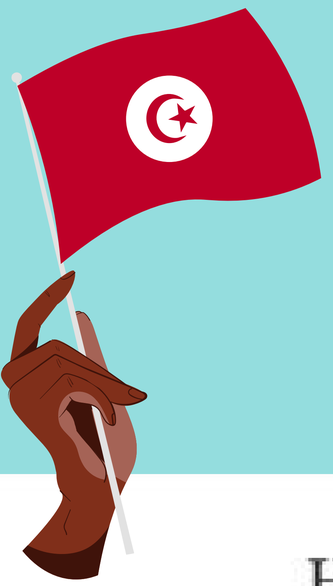


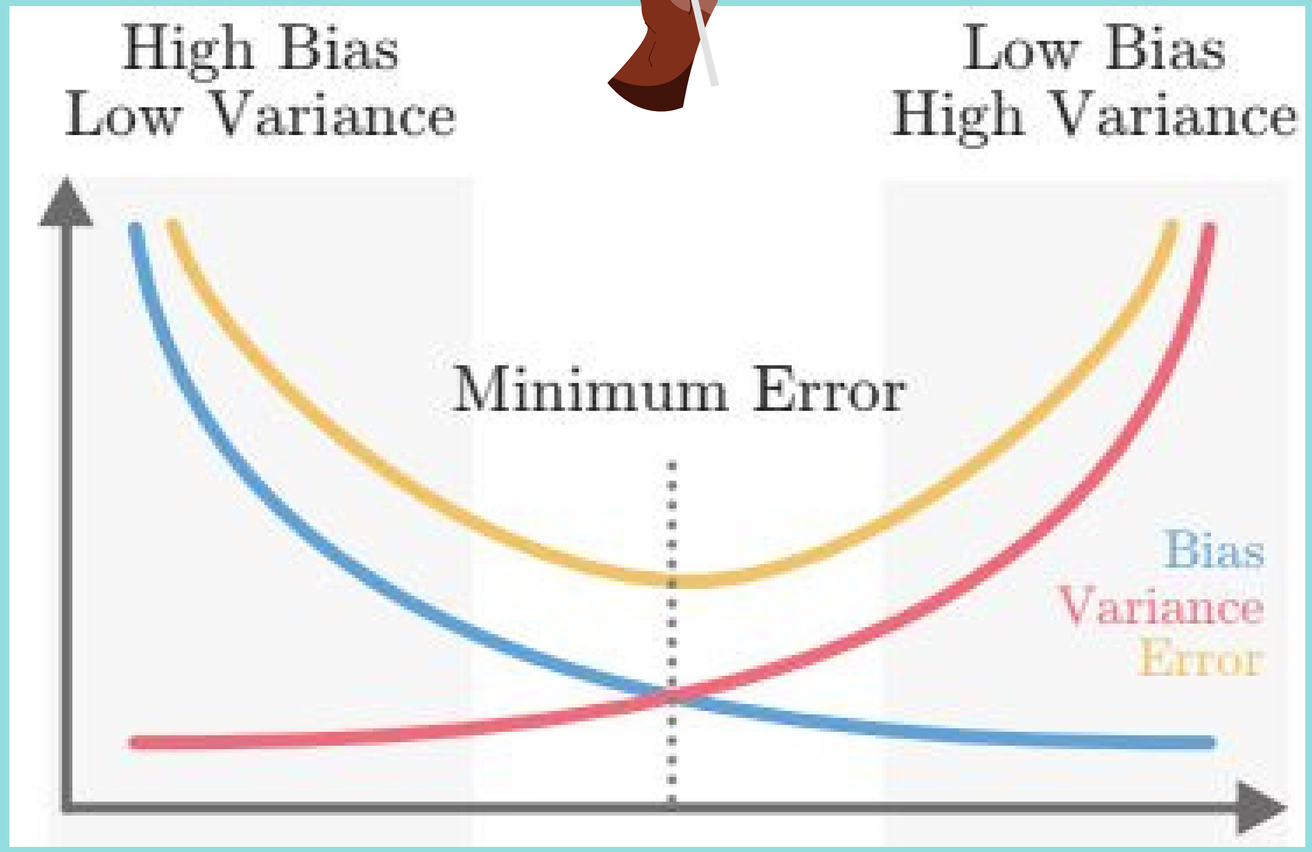
Increasing bias reduces variance and vice-versa

Error =
bias^2 + variance +irreducible error

The best model is where the error is reduced.

Compromise between bias and variance

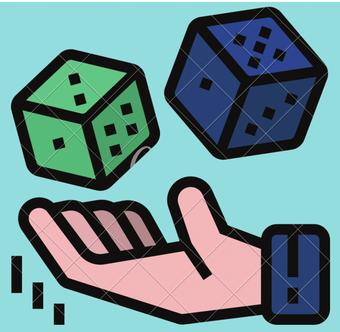


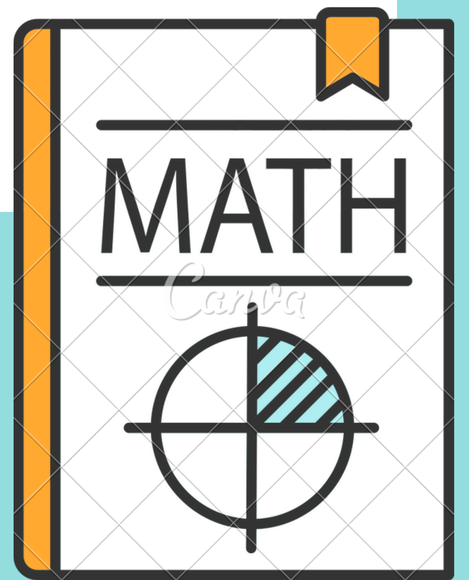


The Bias–Variance Decomposition

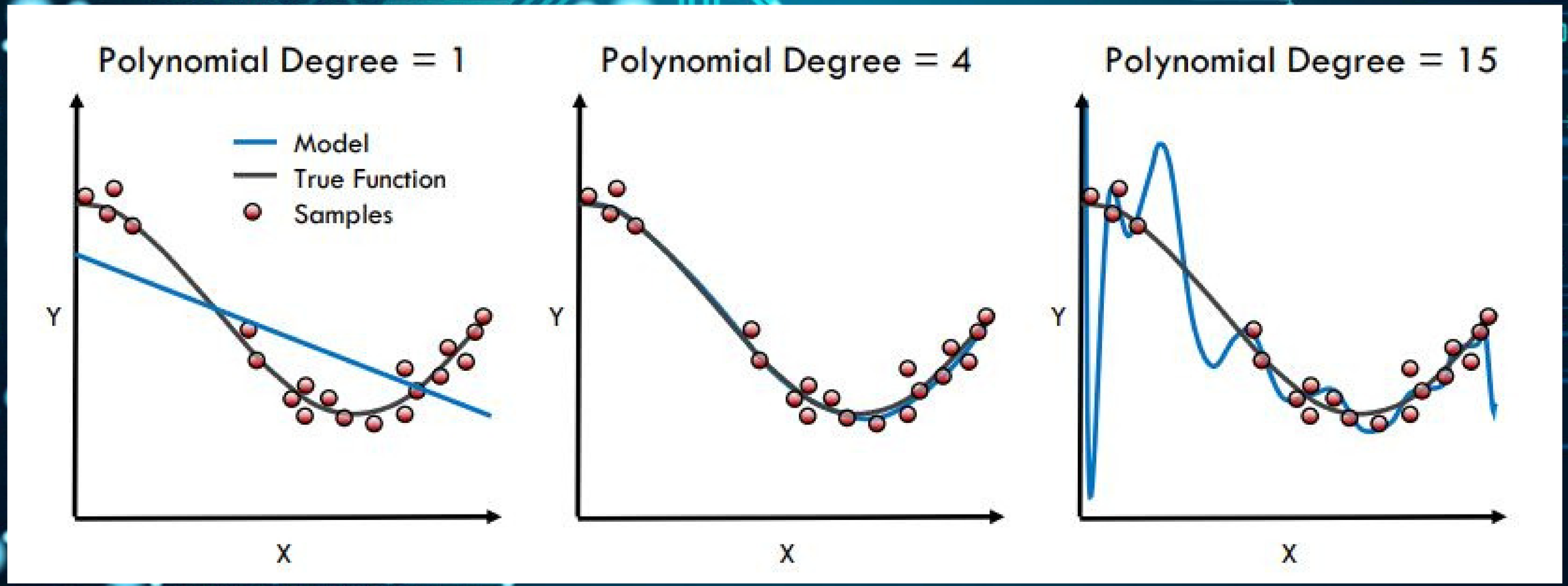
if we assume that $Y = f(X) + \varepsilon$ where $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2$, we can derive an expression for the expected prediction error of a regression fit $\hat{f}(X)$ at an input point $X = x_0$, using squared-error loss:

$$\begin{aligned}
\mathrm{Err}(x_0) &= E[(Y - \hat{f}(x_0))^2 | X = x_0] \\
&= \sigma_\varepsilon^2 + [\mathrm{E}\hat{f}(x_0) - f(x_0)]^2 + E[\hat{f}(x_0) - \mathrm{E}\hat{f}(x_0)]^2 \\
&= \sigma_\varepsilon^2 + \mathrm{Bias}^2(\hat{f}(x_0)) + \mathrm{Var}(\hat{f}(x_0)) \\
&= \text{Irreducible Error} + \mathrm{Bias}^2 + \mathrm{Variance}.
\end{aligned}$$





How Well Does the Model Generalize?



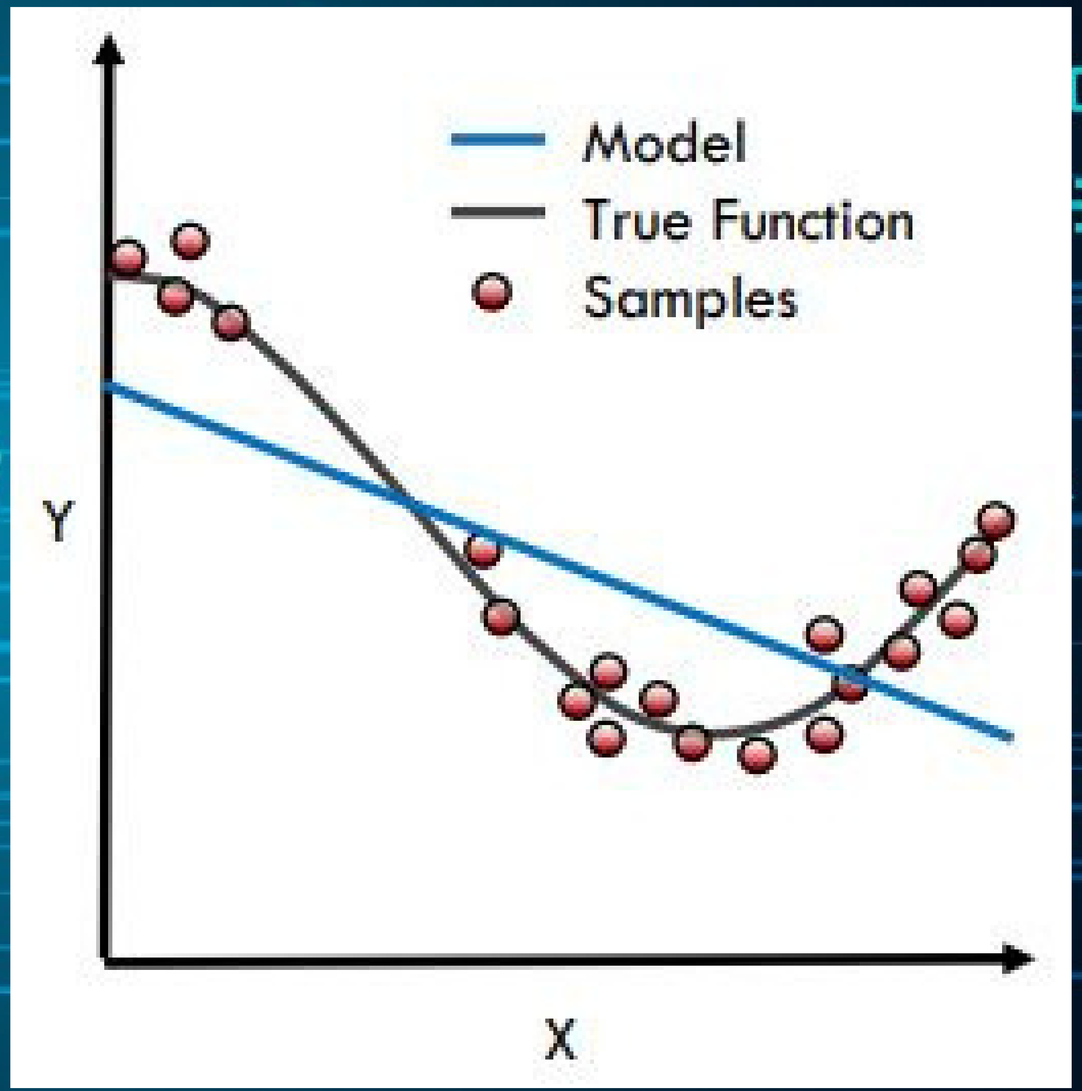
Under-fitting :

poor in training

&

poor at predicting

(High Bias - low variance)



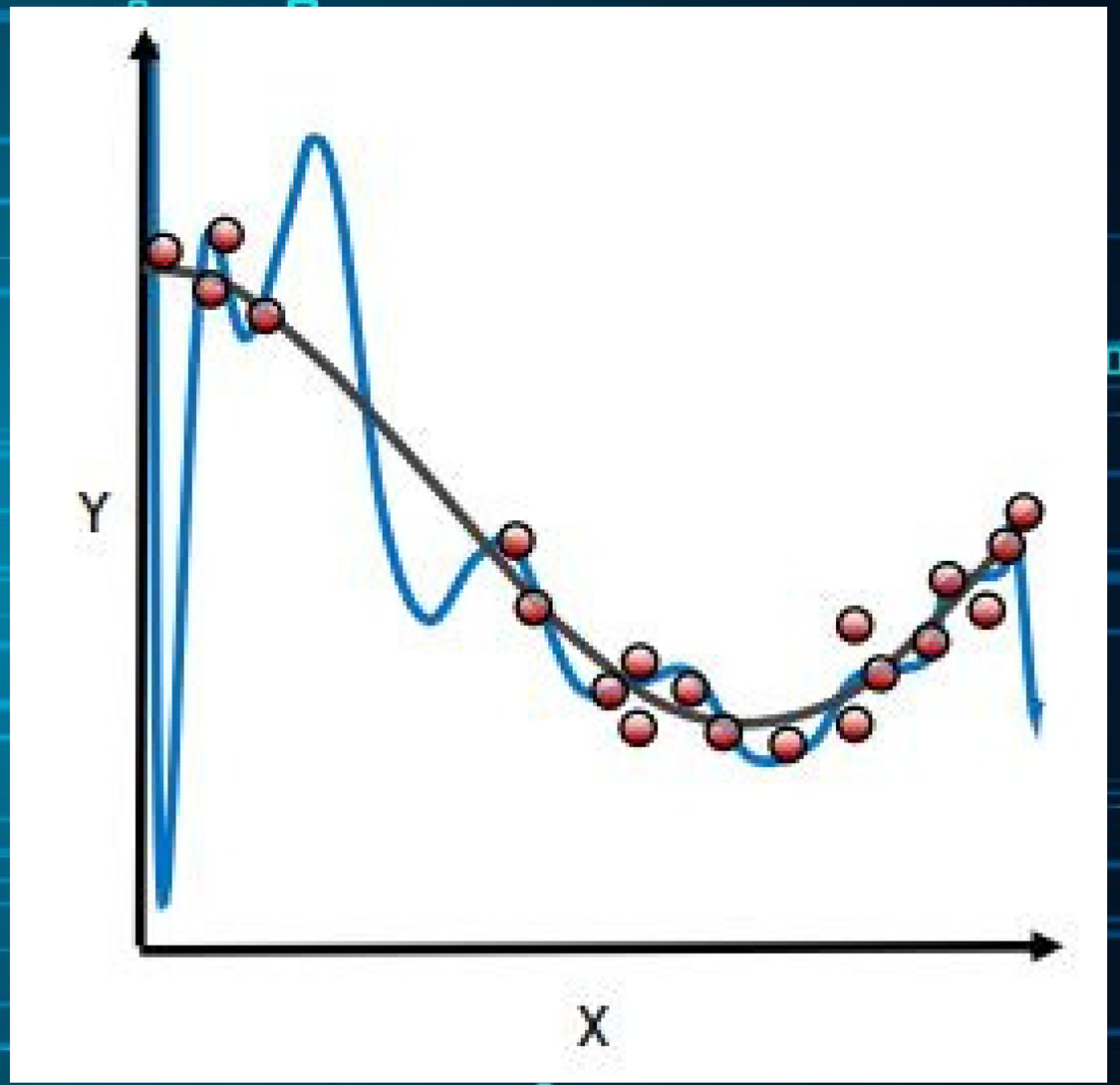
Over-fitting:

Good at training

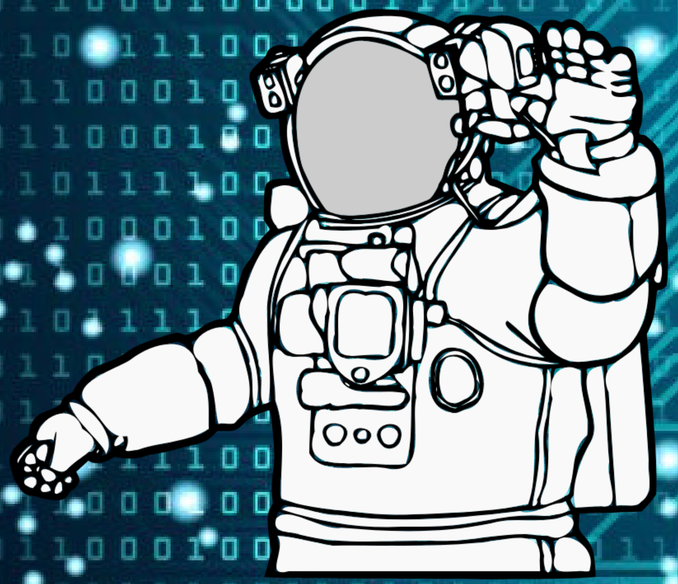
&

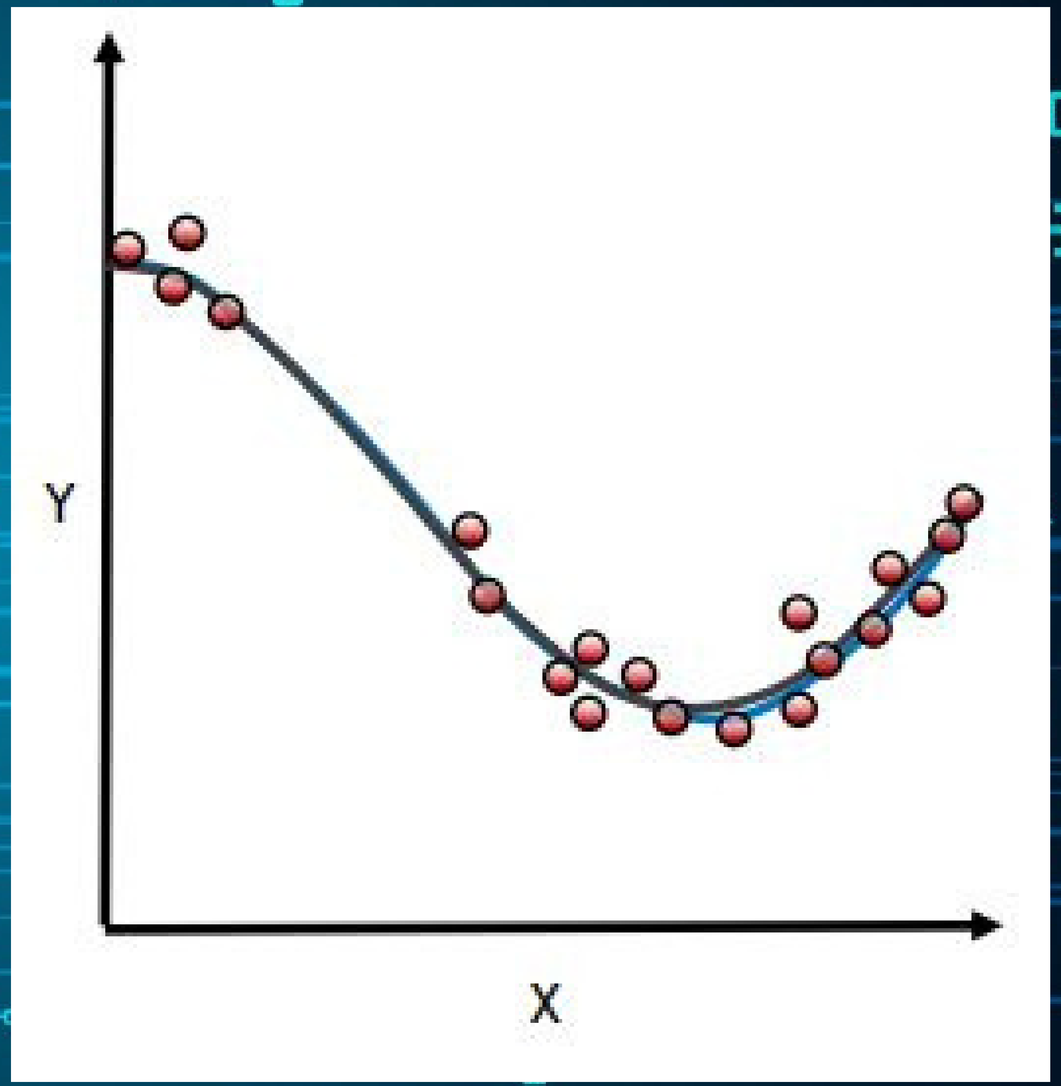
poor at predicting

low bias - high variance

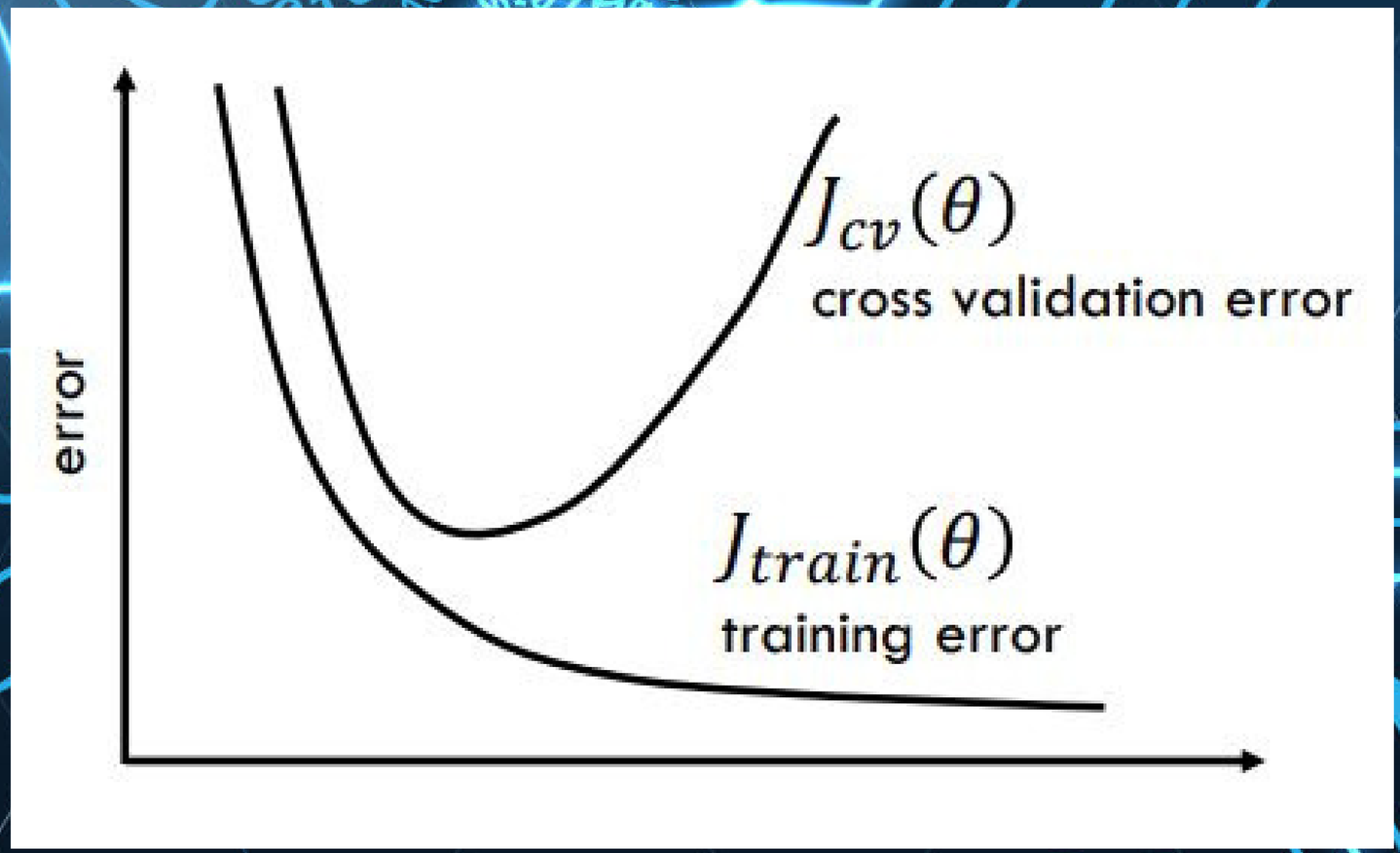


Just Right



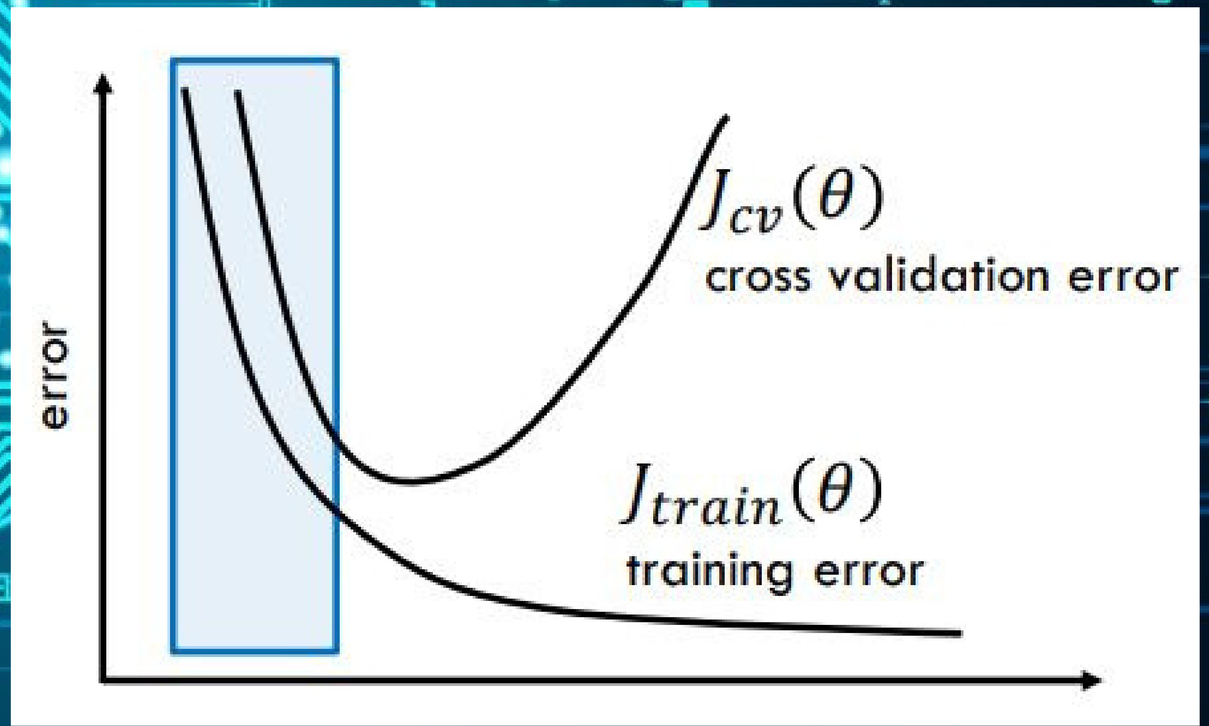


Model Complexity vs Error



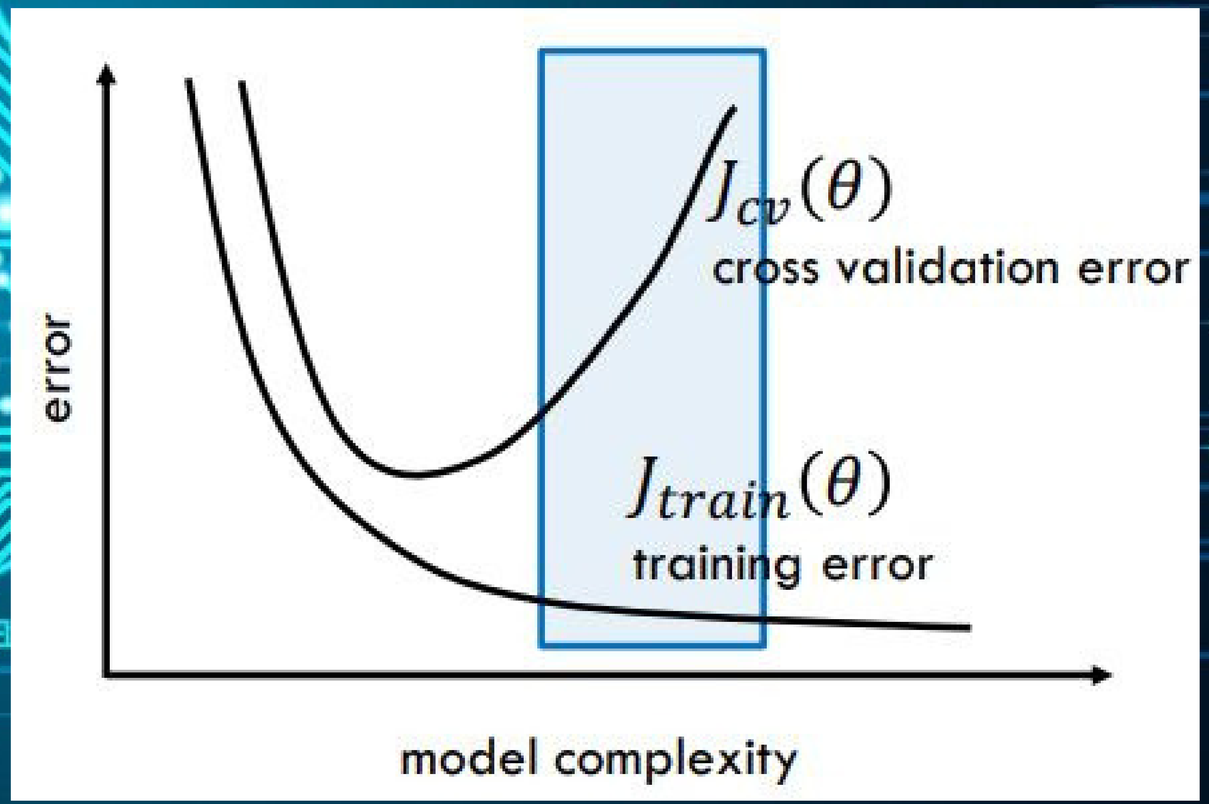
Under-fitting :

training and cross validation error are high

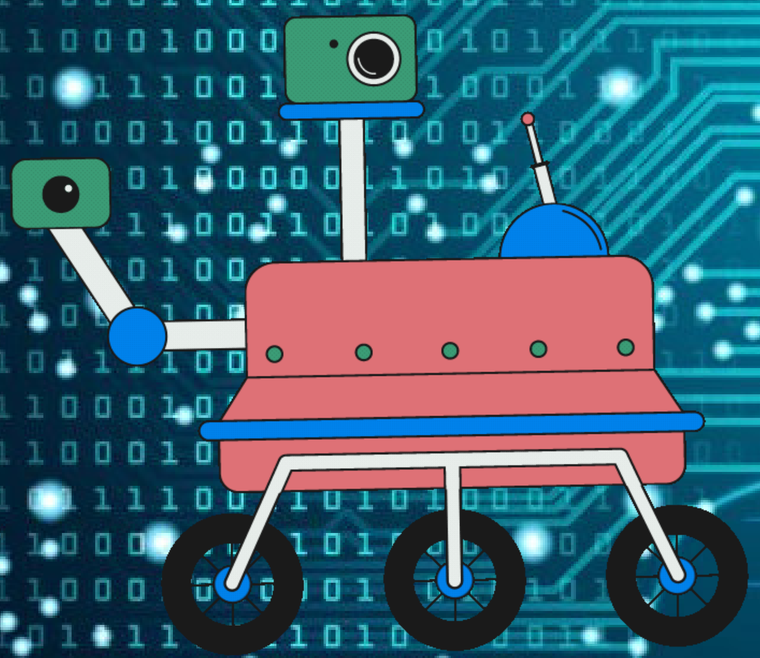


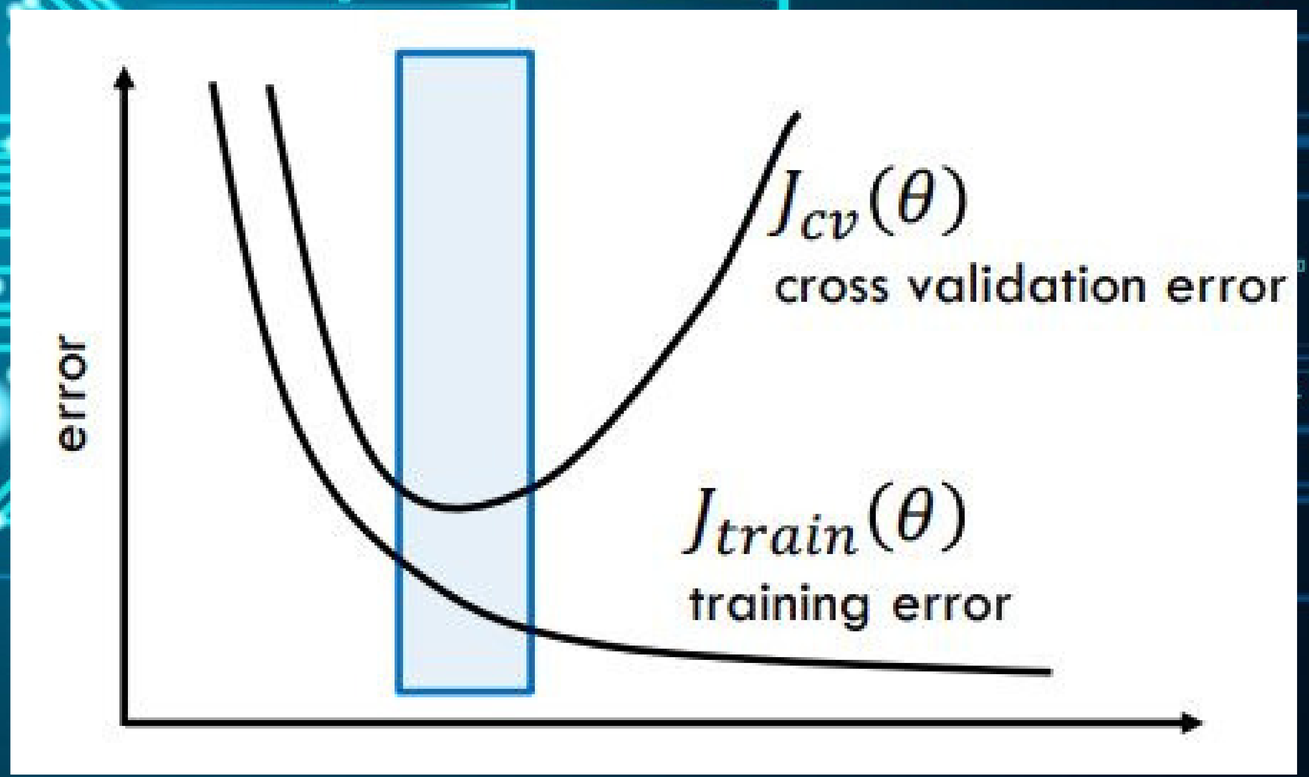
over-fitting :

training error is low, cross validation is high



Just Right




error
$J_{cv}(\theta)$
cross validation error
$J_{train}(\theta)$
training error

"THANK YOU FOR YOUR ATTENTION !"