مقرر التعلم التلقائي | السنة الخامسة – هندسة المعلوماتية / الذكاء الصنعي | Week 7

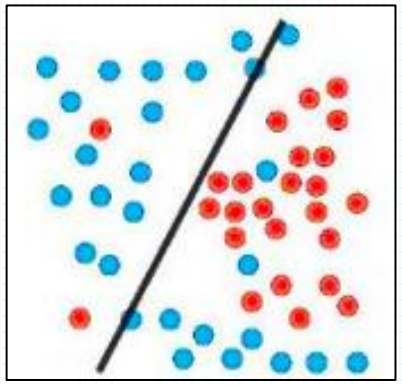
Practical Concerns for Machine Learning 1
Generalization (Bias-Variance Tradeoff)

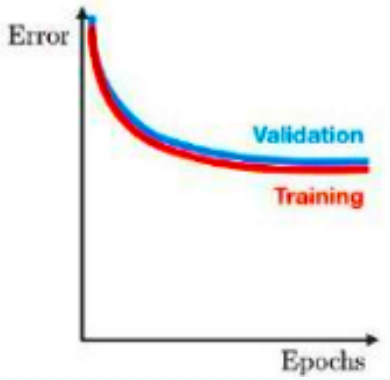
د. رياض سنبل

Access Course Materials

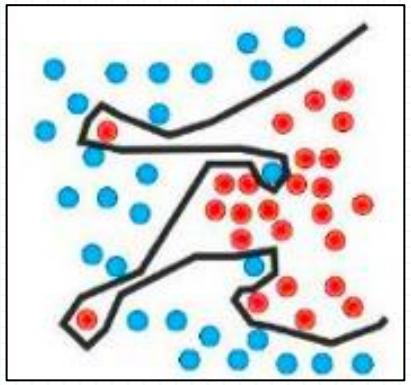
Generalization

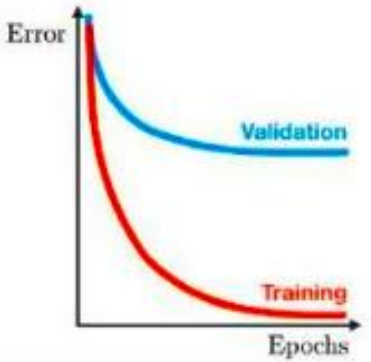
- ***Underfitting***
  - The model is too “simple” to represent all the relevant class characteristics
  - E.g., model with too few parameters produces high error on the training set and high error on the validation set



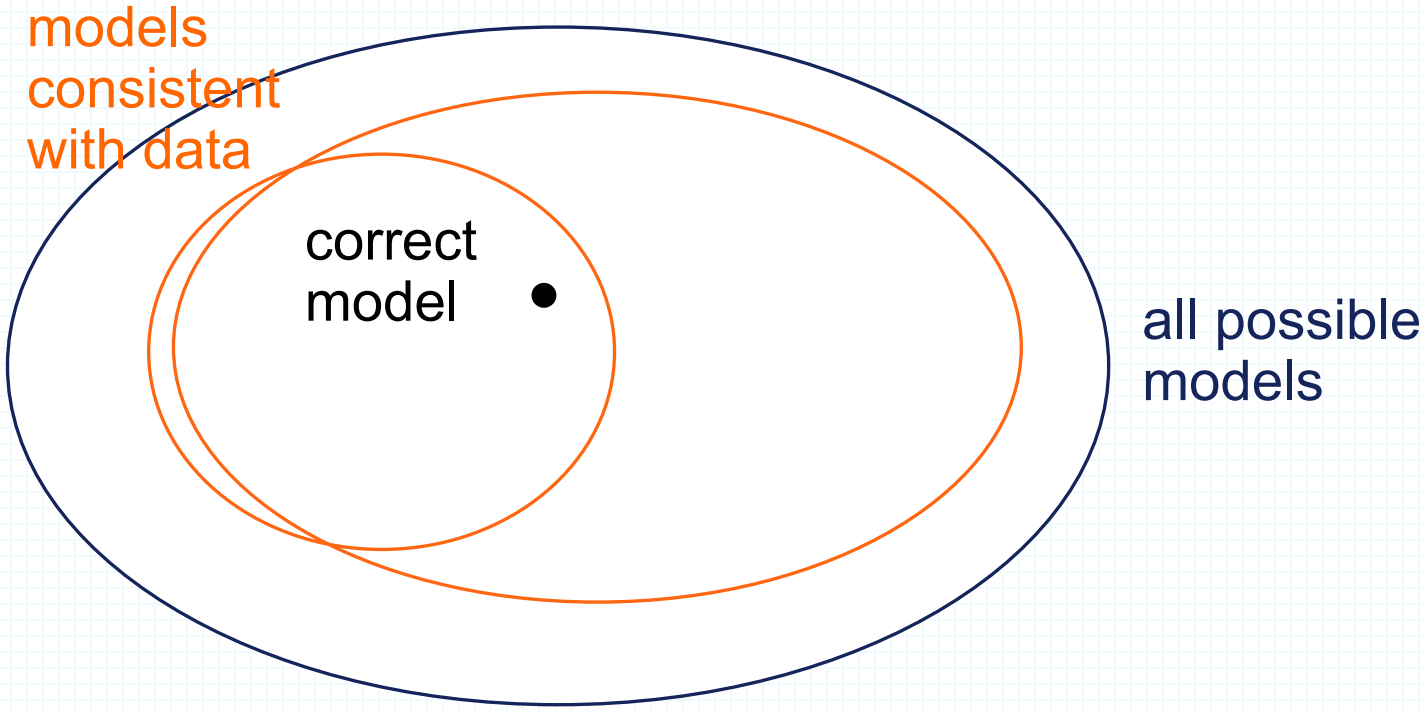


- ***Overfitting***
  - The model is too “complex” and fits irrelevant characteristics (noise) in the data
  - E.g., model with too many parameters produces low error on the training set and high error on the validation set



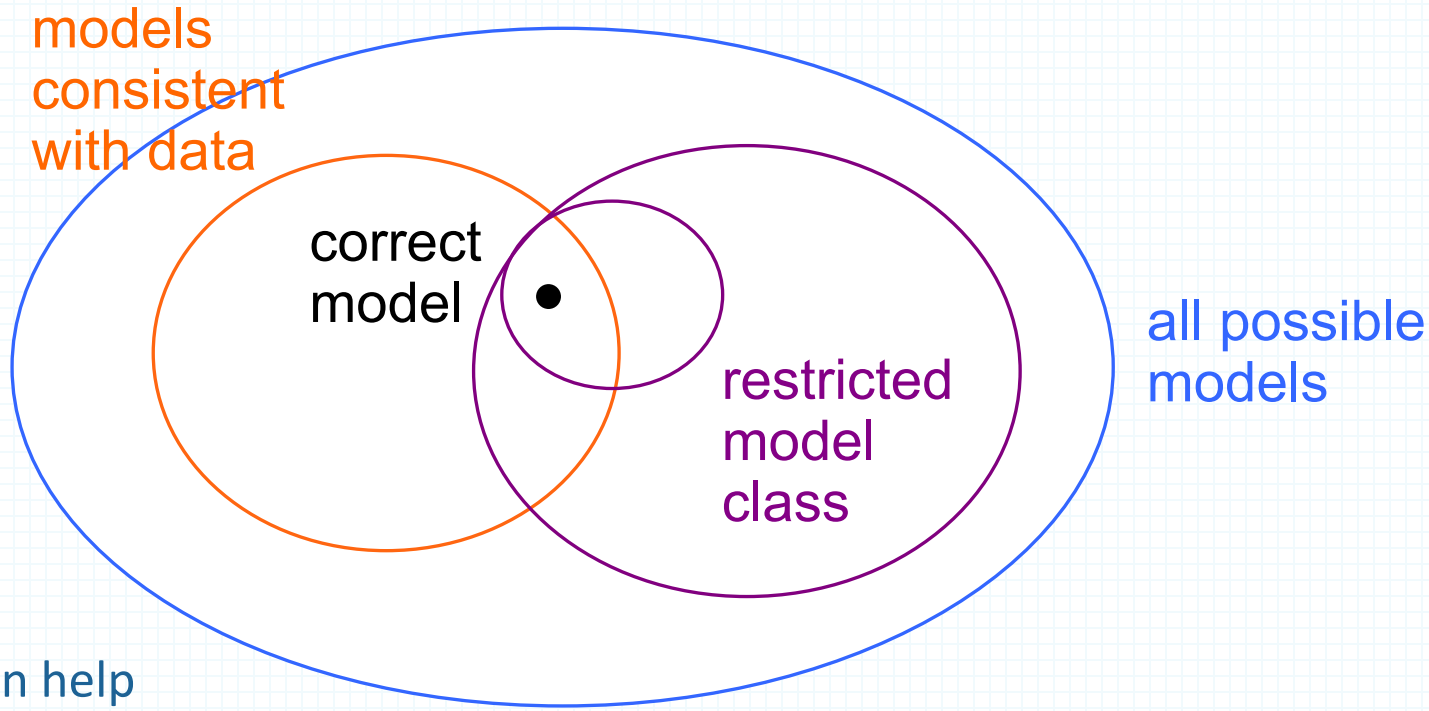


Back to the core idea in ML☺



More data can help!

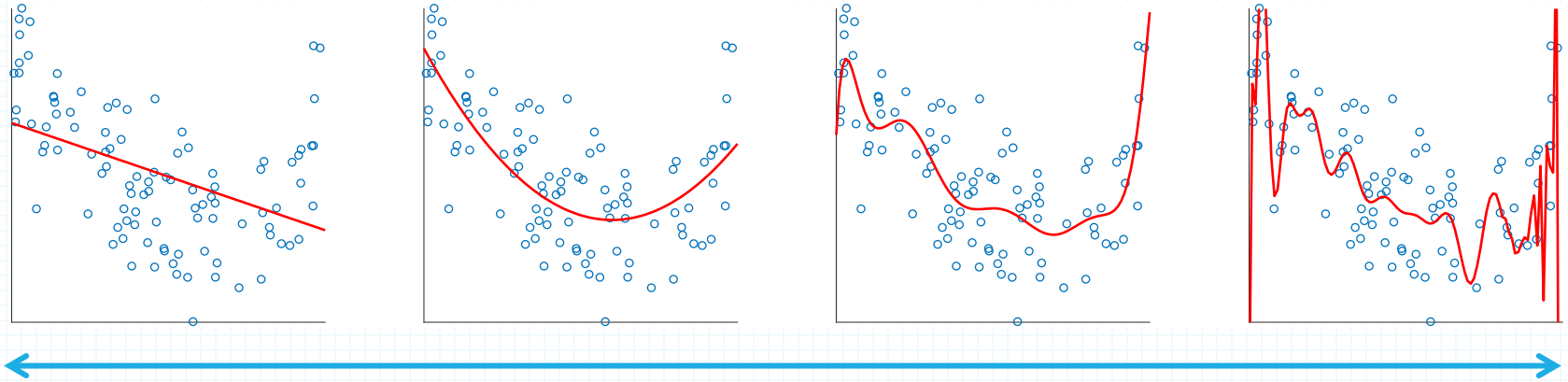
Back to the core idea in ML☺



- Restricting model class can help
- Or it can hurt
- Depends on whether restrictions are domain appropriate

Restricting Models

- Models range in their flexibility to fit arbitrary data



simple model

constrained

small capacity may prevent it from representing all structure in data

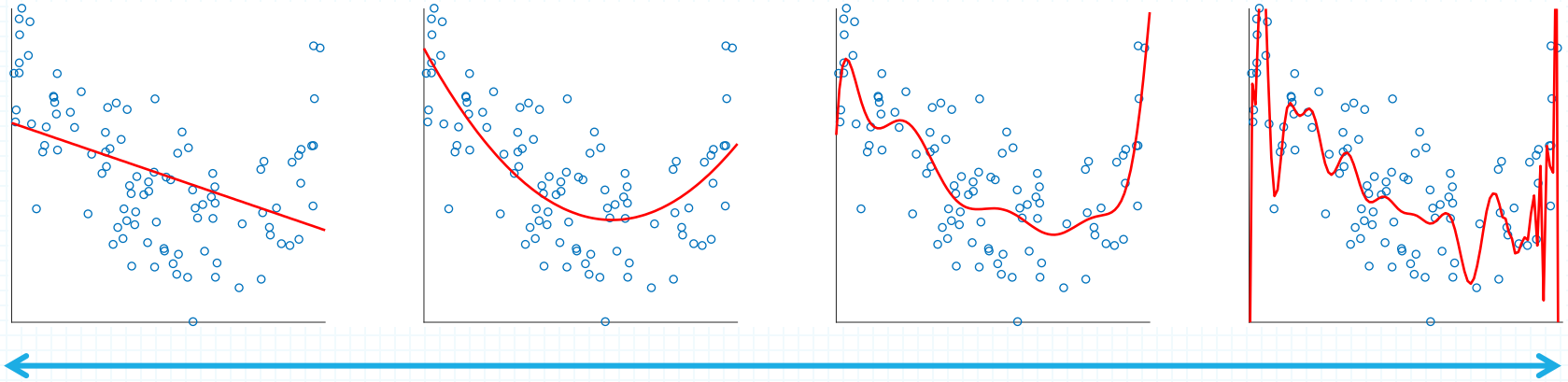
complex model

unconstrained

large capacity may allow it to fit quirks in data and fail to capture regularities

Bias vs Variance

- Models range in their flexibility to fit arbitrary data



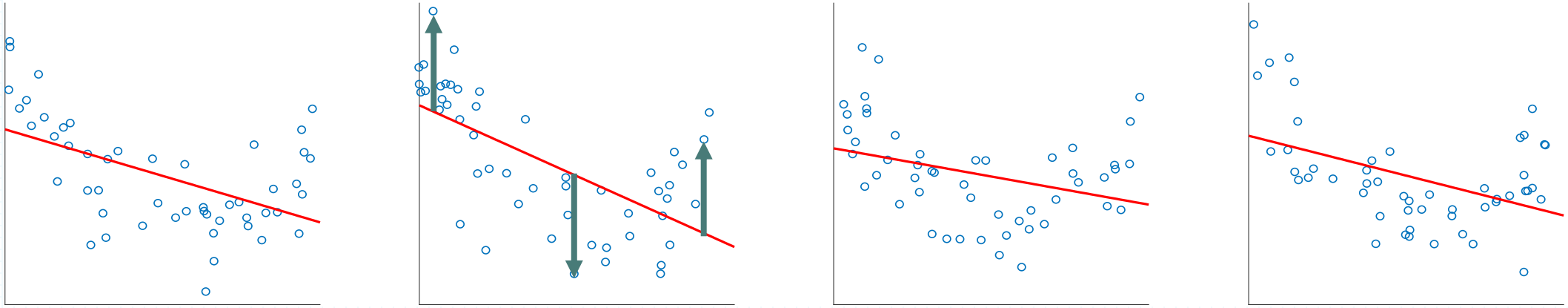
high bias

low variance

low bias

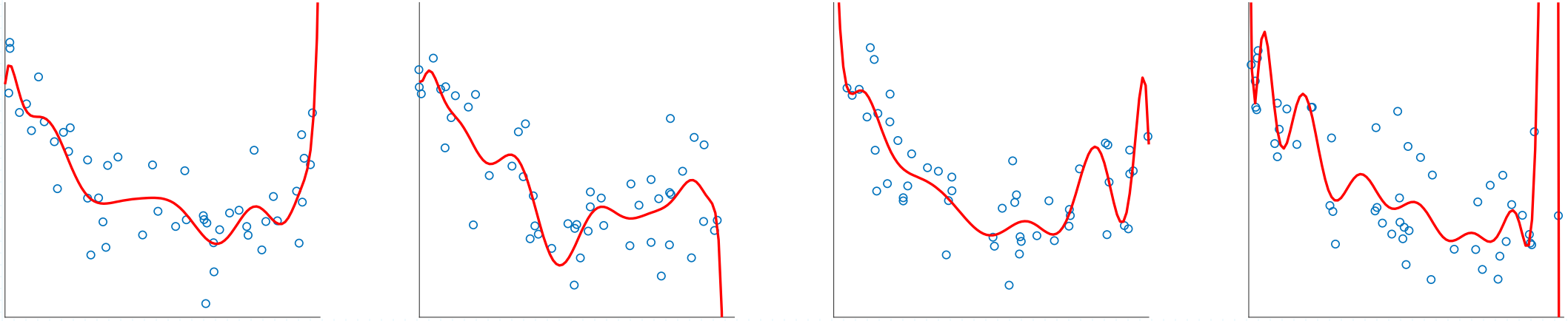
high variance

Bias



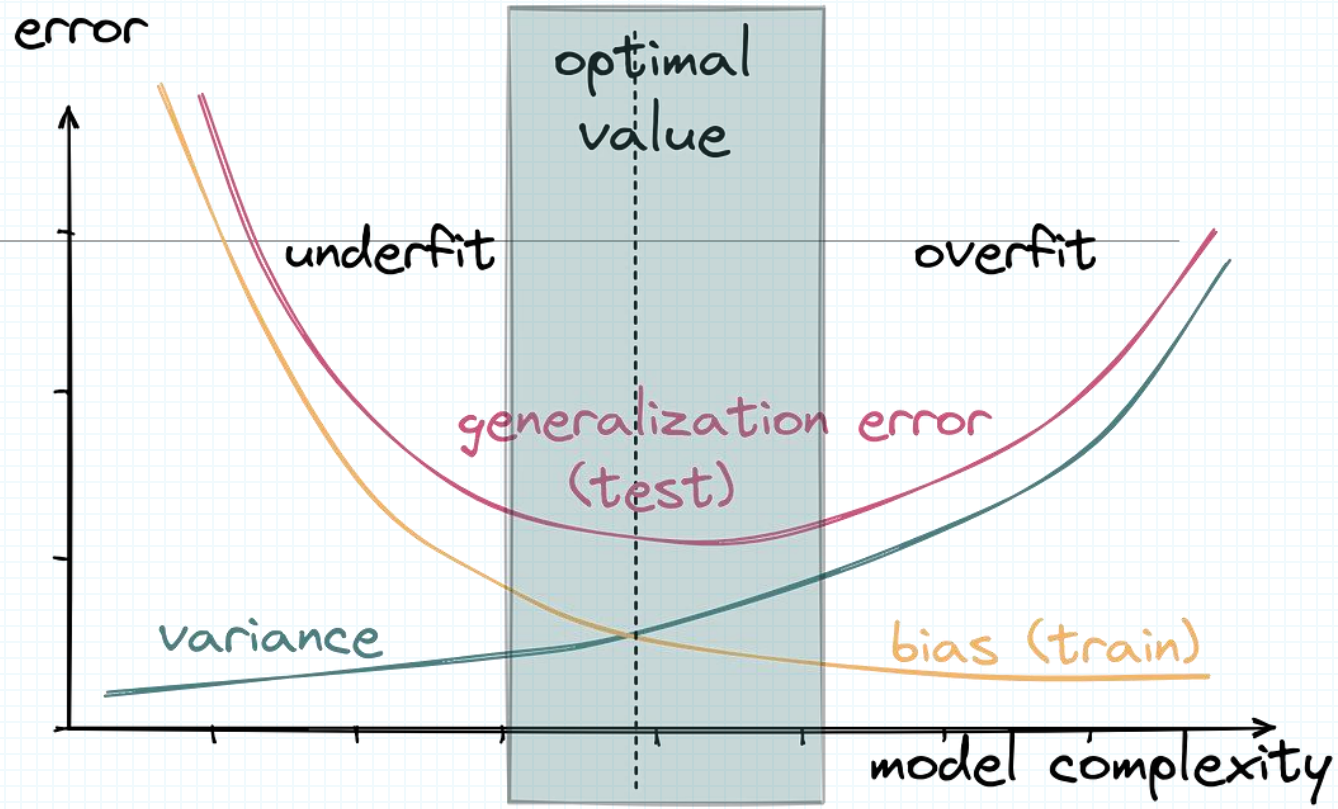
- Bias is the amount that a model's prediction differs from the target value, compared to the **training data**.
- Bias error **results from** simplifying the assumptions used in a model so the target functions are easier to approximate.
- Regardless of training sample, or size of training sample, model will produce consistent errors

Variance



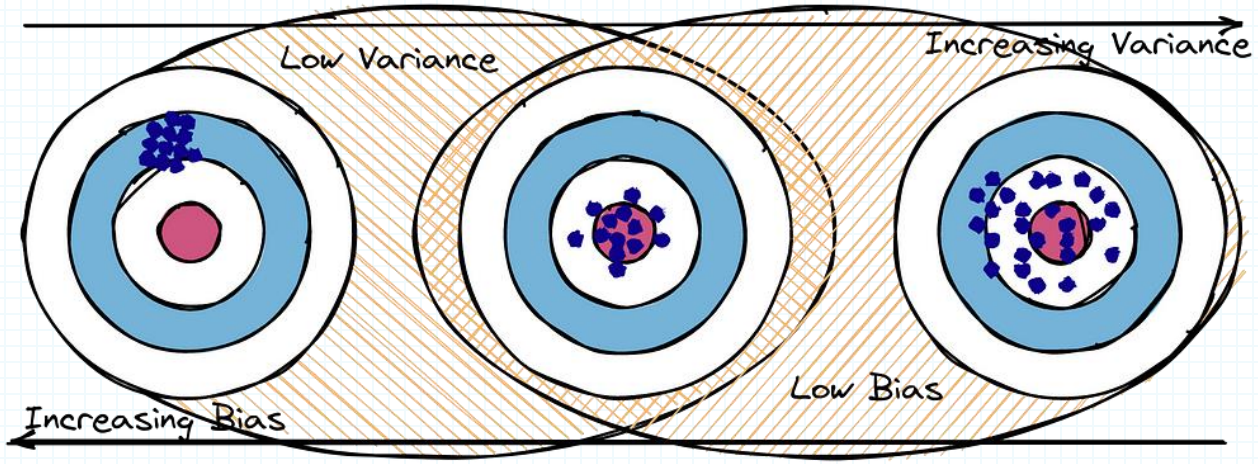
- Variance refers to **the changes in the model** when using different portions of the training data set.
- Simply stated, variance is the **variability in the model prediction**
- Different samples of training data yield different model fits

Bias-Variance Trade off



What does tons of training data do for us in terms of bias-variance trade off?

- ✓ Doesn't affect bias
- ✓ Reduces variance
- ✓ Allows more complex model to be used [shift of optimal complexity to the right]



Bias-Variance Trade off

Training set Error	1%	15%	15%	0.5%
Testing set Error	11%	16%	30%	1%
Bias				
Variance				

DOG vs. CAT

Bias-Variance Trade off

Training set Error	1%	15%	15%	0.5%
Testing set Error	11%	16%	30%	1%
Bias	Low	High	High	Low
Variance	High	Low	High	Low

DOG vs. CAT

Error = Irreducible Error + Bias + Variance

Solutions!

- **High Bias:**
  - Bigger Network?
  - Train Longer?
- **High Variance:**
  - More Data
  - Regularization.

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

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

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