1. Where does the error come from?

Error is due to 'bias' and 'variance'

We don't know the true relationship between input and output (the actual mapping function), we estimate it. When compare the prediction result with testing set, we can calculate the bias and variance to understand the model performance and adjust the model accordingly.

Bias and Variance of Estimator

Estimate the mean of a variable x:

- Assume the mean of x is

$$\mu$$

- Assume the variance of x is

$$\sigma$$

Estimator of mean mu:

- Sample of N points:

$$\{x^1, x^2, x^3, \ldots, x^N\}$$
$$m = \frac{1}{N}\sum_n x^n \neq \mu$$

You have n number of sampling, and calculate m for each sampling.

$$E[m] = E[\frac{1}{N}\sum_n E[x^n]] = \frac{1}{N}\sum_n E[x^n]$$

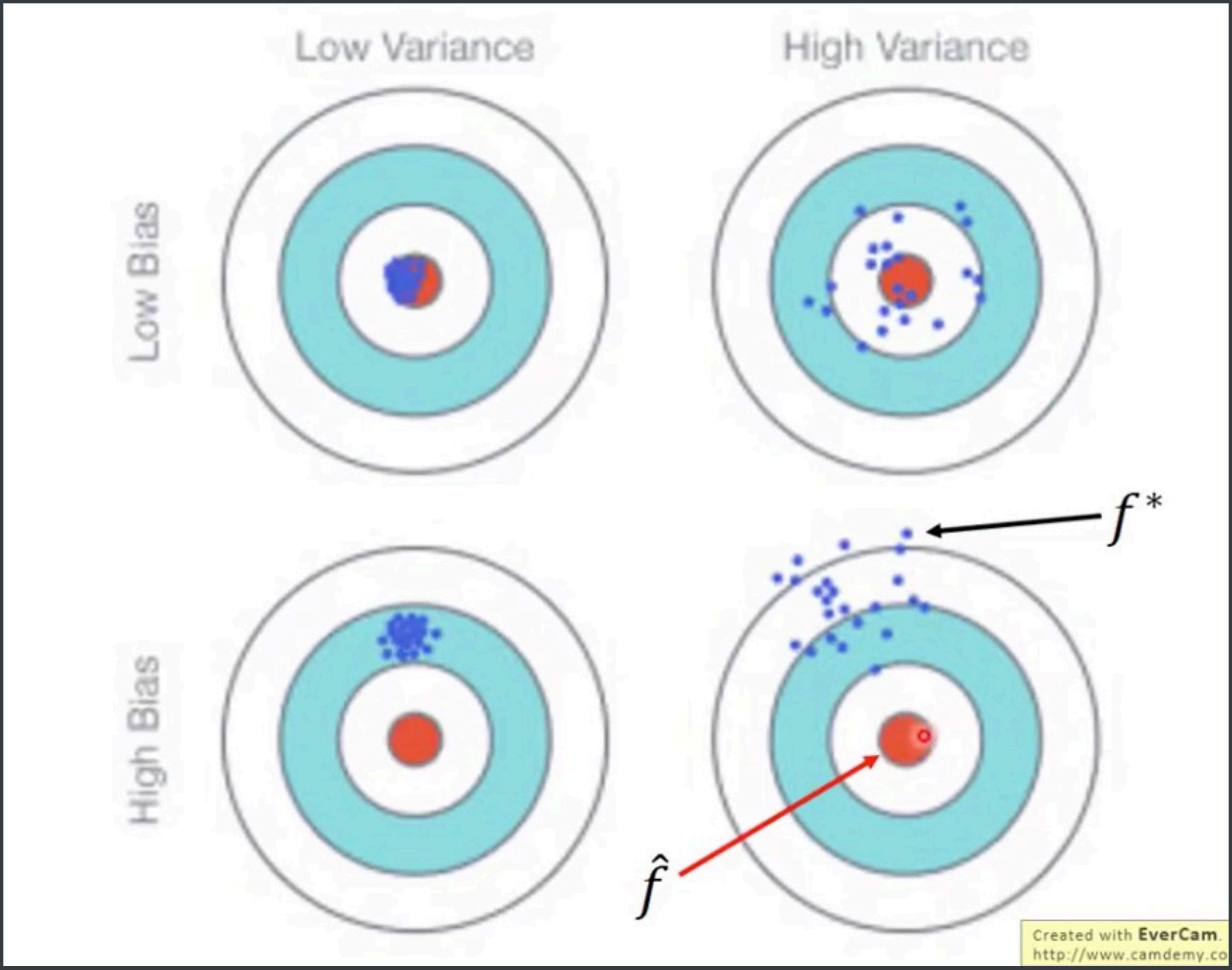
As unbiased estimation is 'equally distributed' around the real mu. But how 'far away' does each m from the real mu? Using variance.

- Variance of m:

$$Var[m] = \frac{\sigma^2}{N}$$
$$s^2 = \frac{1}{N}\sum_n (x^n - m)^2$$

  it depends on the number of samples.

  Larger N: smaller var[m], datapoints of m 'closer'



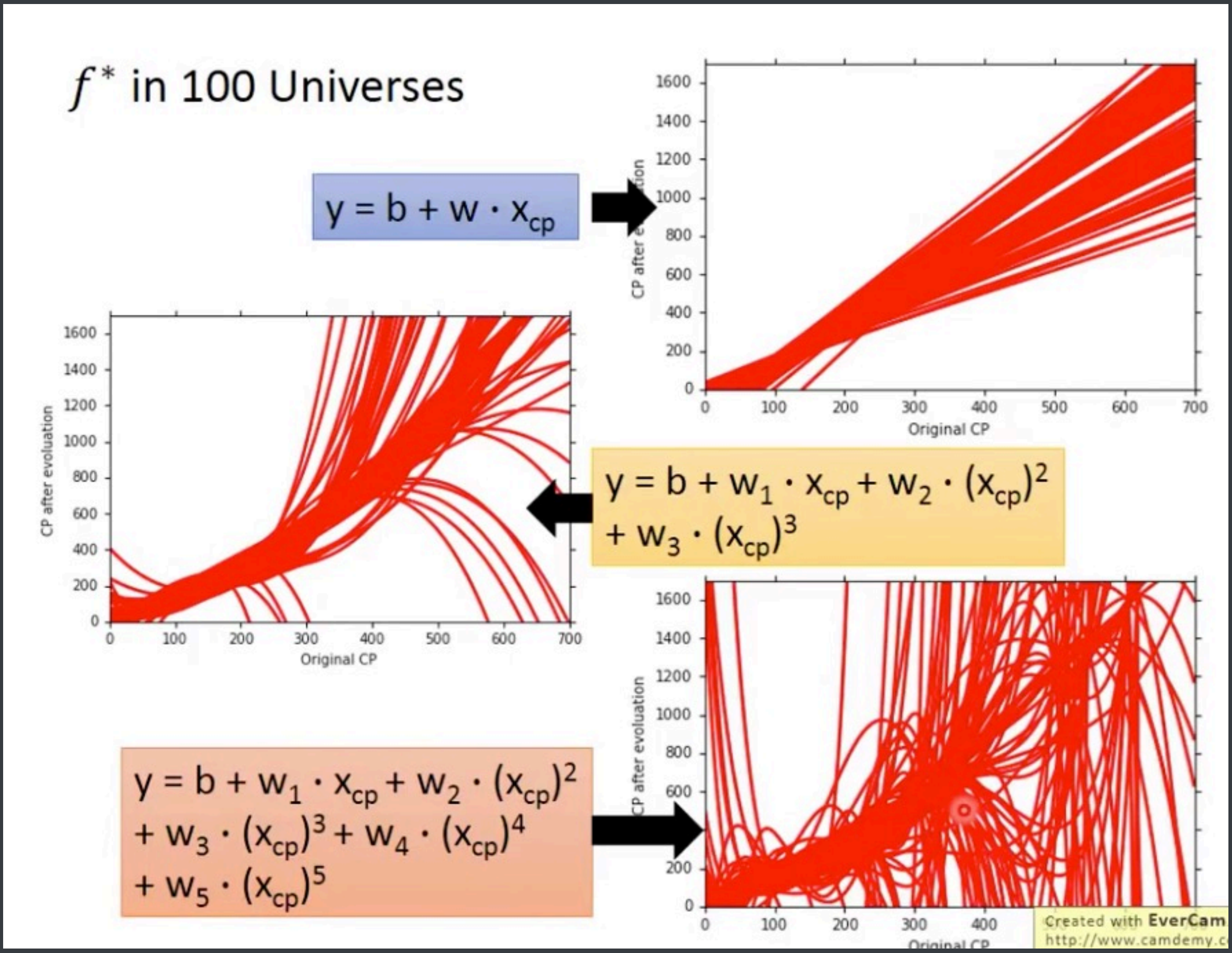
Link the concepts to Regression Question

We wants to estimate f^, we collected some data and have f*

- In different parallel universe (sampling), use same model but obtain different f*

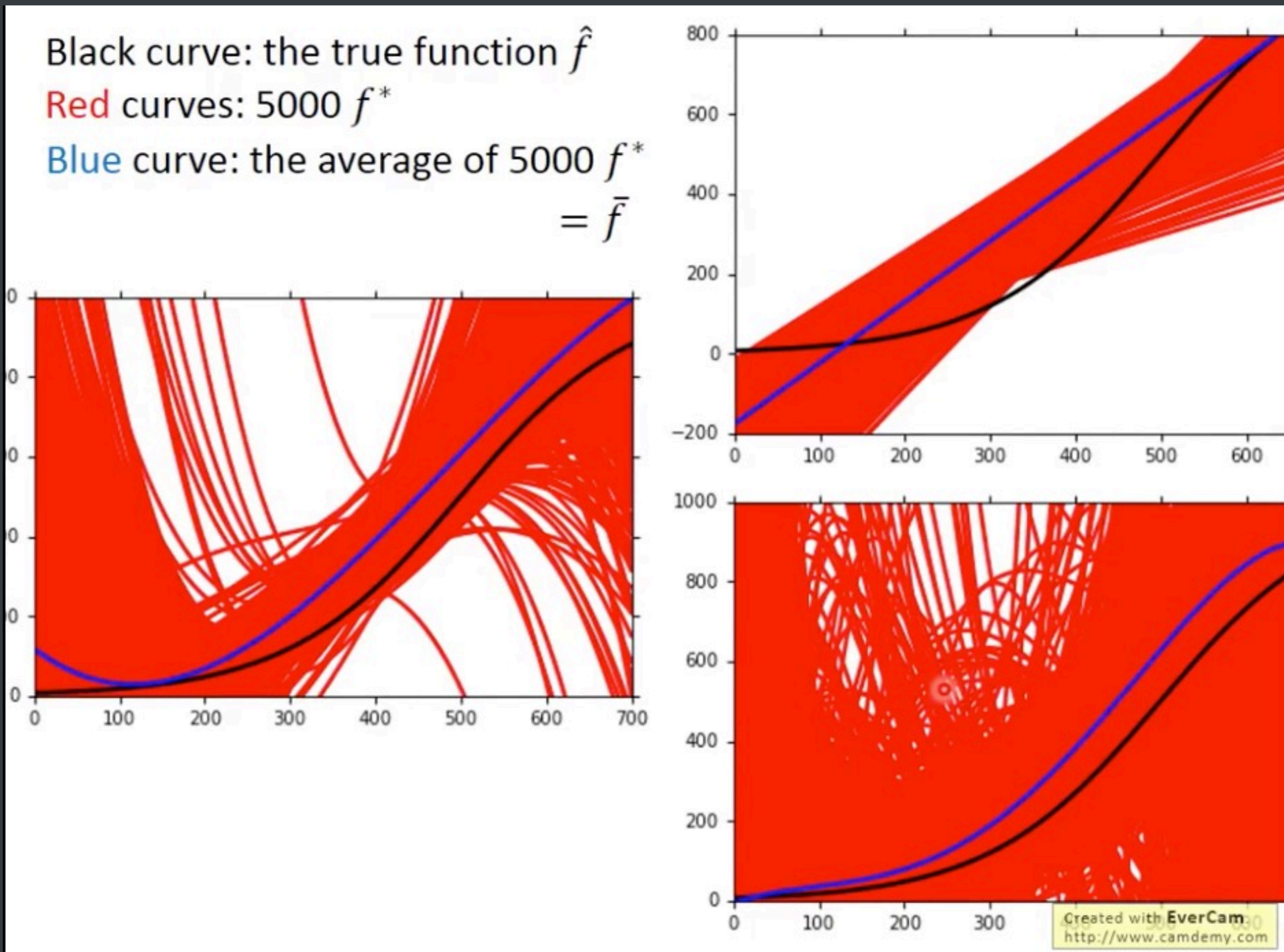
$$E[f^*] = \bar{f}$$

Compare 3 models performance in '100 universes'



Simple model 1, small variance, less sensitive to data

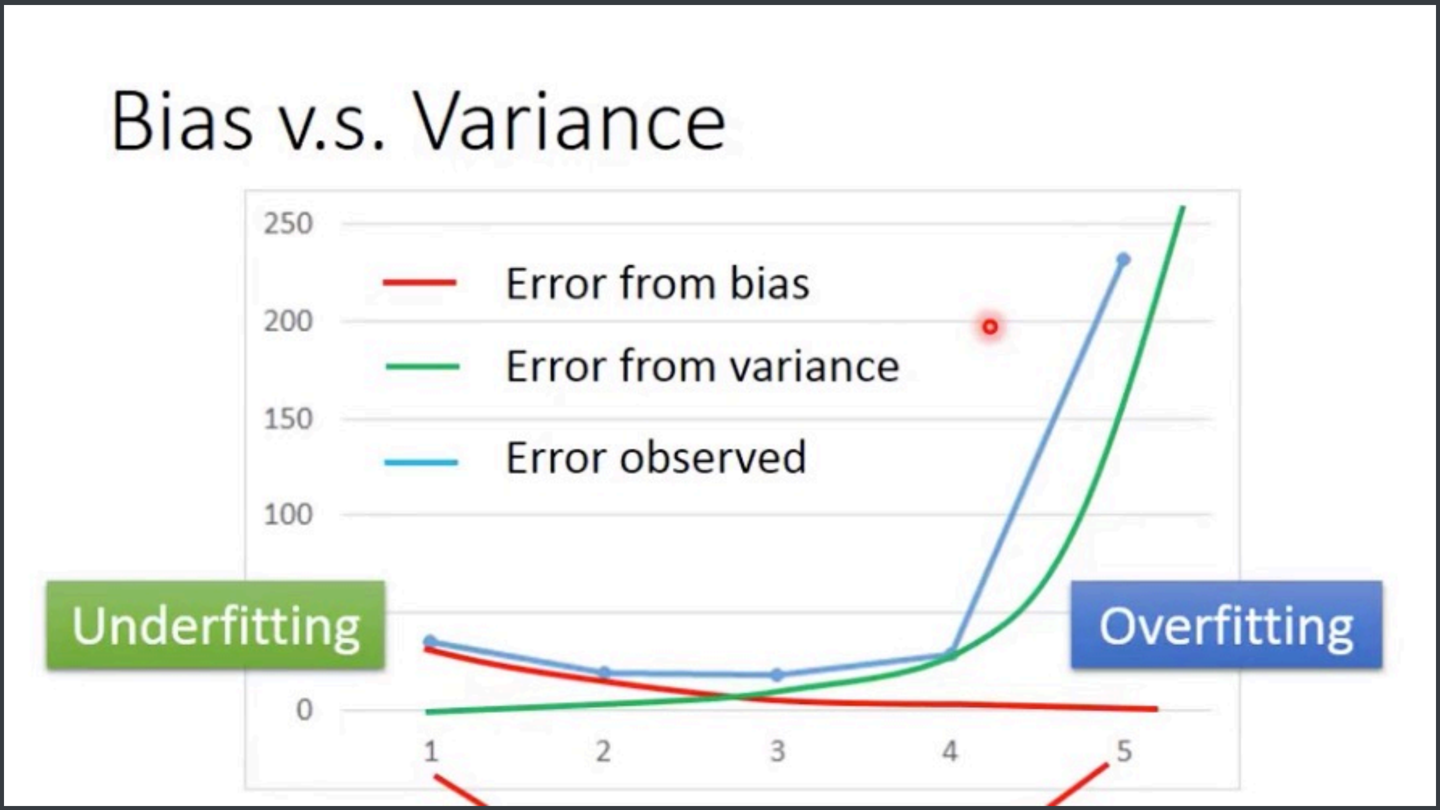
Complex model 3, large variance, more sensitive to data



Simple model 1, larger bias

Complex model 3, smaller bias

Image model as a 'function set', simple model have a smaller function set (space) where the output value is 'restricted' to a certain range; while a complex model function space is large and may cover the target.



Select model 3 (a balance between bias and variance)

What to do with Large bias?

- If model cannot fit training examples. (Underfitting)
- If model can fit training data, but large error on testing data, then probably have large variance (Overfitting)

For bias: redesign the model:

- Add more features as input
- A more complex model

What to do with Large Variance?

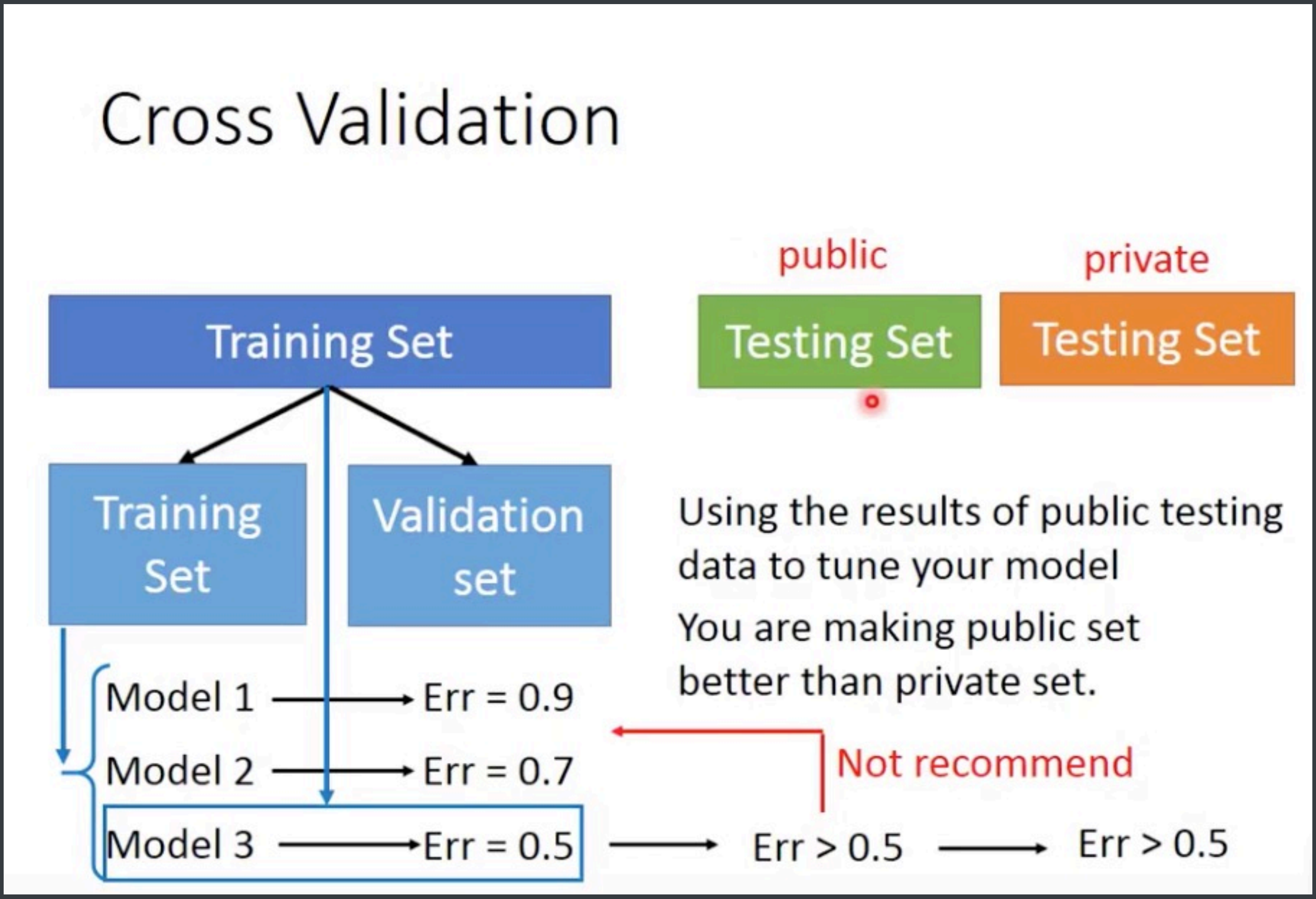
- More data: effective but not always practical
- Regularization

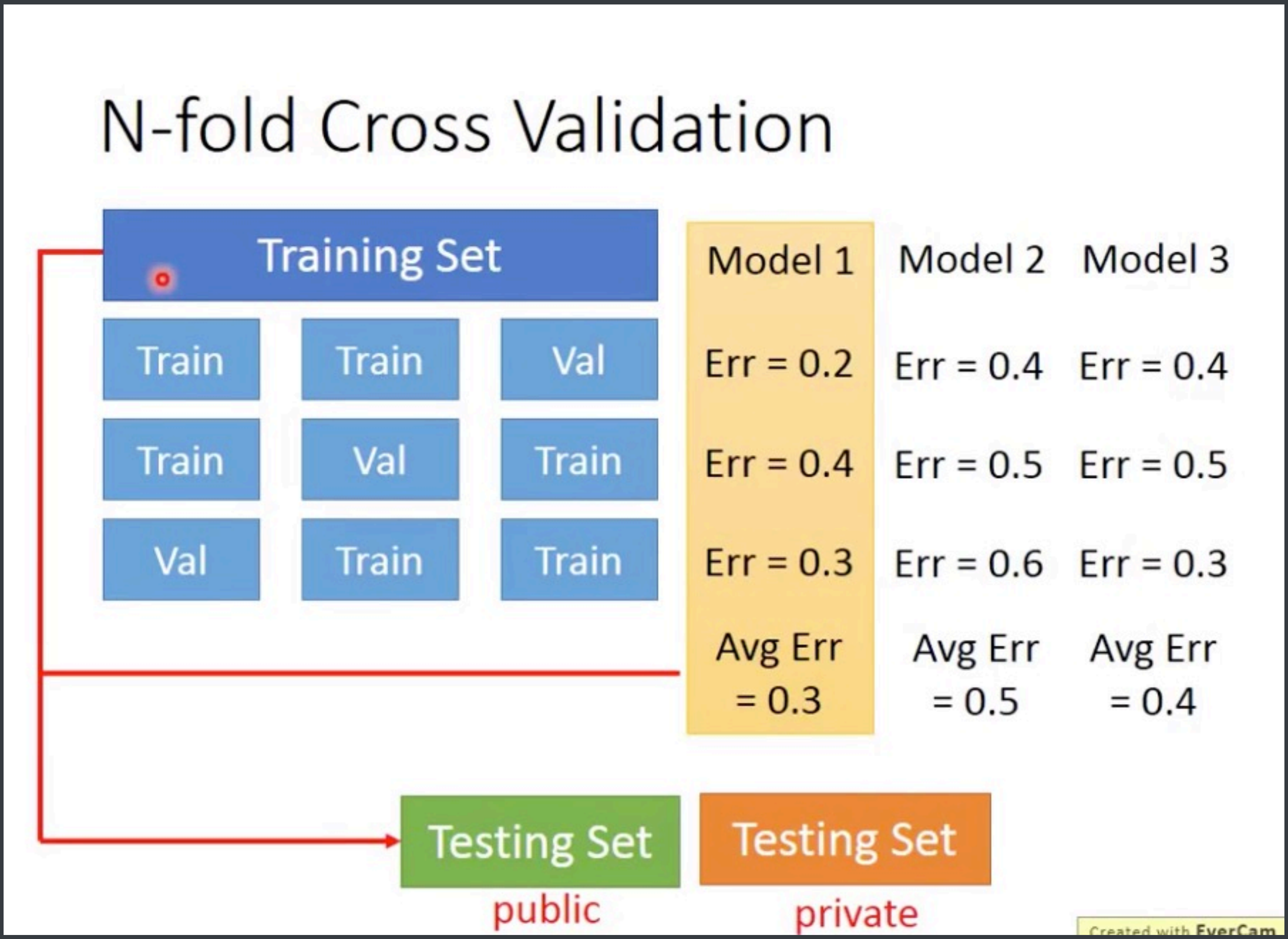
Model Selection

- There is usually a trade-off between bias and variance
- Select a model that balances two kinds of error to minimum total error
- Should NOT do:
  - Only use training set to select model

  (Public test set in Kaggle vs private/hidding test set)

Cross Validation





2. Gradient Descent and Adagrad