Regularization
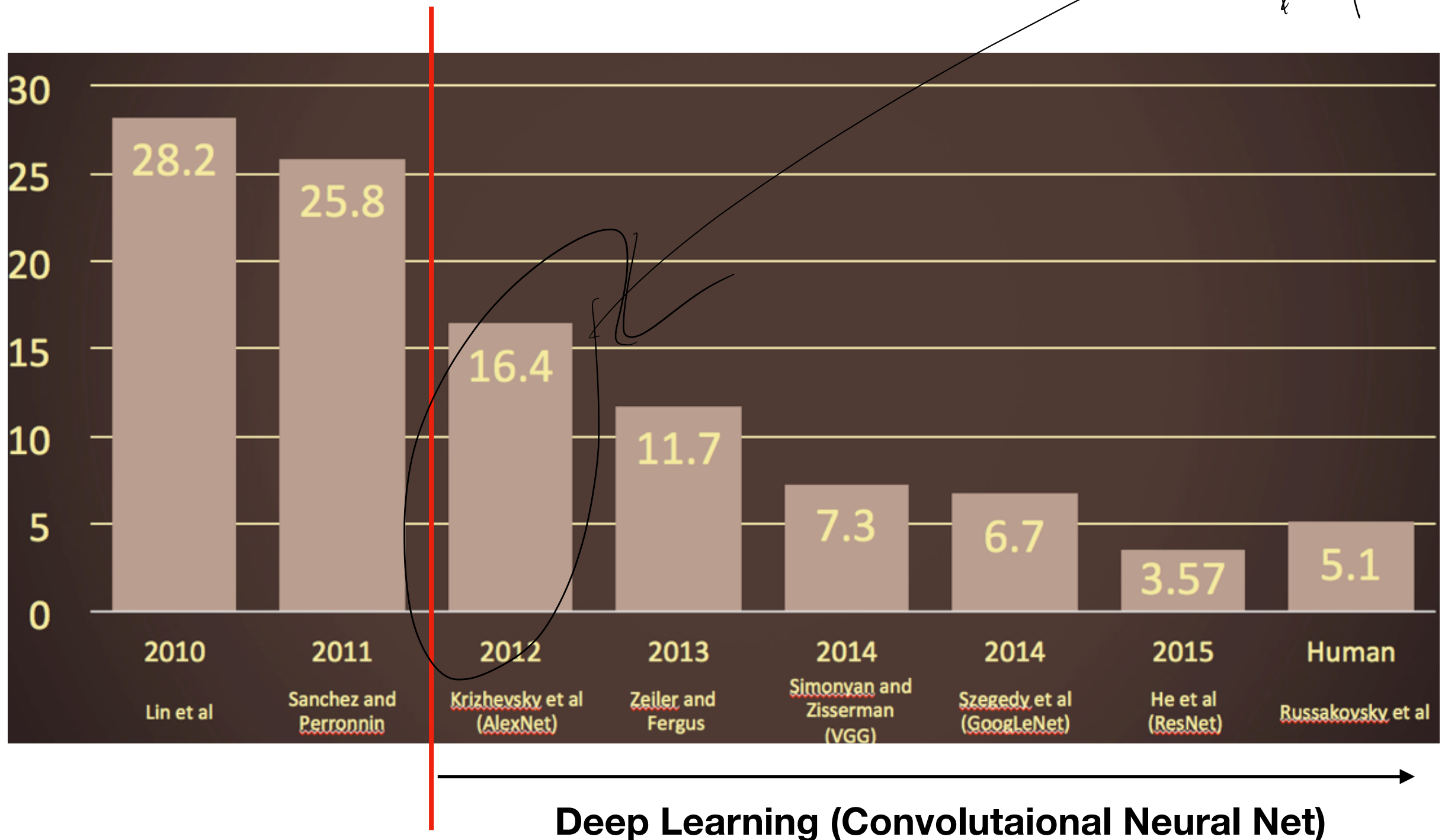
Seyoung Yun

The Breakthrough
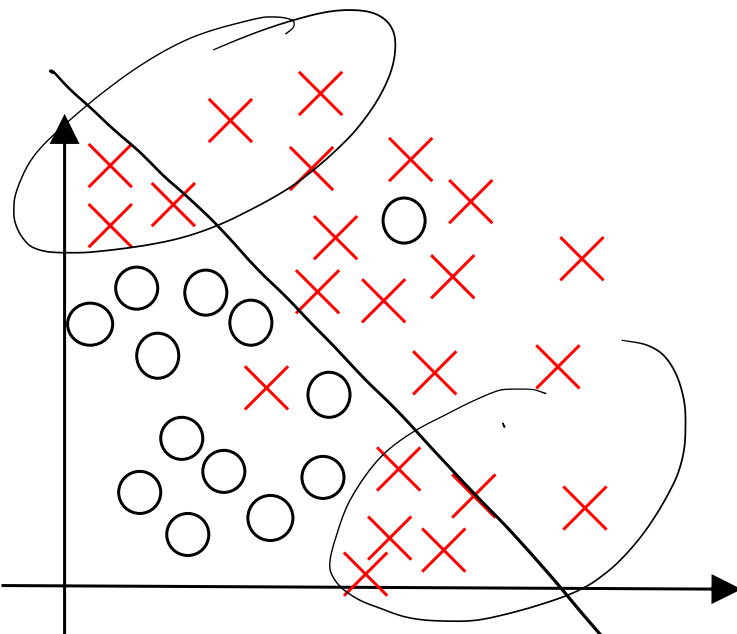

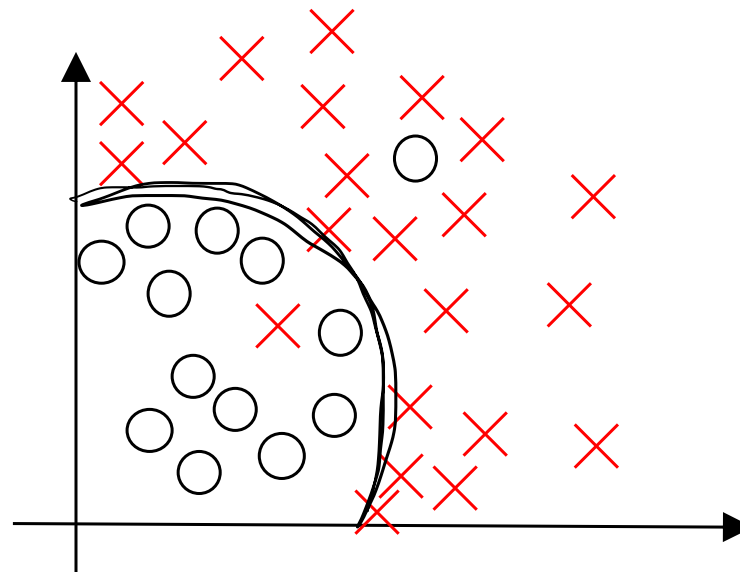- AlexNet : ReLu and Dropout

Bias-Variance

K-NN with $k=1$

high bias   "just right"   high variance
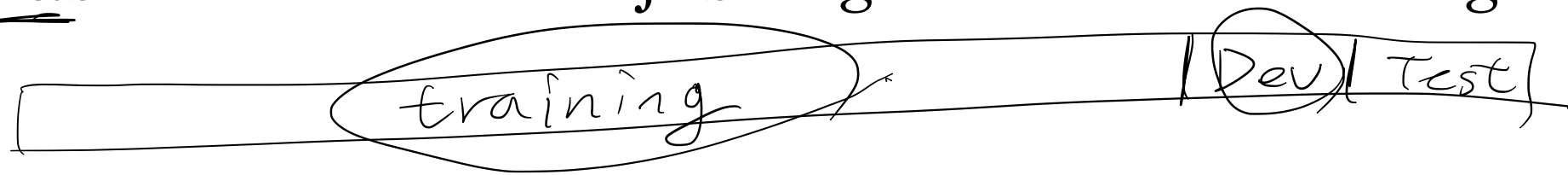
training   [Dev][Test]
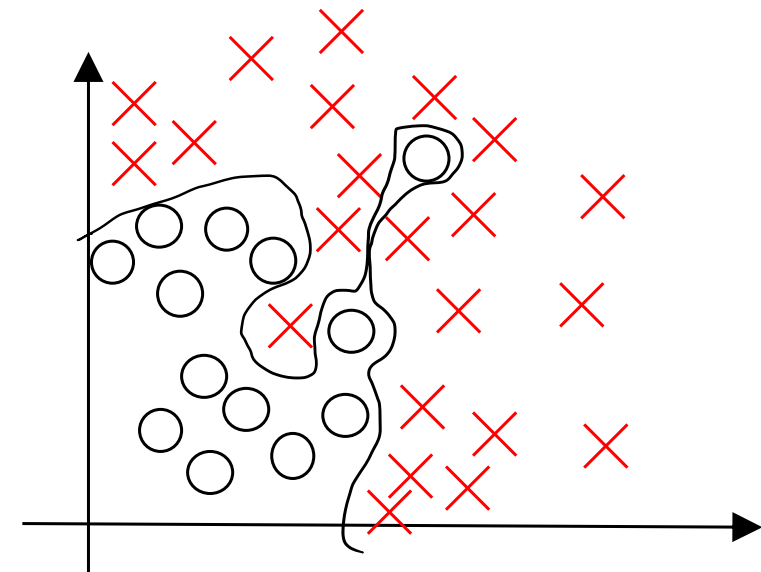
$$Y = f(x) + \varepsilon \quad \Rightarrow \quad \mathbb{E}\left[\left(\hat{f}(x) - f(x)\right)^2\right]$$

$$\mathbb{E}\left[\left(\mathbb{E}[\hat{f}(x)] - f(x)\right)^2\right] \rightarrow \text{bias}$$

$$\mathbb{E}\left[\left(\mathbb{E}[\hat{f}(x)] - \hat{f}(x)\right)^2\right] \rightarrow \text{overfitting.}$$
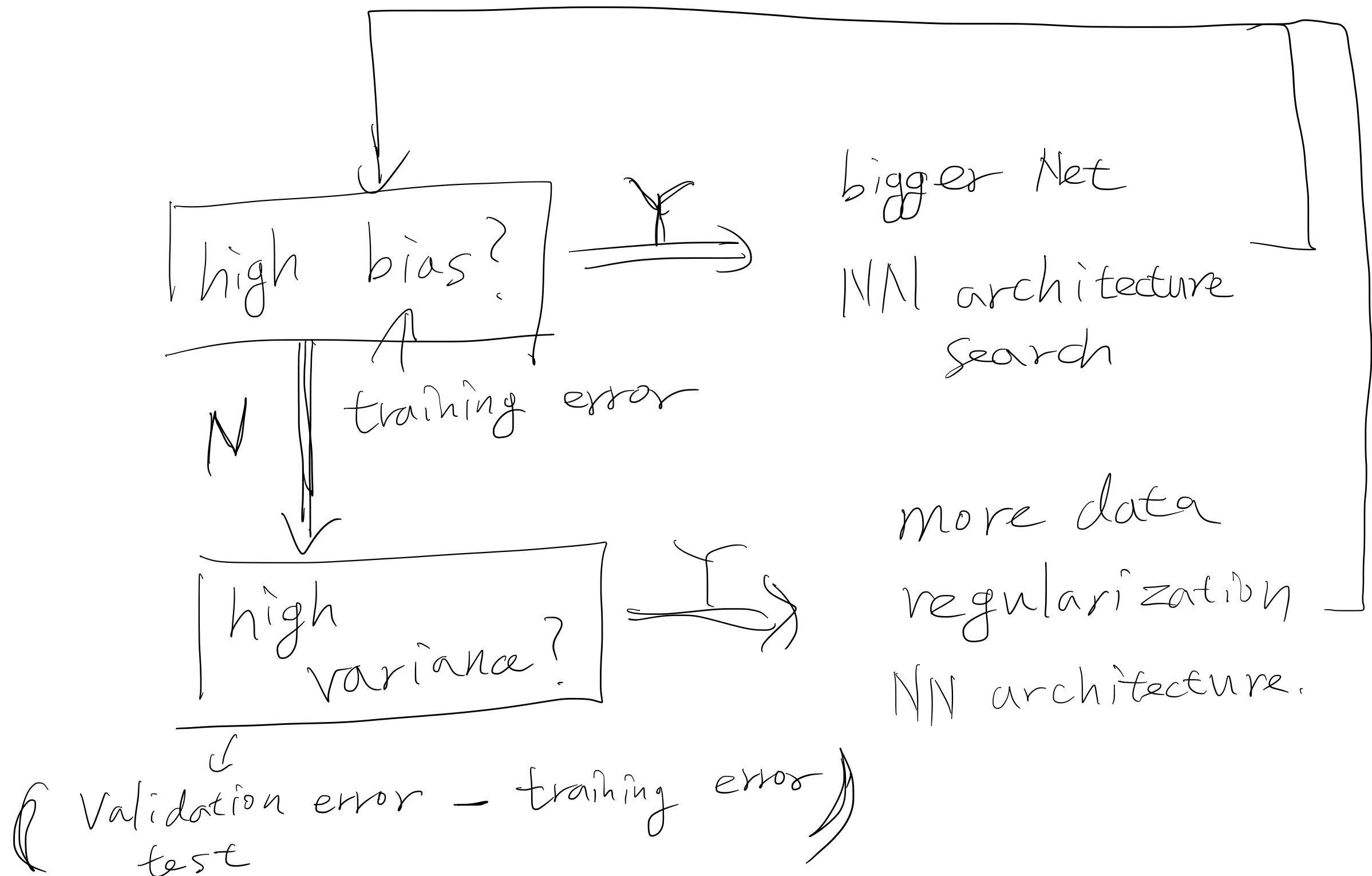
variance.

3

Bias-Variance tradeoff

Flexibility $\uparrow$ $\Rightarrow$ bias $\downarrow$, variance $\uparrow$
#parameters

#data point $\uparrow$ $\Rightarrow$ bias $-$ variance $\downarrow$

regularization $\Rightarrow$ bias $\nearrow$ variance $\downarrow$

Basic recipe for deep learning
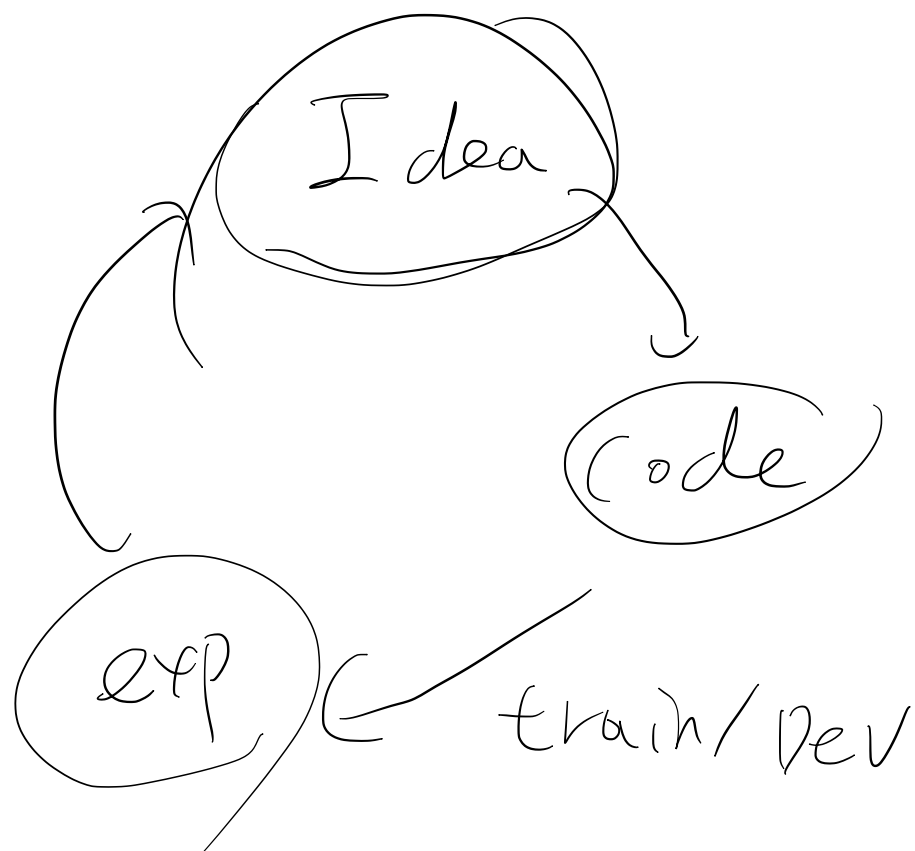
high bias?

$\longrightarrow Y \longrightarrow$

bigger Net
NN architecture
search

N

training error

high variance?

$\longrightarrow Y \longrightarrow$

more data
regularization
NN architecture.

(Validation error — training error)
test

Train/Dev/Test set

6   2   2

Training Set	Dev. Set	Test

98%   1%   1%

Idea

Code

exp   train/Dev

Bias/Variance and Train/Dev

Cat classification

	Algorithm1	Algorithm2	Algorithm3	Algorithm4
Train Error	1	15	15	0.5
Dev Error	11	16	30	1
Bias	L	H	H	L
Variance	H	L	H	L