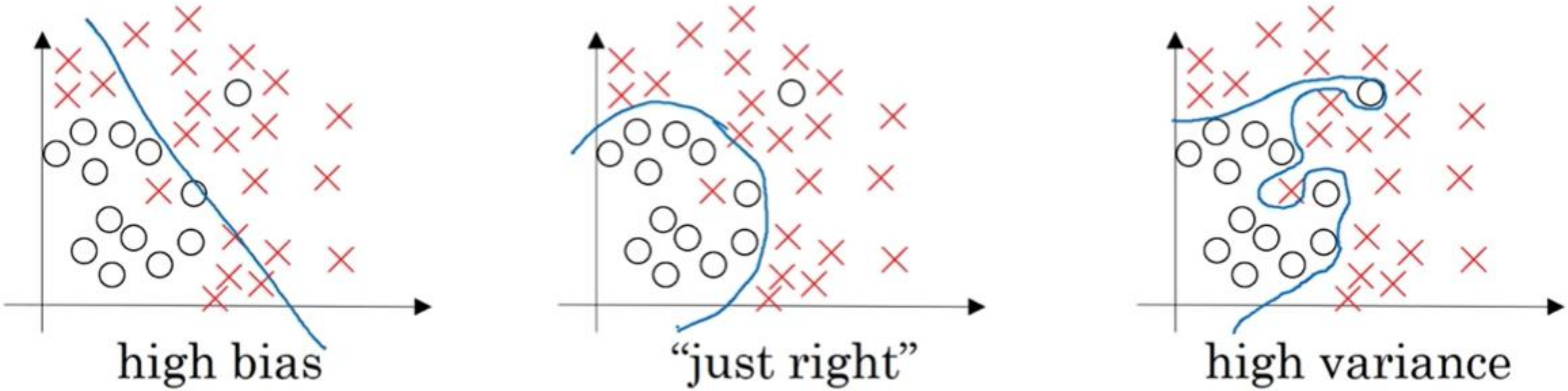
Regularization

Dr. Chandranath Adak

Dept. of CSE, IIT Patna

High Bias vs. High Variance

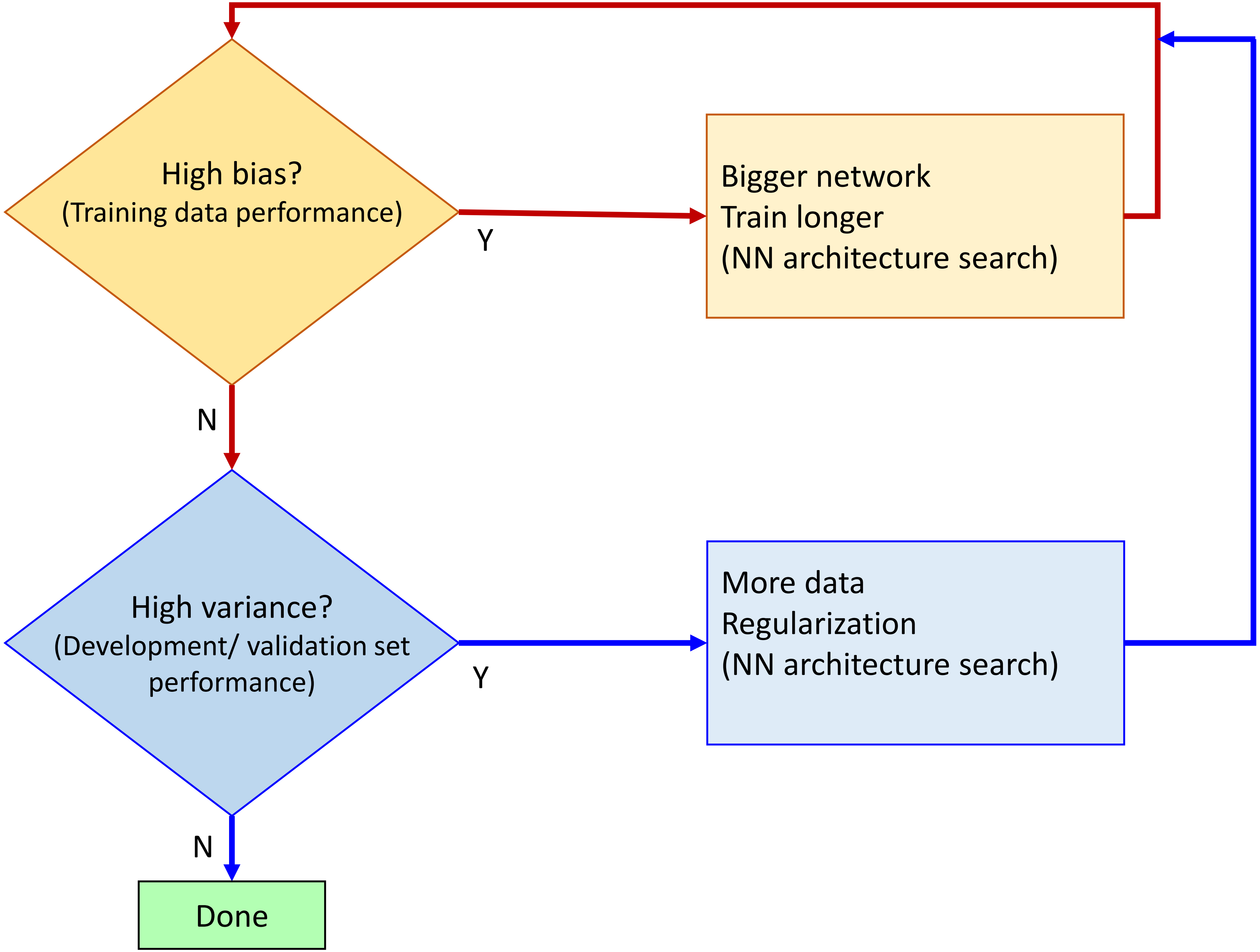


(Underfitting)

(Overfitting)

Training error	1%	15%	15%	0.5%
Testing error	11%	16%	30%	1%
Infer	High variance	High bias	High bias & High variance	Low bias & Low variance

Assumption: Human/ optimal (Bayes’) error ≈ 0%


High bias?
(Training data performance)
Y
Bigger network
Train longer
(NN architecture search)
N
High variance?
(Development/ validation set performance)
Y
More data
Regularization
(NN architecture search)
N
Done

Regularization

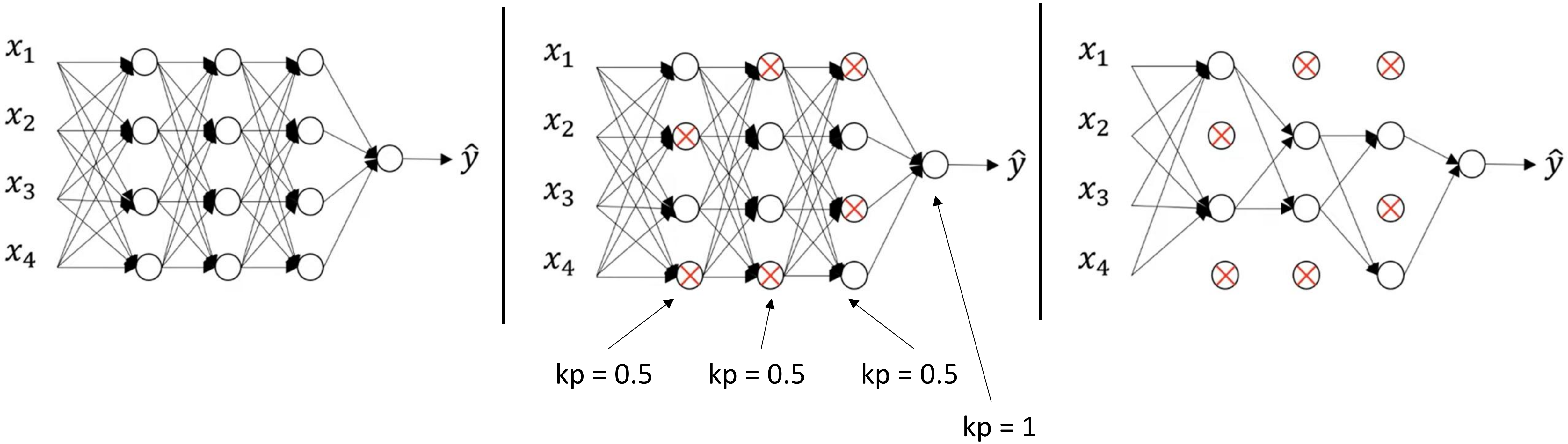
$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\mathcal{L}\left(\hat{y}^{(i)}, y^{(i)}\right) + \frac{\lambda}{2m}\|w\|^2 + \frac{\lambda}{2m}\|b\|^2$$

L2 regularization = $\|w\|^2 = \sum_{j=1}^{n_x} w_j^2 = w^T w$

L1 regularization = $\|w\|_1 = \sum_{j=1}^{n_x} |w_j|$

λ increases: underfitting
λ decreases: overfitting

Dropout Regularization



keep_prob (kp) = 1 → No Dropout

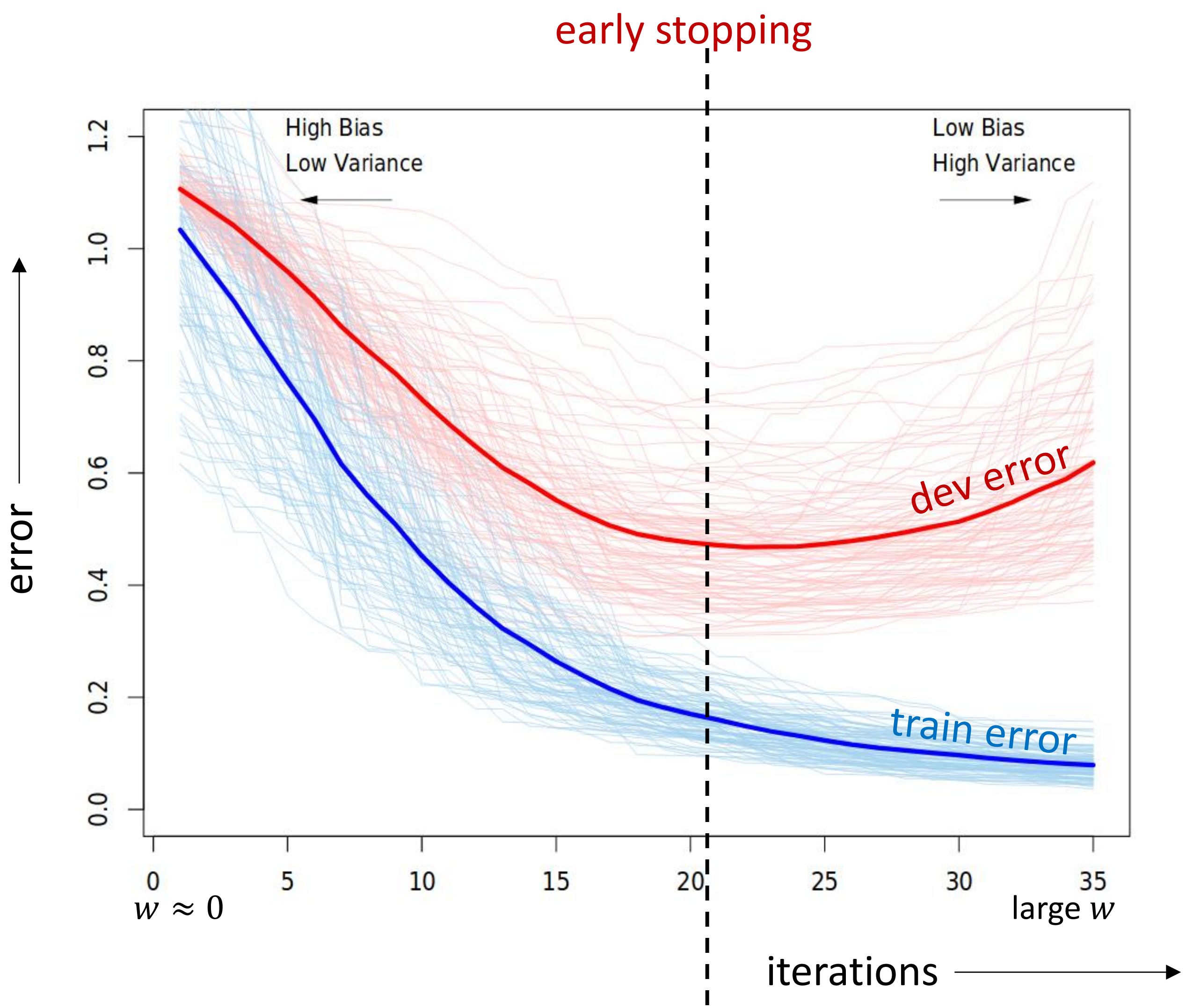
Intuition: Can’t rely on any one set of features, so have to spread-out weights

Data Augmentation

Data Augmentation

https://arxiv.org/pdf/1809.06839v1.pdf

Early Stopping



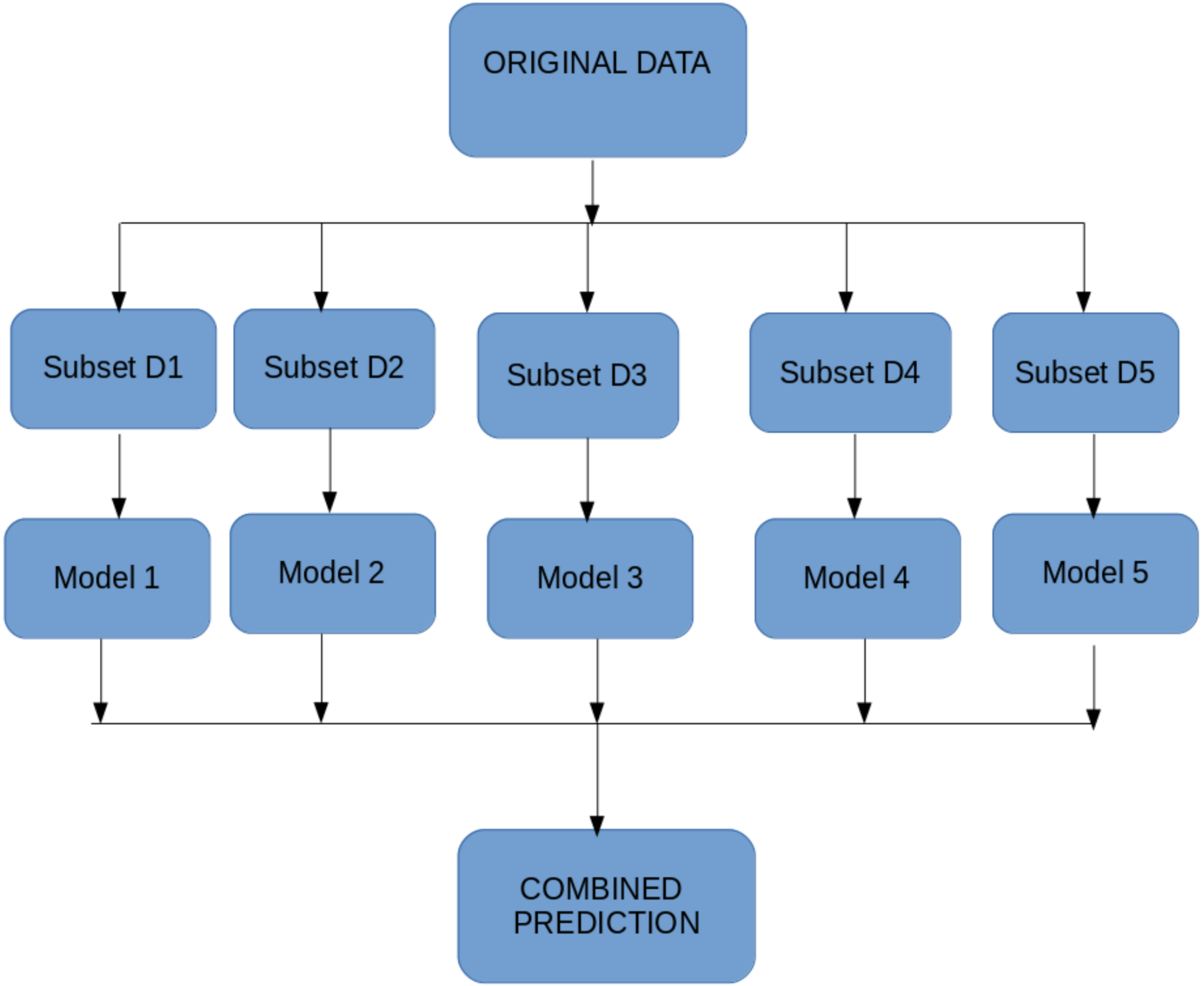
Ensemble for Generalization

- Train on random subsets of the full training data. This procedure is known as **bagging**.

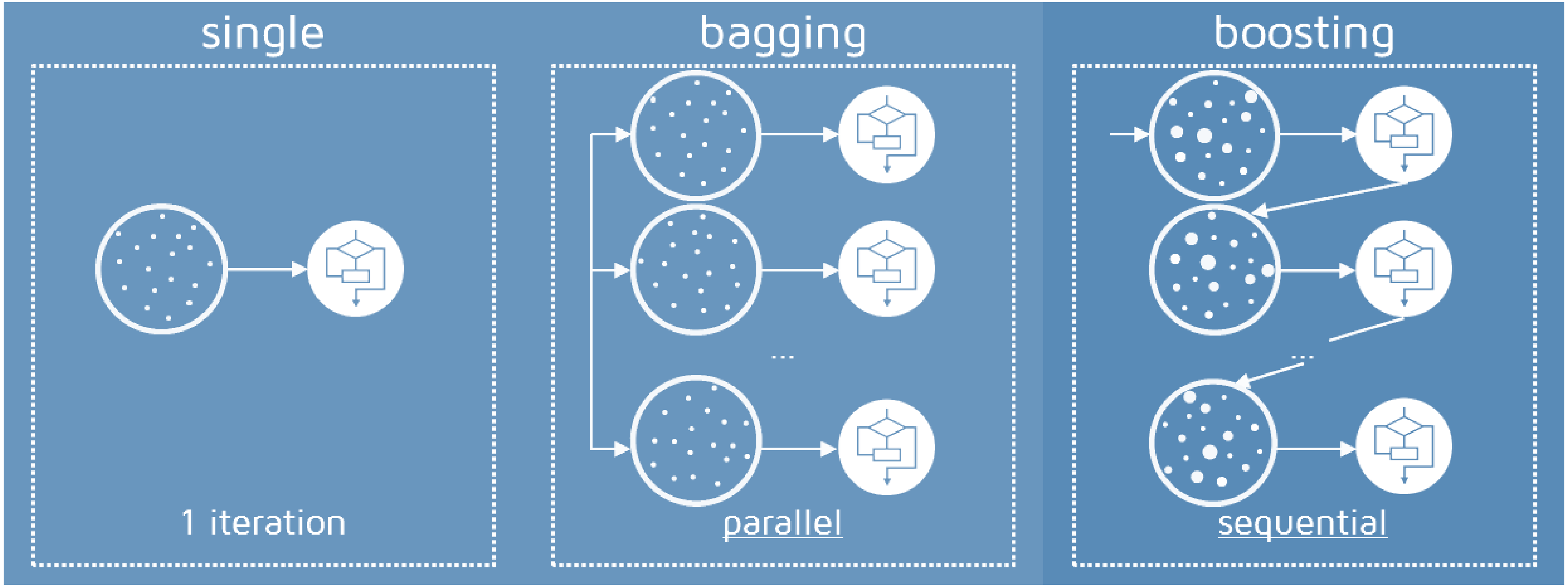
- Train networks with different architectures (e.g. different numbers of layers or units, or different choice of activation function).

- Use entirely different models or learning algorithms.

Bagging


ORIGINAL DATA
Subset D1
Subset D2
Subset D3
Subset D4
Subset D5
Model 1
Model 2
Model 3
Model 4
Model 5
COMBINED PREDICTION

Bagging vs. Boosting



Bagging tries to solve the over-fitting problem.
Boosting tries to reduce bias.

Thank You!