Machine Learning

- Gradient Descent Algorithm
- Linear Regression
- Non-Linear Regression
- Logistic Regression
- Decision Trees
    - Regression Trees
    - Classification Trees
- Clustering Algorithms
    - K-Means
    - Hierarchical clustering
    - DB-Scan
    - Mean Shift
    - GMM
- Support Vector Machine

Deep Learning

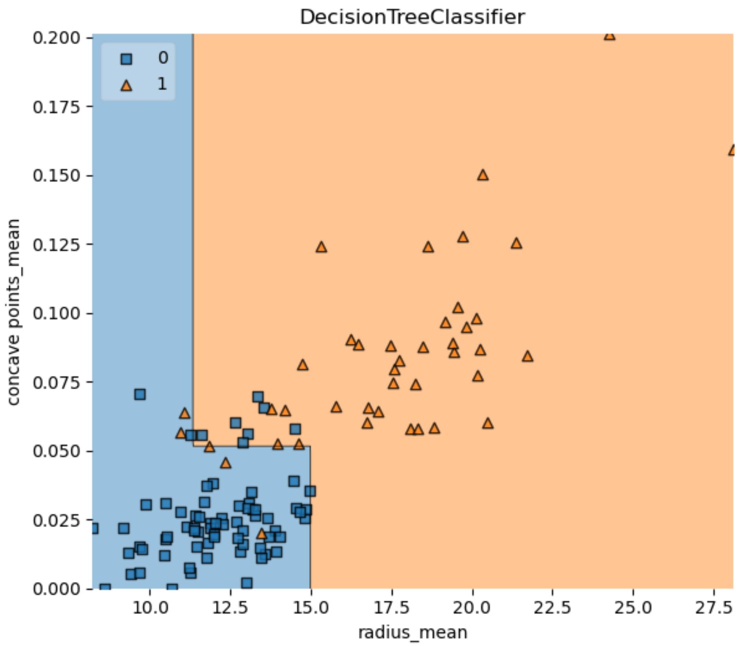
- MLP
- CNN

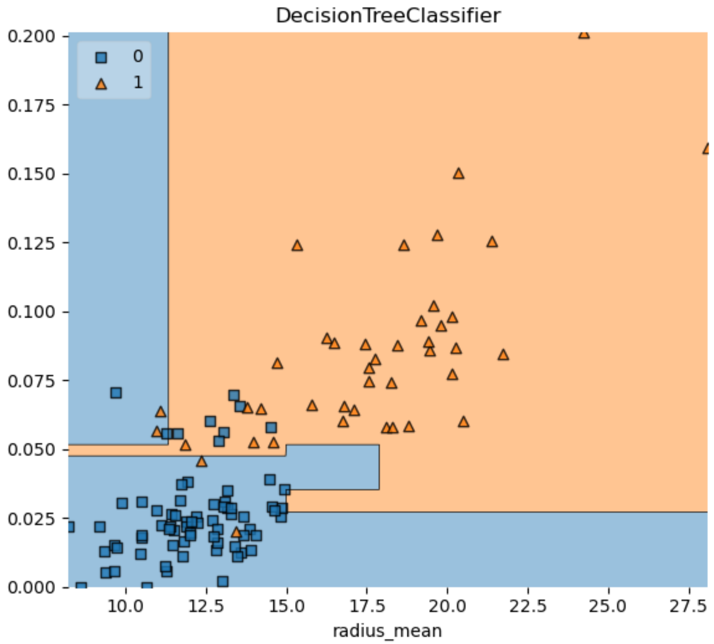
Datasets

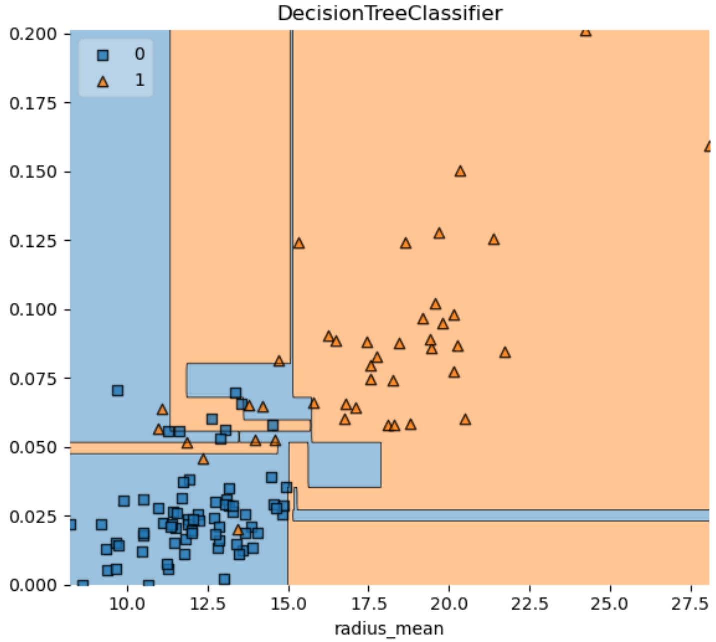
- Breast Cancer Wisconsin
- MIMIC-III
- Framingham Heart Study
- Alzheimer’s Disease Neuroimaging Initiative
- Drug discovery
- Microbiome

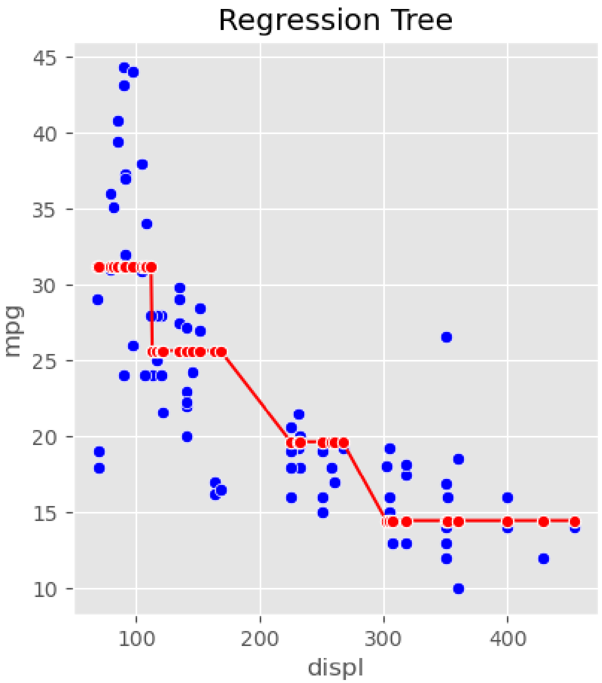
Model complexity

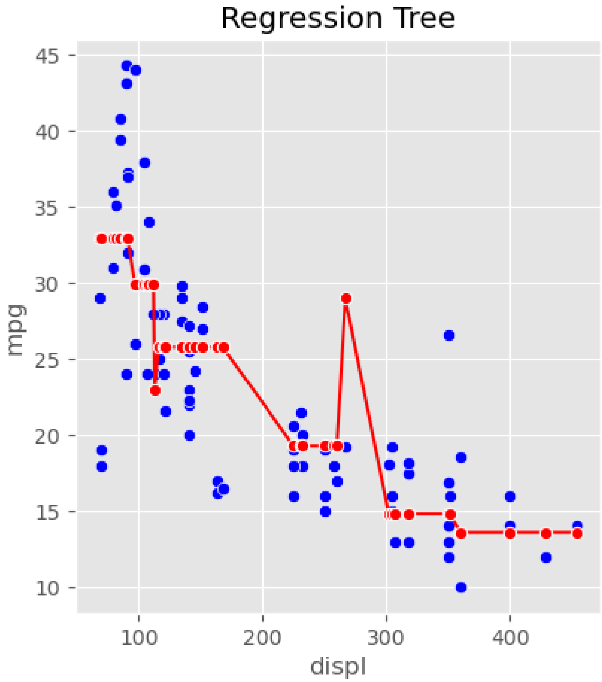
my humble ML course

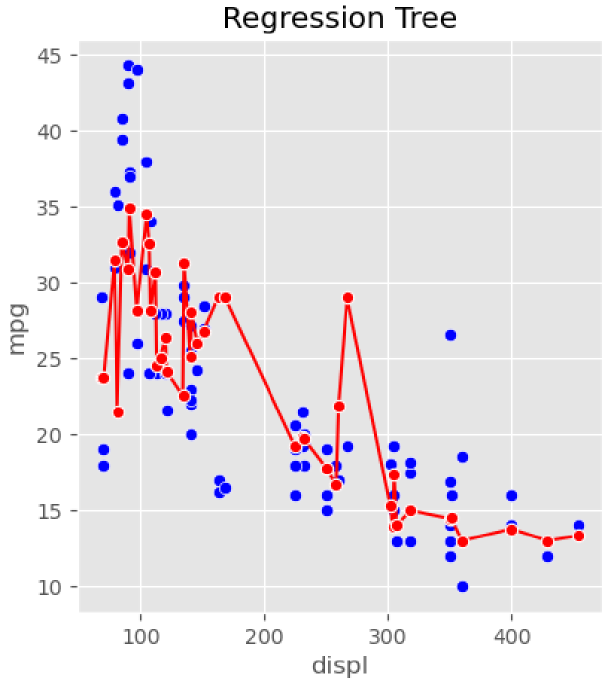






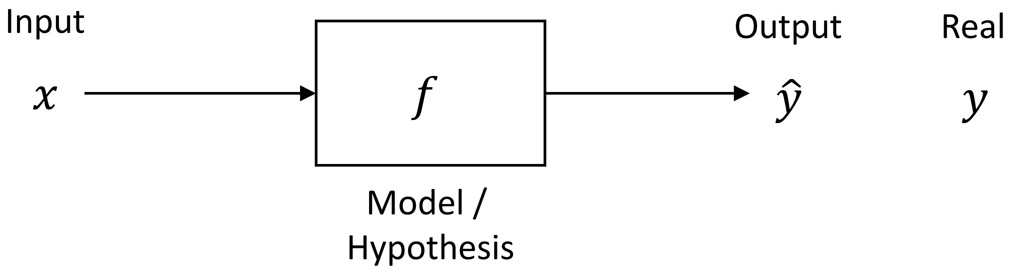






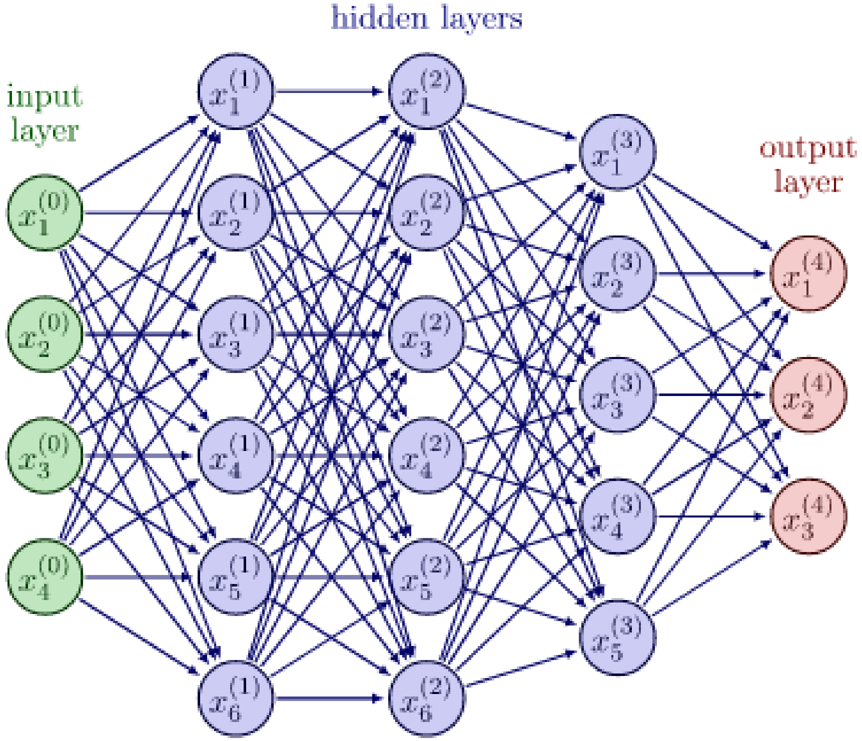
Model complexity

- supervised learning
- find a model $\hat{f}$ that best approximates f : $\boxed{\hat{f}} \approx \boxed{f}$
- $\hat{f}$ should achieve a low predictive error on unseen datasets.



7 → 10 → 100

2D3D

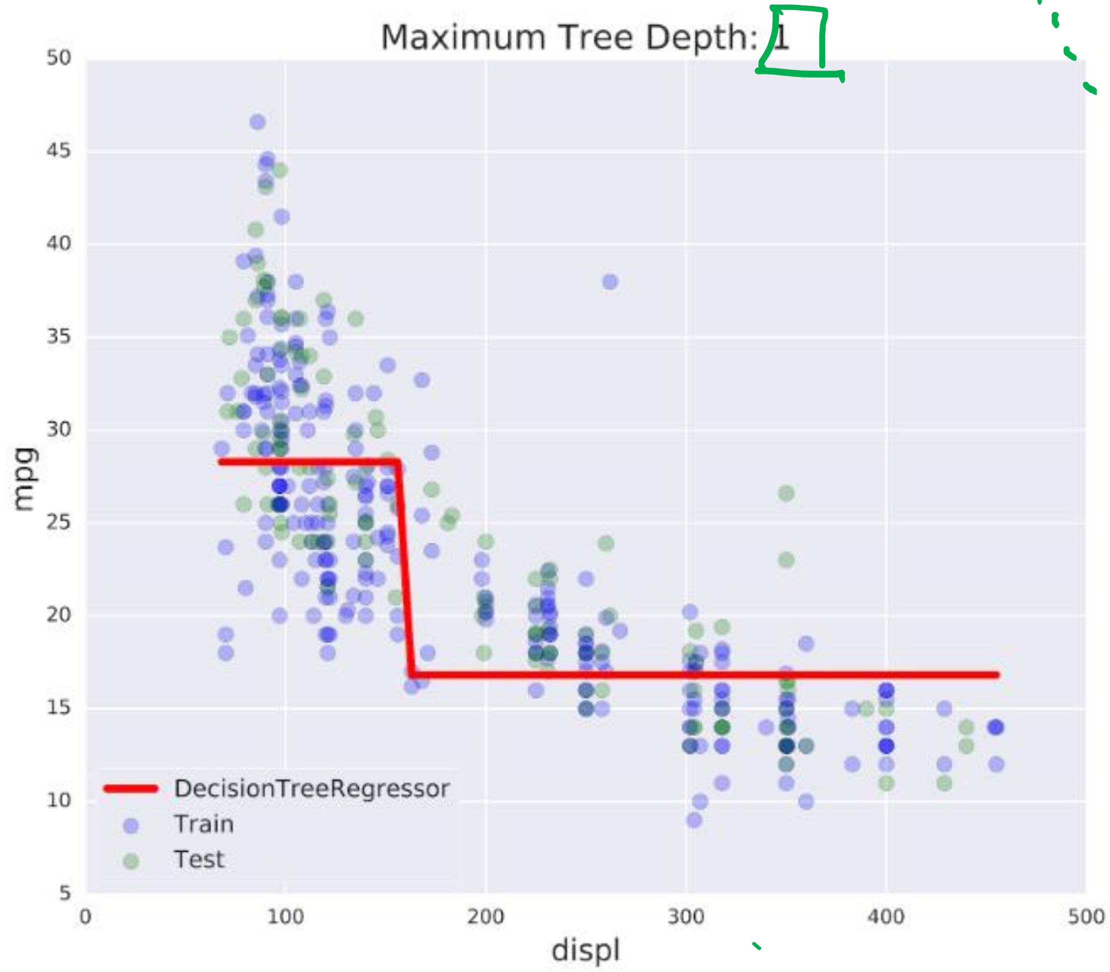


.2

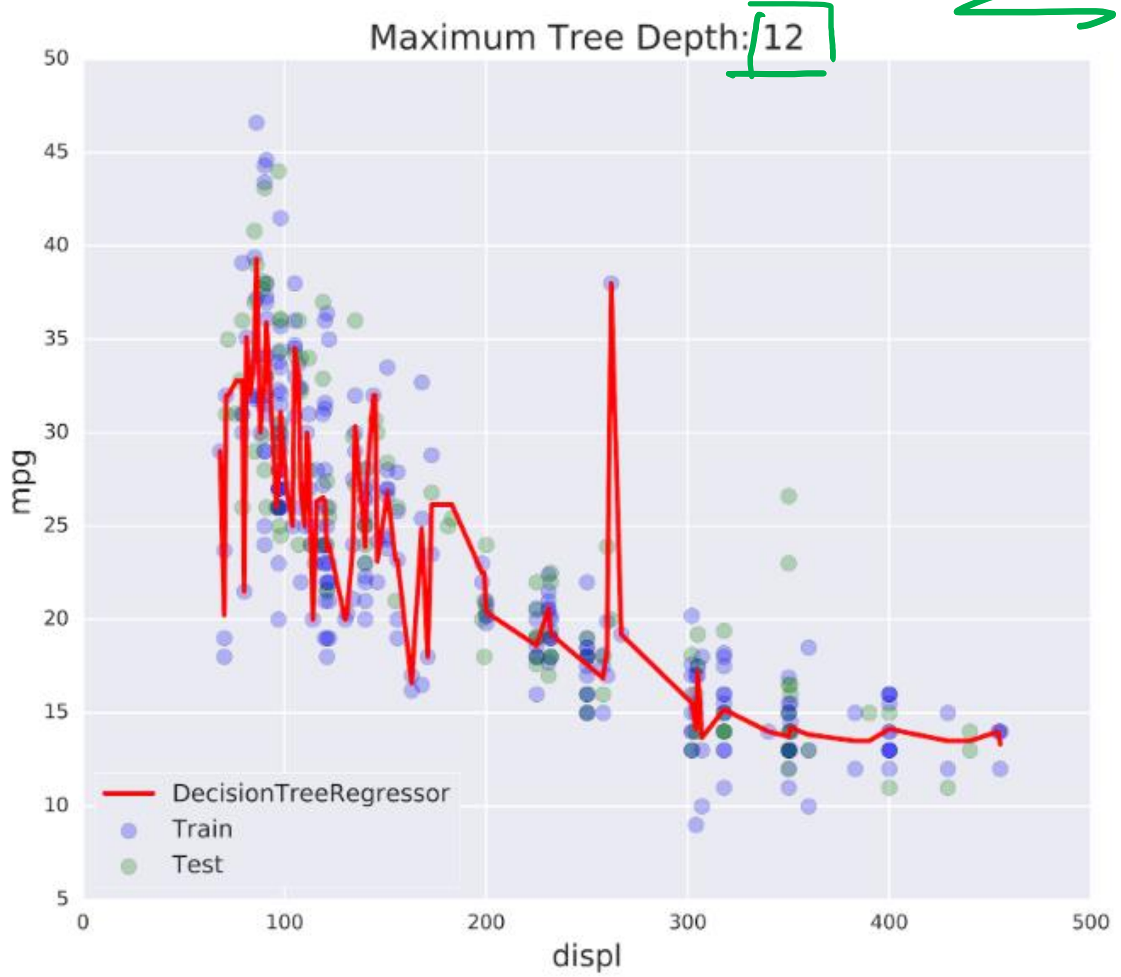
https://www.deep-mind.org/2023/03/26/the-universal-approximation-theorem/

Model complexity

learning rate



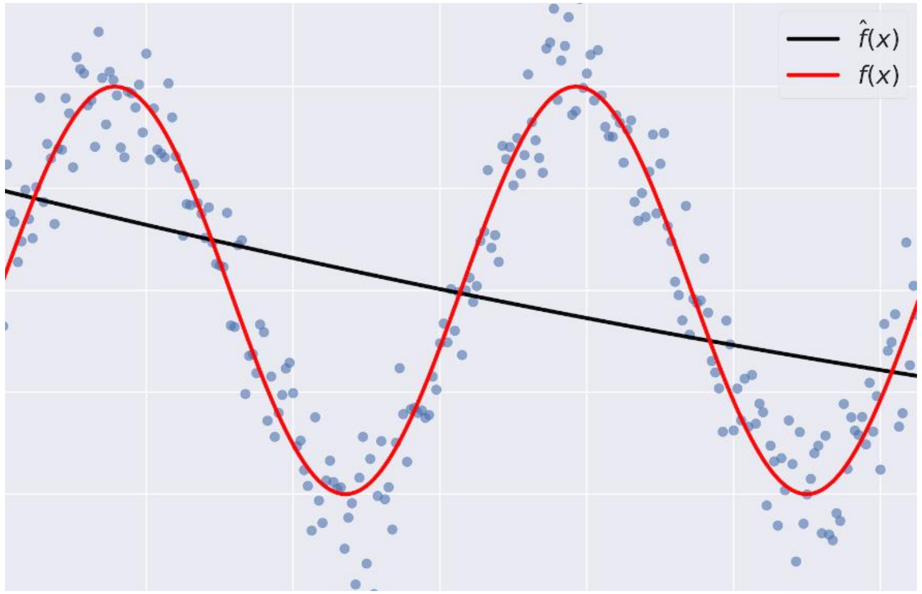
$\hat{f}$ is not flexible enough to approximate f



$\hat{f}$ fits the training set noise

Bias

- bias is the difference between the average prediction of our model and the correct value which we are trying to predict.
- how much $\hat{f} \neq f$



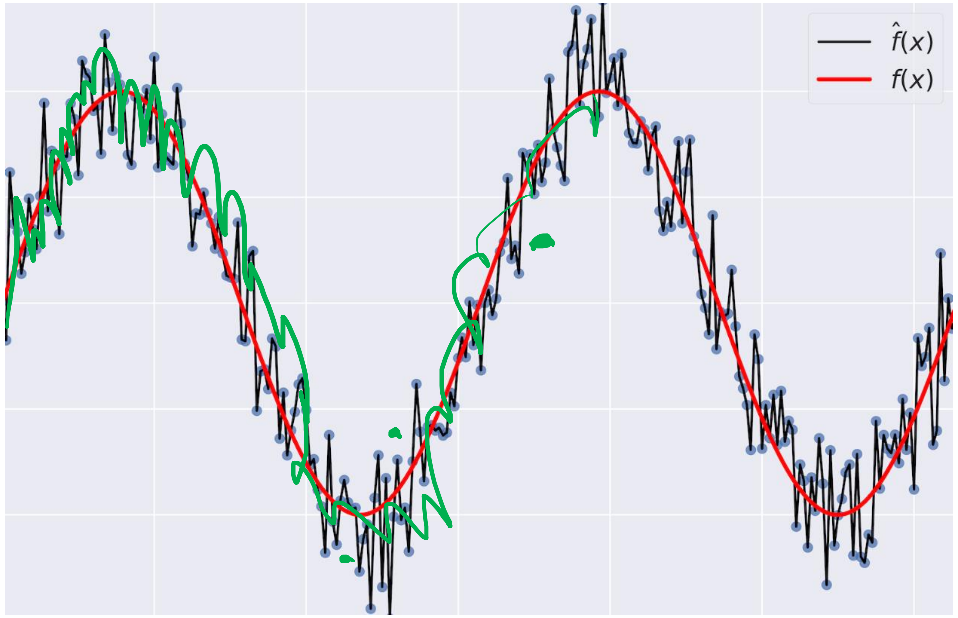
high bias: very little attention to the training data and oversimplifies the model

underfitting

low bias: …

Variance

- variability of model prediction for a given data point or a value which tells us spread of our data.
- how much $\hat{f} \neq f$



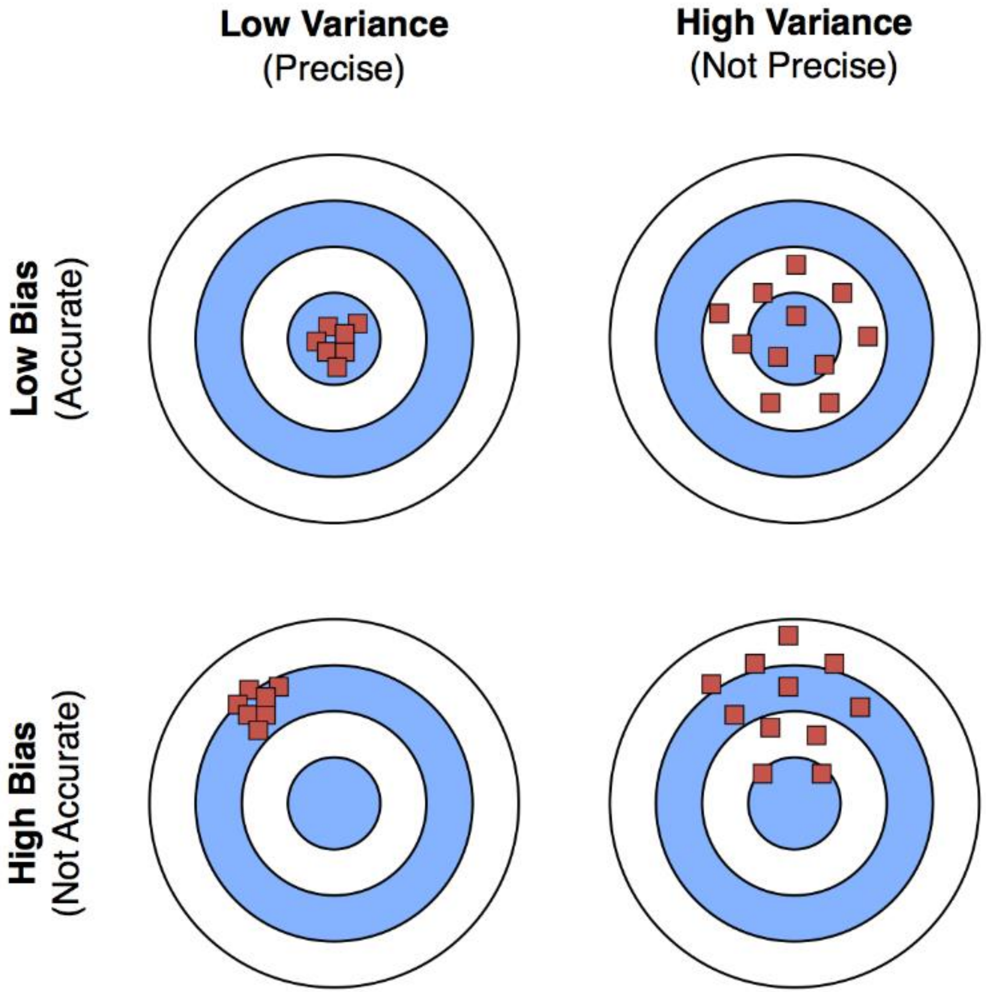
high variance: pays a lot of attention to training data and does not generalize on the data which it hasn't seen before

overfitting

low variance: …

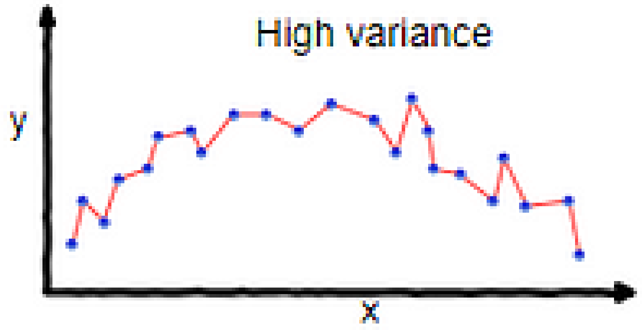
bias-variance tradeoff

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

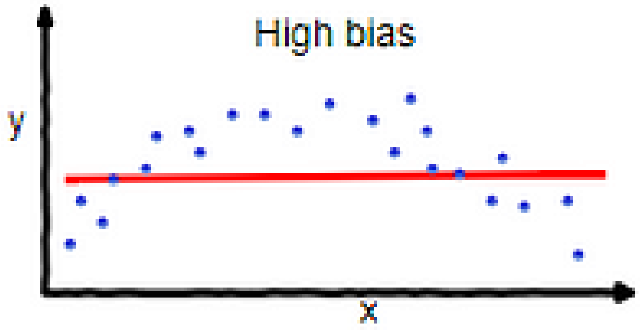


bias-variance tradeoff

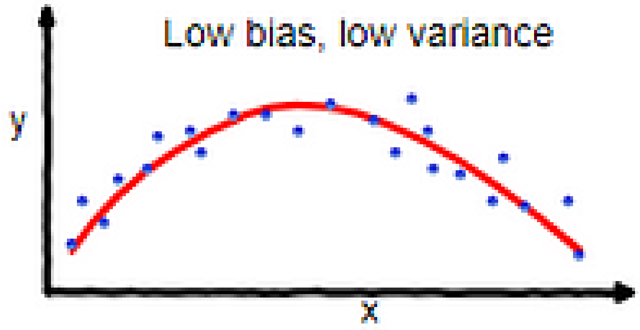
$$Err(x) = \text{Bias}^2 + \text{Variance} + \boxed{\text{Irreducible Error}}$$



overfitting

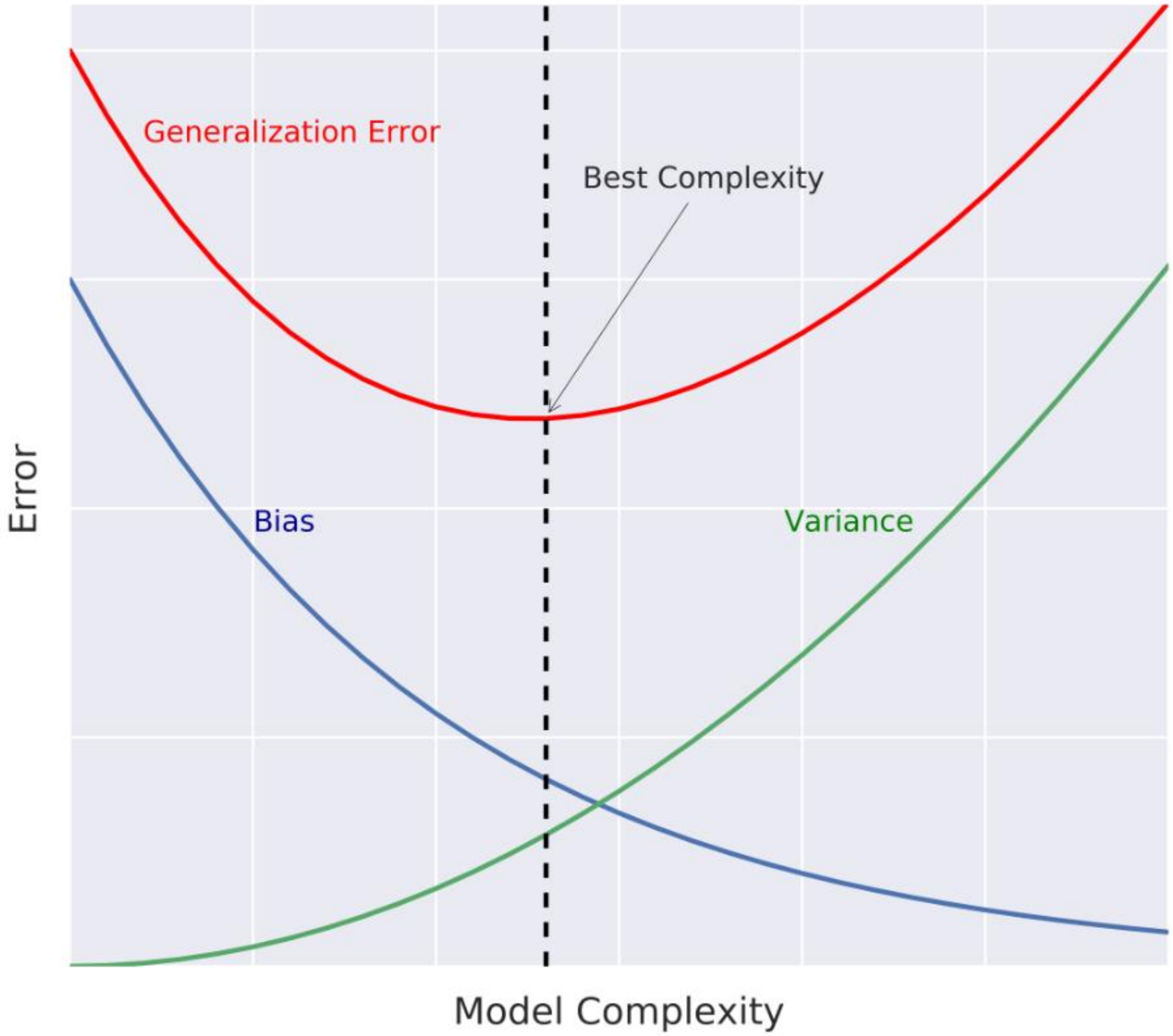


underfitting



Good balance

bias–variance tradeoff



bias-variance tradeoff

When the model complexity increases, the variance increases while the bias decreases

When the model complexity decreases, the variance decreases while the bias increases.

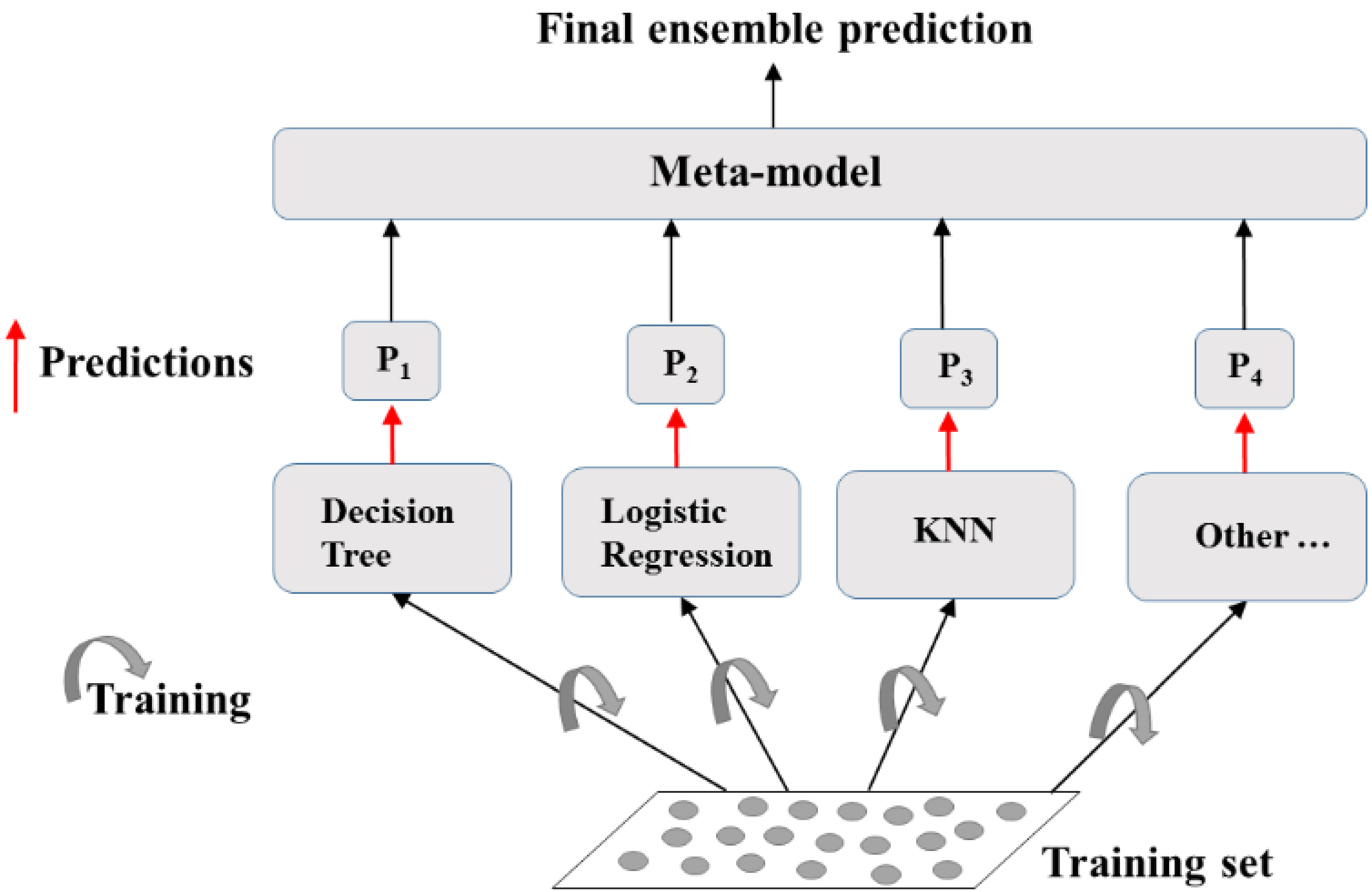
$$Err(x) = \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{Irreducible\ Error}$$

Since this error is the sum of three terms with the irreducible error being constant, you need to find a balance between bias and variance because as one increases the other decreases. This is known as the bias-variance trade-off.

in context …

CART in practice:

- Can easily overfit
- How do we regularize?
  - Can stop if gain is small
  - But this can be short-sighted
- Can use tree pruning
- Can use ensemble methods.



in context …

Algorithm	Bias	Variance
Linear Regression	High Bias	Less Variance
Decision Tree	Low Bias	High Variance
Bagging	Low Bias	High Variance (Less than Decision Tree)
Random Forest	Low Bias	High Variance (Less than Decision Tree and Bagging)

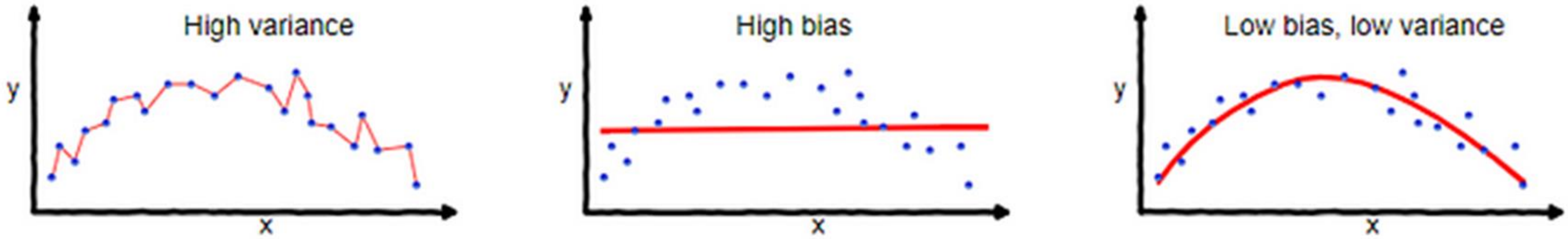
Diagnose bias and variance problems

Generalization error

How do we estimate the generalization error of a model?

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

- ML algorithms -> data-driven models
- Usually you only have one dataset

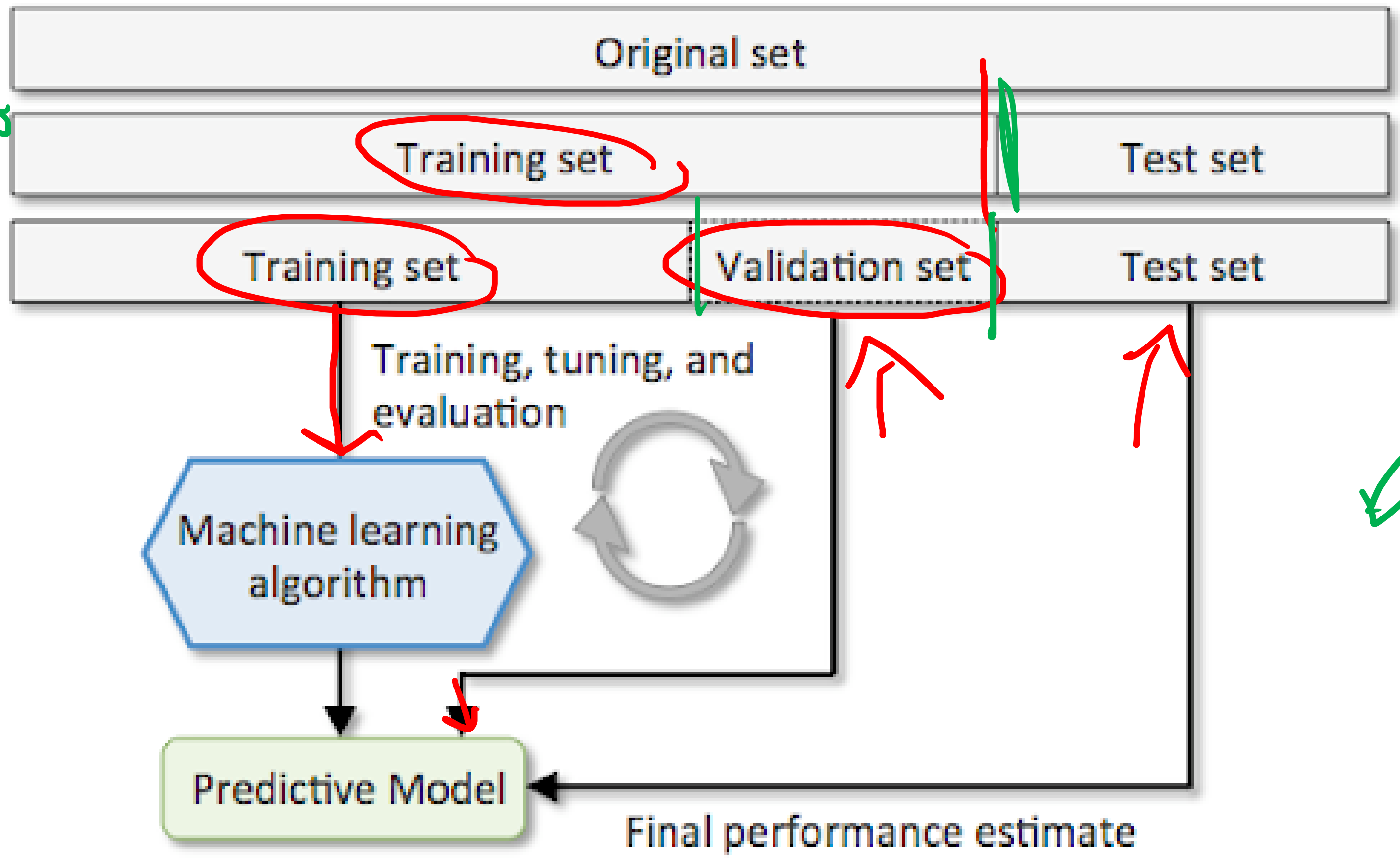


Hold Out CV

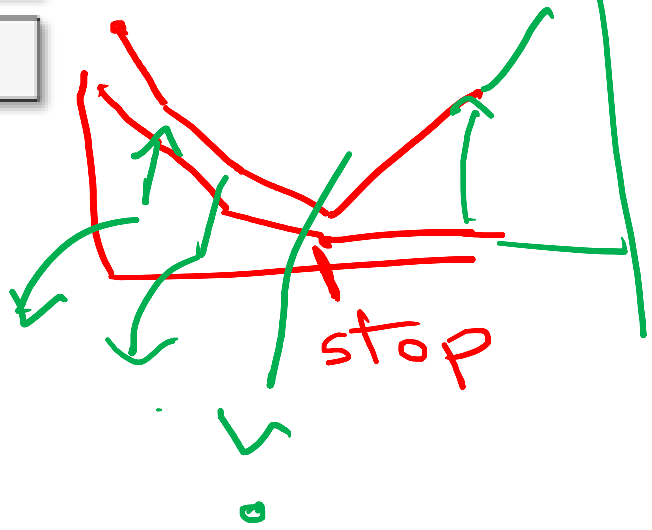
ingeniería de datos

500

estratificado

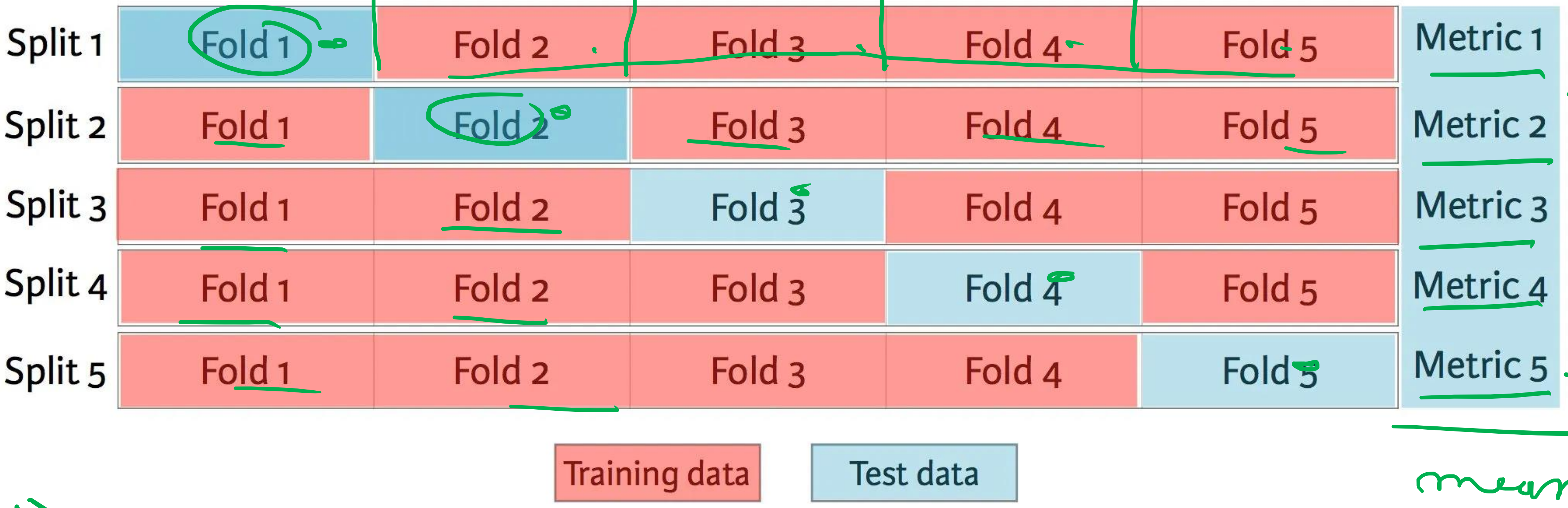


overfitting



K Fold CV

5



mean(   )

Diagnose problems

Bias problems

High bias (underfitting)

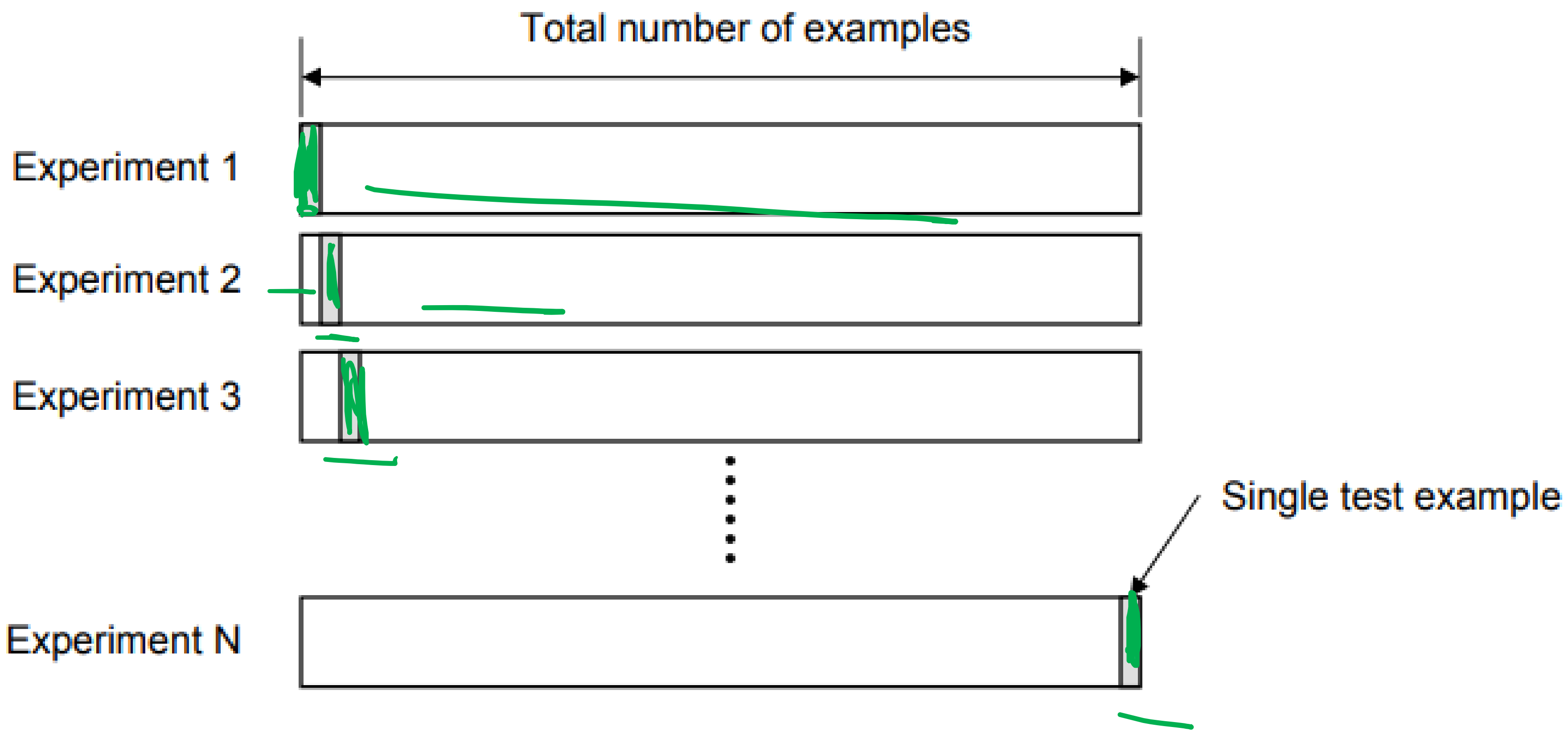
$$\text{CV error} \approx \text{training set error} \gg \boxed{\text{desired error}}$$

Variance problems

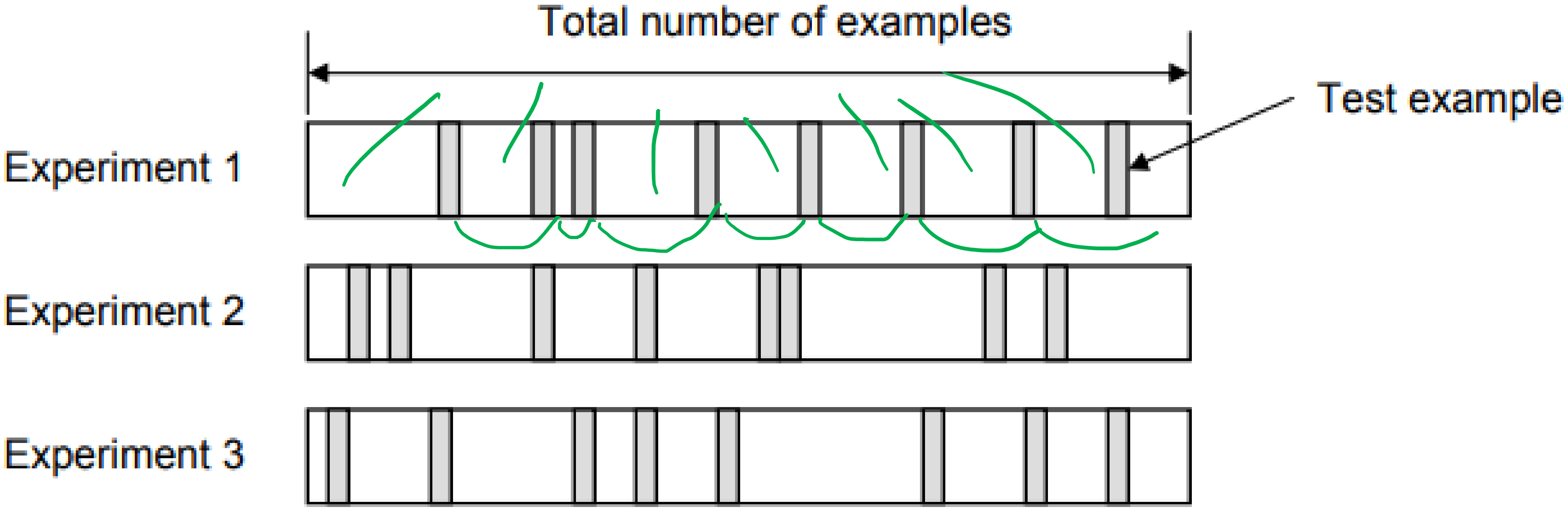
High variance (overfitting)

$$\text{CV error} > \text{training set error}$$

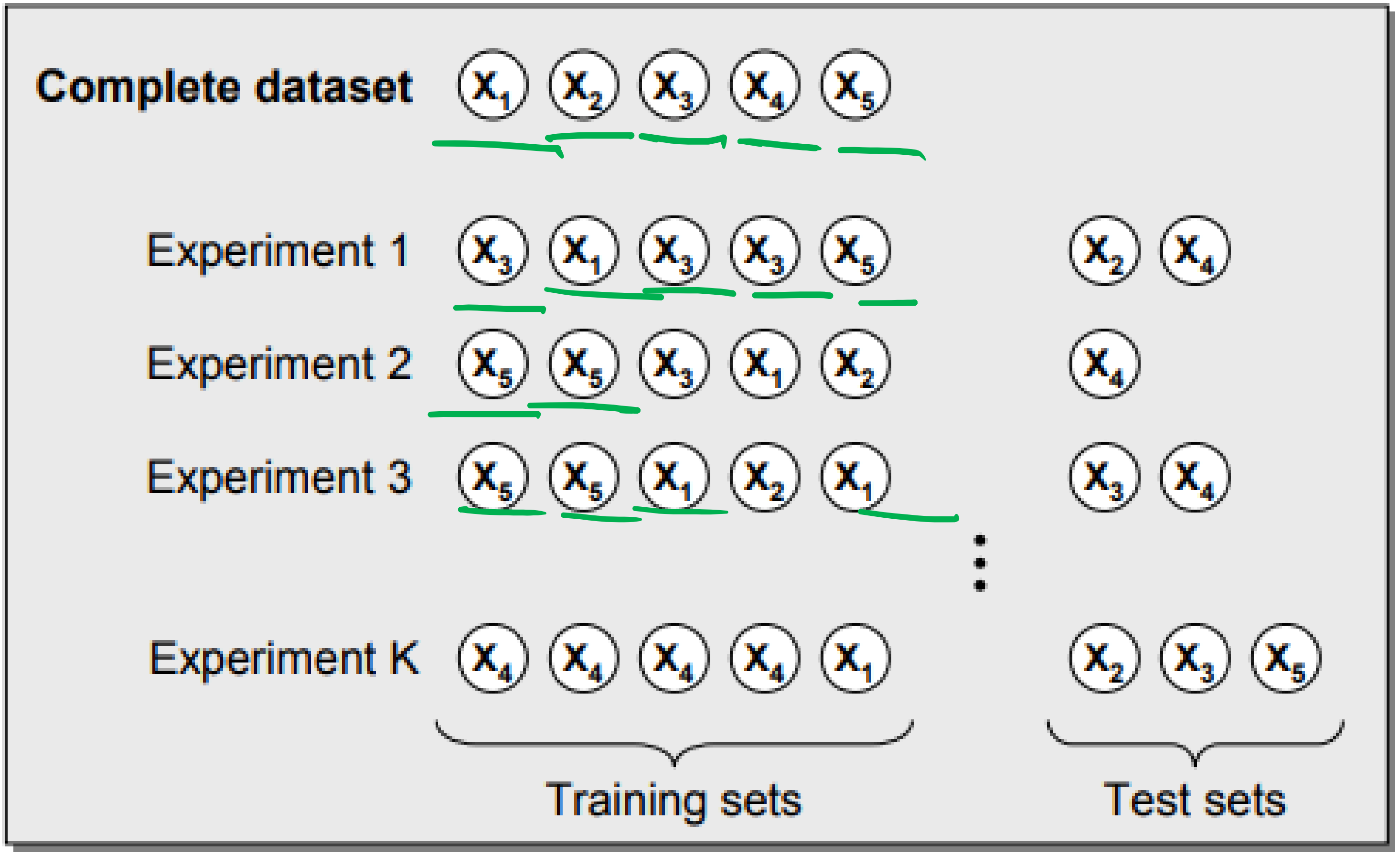
Leave-One-Out CV (LOOCV)



Random Subsampling



Bootstrapping



Bibliography

https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://app.datacamp.com/learn/courses/machine-learning-with-tree-based-models-in-python
https://www.bmc.com/blogs/bias-variance-machine-learning/
https://medium.com/@eijaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f
https://moredvikas.wordpress.com/2018/10/10/machine-learning-model-validation-techniques/