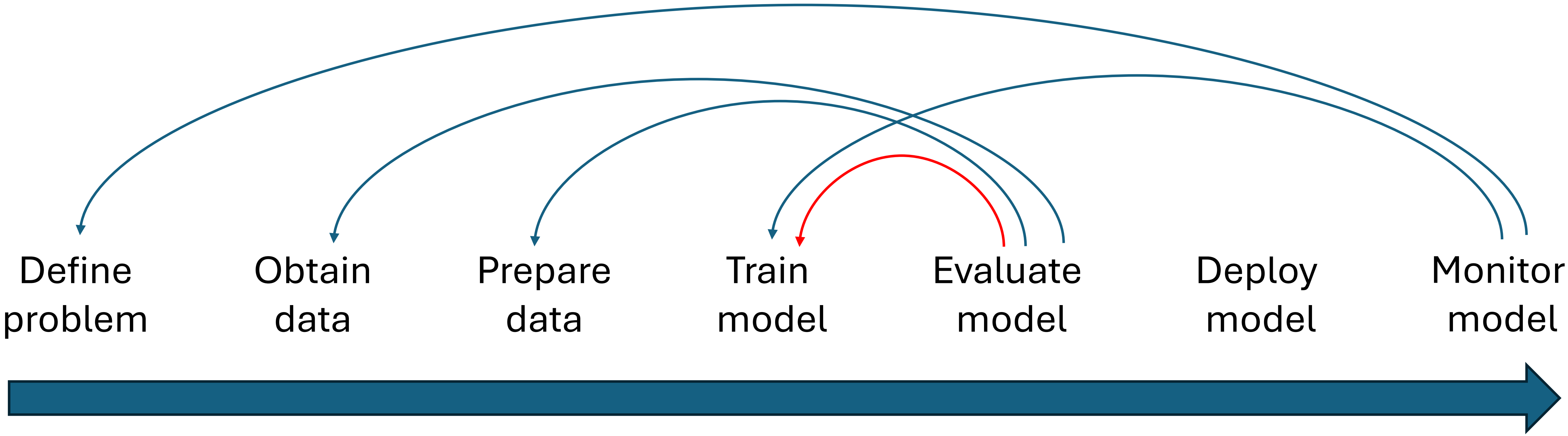
AI technology – part 2

Sonja Aits

Lund University

2025-09-23

AI workflow


Define problem
Obtain data
Prepare data
Train model
Evaluate model
Deploy model
Monitor model

Define the problem

- What is the goal?

Define the problem

- What is the goal?
- What is the scope?

Define the problem

- What is the goal?
- What is the scope?
- How will you <u>measure</u> success?

Understanding project context

- How have others approached this?
- What resources and skills do you have?
- Operational requirements?

Making an initial plan

- What AI approach will you use?
  - AI task
  - Learning paradigm
  - Model types and architectures
  - Evaluation strategy

Setting clear goals

Bad

AI model for diagnosing pancreatic cancer

Good

AI model for classifying CT images from XZY machine as pancreatic cancer vs healthy with 99% sensitivity and 95% specificity

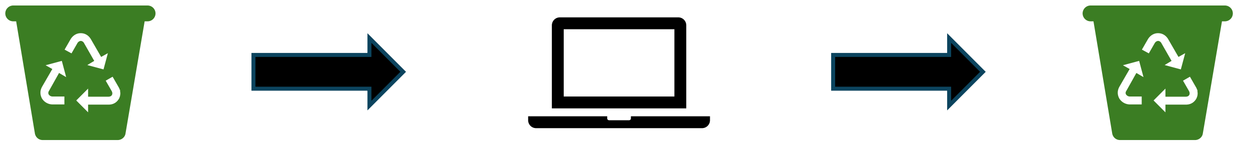
Do you have ideas for your project goal?

Discuss with neighbor

Obtain the data

- What type of data will be needed?
- Is the data sensitive?
- How much data will be needed?
- What problems do you foresee?
- Are there commonly used benchmarking datasets?

The Golden Rule of Data Science



"Garbage in – Garbage out"

What kind of problems have you encountered in data?

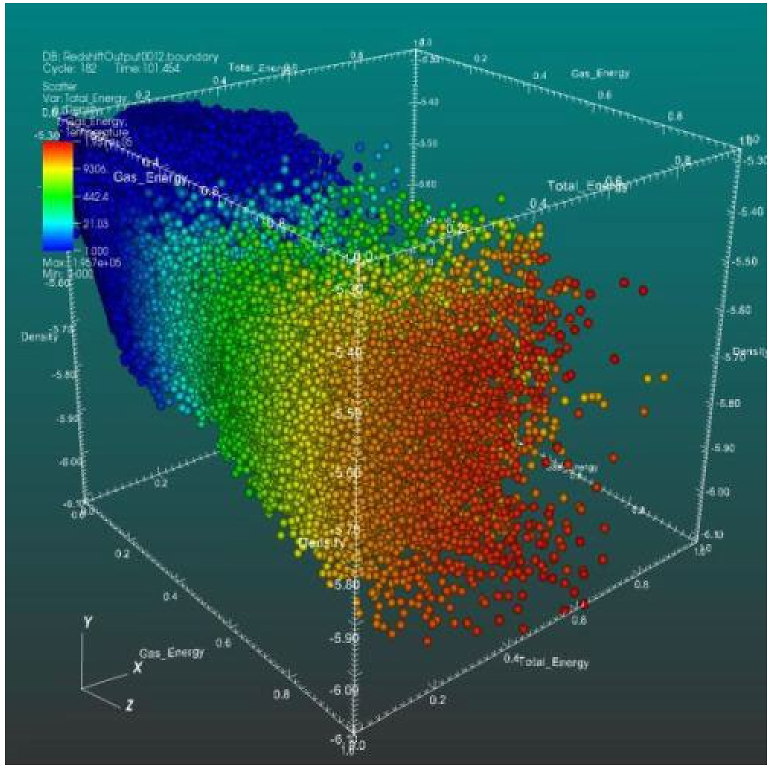
Before you prepare the data, you need to understand it!

Exploratory Data Analysis

- **Manual inspection of raw data!**
- Visualization
- Statistical analysis

→ Missing values

→ Flaws



How do you examine your data?

Data preprocessing

- Data exclusion
- Normalization (0 to 1, -1 to 1)
- Imputation of missing values
- Dimensionality reduction
- Encoding
- Augmentation
- Splitting

Encoding

- All data must be converted to numerical data
- Many ways to do this!

Binary encoding

Cancer	Healthy
1	0

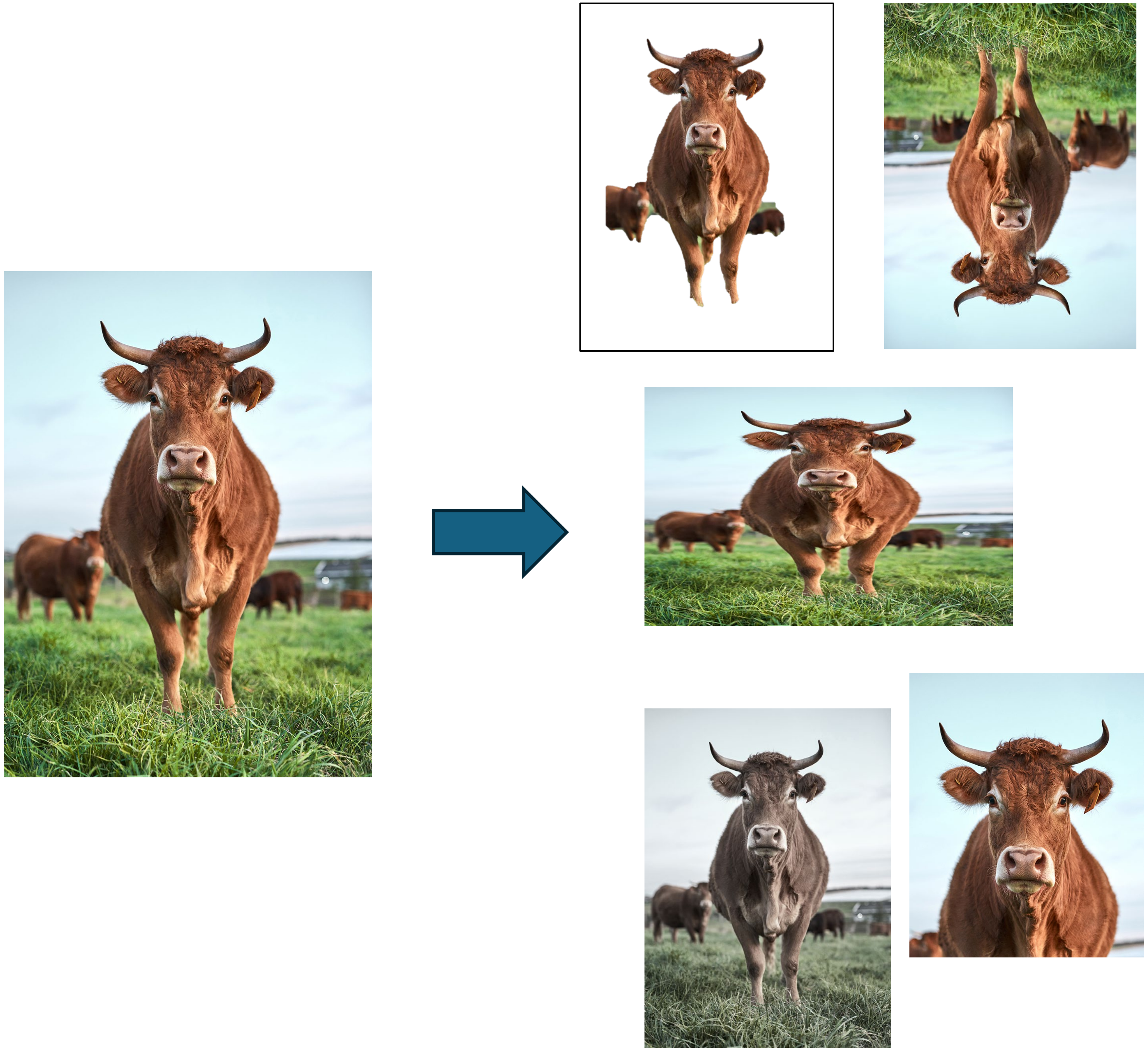
One-hot encoding

Sample	Color
1	red, blue
2	red, green
3	green
4	blue, green

Sample	R	B	G
1	1	1	0
2	1	0	1
3	0	0	1
4	0	1	1

Augmentation

- Introduces variation
- Increases dataset size
- Can balance data



How could you augment your data?

Ultimate training goal

Train model that performs well on previously **unseen data**

→ Model can **generalize**

Splitting data

- **Training set**: to train models
- **Validation set (Development set)**: to evaluate trained models during development
- **Test set**: to assess generalization after development

Train	Validate	Test

- Train, validation and test set must be independent (no related examples)
- Similar data distribution in train, validation and test set
- Balance sample groups

Splitting data

- **Training set**: to train models
- **Validation set (Development set)**: to evaluate trained models during development
- **Test set**: to assess generalization after development
- Additional independent test data is desirable

K-fold cross validation can be used when datasets are small

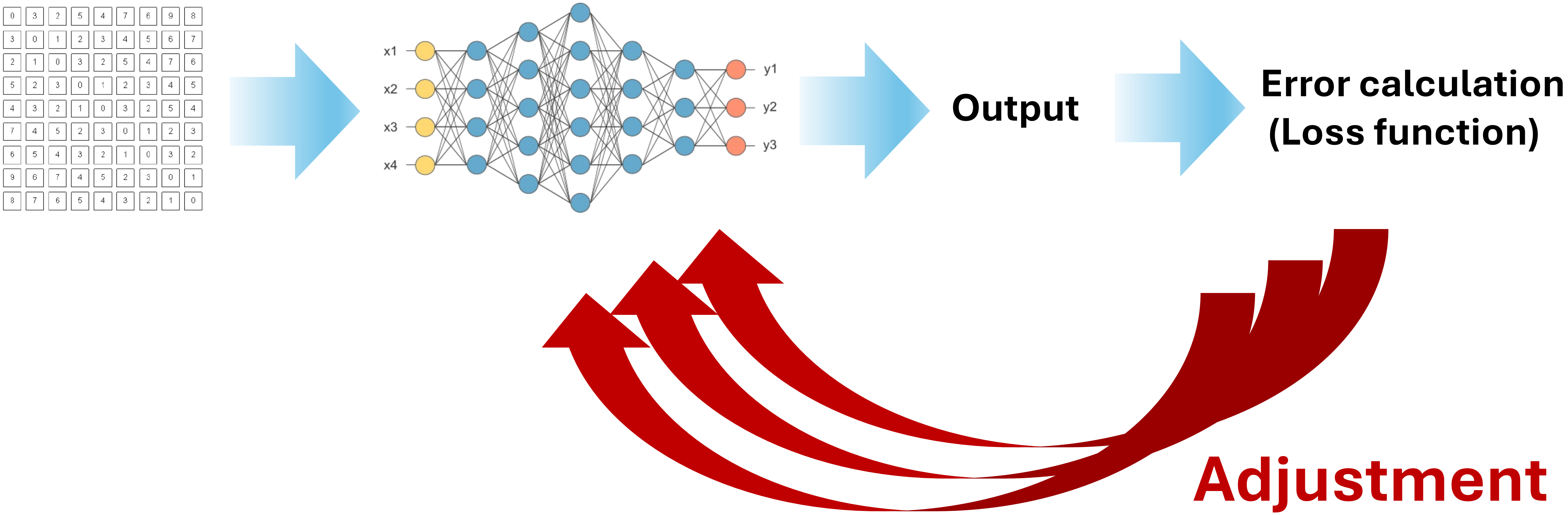
Train	Train	Train	Validate	Test
Validate	Train	Train	Train	Test
Train	Validate	Train	Train	Test
Train	Train	Validate	Train	Test
				Independent test data

How would you split this data for 5-fold cross validation?

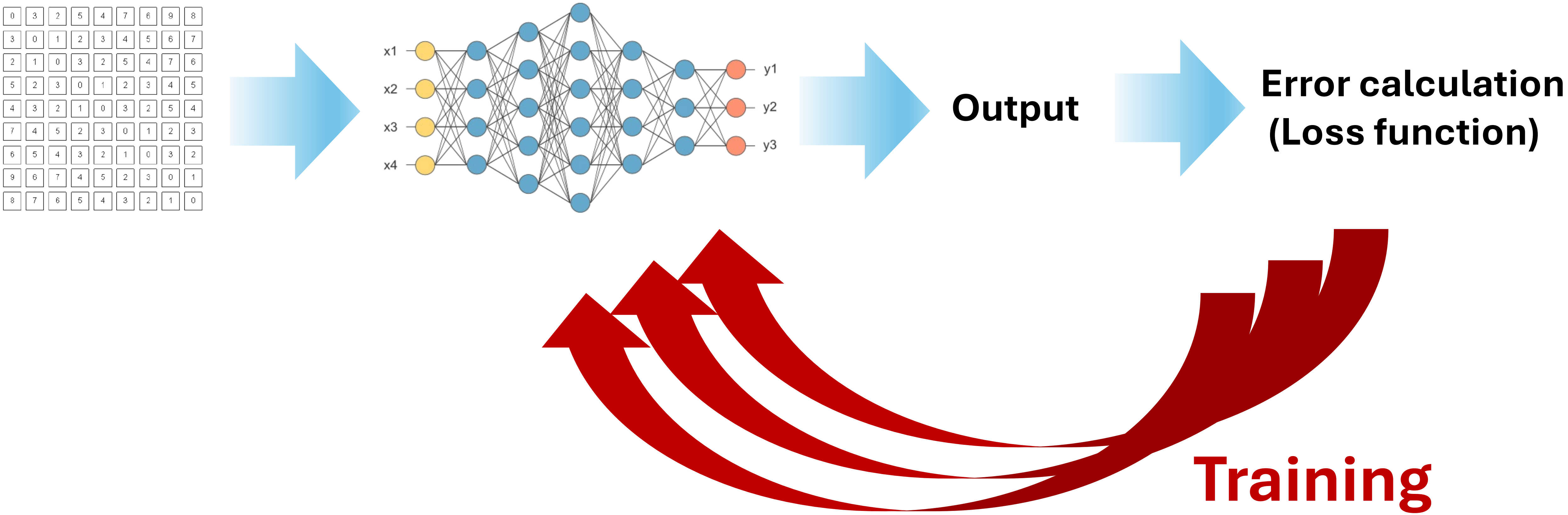
- 4 patient groups, 5 patients each
- 3 time points per patient
- 2 sample taken each time point from same patient

Model training

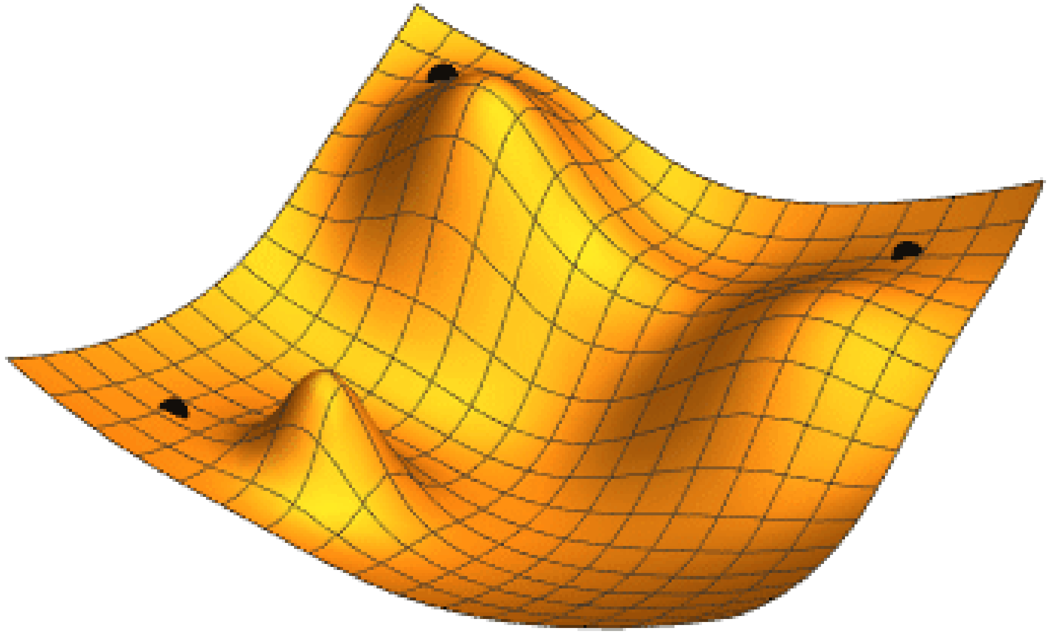
Weights and other parameters are <u>repeatedly</u> adjusted to reduce errors



Weights and other parameters are repeatedly adjusted to reduce errors

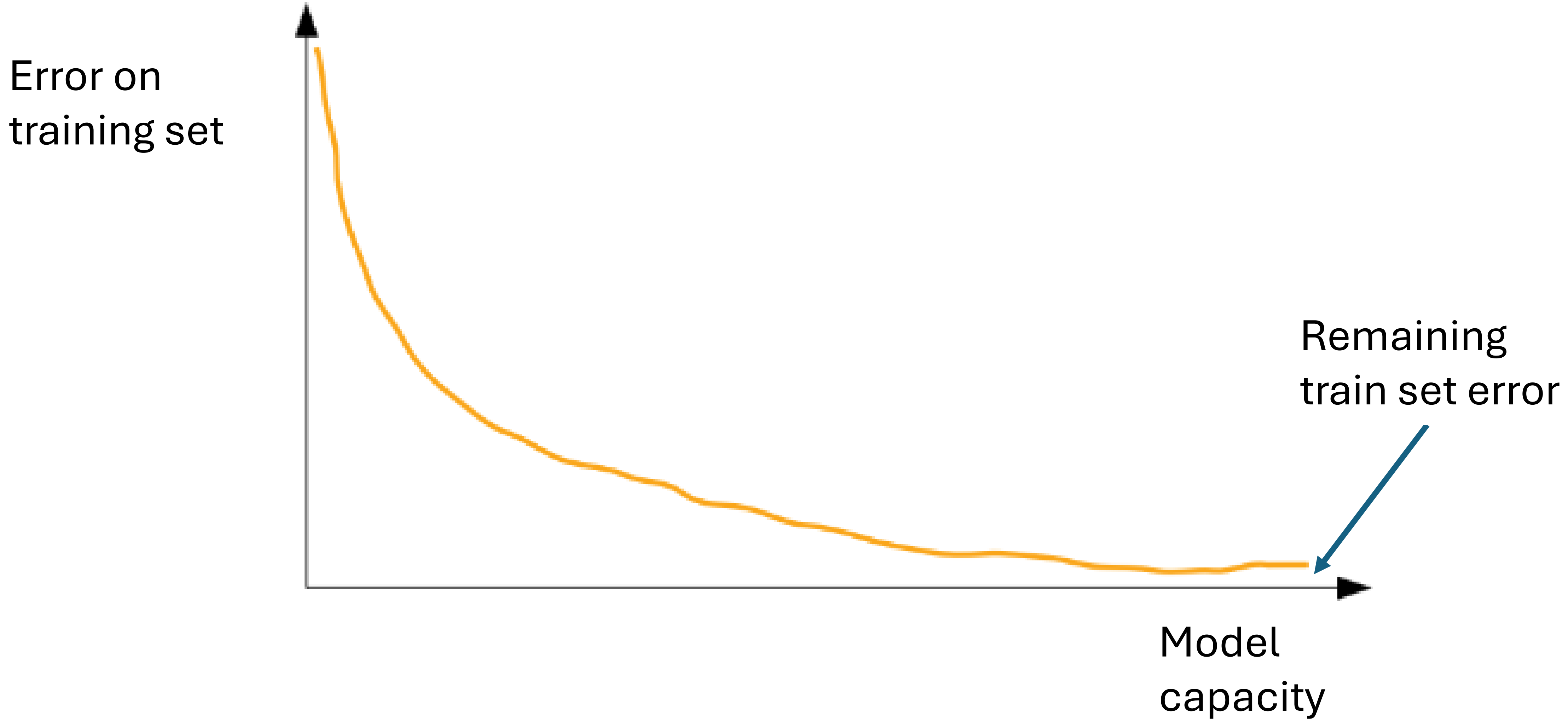


Parameters are adjusted by gradient descent



https://commons.wikimedia.org/wiki/File:Gradient_descent.gif

Training goal:
Minimizing the training set error



Model capacity is dependent on network architecture

- Number of hidden layers
- Number of nodes/layer
- Network components

→ “**Hyperparameters**”: optimized by experimentation

Time to play

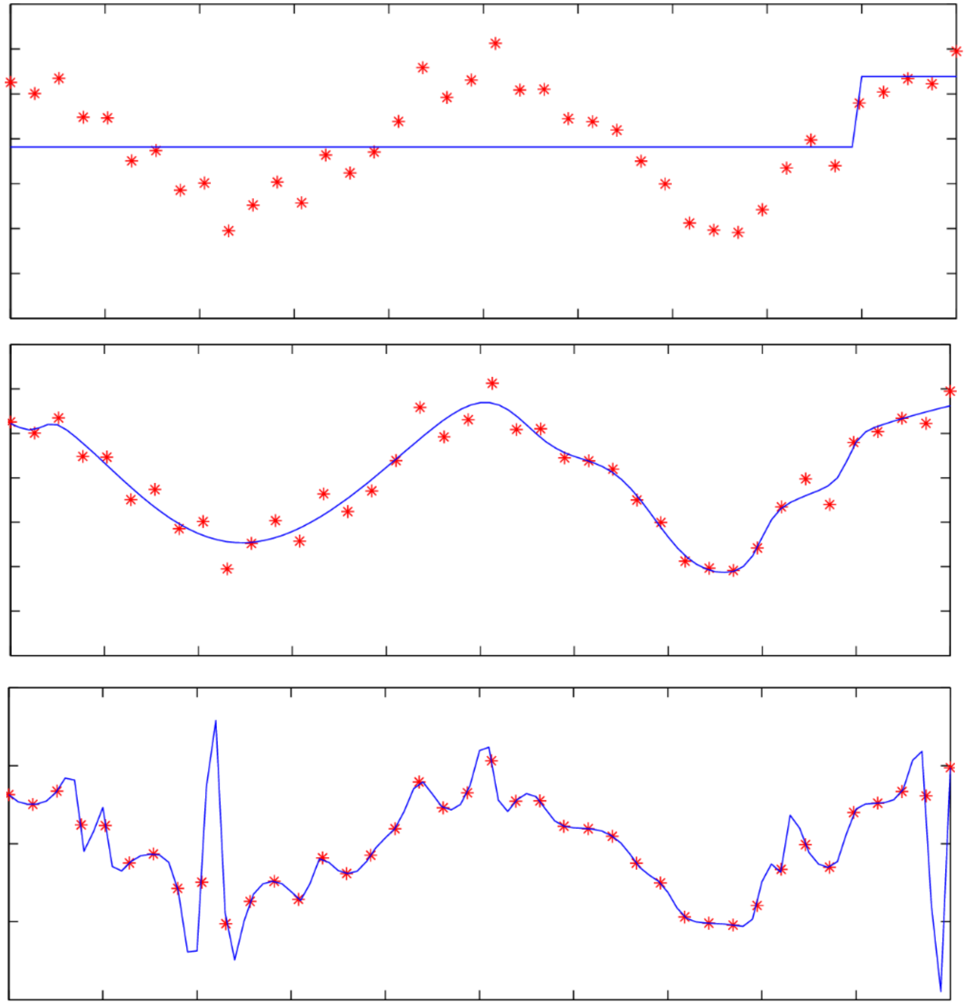
Increasing model capacity improves data fitting but eventually leads to overfitting

Model capacity

Low

Medium

High

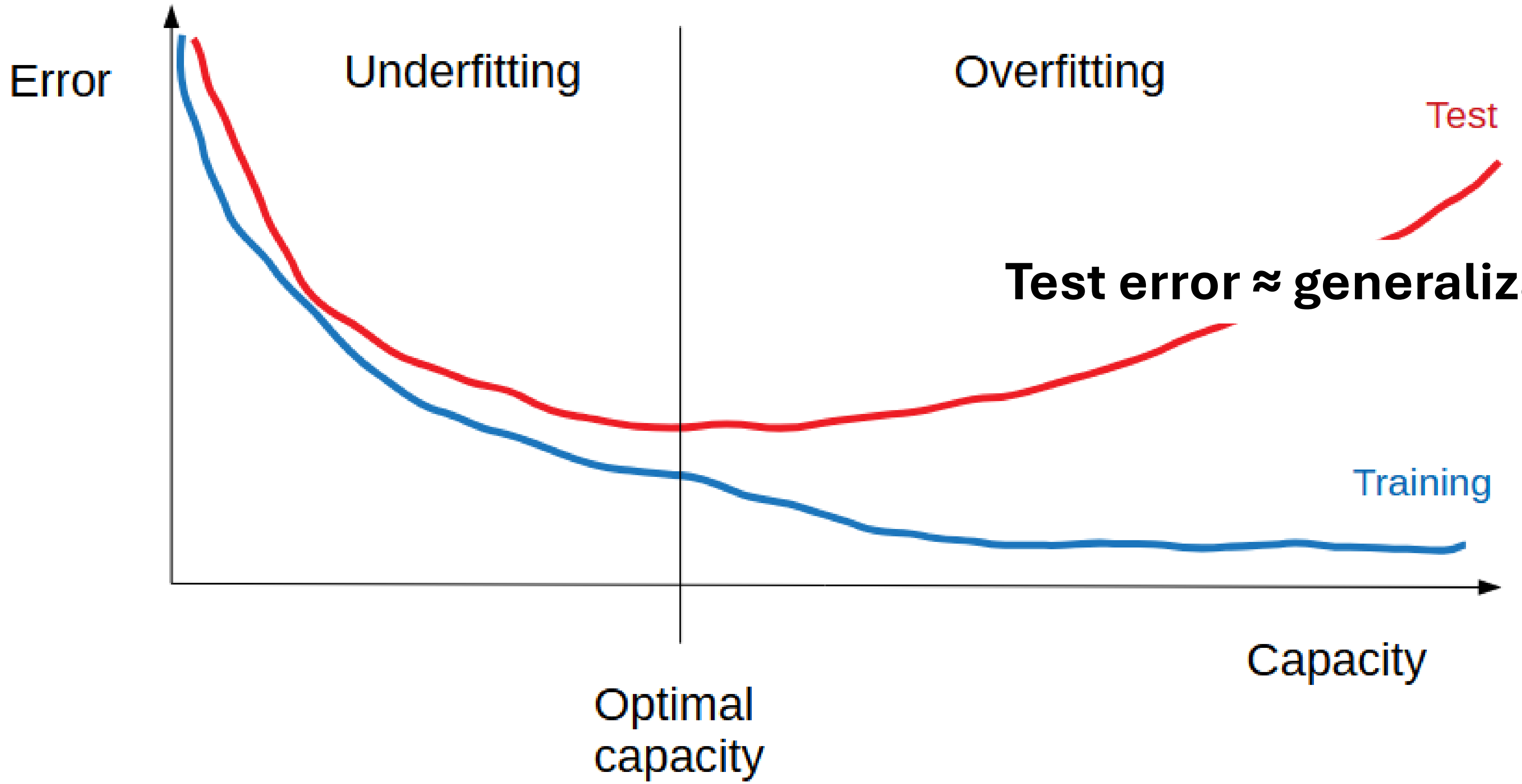


Underfitting

Good fit

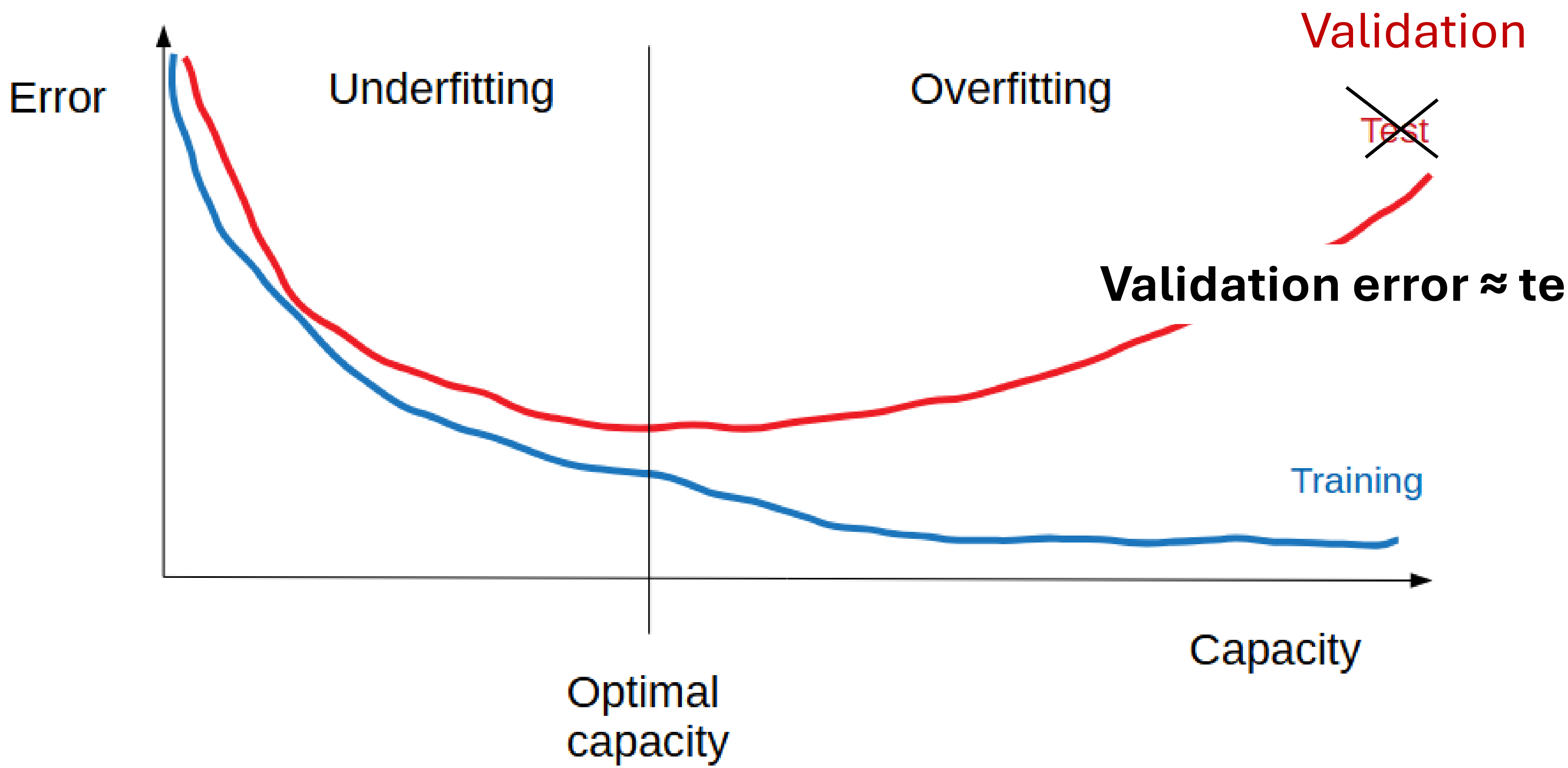
Overfitting

Training goal:
Minimizing test error



Test error ≈ generalization error

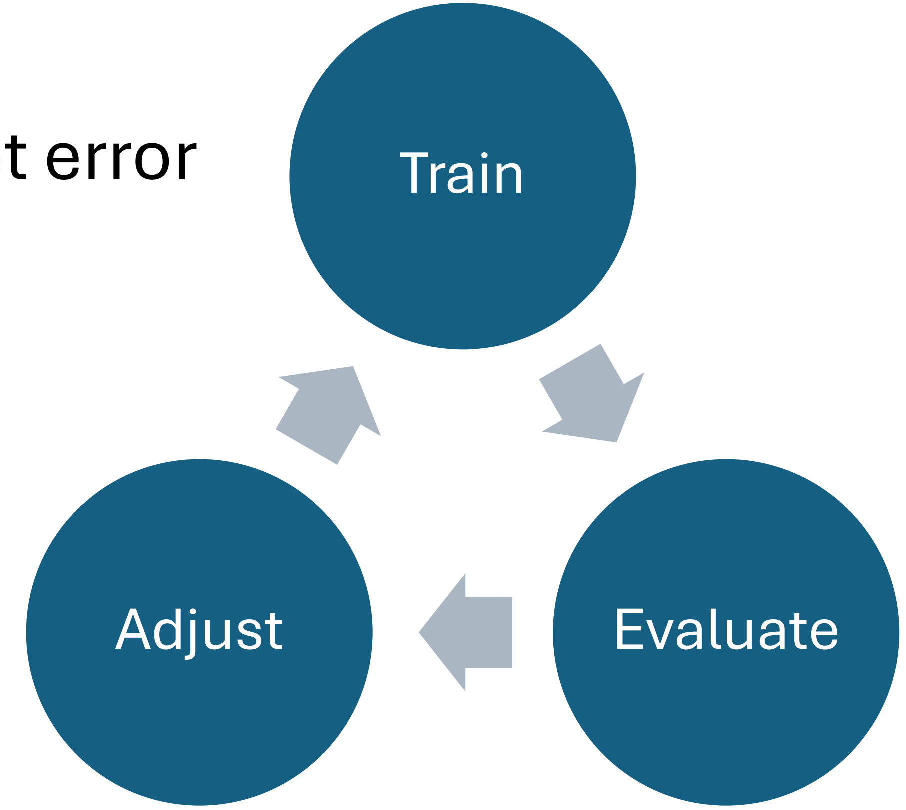
Training goal:
Minimizing validation error



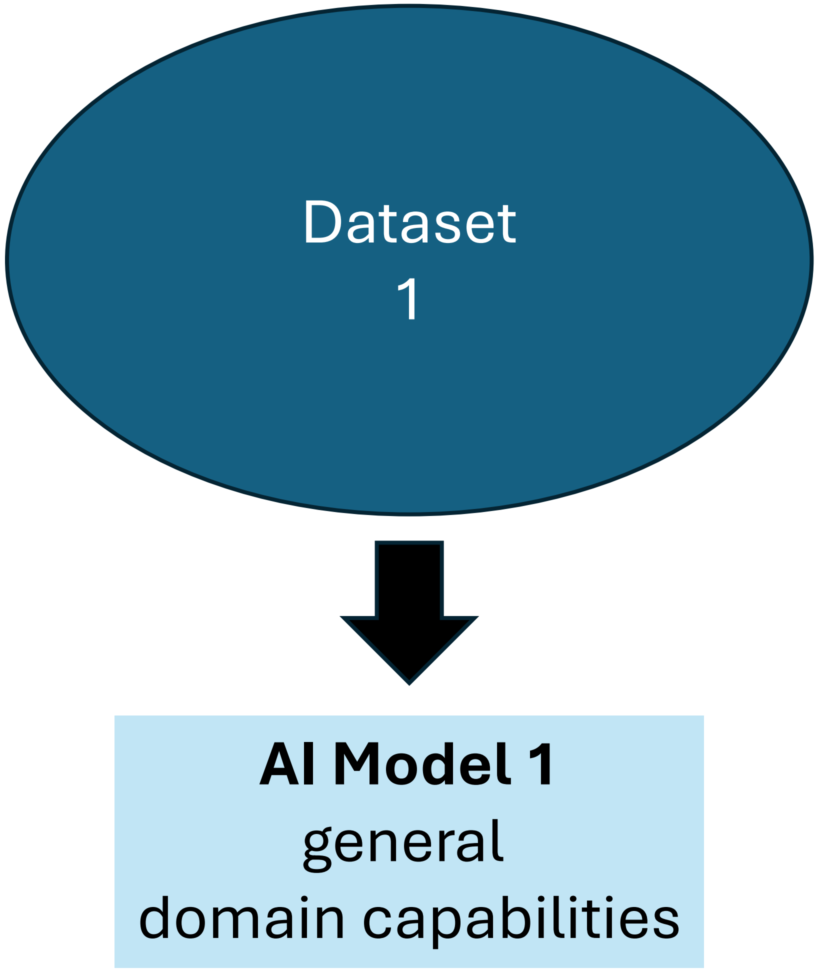
Validation error ≈ test error

How to find the best model

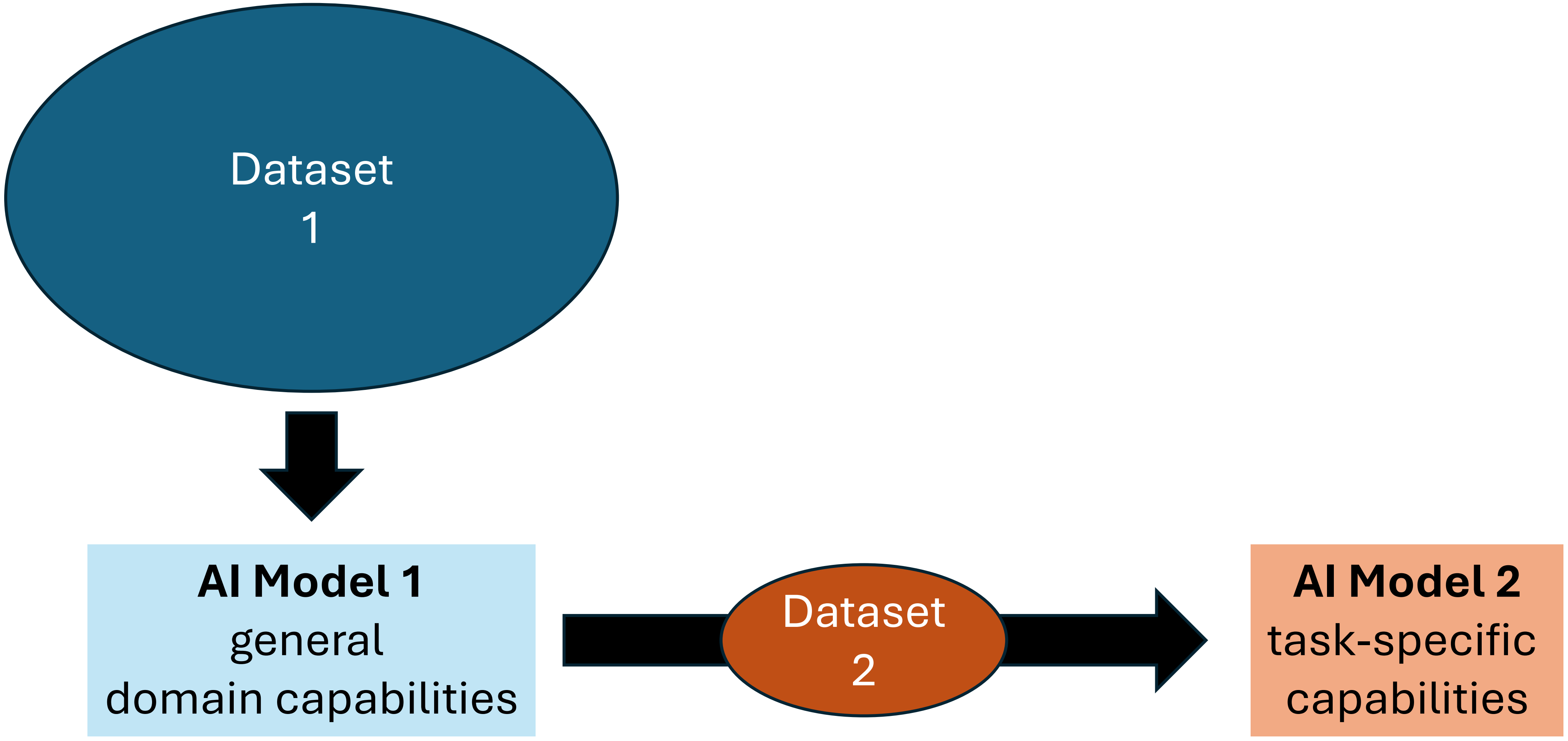
- Simple model first: baseline
- Test multiple model architectures
- Vary hyperparameters
- Gradual optimization to lower validation set error



Transfer learning enables more efficient AI training



Transfer learning enables more efficient AI training



Dataset requirements

- Very difficult to predict
- Depends on task difficulty
- Depends on model
- Adding more data often easiest way to improve model

Computational resources

- Models usually trained on GPUs
- Small models can be trained on CPUs (even on laptop)
- LU students/staff have access to free distributed supercomputing infrastructure: NAISS
- SOTA models often require resources only the largest tech companies have

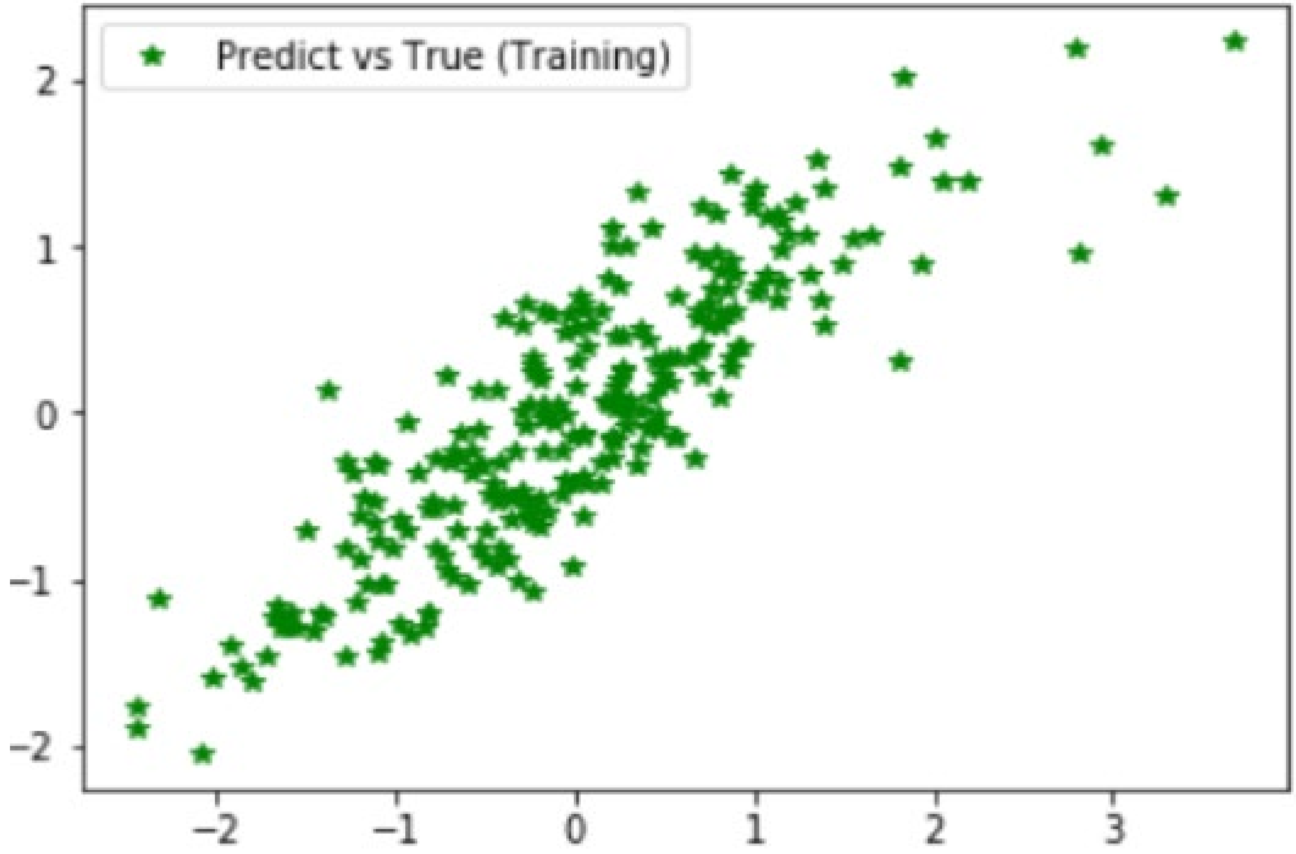
Model evaluation

Evaluation metrics

- Primary performance metric
  - Must match research goal
  - Used to rank models
- Secondary performance metric(s)
  - May serve to exclude models
  - May be used to select between models similar in primary metric

Many different metrics and evaluation tools exist!

Model evaluation:
Scatter plots – True vs Predicted



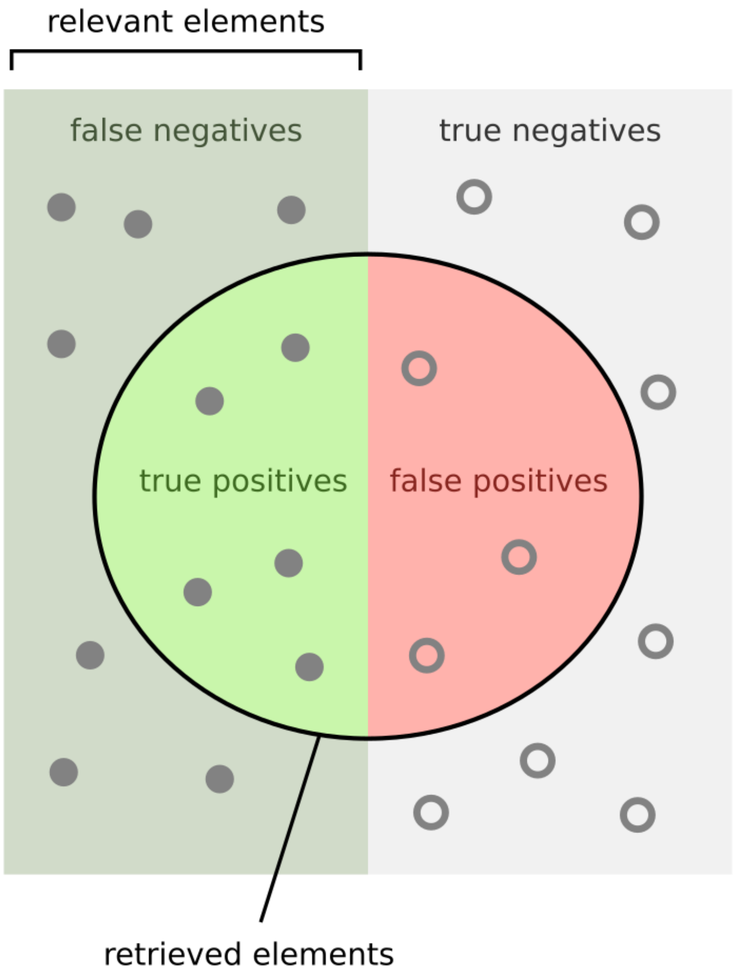
Calculate correlation!

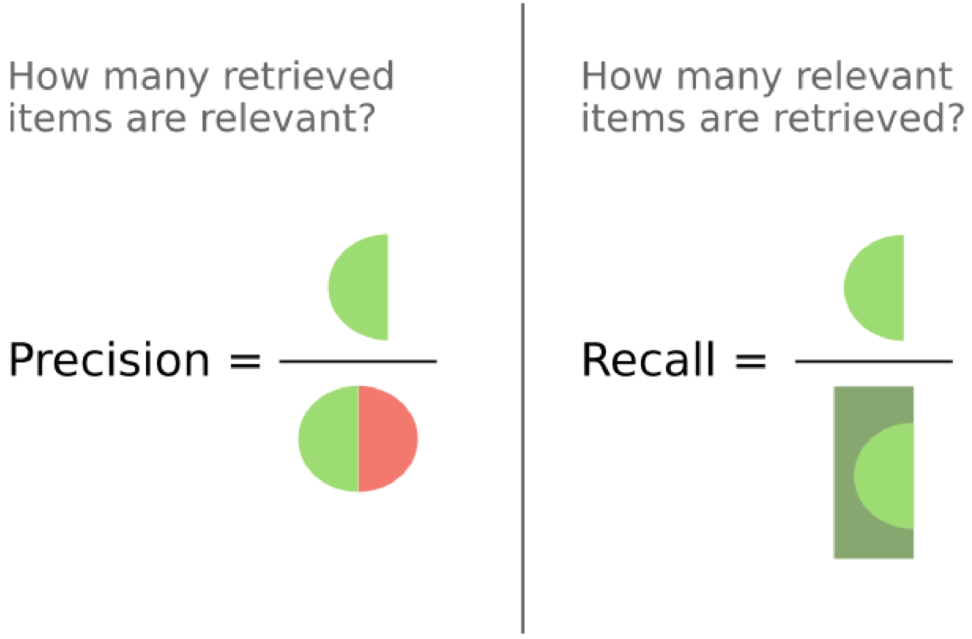
Model evaluation: Confusion matrix

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

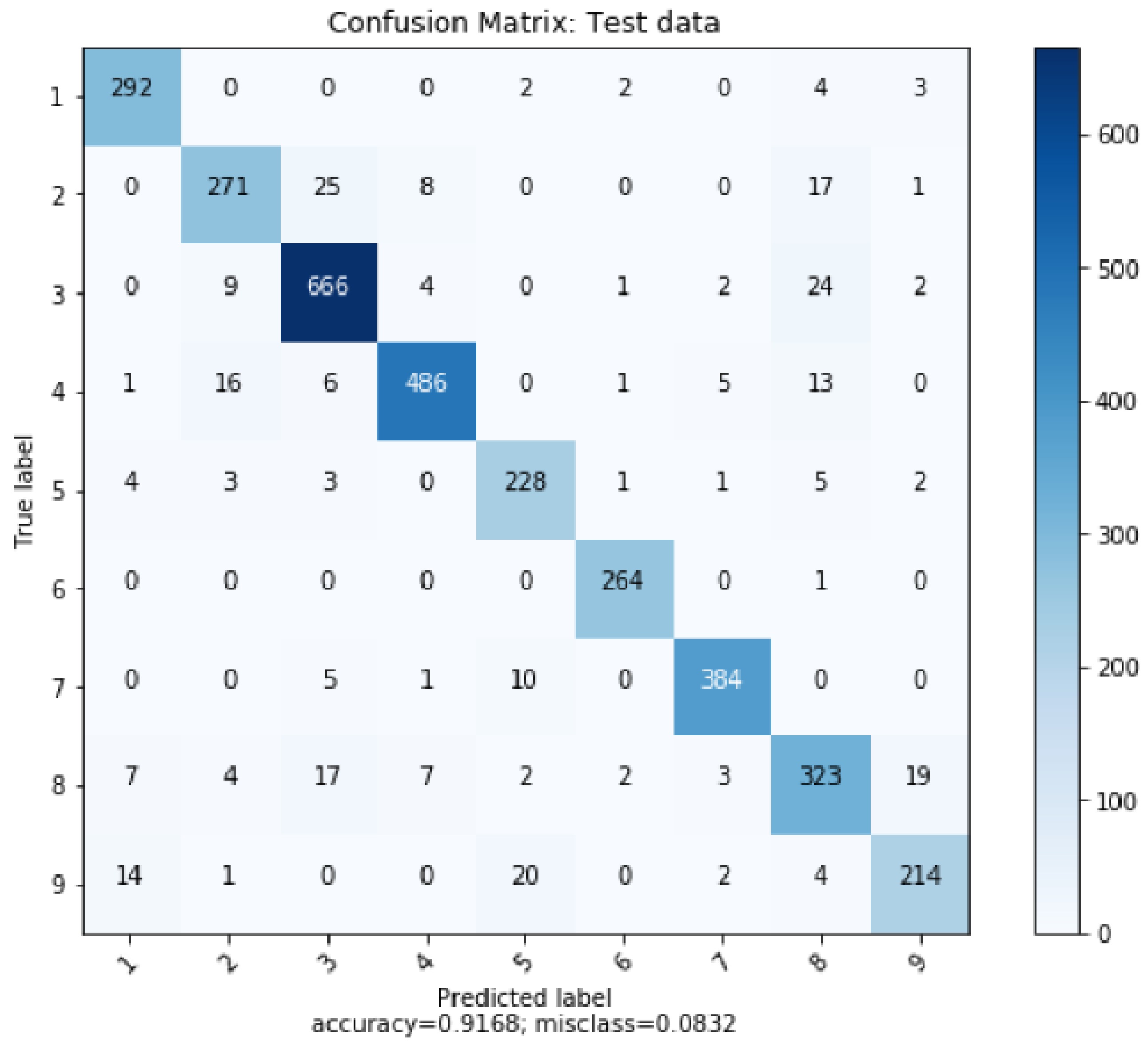
		Predicted condition	
	Total 8 + 4 = 12	Cancer 7	Non-cancer 5
Actual condition	Cancer 8	6	2
	Non-cancer 4	1	3

Model evaluation: Precision and Recall





https://en.wikipedia.org/wiki/Precision_and_recall


Confusion Matrix: Test data
True label
Predicted label
accuracy=0.9168; misclass=0.0832
1 2 3 4 5 6 7 8 9
292 0 0 0 2 2 0 4 3
0 271 25 8 0 0 0 17 1
0 9 666 4 0 1 2 24 2
1 16 6 486 0 1 5 13 0
4 3 3 0 228 1 1 5 2
0 0 0 0 0 264 0 1 0
0 0 5 1 10 0 384 0 0
7 4 17 7 2 2 3 323 19
14 1 0 0 20 0 2 4 214
0 100 200 300 400 500 600

Model evaluation:
F1 score

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Done training?
Final evaluation

- Evaluate on test set to assess generalization
- Use additional independent test set(s), if possible established benchmarking sets (compare to current SOTA)
- Compare to human performance (if relevant)
- Evaluate additional operational requirements

Expectation:

Test error > Validation error > Training error

(but ideally as little difference as possible)

Additional operational requirements

- Runtime
- Hardware requirements (e.g. memory, GPU)
- Explainability
- …

Deployment

- Embedding AI model into a larger system which includes all steps from data input to final output and attractive user interface
- Incorporation into existing workflows/infrastructure
- Documentation
- External evaluation/approval
- Training of users

Continuous monitoring

- Changing data patterns will affect model performance
- Examples not represented in training data may not be handled as desired
- New demands/technology can arise

→ Continuous monitoring required

→ Model replacement/retraining when needed

Summary

- Clear goals and high-quality data are essential for success
- Evaluation metrics need to fit research goal
- Avoid “data leakage” between split subsets: should be independent
- Test set needs to be held back to assess generalization
- Model training is an iterative process
- Development continues after model training for deployment and monitoring