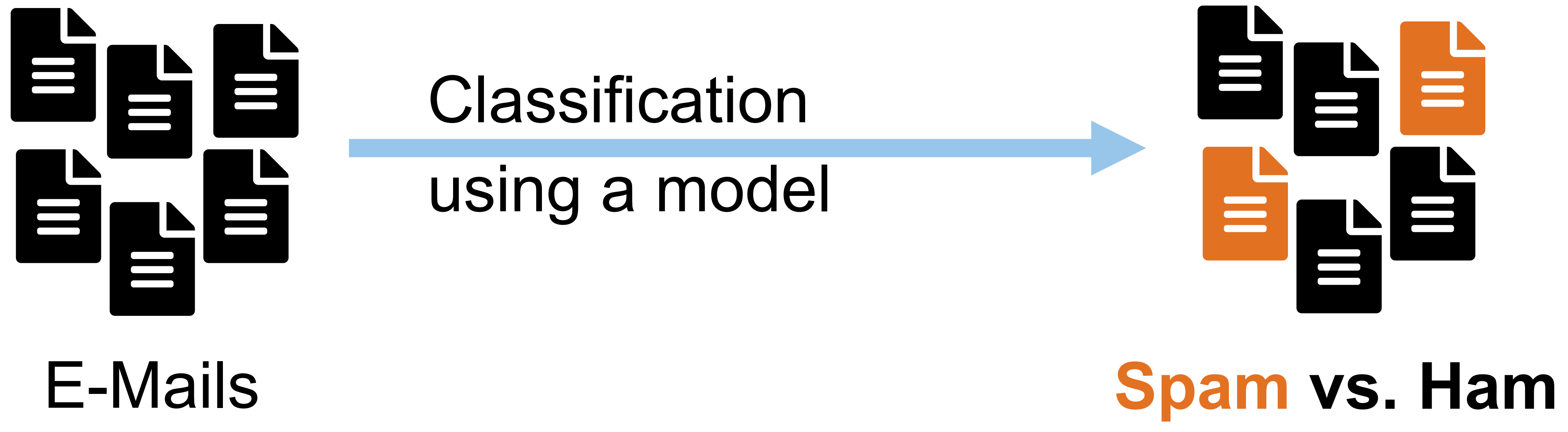
Supervised Learning: Choosing the right model

Mirco Schönfeld
mirco.schoenfeld@uni-bayreuth.de

How to Design a Model: Feature Selection

Formulate characteristics that help distinguishing between classes.

For spam-detection: find words or combinations of words that indicate a mail being spam.



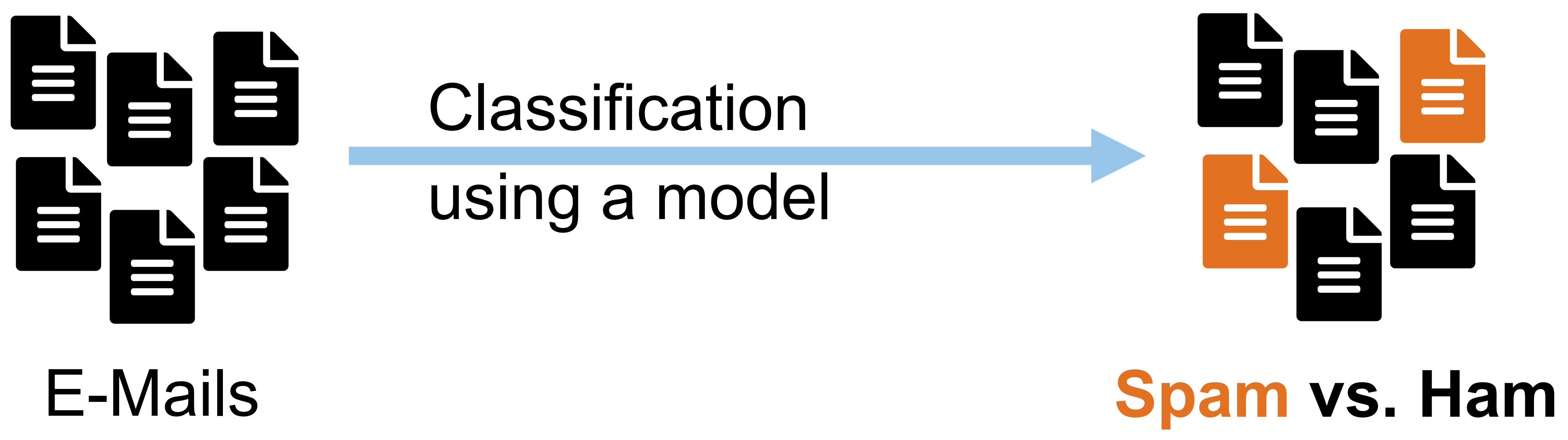
Spam: Wholesale Fashion Watches -57% today. Designer watches for cheap ...
Spam: You can buy ViagraFr$1.85 All Medications at unbeatable prices! ...
Spam: WE CAN TREAT ANYTHING YOU SUFFER FROM JUST TRUST US ...
Spam: Sta.rt earn*ing the salary yo,u d-eserve by o’btaining the prope,r crede’ntials!

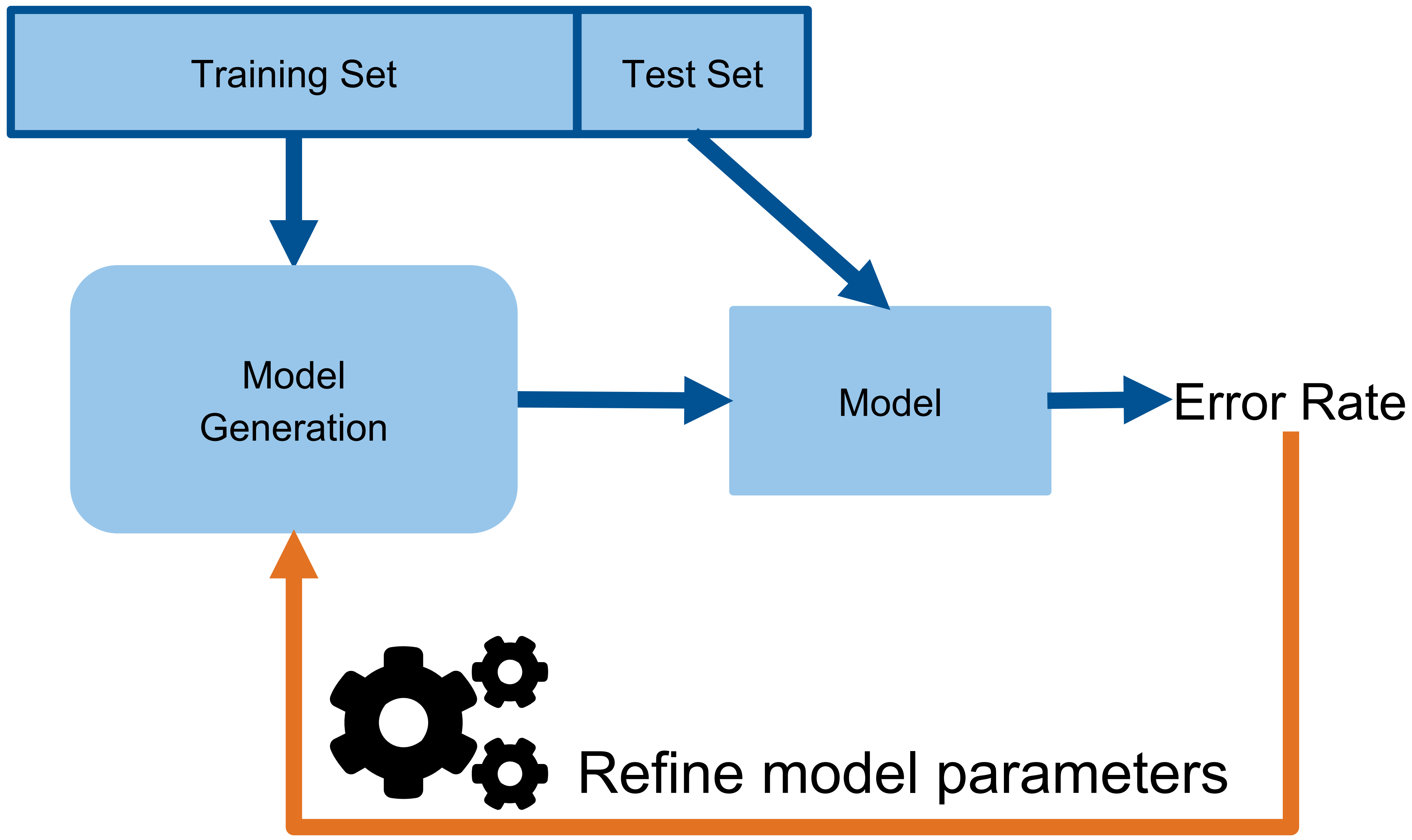
Ham: The practical significance of hypertree width in identifying more ...
Ham: Abstract: We will motivate the problem of social identity clustering: ...
Ham: Good to see you my friend. Hey Peter, It was good to hear from you. ...
Ham: PDS implies convexity of the resulting optimization problem (Kernel Ridge ...

Curse of Dimensionality:

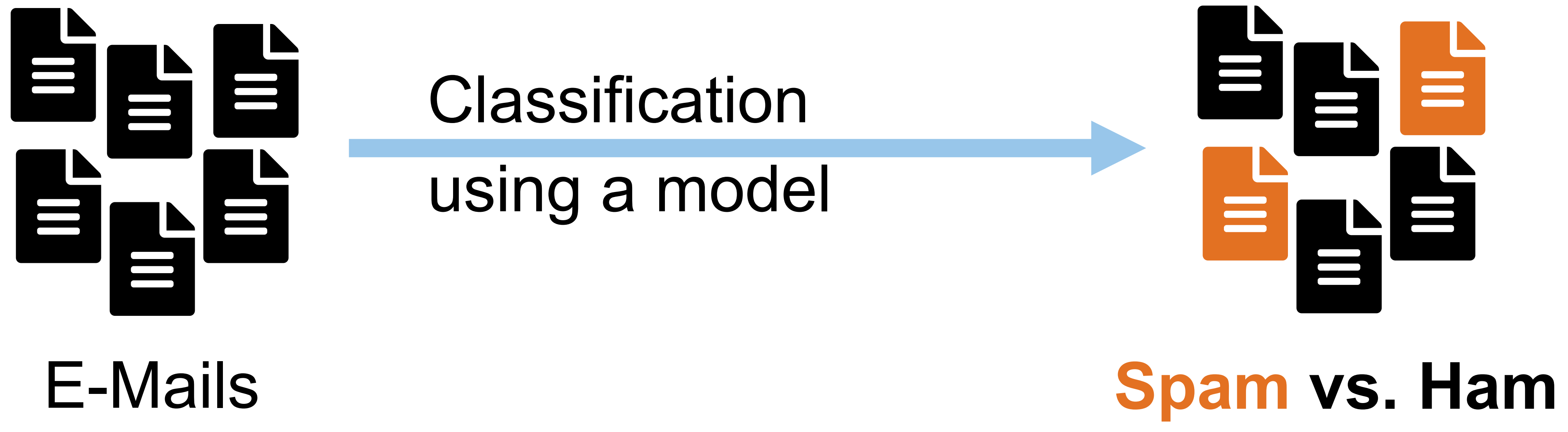
Including more features will improve classification *conceptually* but will render computation increasingly difficult.

Training the Model





How to Choose a Model



Model A:

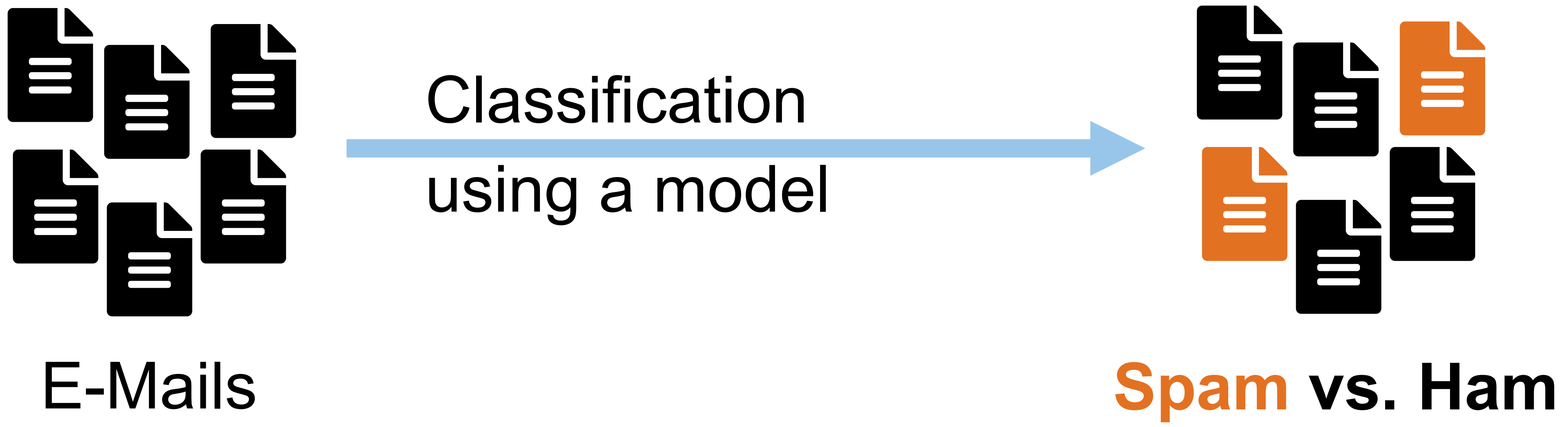
		Correct class	
		Ham	Spam
Spam-Filter reports	Ham	189	1
	Spam	11	799

?

Model B:

		Correct class	
		Ham	Spam
Spam-Filter reports	Ham	200	38
	Spam	0	762

Good Classifications: The Confusion Matrix



Model A:

		Correct class	
		Ham	Spam
Spam-Filter reports	Ham	189	1
	Spam	11	799

Model B:

		Correct class	
		Ham	Spam
Spam-Filter reports	Ham	200	38
	Spam	0	762

		Correct	
		Positive (P)	Negative (N)
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

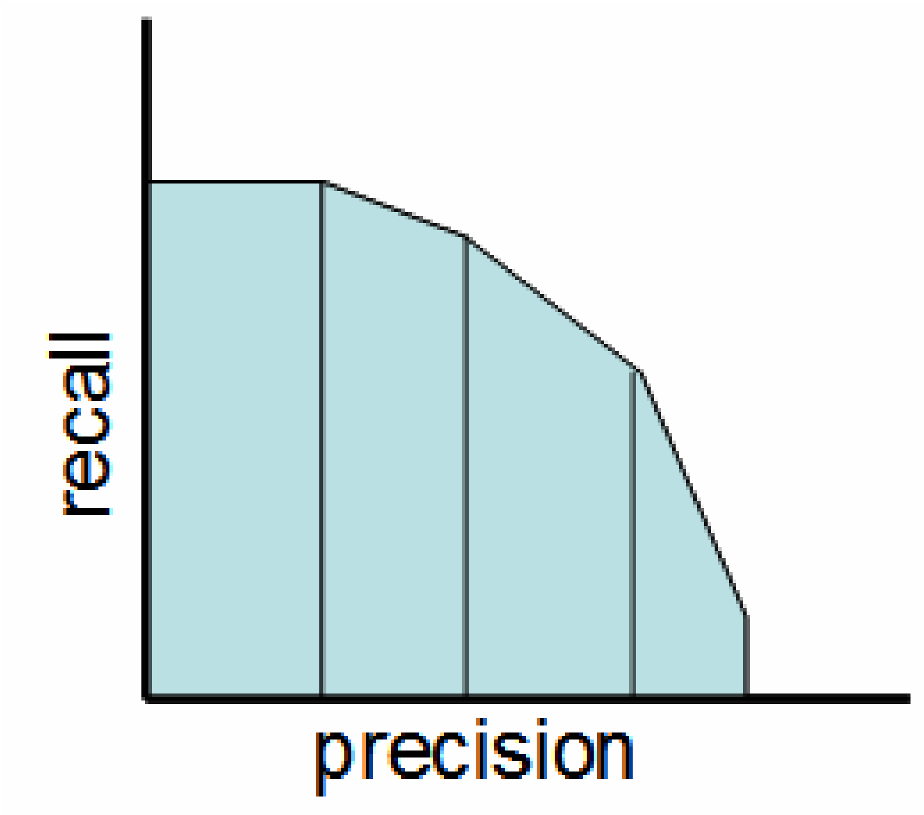
Measuring Goodness

- **Precision**: Proportion of predicted positives that are truly positive
  good choice when we need to be very sure of prediction

$$\frac{TP}{TP + FP}$$

- **Recall**: Proportion of actual positives that are correctly classified
  good choice when as many positives as possible should be captured

$$\frac{TP}{TP + FN}$$



		Correct	
		Positive (P)	Negative (N)
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226

Measuring Goodness

- **Accuracy**: Proportion of true results among total number of cases
  good choice when classes are balanced

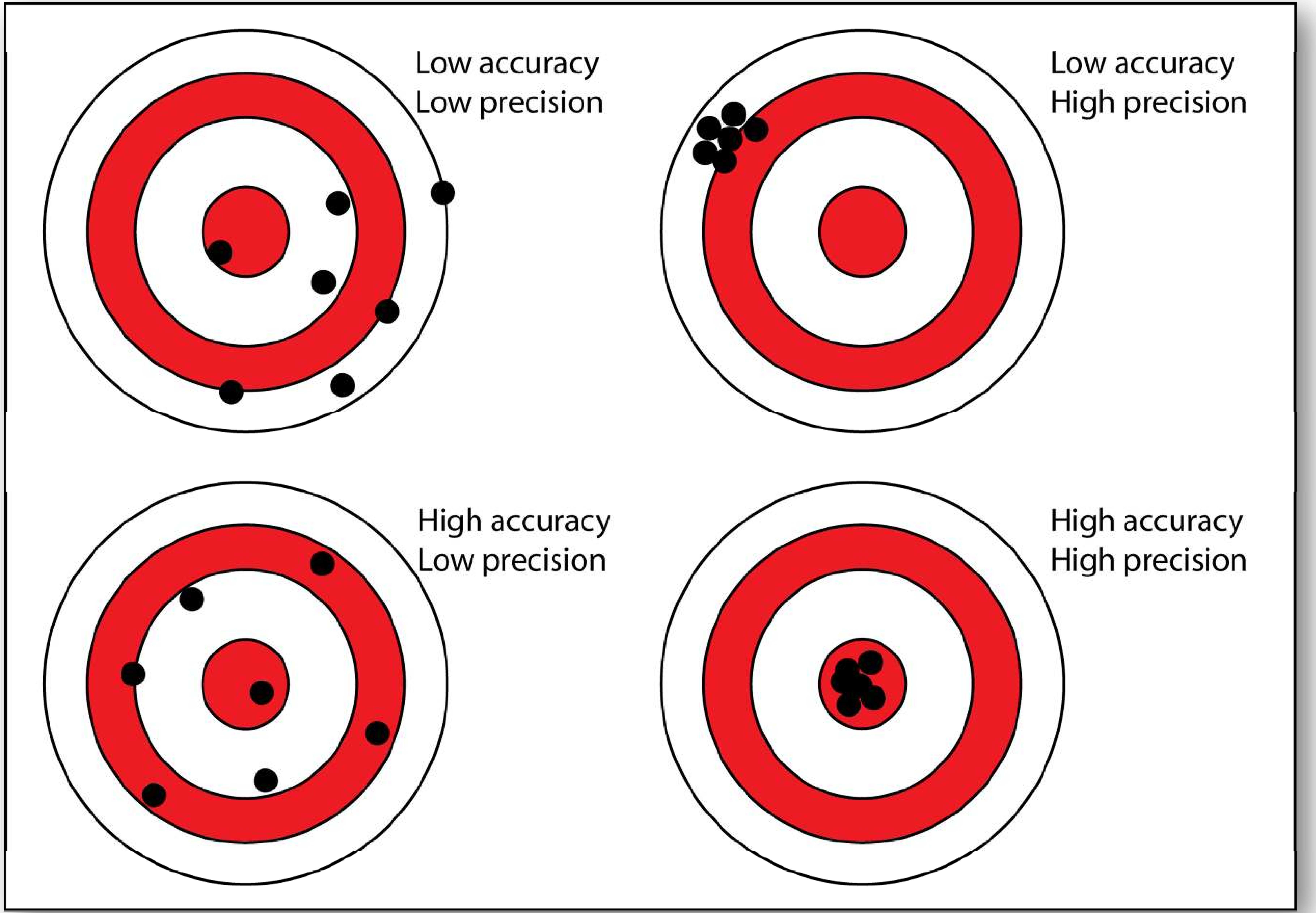
$$\frac{TP + TN}{TP + FP + FN + FN}$$

- $\boldsymbol{F_1}$ **Score**: harmonic mean between precision & recall – a number between 0 and 1
  good choice when we want a model with both good precision and recall

$$2 * \frac{precision * recall}{precision + recall}$$

Important variant F_β allows to apply a custom weight to precision & recall

		Correct	
		Positive (P)	Negative (N)
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative



https://www.antarcticglaciers.org/glacial-geology/dating-glacial-sediments-2/precision-and-accuracy-glacial-geology/

Measuring Goodness & more

Prevalence

$$\frac{P}{P+N}$$

accuracy (ACC)

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{P}+\mathrm{N}} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

balanced accuracy (BA)

$$\mathrm{BA} = \frac{TPR+TNR}{2}$$

F1 score

is the harmonic mean of precision and sensitivity:

$$\mathrm{F}_1 = 2 \times \frac{\mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV}+\mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

phi coefficient (φ or r_φ) or Matthews correlation coefficient (MCC)

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

Fowlkes-Mallows index (FM)

$$\mathrm{FM} = \sqrt{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}} = \sqrt{PPV \times TPR}$$

informedness or bookmaker informedness (BM)

$$\mathrm{BM} = \mathrm{TPR}+\mathrm{TNR}-1$$

markedness (MK) or deltaP (Δp)

$$\mathrm{MK} = \mathrm{PPV}+\mathrm{NPV}-1$$

Diagnostic odds ratio (DOR)

$$\mathrm{DOR} = \frac{\mathrm{LR+}}{\mathrm{LR-}}$$

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = 1-\mathrm{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} = 1-\mathrm{FPR}$$

precision or positive predictive value (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = 1-\mathrm{FDR}$$

negative predictive value (NPV)

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}} = 1-\mathrm{FOR}$$

miss rate or false negative rate (FNR)

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{P}} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} = 1-\mathrm{TPR}$$

fall-out or false positive rate (FPR)

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{N}} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = 1-\mathrm{TNR}$$

false discovery rate (FDR)

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}} = 1-\mathrm{PPV}$$

false omission rate (FOR)

$$\mathrm{FOR} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TN}} = 1-\mathrm{NPV}$$

Positive likelihood ratio (LR+)

$$\mathrm{LR+} = \frac{\mathrm{TPR}}{\mathrm{FPR}}$$

Negative likelihood ratio (LR-)

$$\mathrm{LR-} = \frac{\mathrm{FNR}}{\mathrm{TNR}}$$

prevalence threshold (PT)

$$\mathrm{PT} = \frac{\sqrt{\mathrm{TPR}(-\mathrm{TNR}+1)}+\mathrm{TNR}-1}{(\mathrm{TPR}+\mathrm{TNR}-1)} = \frac{\sqrt{\mathrm{FPR}}}{\sqrt{\mathrm{TPR}}+\sqrt{\mathrm{FPR}}}$$

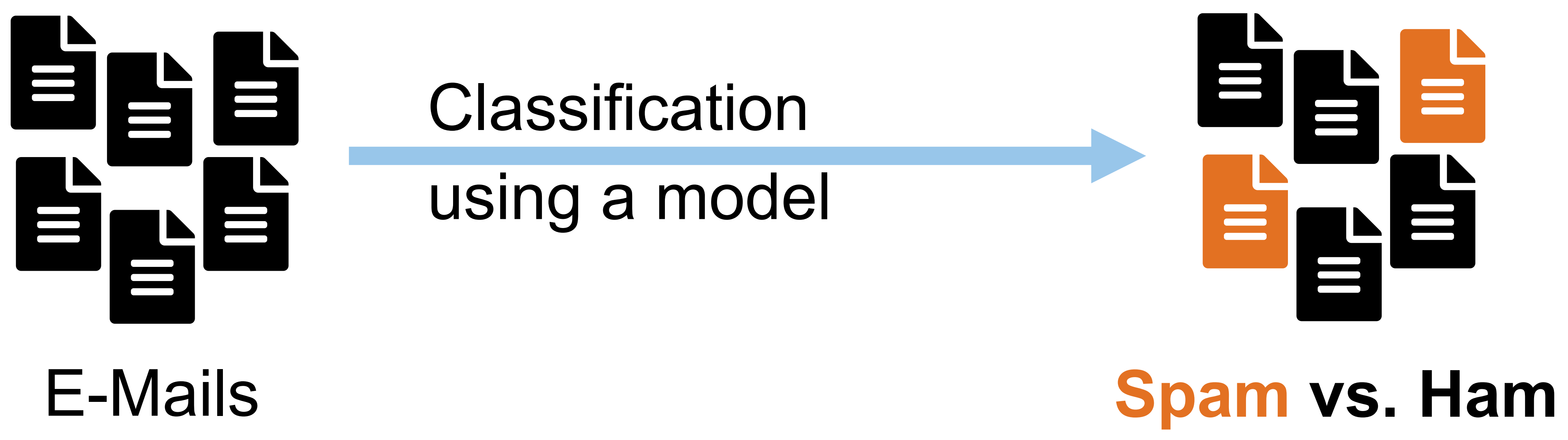
threat score (TS) or critical success index (CSI)

$$\mathrm{TS} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}}$$

		Correct	
		Positive (P)	Negative (N)
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

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

How to Choose a Model



Minimize error rate

Model A:

		Correct class	
		Ham	Spam
Spam-Filter reports	Ham	189	1
	Spam	11	799

Maximize *utility*

Model B:

		Correct class	
		Ham	Spam
Spam-Filter reports	Ham	200	38
	Spam	0	762

Interpretability

Large project set out to evaluate ML application to problems in healthcare

Scenario: predicting pneumonia risk

Goal: predict probabilty of death for patients with pneumonia

The most accurate model of the study was a multitask neural net.

Outperformed other models by wide margin but was still dropped. Why?

One rule-based system learned the rule „patient has asthma ➔ lower risk“

Reflected a true pattern in training data

The best model was the least intelligible one – was deemed to risky

No way of checking the features that were picked up

Cooper, Gregory F., et al. "An evaluation of machine-learning methods for predicting pneumonia mortality." Artificial intelligence in medicine 9.2 (1997): 107-138.
Caruana, Rich, et al. "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission." Proceedings of the 21th ACM SIGKDD. 2015.

Goals of Interpretable Models

- Trust: Identify and mitigate *bias*

  Recognizing bias in a black-box algorithm is *very* hard
- Causality: Account for context

  Helps you understand how the factors included in the model led to the prediction
- Informativeness: Extract knowledge

  Helps you determine if patterns that appear to be present in the model are really there.

  Rather learning from the model than evaluating it (compared to identifying bias)
- Transferability: Generalize

  Models are trained on carefully collected datasets to solve narrowly defined problems. Interpretable models should help you determine if and how they can be generalized
- Fair and Ethical Decision-Making

  algorithmic decision-making mediates more and more of our interactions. Need a way to make sure that decisions conform to ethical standards

Lipton, Zachary C. "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery." Queue 16.3 (2018): 31-57.
https://shapeofdata.wordpress.com/2016/11/17/goals-of-interpretability/

Properties of Interpretable Models

1. Transparency
   - Simulatability

     Transparency at the level of the entire model
   - Decomposability / Intelligibility

     Transparency at the level of the individual components, e.g. parameters
   - Algorithmic Transparency

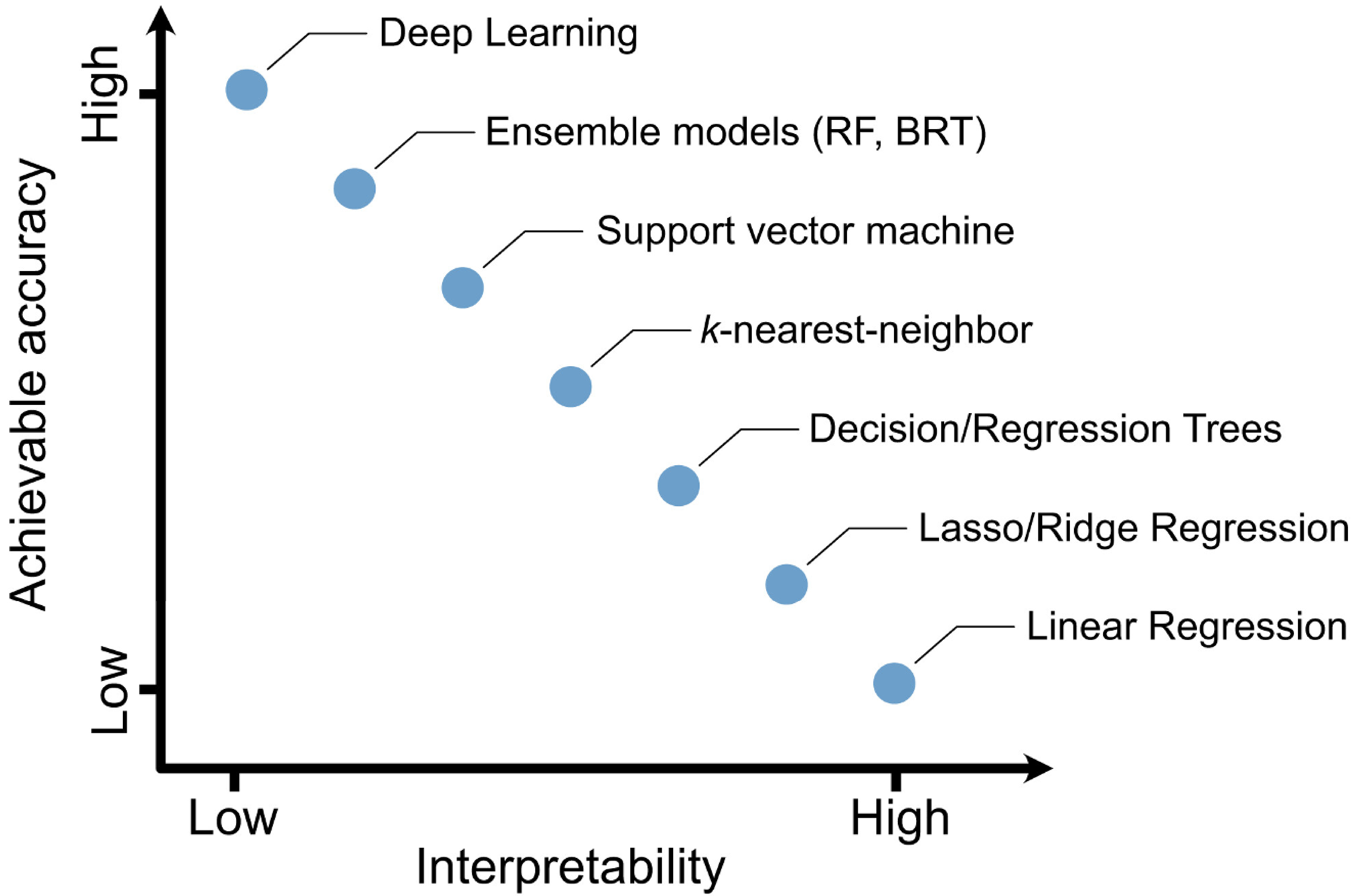
     Transparency at the level of the training algorithm

Lipton, Zachary C. "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery." Queue 16.3 (2018): 31-57.

Properties of Interpretable Models

2. Post-hoc Interpretability
   - Text Explanations
   - Visualization
   - Local Explanations
   - Explanation by Example

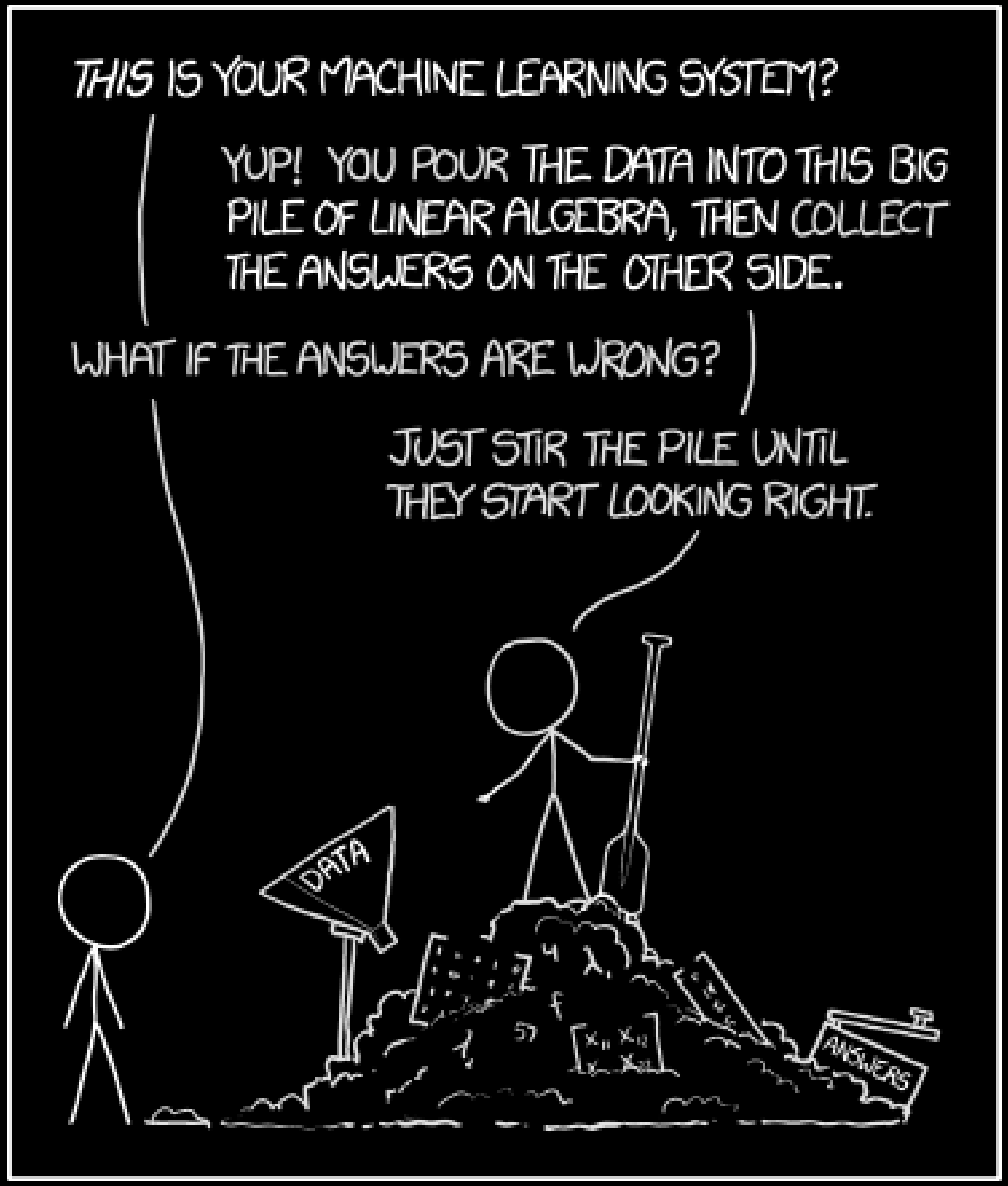
Lipton, Zachary C. "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery." Queue 16.3 (2018): 31-57.



Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, 14, 994–1016. https://doi.org/10.1111/2041-210X.14061

How to select a model?

- Quality of predictions
  i.e. performance in terms of a quality metric
- Speed
  i.e. training time, prediction time
- Robustness
  i.e. handling noise or missing values and still classify correctly
- Scalability
  i.e. computational efficiency
- Interpretability
  subjective means
- Other



https://xkcd.com/1838/

Thanks.

mirco.schoenfeld@uni-bayreuth.de