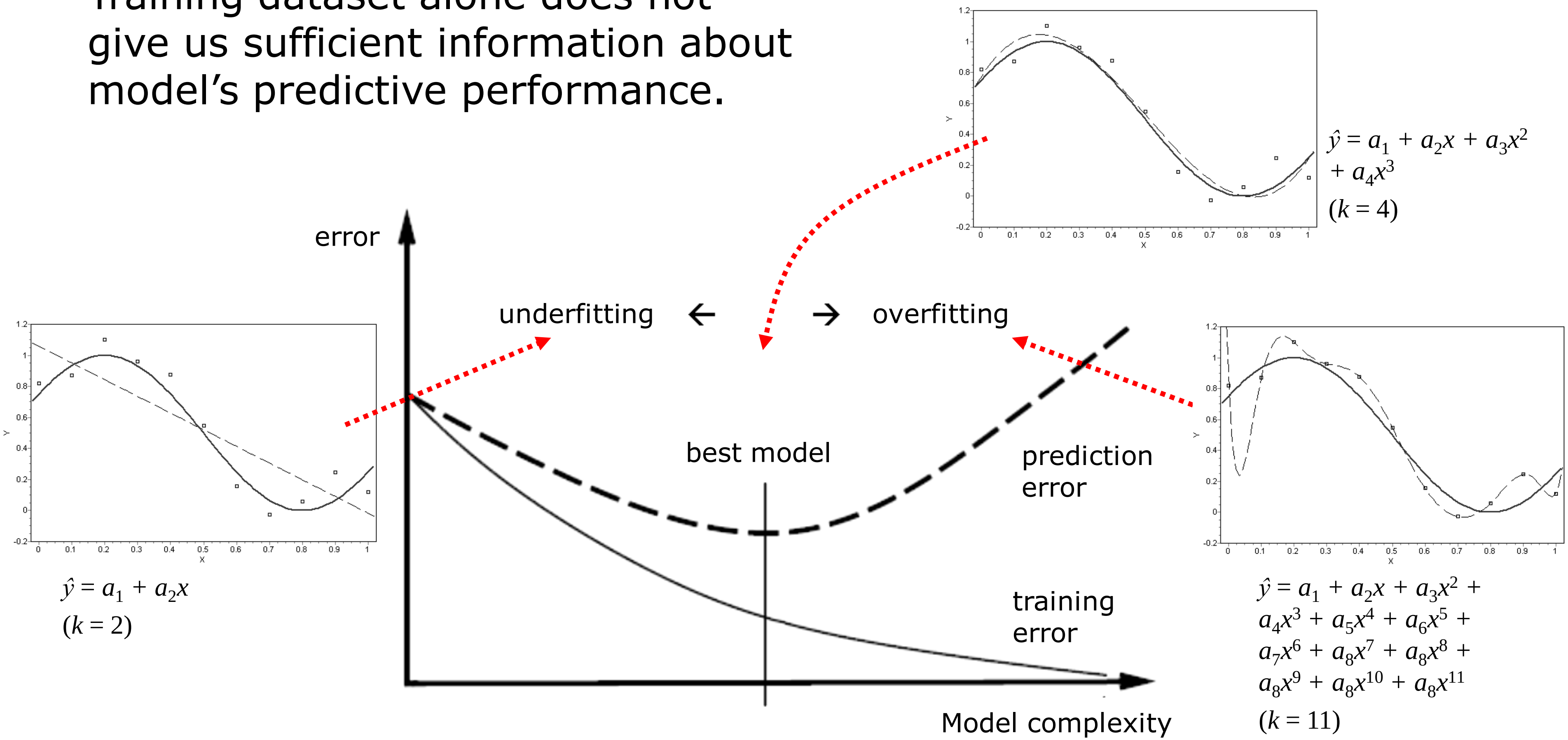
5

Alternative to the resampling methods

General scheme of the whole process

Underfitting and overfitting

Training dataset alone does not give us sufficient information about model's predictive performance.



Here "error" means, e.g., SAE, MAE, SSE, MSE, RMSE.
For R^2, this picture should be vertically flipped as R^2 is the opposite of error.

Validation set and test set

- The aims of model selection and final evaluation are different:
  - Model selection aims to select the best model by evaluating predictive performance of models-candidates

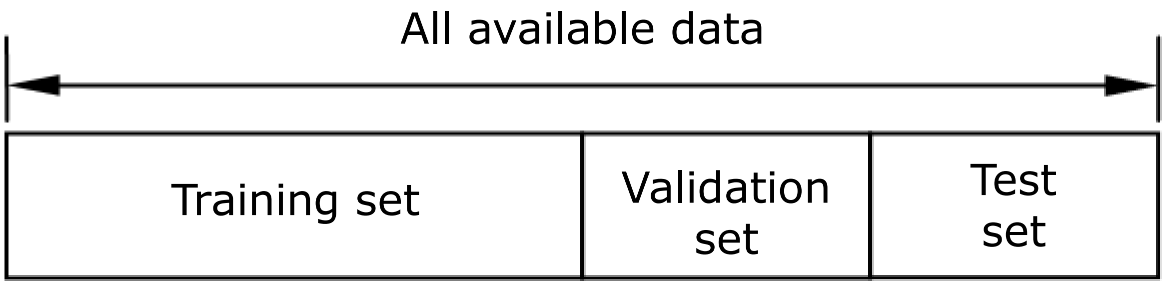
    (here we mostly pay attention to the relative differences between evaluations of different models)
  - Final estimation of the true prediction error aims to estimate the true prediction error as closely as possible in this way giving information about the expected error of the model in its future applications

Resampling methods

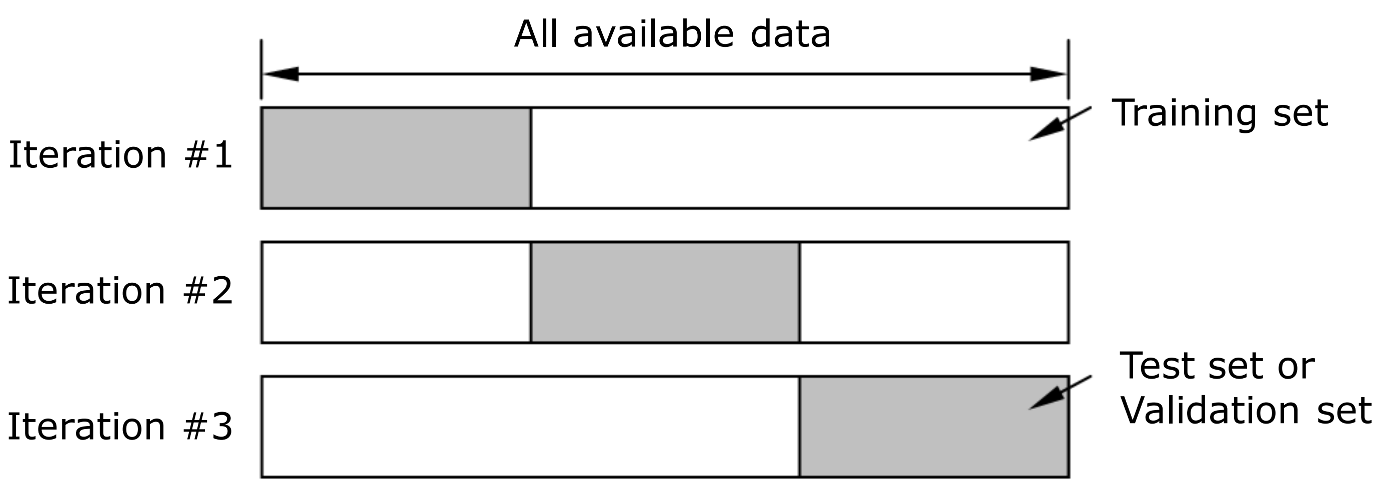
- In both cases (model selection and true prediction error estimation) we can use the so-called resampling methods
- The idea: evaluate the model on (additional) data that was not included in the training set
- The data may be
  - Additionally generated (but this can be very expensive or even impossible to do)
  - Simply subtracted from the already existing full dataset and set aside
- It's important that these (additional) data points would not be included in the training set, i.e., the data points would not be used for building the models (estimation of model parameters etc.)

Hold-Out & Cross-Validation

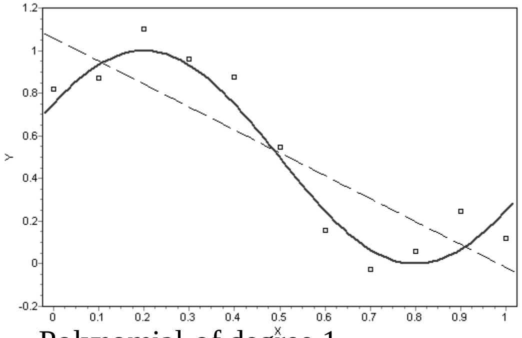
- Hold-Out



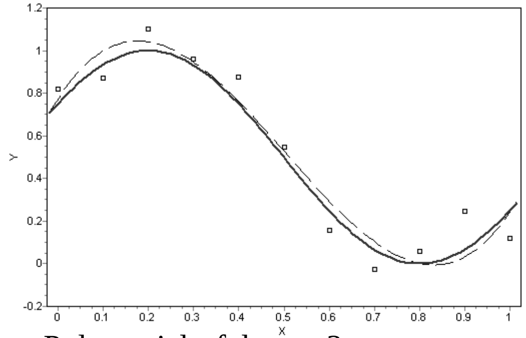
- Cross-Validation



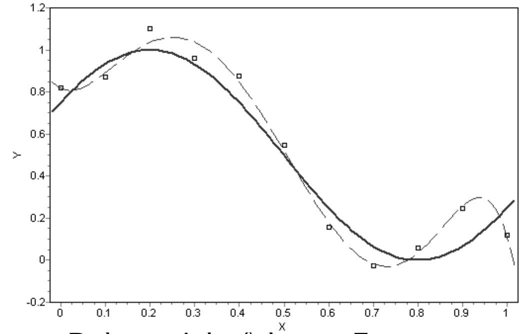
Example



Polynomial of degree 1



Polynomial of degree 3



Polynomial of degree 7

MSE = 0.0439 R^2 = 0.72	MSE = 0.0123 R^2 = 0.92	MSE = 0.0012 R^2 = 0.99	Evaluation in training dataset
MSE = 0.0307 R^2 = 0.81	MSE = 0.0188 R^2 = 0.88	MSE = 0.1199 R^2 = 0.25	Evaluation using Hold-out
MSE = 0.0647 R^2 = 0.59	MSE = 0.0540 R^2 = 0.66	MSE = 1.5395 R^2 = -8.66	Evaluation using Cross-Validation

Alternative to resampling methods

- Complexity-penalization criteria
  - Sometimes also called “information criteria” or “analytical criteria”
  - Founded in information theory and other fields
- Can be used for model selection in place of resampling methods
- These criteria don’t require separate validation data set – instead they use just the training set and evaluate model using training error, training set size, and model complexity
- Two of the most known and popular criteria:
  - Akaike’s Information Criterion (AIC)
  - Bayesian Information Criterion (BIC), also called Minimum Description Length (MDL)

BIC and MDL (and a possible interpretation)

- BIC and MDL simplest form:

$$MDL = \underbrace{n\ln(MSE)}_{} + \underbrace{k\ln(n)}_{}$$

The amount of information required to describe errors that the model makes

(larger errors = more information)

[*combating underfitting*]

The amount of information required to describe the model

(more complex model = more information)

[*combating overfitting*]

n is training set size; k is model complexity (*for linear regression it is model's number of parameters*)

BIC and MDL must be minimized, same as SAE, MAE, SSE, MSE, RMSE etc.

Akaike's Information Criterion

- Akaike's Information Criterion, AIC:

$$AIC = n\ln(MSE) + 2k$$

- AIC version corrected for small data – Corrected Akaike's Information Criterion (AICc).
  AICc simplest form:

$$AICc = n\ln(MSE) + 2k + \frac{2k(k+1)}{n-k-1}$$

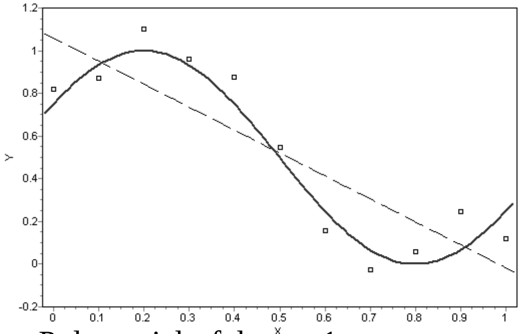
AIC and AICc must be minimized, same as SAE, MAE, SSE, MSE, RMSE etc.

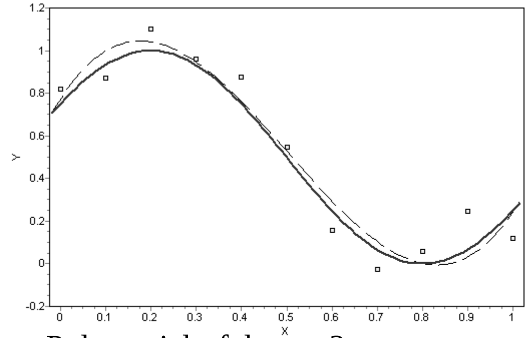
Advantages and disadvantages

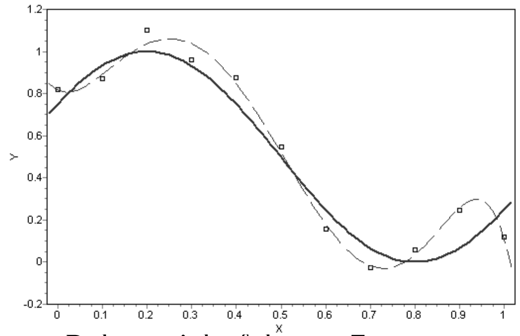
- Advantages
  - Easy to use and implement
  - Efficiency of computation
- Disadvantages
  - The computed evaluation measurements are usually interpretable only relatively when comparing models (during model selection). They don't offer direct and reliable estimation of model's prediction error. They are usually used only for model selection, not for final estimation of prediction error.
  - Simple mathematical forms of these criteria are known only for some types of models (and data types).

    (For example, linear regression, nearest-neighbors method, spline methods, tree methods etc.)

Example

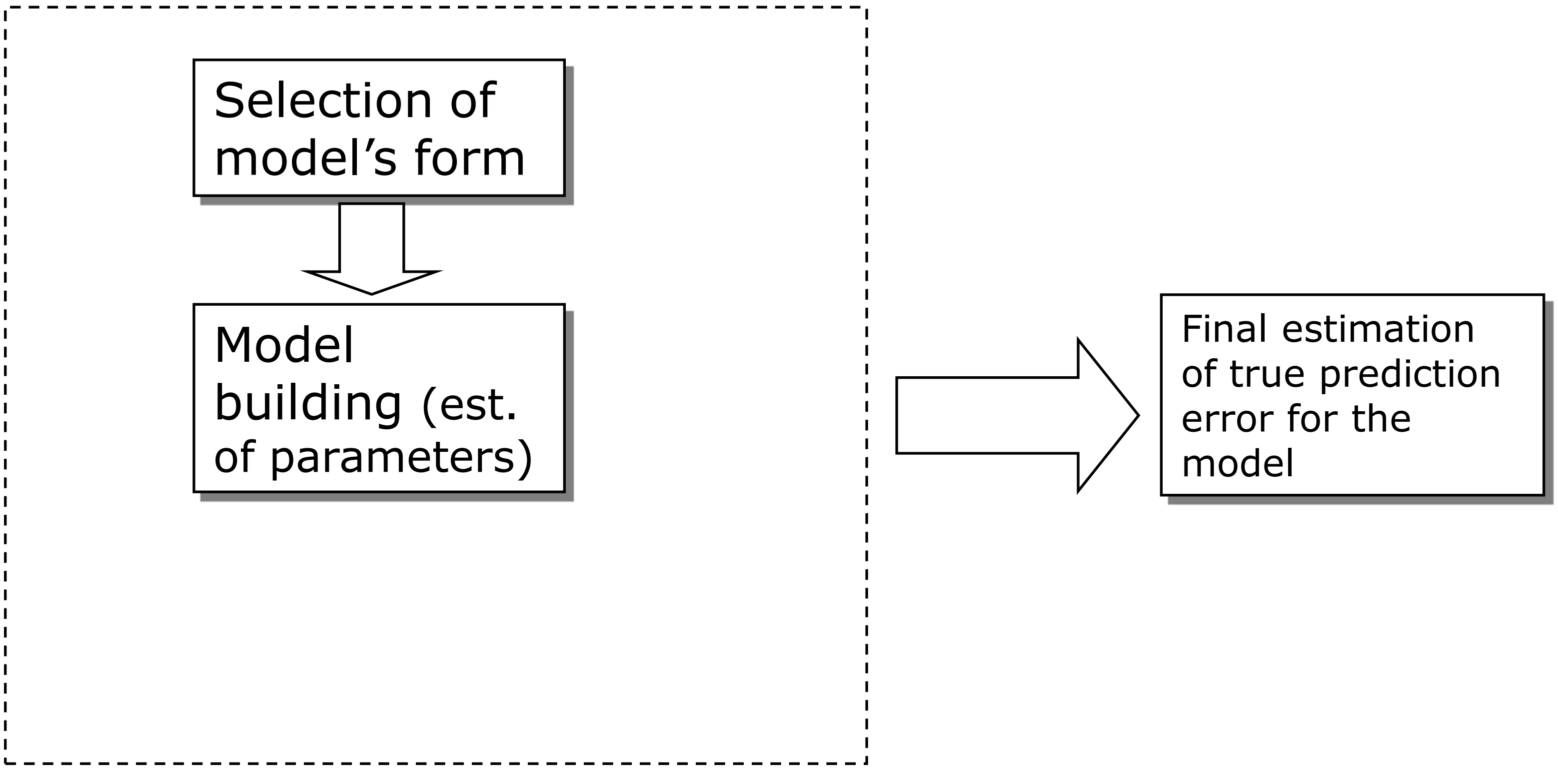

Polynomial of degree 1


Polynomial of degree 3

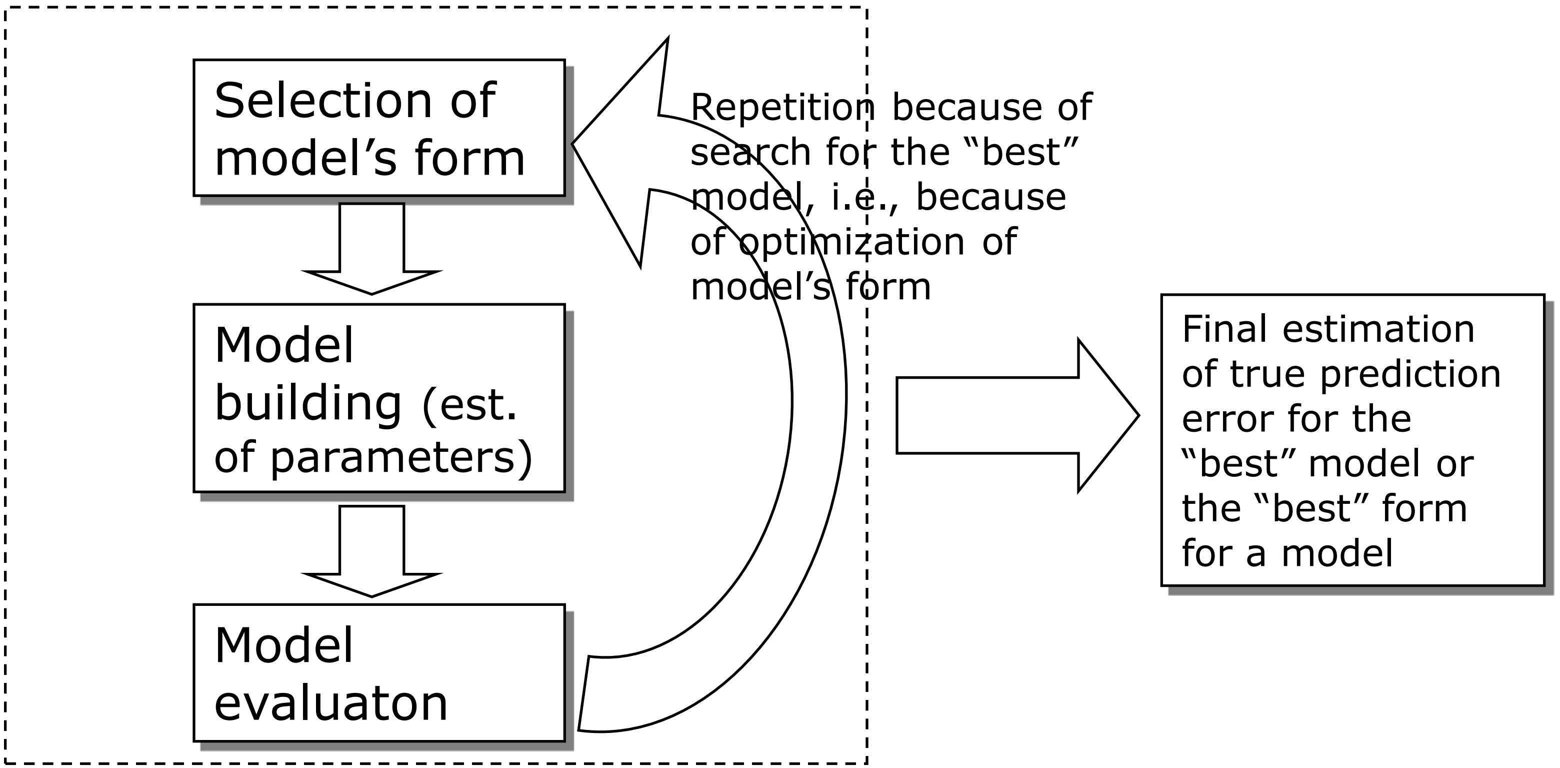

Polynomial of degree 7

MSE = 0.0439 R^2 = 0.72	MSE = 0.0123 R^2 = 0.92	MSE = 0.0012 R^2 = 0.99	Evaluation in training dataset
MSE = 0.0307 R^2 = 0.81	MSE = 0.0188 R^2 = 0.88	MSE = 0.1199 R^2 = 0.25	Evaluation using Hold-out
MSE = 0.0647 R^2 = 0.59	MSE = 0.0540 R^2 = 0.66	MSE = 1.5395 R^2 = -8.66	Evaluation using Cross-Validation
MDL = -29.59	MDL = -38.76	MDL = -54.39	Evaluation using MDL
AICc = -28.89	AICc = -33.68	AICc = 14.43	Evaluation using AICc

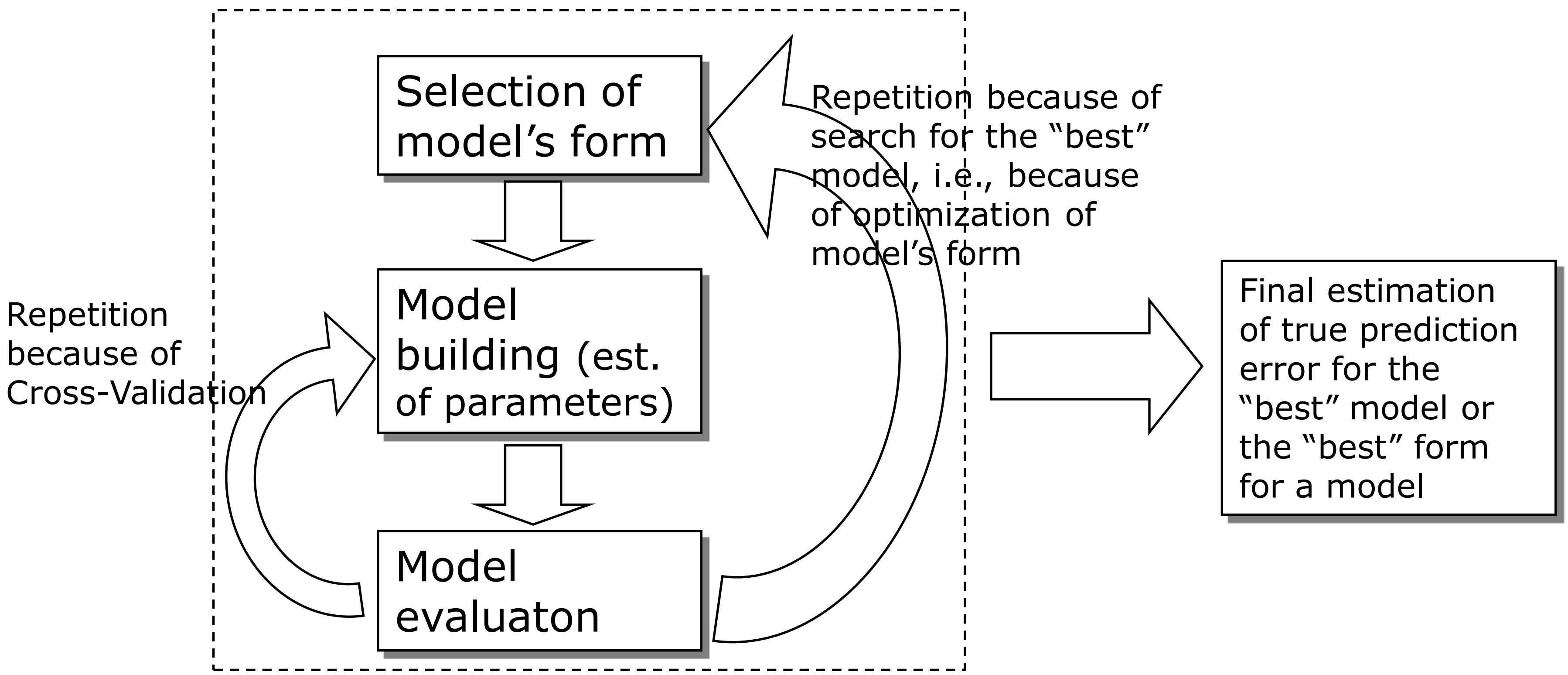
General scheme



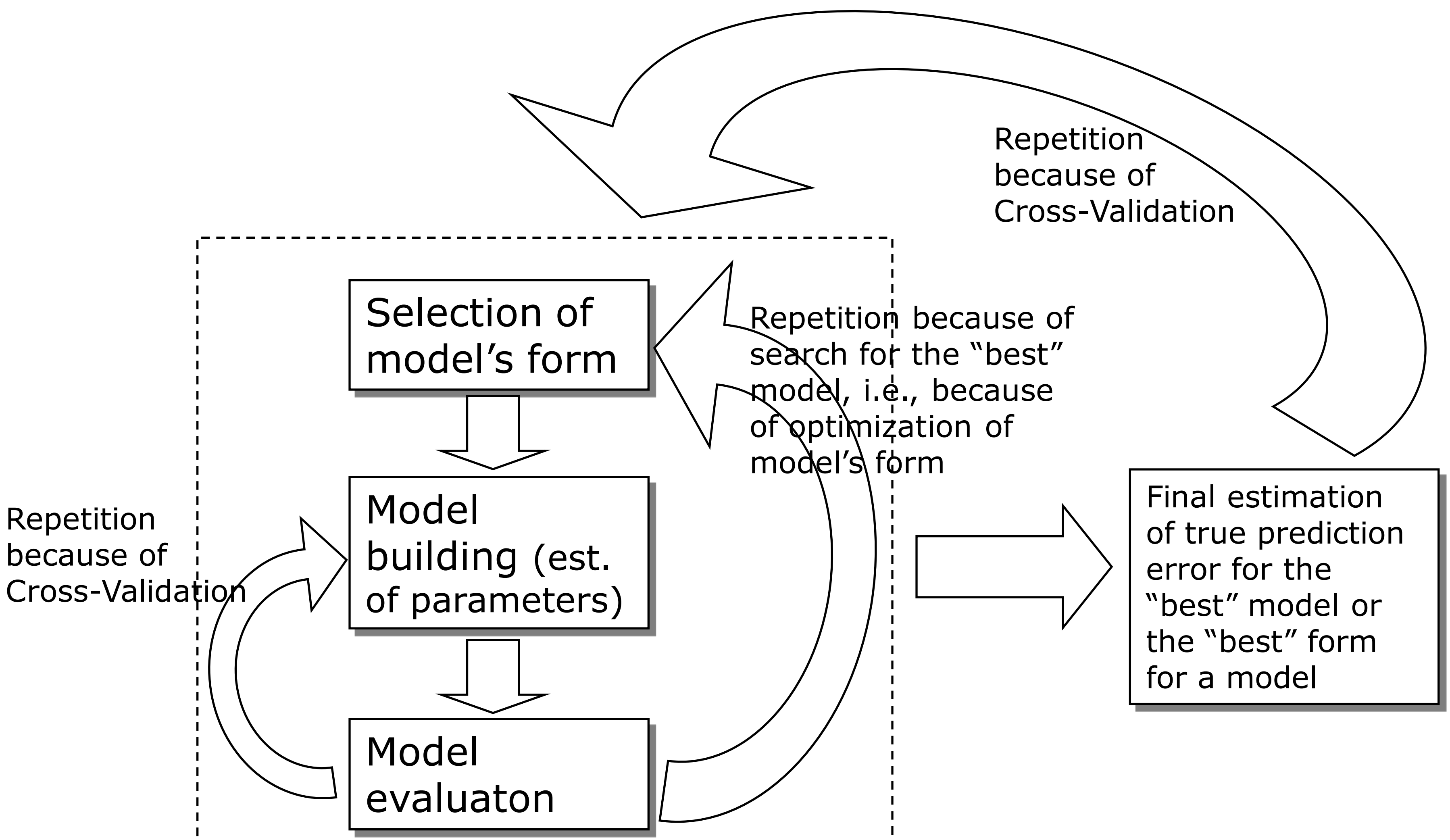
General scheme



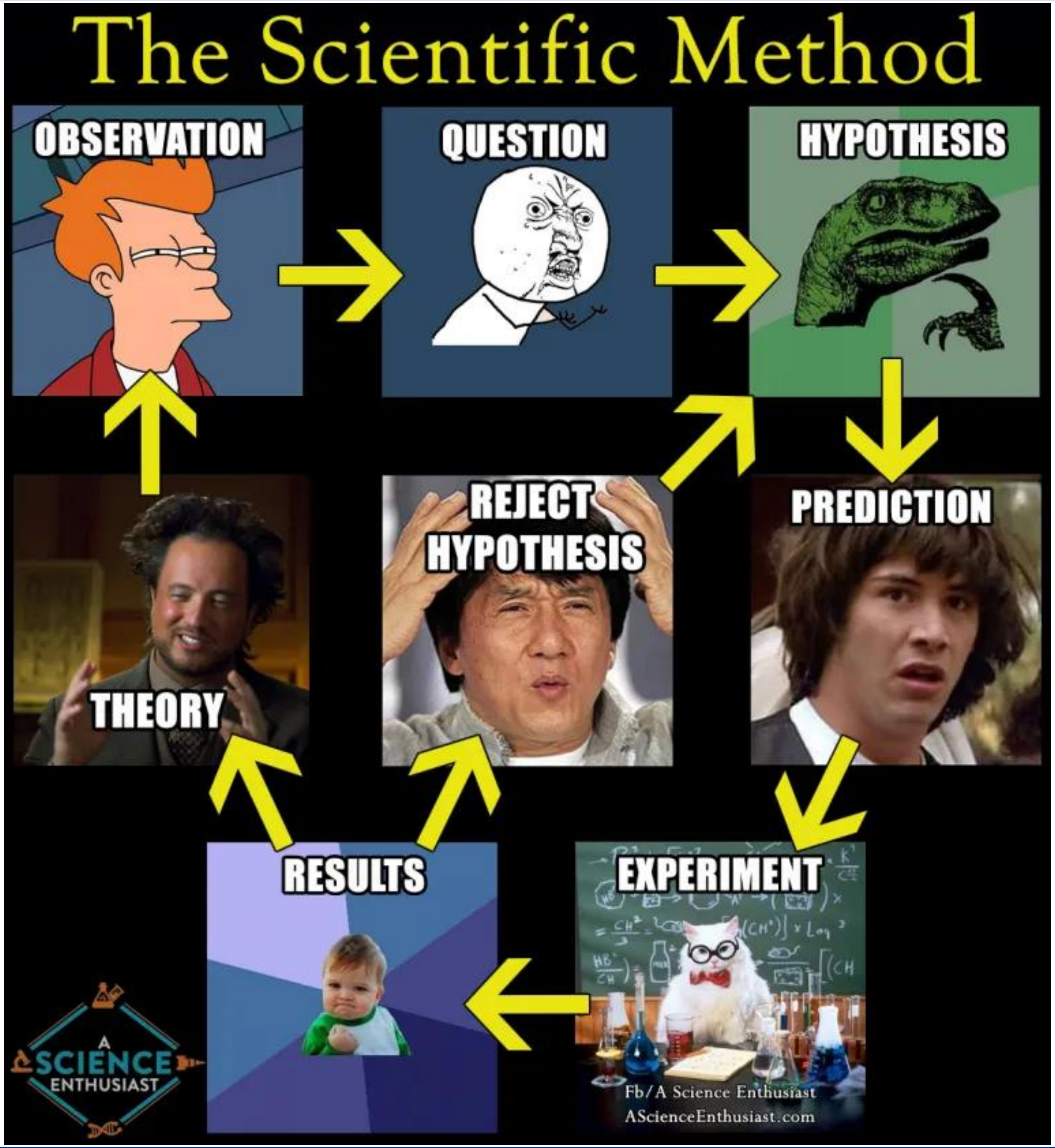
General scheme



General scheme



General scheme: the scientific method



https://ascienceenthusiast.com/steps-of-the-scientific-method/