NIO

Exercise 12: Bias &Variance

Duc Duy Pham, M.Sc.

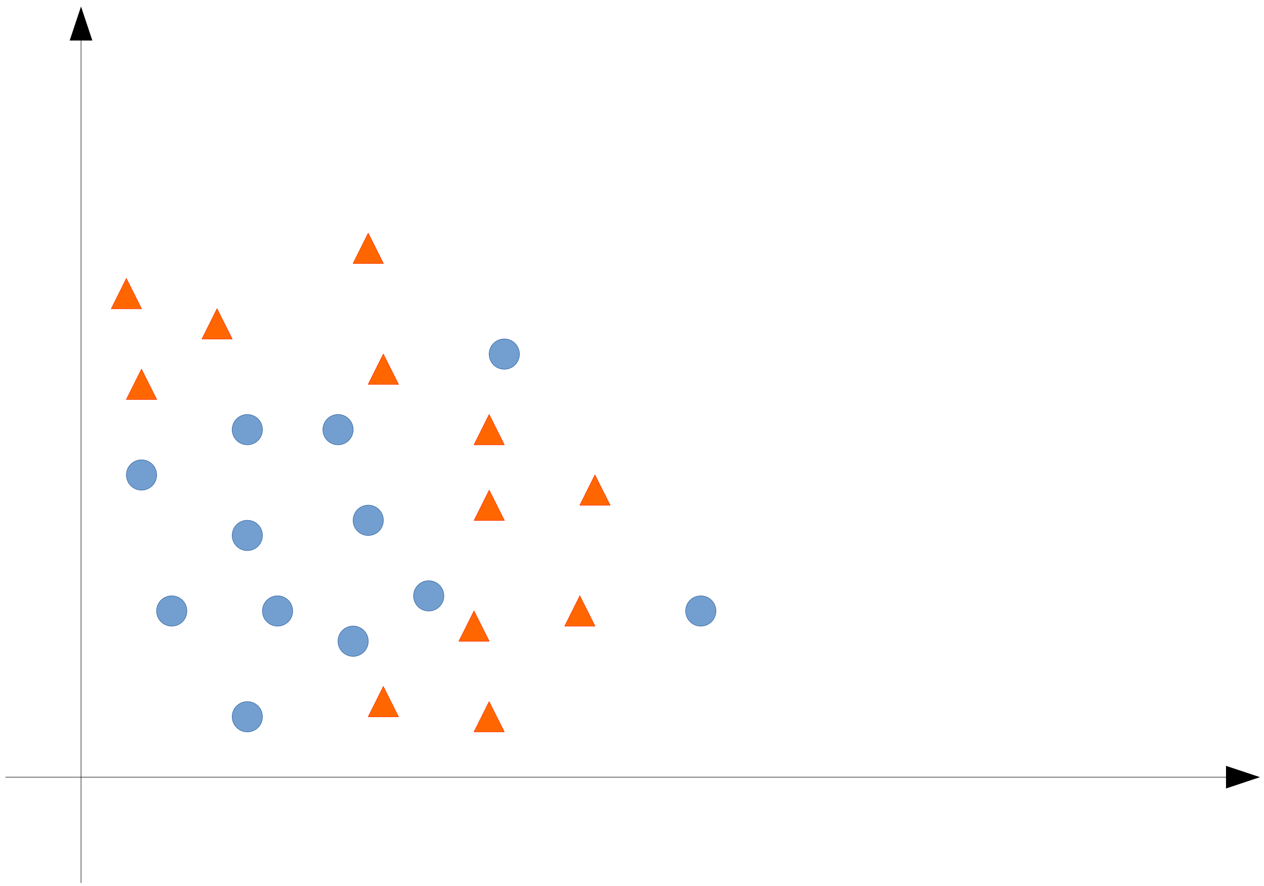
Raum: BC 410
Tel.: 0203-379-3734
Email: duc.duy.pham@uni-due.de

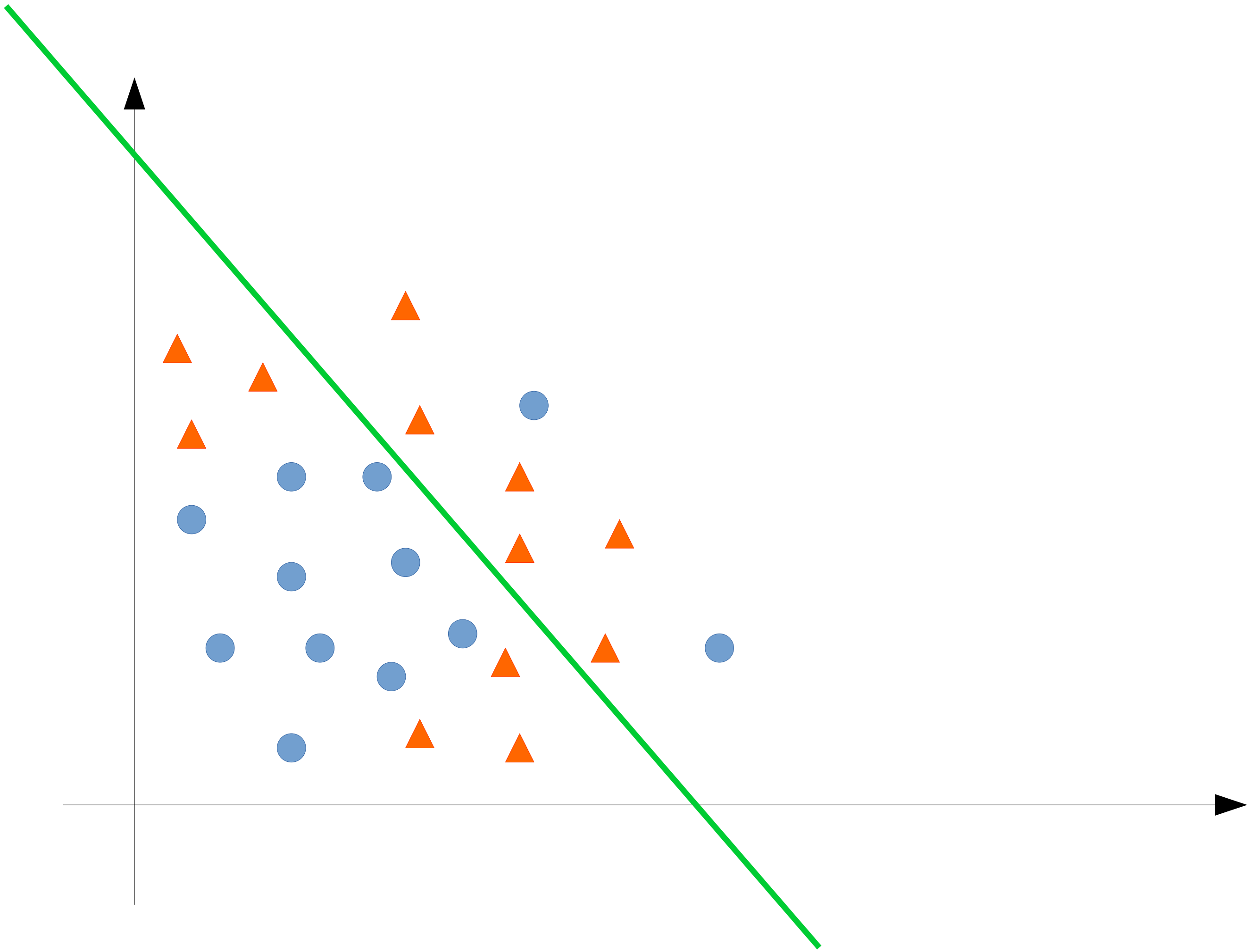
Contents

Contents

Model Flexibility

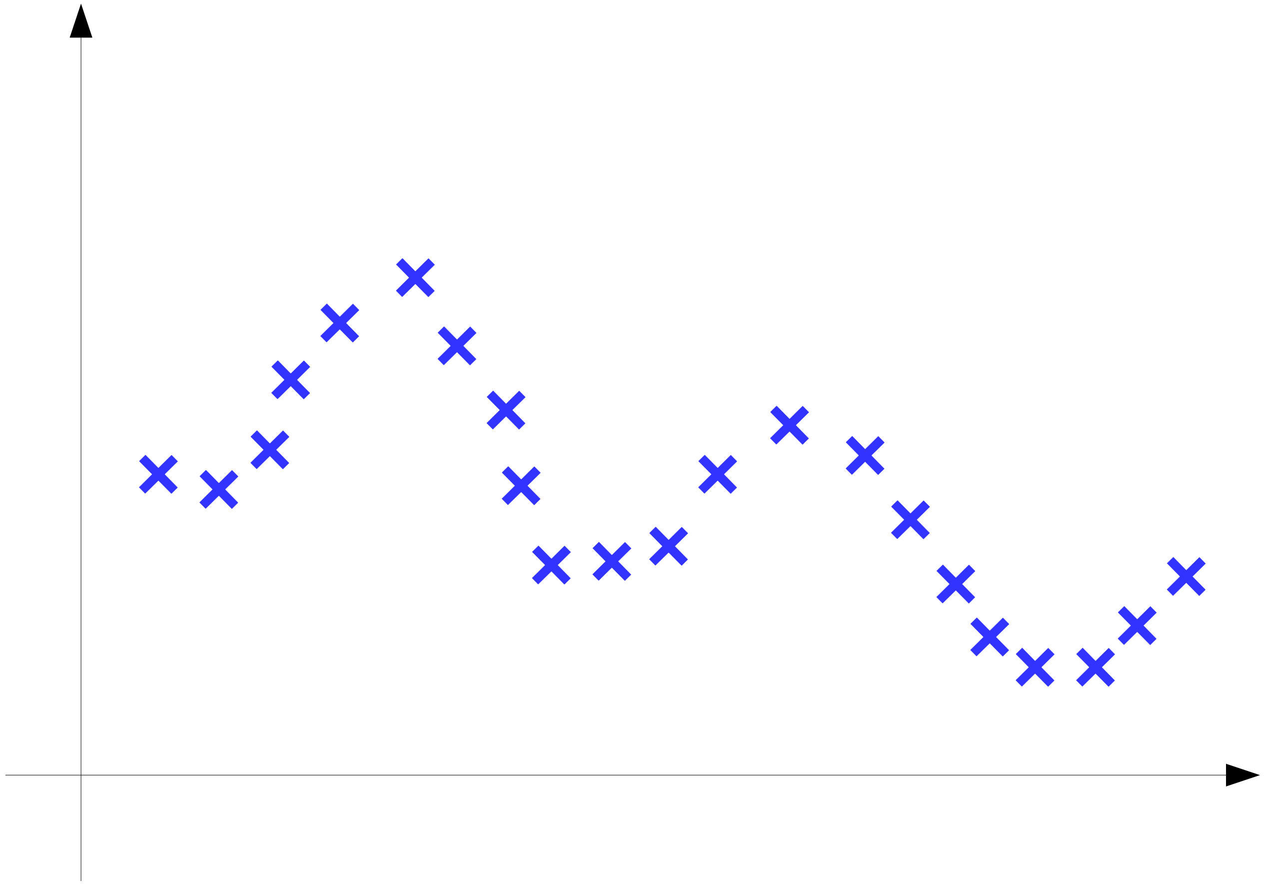
- ML models can be:
  - Not flexible enough
  - Too flexible
  - (Just right)

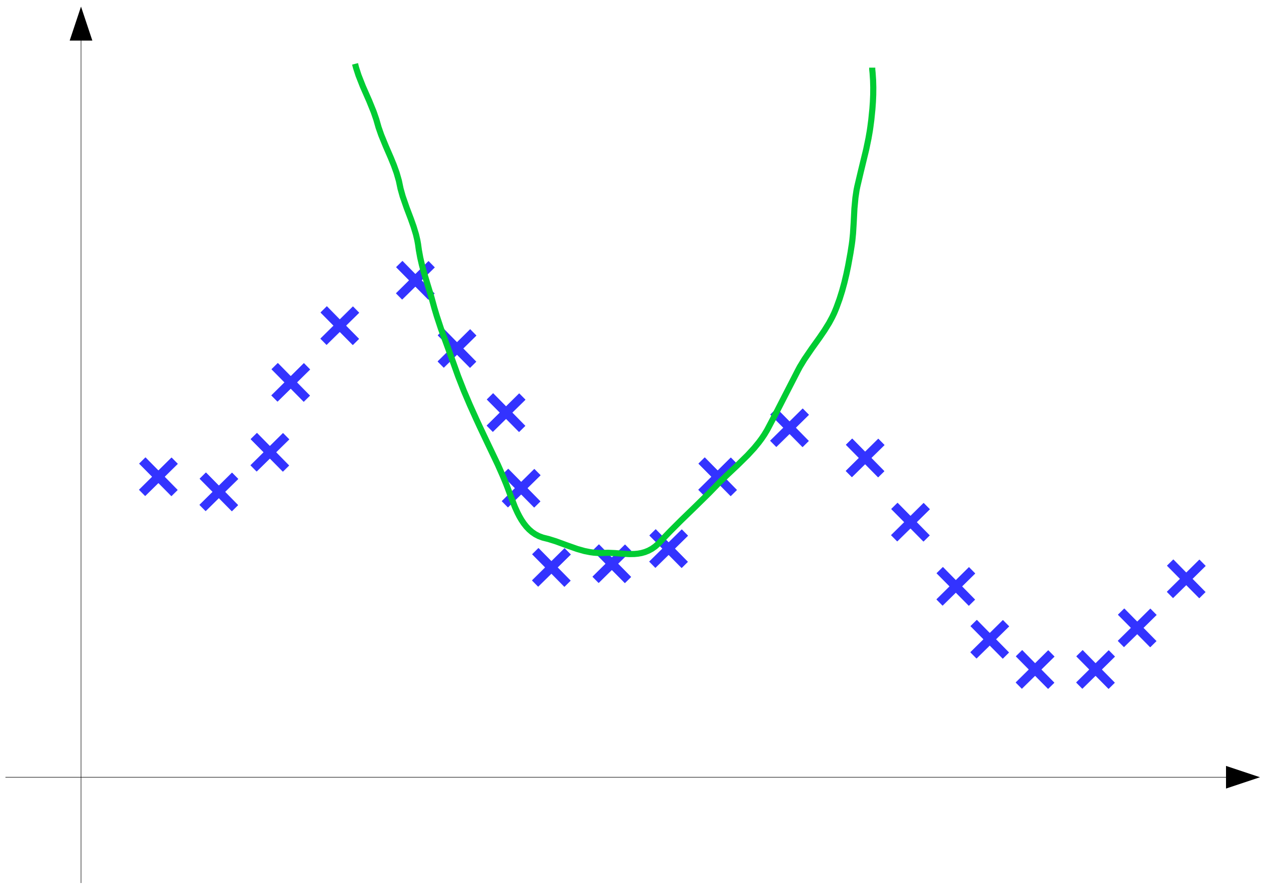


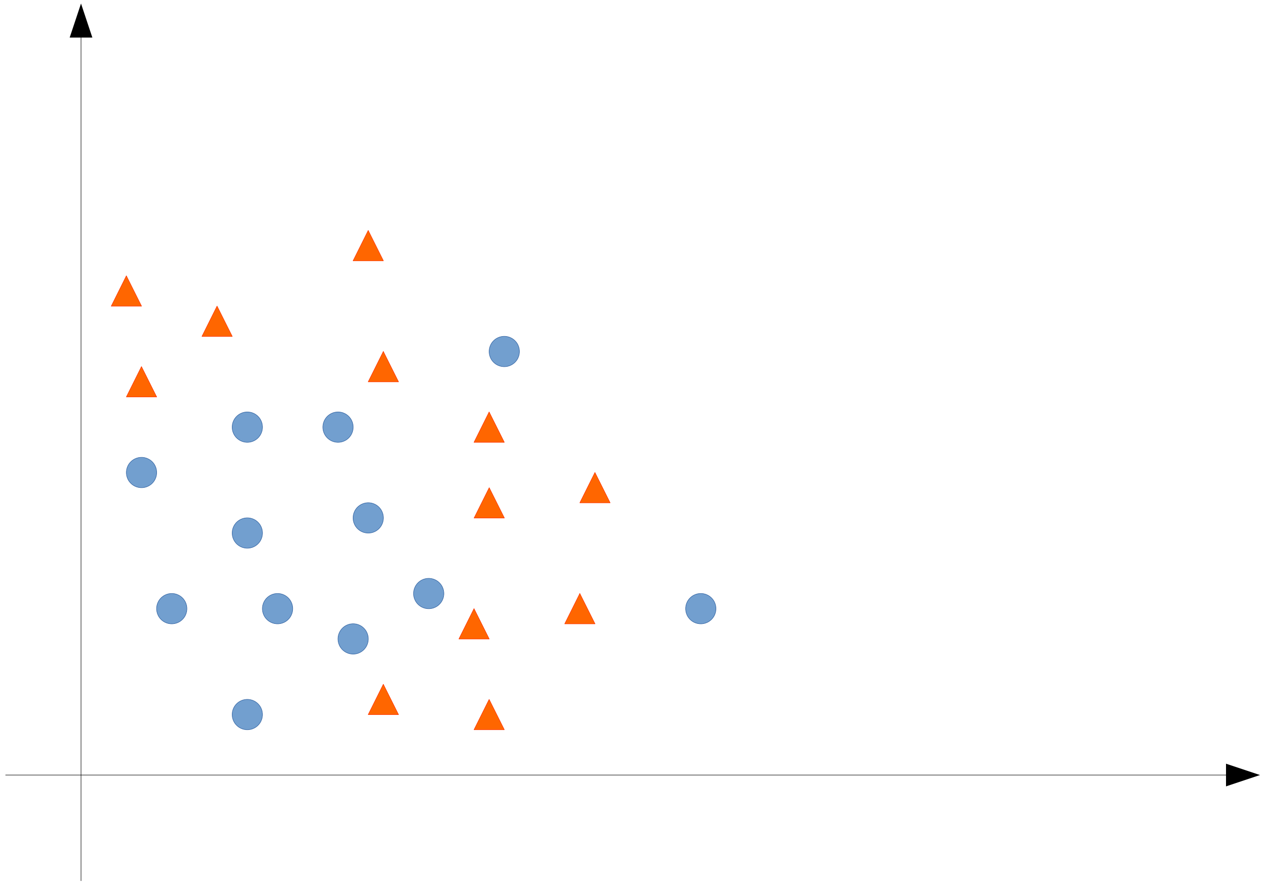


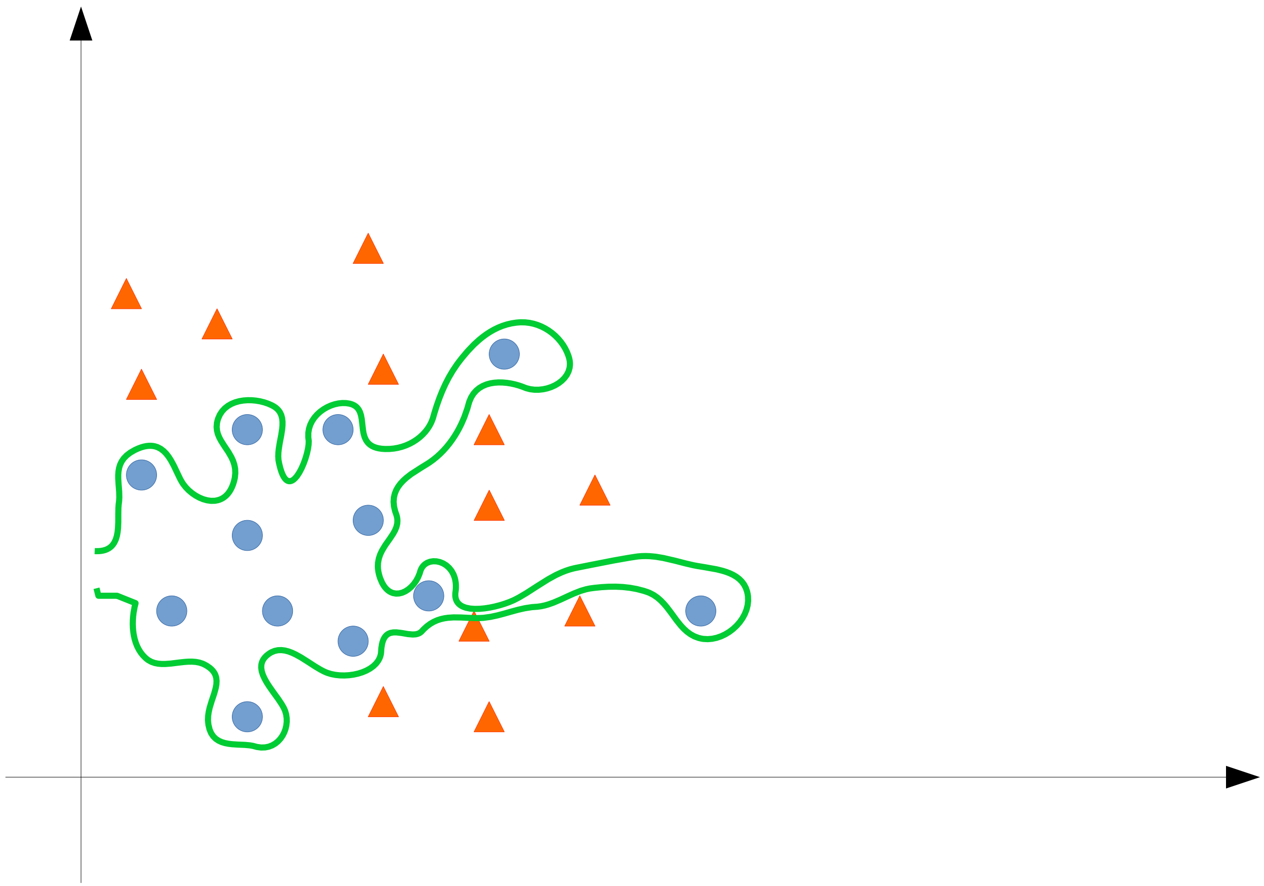
Model Flexibility

- Other example of lack of flexibility:
  - Too low assumption of polynomial degree for regression problem
  - Leads to high error rate even in training data



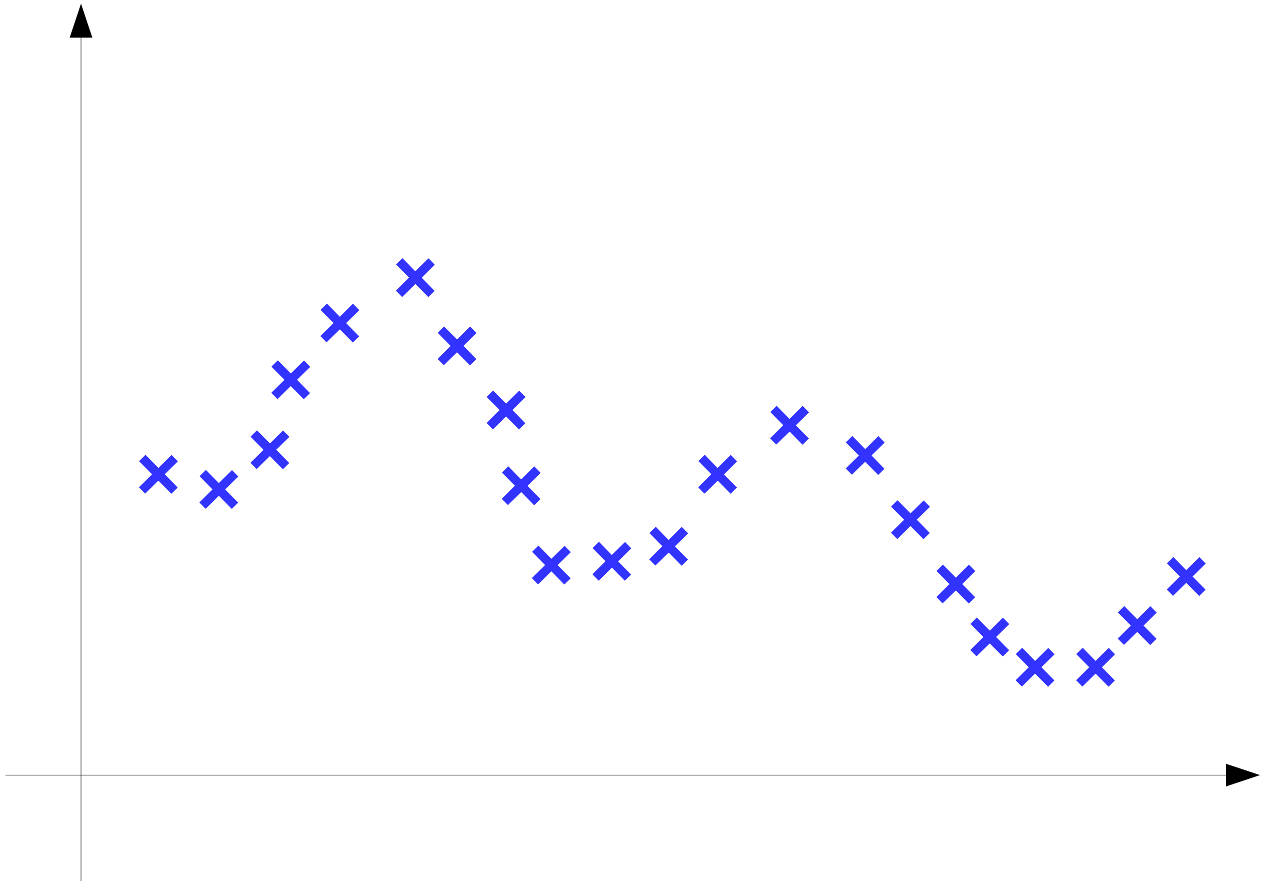


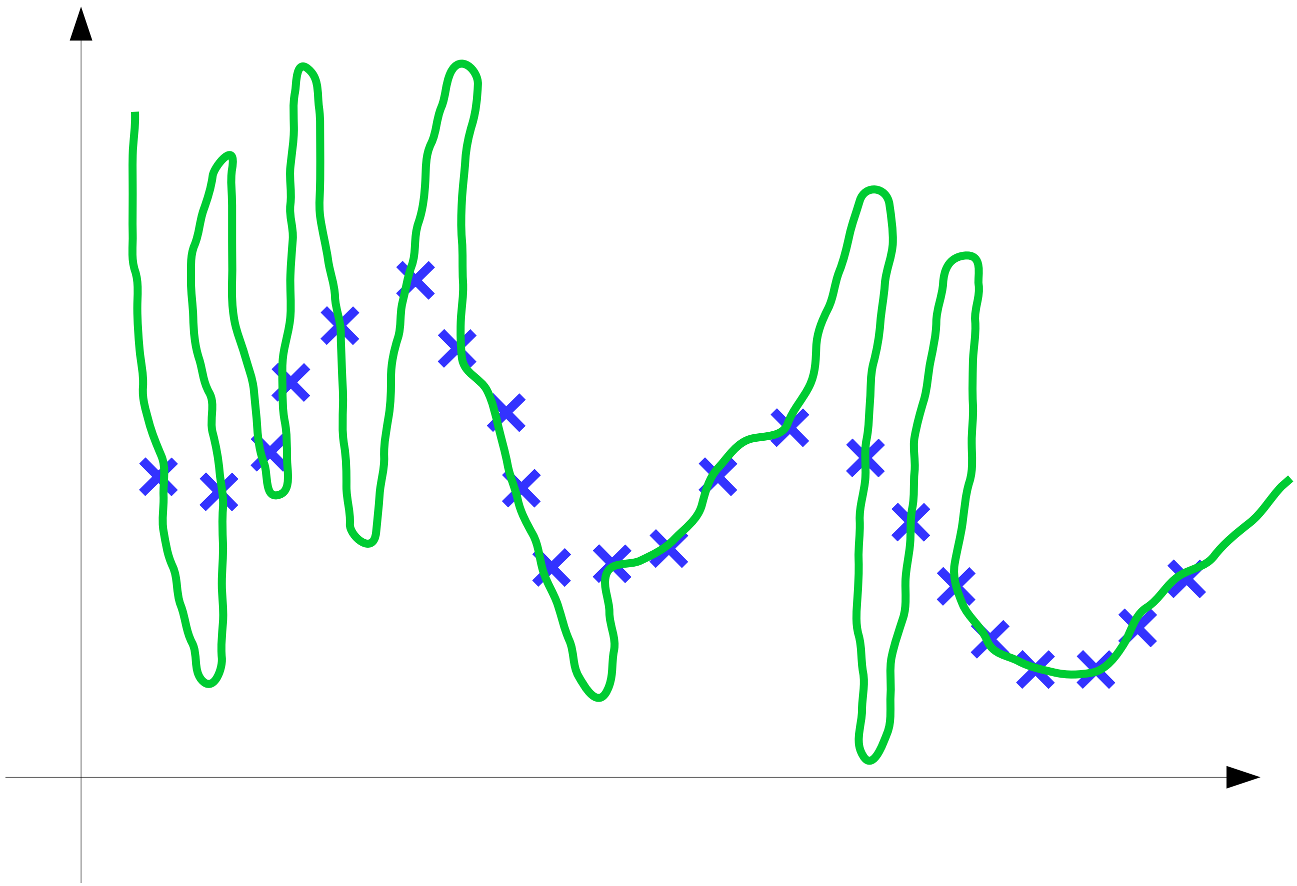




Model Flexibility

- Other example of too flexible models:
  - Too high assumption of polynomial degree for regression problem
  - Leads to low error rate in training data
  - Bad generalization (high error rate in validation and test data)



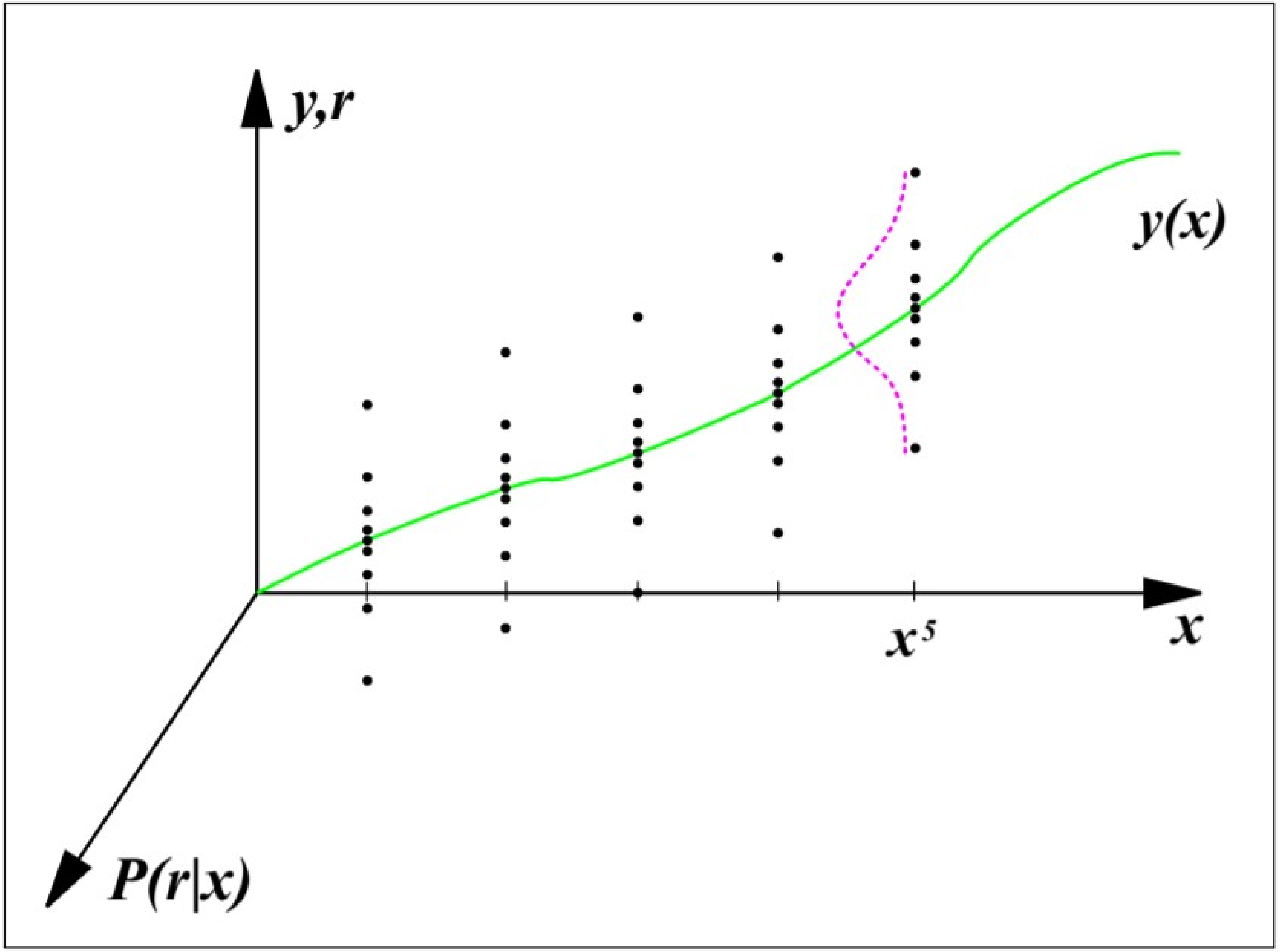


Contents

- Model Flexibility
- Bias & Variance

$$D(w) = \lim_{M \to \infty} \frac{1}{2M} \sum_{m=1}^{M} \sum_{i=1}^{N_L} (y_i(x^m, w) - r_i^m)^2$$

$$\begin{aligned} D(w) \;=\; & \underbrace{\frac{1}{2}\sum_{i=1}^{N_L}\int (y_i(x,w) - \langle r_i|x\rangle)^2 P(x)dx}_{\text{Term I}} + \\ & \underbrace{\frac{1}{2}\sum_{i=1}^{N_L}\int (\langle r_i^2|x\rangle - \langle r_i|x\rangle^2) P(x)dx}_{\text{Term II}} \end{aligned}$$

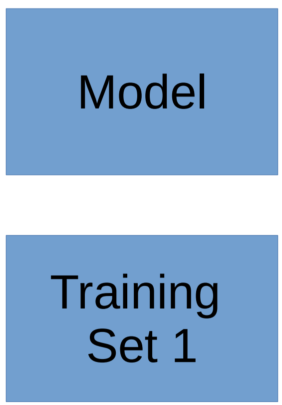

y,r
y(x)
x
x^5
P(r|x)

Revision of Lecture

- What is the „expected“ model output?

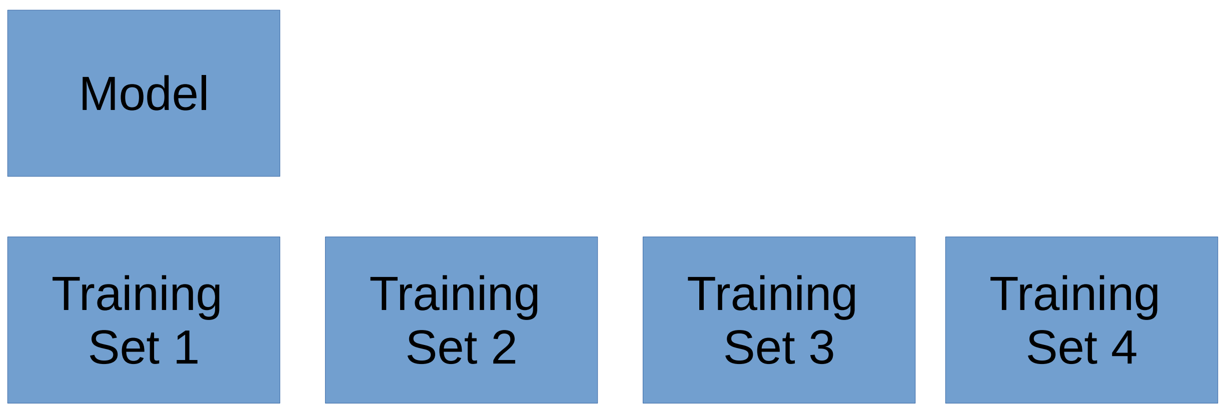
Revision of Lecture

- What is the „expected“ model output?



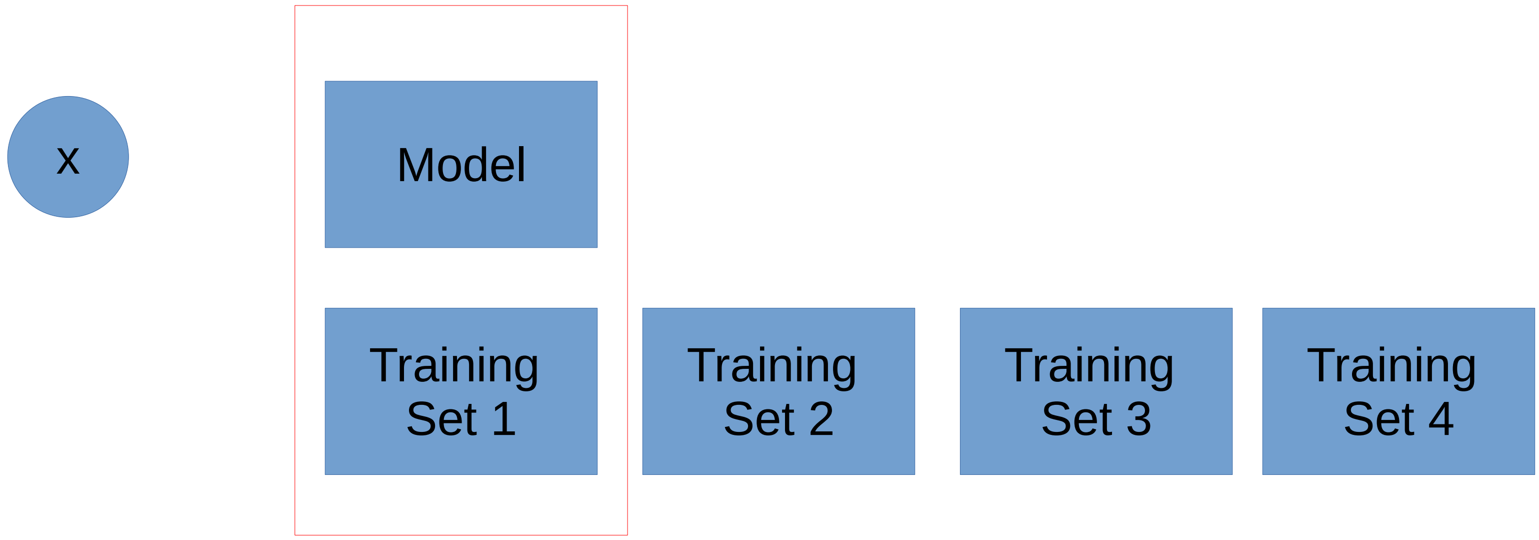
Revision of Lecture

- What is the „expected“ model output?



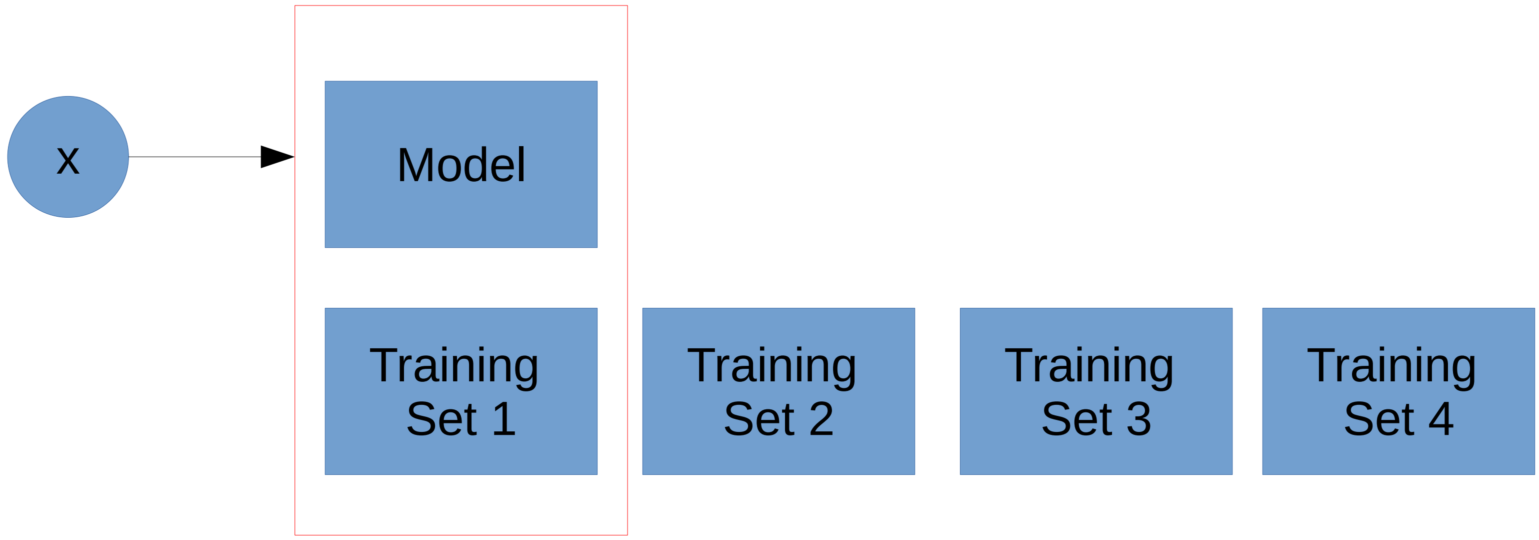
Revision of Lecture

- What is the „expected“ model output?



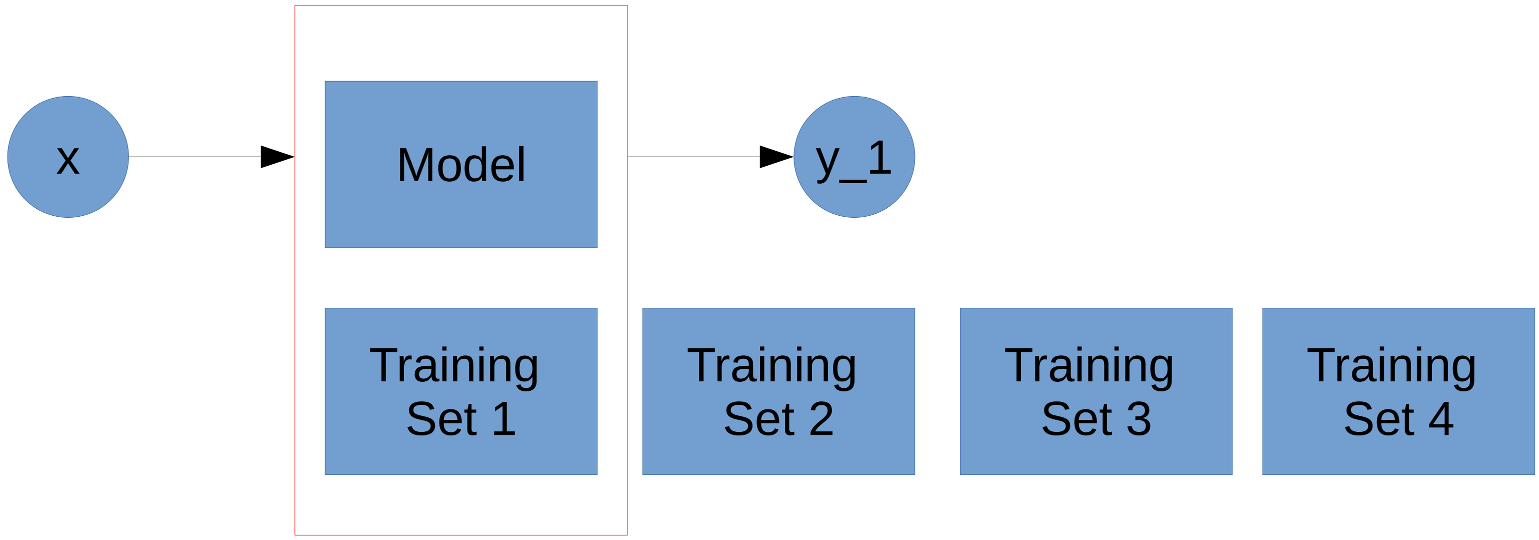
Revision of Lecture

- What is the „expected“ model output?



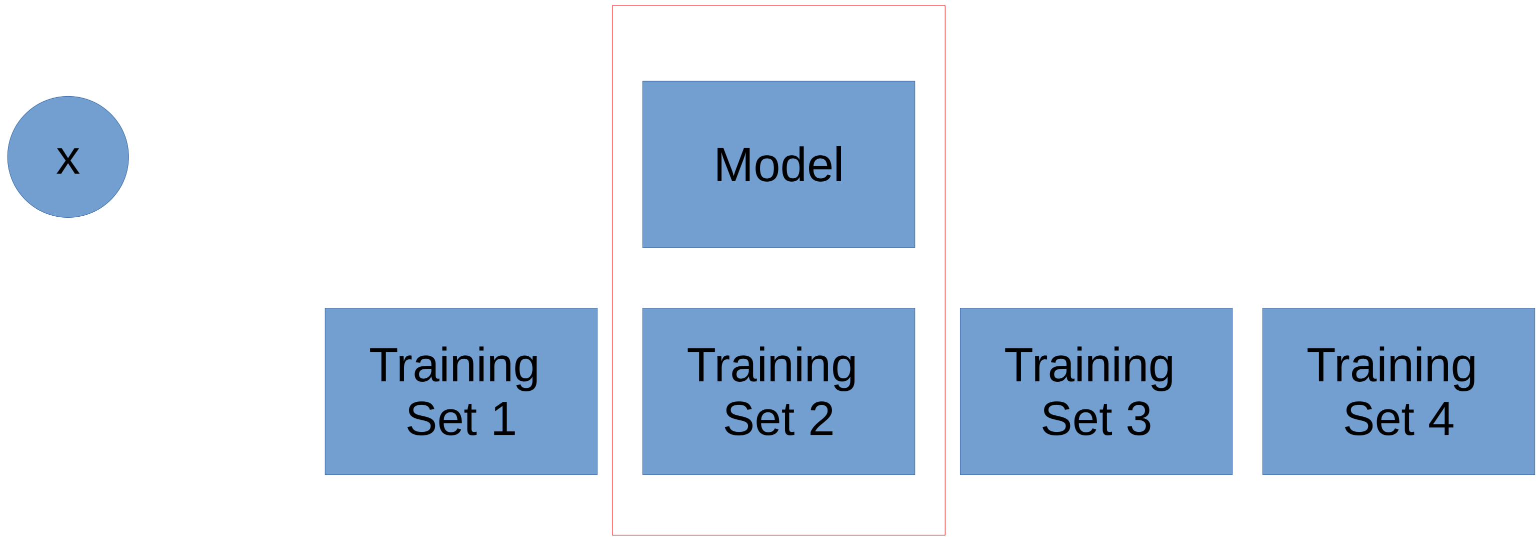
Revision of Lecture

- What is the „expected“ model output?



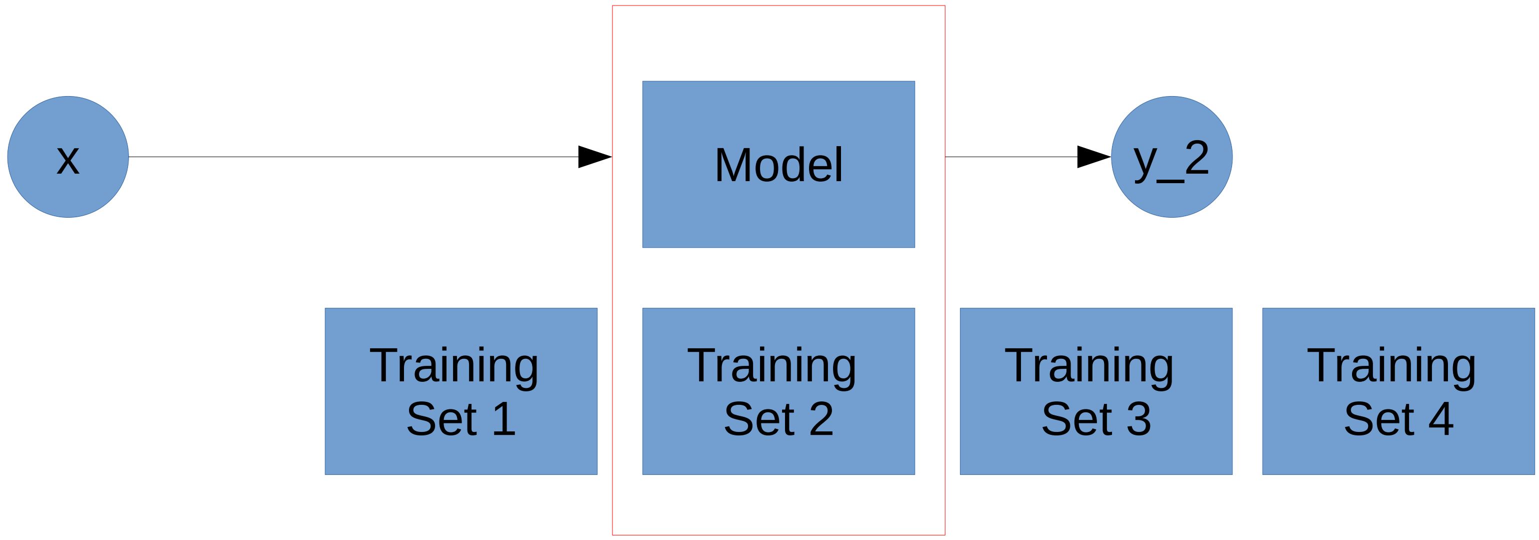
Revision of Lecture

- What is the „expected“ model output?



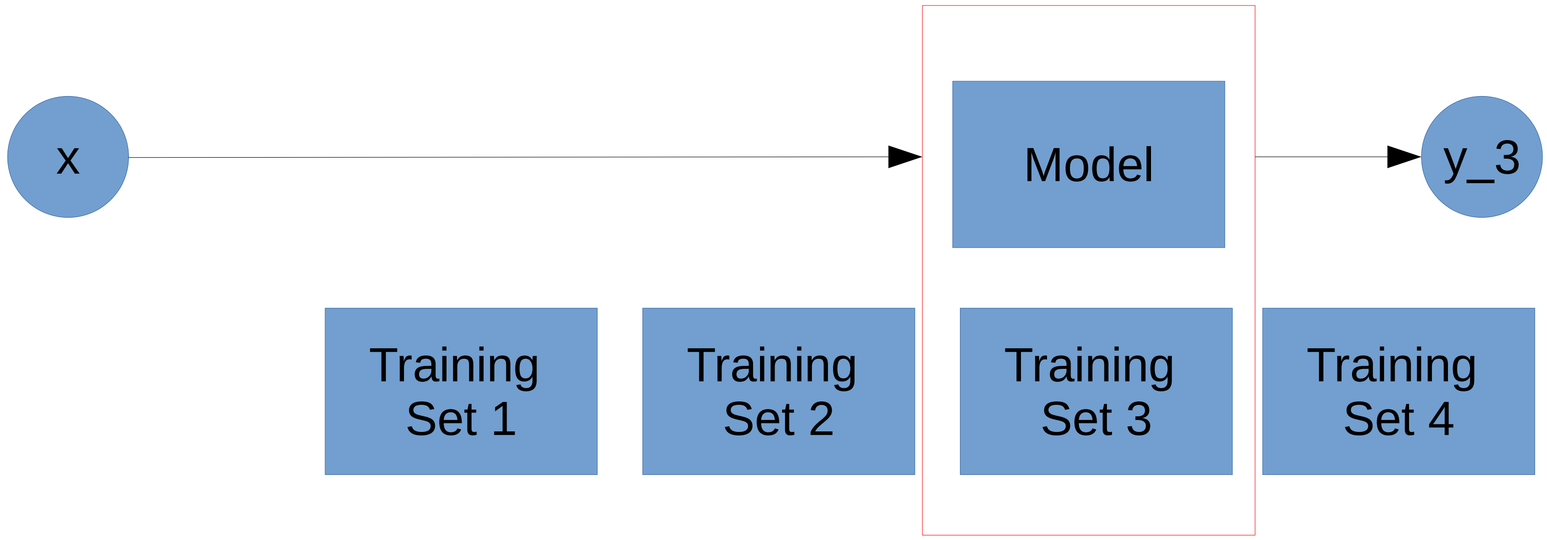
Revision of Lecture

- What is the „expected“ model output?



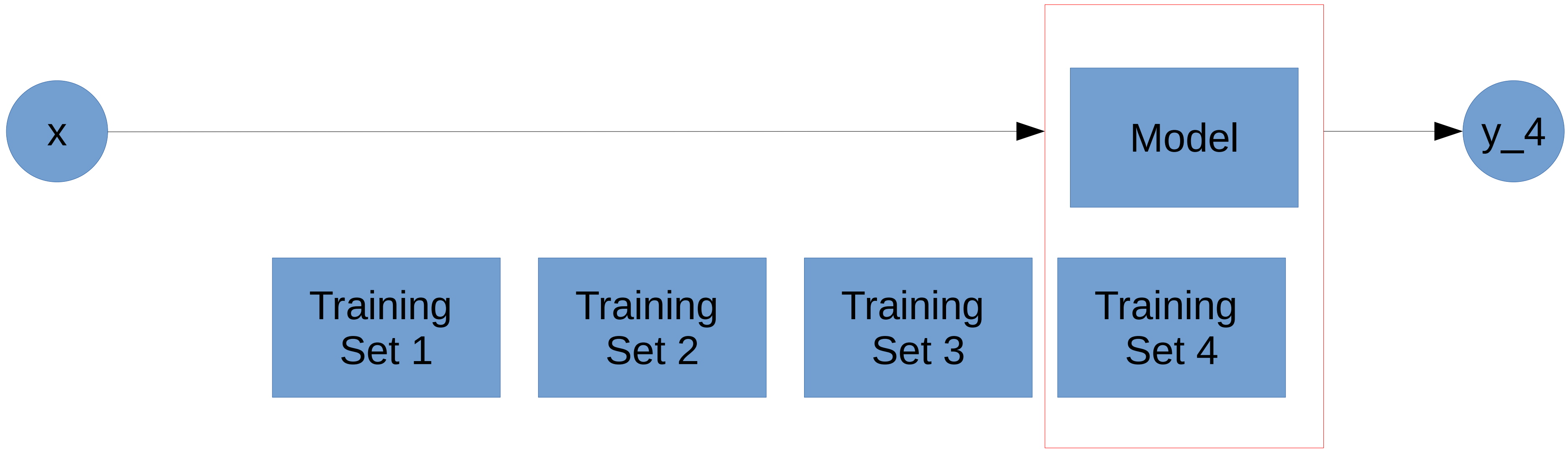
Revision of Lecture

- What is the „expected“ model output?



Revision of Lecture

- What is the „expected“ model output?



Revision of Lecture

- What is the „expected“ model output?
- The expected model output over multiple training sets!

Revision of Lecture

- What is Bias?

Revision of Lecture

- What is Bias?

$$\boldsymbol{E_T}\Big(\underbrace{\left(\boldsymbol{E_T}(\boldsymbol{y}(\boldsymbol{x})) - \langle \boldsymbol{r}|\boldsymbol{x}\rangle\right)^{2}}_{\text{Bias}_{\boldsymbol{y}}^{\boldsymbol{2}}(\boldsymbol{x})}\Big)$$

Offen im Denken

Revision of Lecture

- What is Bias?

$$E_T\Big(\underbrace{\left(E_T(y(x)) - \langle r|x\rangle\right)^2}_{\mathrm{Bias}_y^2(x)} \Big)$$

$$(\mathrm{Bias}_y)^2 \;=\; \frac{1}{2}\int \mathrm{Bias}_y^2(x)P(x)dx$$

Offen im Denken

Revision of Lecture

- What is Bias?
  - Deviation from expected output of (trained) model and expected target given some observation

Revision of Lecture

- What is Bias?
  - Deviation from expected output of (trained) model and expected target given some observation
  - High Bias
    => strong deviation from expected output of model to desired output

CAUTION

- Model not flexible enough => high bias
- BUT: High bias does not automatically mean, that model is not flexible enough!

Revision of Lecture

- What is Variance?

Revision of Lecture

- What is Variance?

$$\underbrace{E_T\Big((y(x) - E_T(y(x)))^2\Big)}_{\text{Varianz}_y(x)}$$

Revision of Lecture

- What is Variance?

$$\underbrace{E_T\left(\left(y(x) - E_T(y(x))\right)^2\right)}_{\text{Varianz}_y(x)}$$

$$\text{Varianz}_y = \frac{1}{2}\int \text{Varianz}_y(x) P(x) dx$$

Revision of Lecture

- What is Variance?
  - Expected (squared) deviation from model output and *expected model output*

Revision of Lecture

- What is Variance?
  - Expected (squared) deviation from model output and *expected model output*
  - High Variance
=> Model tends to strong deviations from expected output in training data

Revision of Lecture

- What is Variance?
  - Expected (squared) deviation from model output and *expected model output*
  - High Variance
    => Model tends to strong deviations from expected output in training data
    => Model is strongly dependent on choice of training data!

Revision of Lecture

- What is Variance?
  - Expected (squared) deviation from model output and *expected model output*
  - High Variance
    => Model tends to strong deviations from expected output in training data
    => Model is strongly dependent on choice of training data!
    => Generalization is bad!
  - Often caused by too flexible model

CAUTION

- Model too flexible => high variance (often high bias on unseen data)
- Also: High variance does not automatically mean, that model is too flexible!

$$E_T\Big((y(x) - \langle r|x\rangle)^2\Big) = \underbrace{E_T\Big((y(x) - E_T(y(x)))^2\Big)}_{\mathsf{Varianz}_y(x)} +$$

$$E_T\Big(\underbrace{(E_T(y(x)) - \langle r|x\rangle)^2}_{\mathsf{Bias}_y^2(x)} \Big)$$

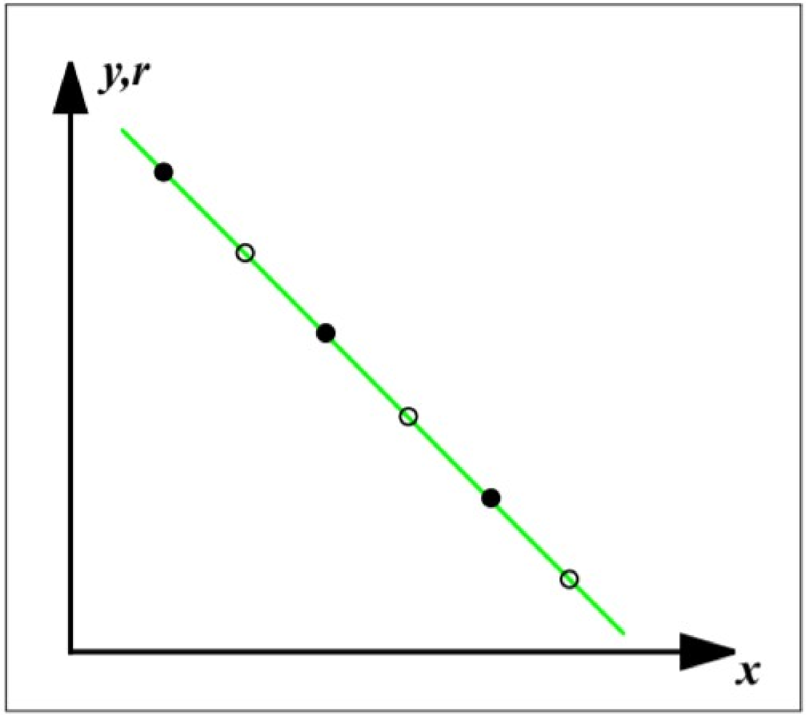
Mittelung über alle x:

$$(\mathsf{Bias}_y)^2 = \frac{1}{2} \int \mathsf{Bias}_y^2(x) P(x) dx$$

$$\mathsf{Varianz}_y = \frac{1}{2} \int \mathsf{Varianz}_y(x) P(x) dx$$

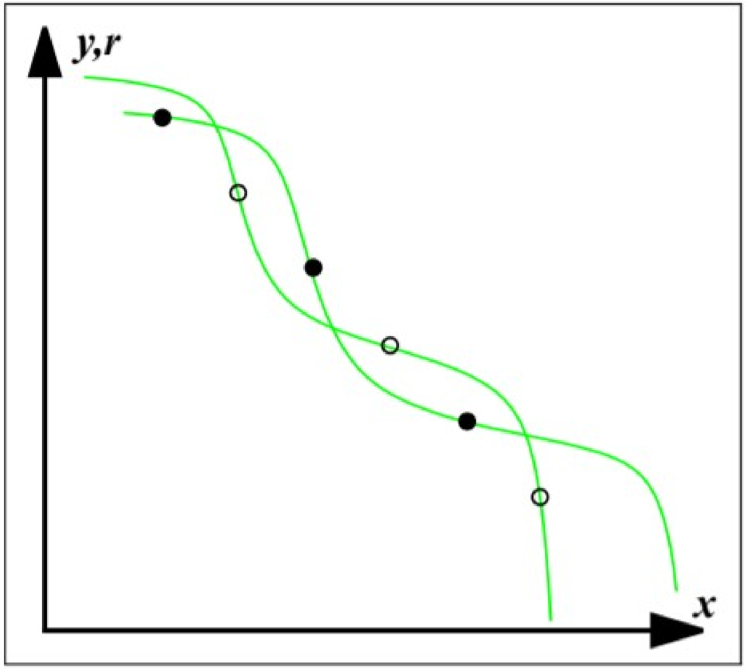
Beispiele für Bias und Varianz

a) Beispiel mit linearer Modellfunktion: keine Varianz und kein Bias, wenn Daten tatsächlich linear verteilt sind.



Beispiele für Bias und Varianz

b) Beispiel mit polynominaler Modellfunktion: Varianz und Bias abhängig von tatsächlicher Verteilung der Daten und verwendetem Grad des Polynoms.



Beispiele für Bias und Varianz

c) Beispiel mit einer festen Funktion $\overline{\boldsymbol{f}}(\boldsymbol{x})$ als Abbildungsfunktion $\boldsymbol{y}(\boldsymbol{x})$, wobei $\overline{\boldsymbol{f}}(\boldsymbol{x})$ unabhängig von $\boldsymbol{\Omega_T}$ sei. $\overline{\boldsymbol{f}}$ wird nicht gelernt, sondern a priori vorgegeben.

Varianz verschwindet, da $\boldsymbol{y}(\boldsymbol{x}) = \overline{\boldsymbol{f}}(\boldsymbol{x})$, und somit $\boldsymbol{E_T}(\boldsymbol{y}(\boldsymbol{x})) = \overline{\boldsymbol{f}}(\boldsymbol{x})$.

Bias ist i.A. hoch, da Abhängigkeit von Trainingsmenge nicht berücksichtigt wurde.

Beispiele für Bias und Varianz

d) Beispiel mit auswendig gelernten Funktionen $\boldsymbol{f_j(x)}$, die die Trainingsmengen $\boldsymbol{\Omega_{T_j}}$ jeweils perfekt widerspiegeln.

Für die Trainingselemente, die in der Schnittmenge der herangezogenen $\boldsymbol{\Omega_{T_j}}$ liegen, $\boldsymbol{j \in \{1, \cdots\}}$, ist Bias und Varianz gleich 0.

Für die übrigen Trainingselemente ist Bias und Varianz i.A. sehr hoch.

Revision of Lecture

- What is the Bias-Variance-Dilemma?

Revision of Lecture

- What is the Bias-Variance-Dilemma?
  - Want to minimize Bias and Variance at the same time
  - Tradeoff between Bias and Variance
  - Low Bias => (often) High Variance
  - Low Variance => High Bias
  - Need to find balance between both

Revision of Lecture

- Solution attempt?

Revision of Lecture

- Solution attempt?
  - Use flexible model
    - Low expected Bias
    - High expected Variance
  - Reduce Variance by providing large amount of training data
    - Training data should cover a wide range of different samples
    - This strategy is often applied for CNNs (which tend to overfit!)

Revision of Lecture

- What is Cross Validation?

Revision of Lecture

- What is Cross Validation?
  - Variable separation of (learning) data set into training and validation (evaluation) sets, i.e.

$$\Omega_L := \Omega_{T_j} \cup \Omega_{E_j} \quad , j \in \{1, 2, 3, \ldots, K\}$$

  - Systematic or random change of separation strategy
  - Training of machine learning model for each training set Ω_{T_j} and validation (evaluation) with respective Ω_{E_j}
  - K separations leads to K differently trained models!

Systematic Cross Validation

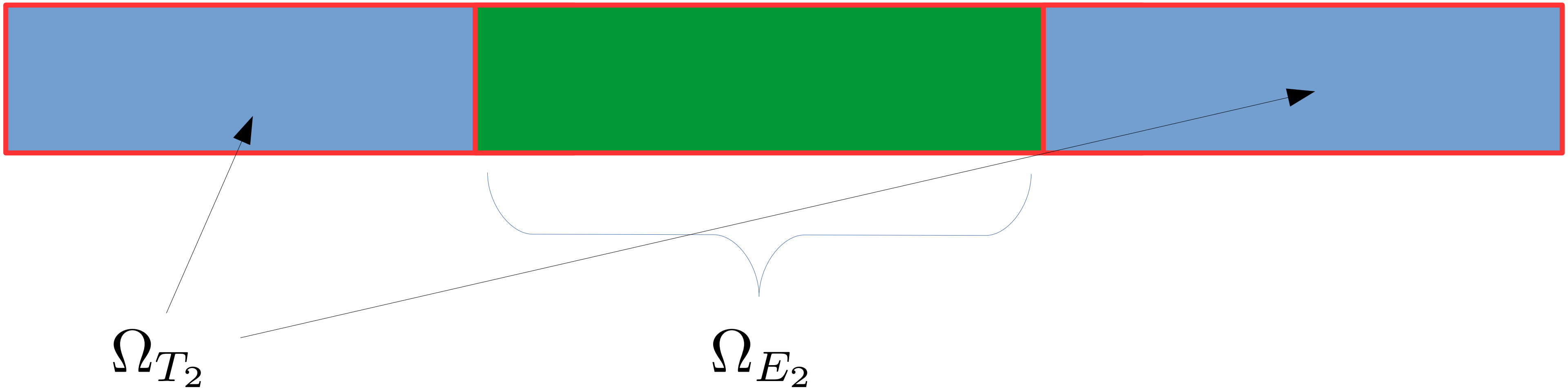
Learning data set Ω_L

Systematic Cross Validation

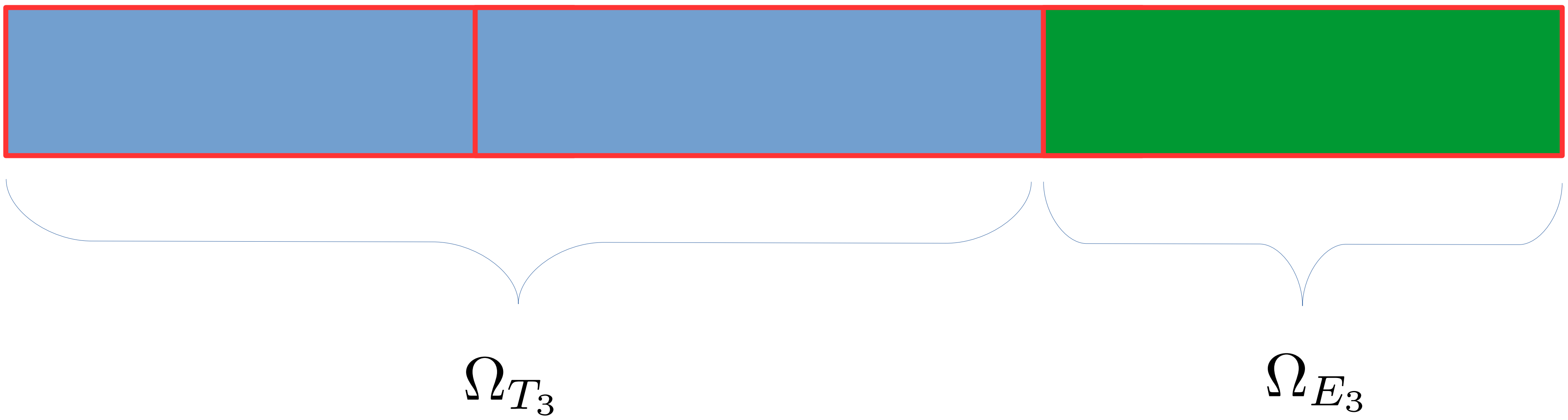
Make K=3 Separations of Learning Set

Systematic Cross Validation

Systematic Cross Validation



Systematic Cross Validation



Revision of Lecture

- What is Cross Validation used for?

Revision of Lecture

- What is Cross Validation used for?
  - 1. Use case:
    - Check chosen machine learning model for its generalization capacities:
      - Question of interest: Are the chosen hyperparameters for model good? / Is chosen model good?
      - K separations => K validations (of chosen and fixed hyper parameters)
      - If validation error rates are all good
        => assumed good generalization

Revision of Lecture

- What is Cross Validation used for?
  - 2. Use case:
    - Construct more robust (ensemble) model from trained models:
      - Question of interest: How to create more robust model (with current hyperparameters/ with current model)?
      - K separations = > K differently trained models (with same hyper parameters)
      - Combine K differently trained models into one ensemble model