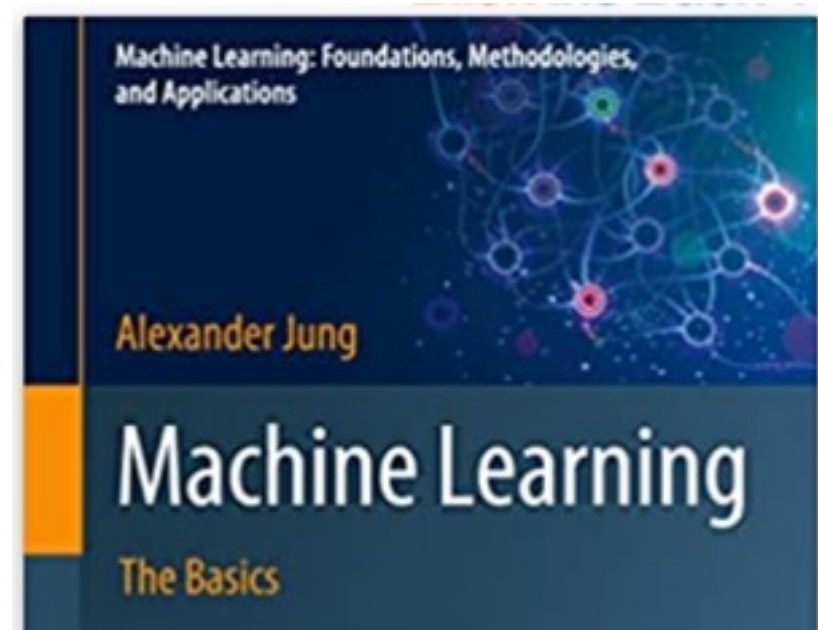



# ML Diagnosis and Generalization

Alex(ander) Jung  
Assistant Professor for Machine Learning  
Department of Computer Science  
Aalto University

# Reading.

Ch. 6.6 of <https://mlbook.cs.aalto.fi>



[Install](#) [User Guide](#) [API](#) [Examples](#) [Community](#) [More ▾](#)

[Prev](#) [Up](#) [Next](#)

**scikit-learn 1.1.1**  
[Other versions](#)

Please [cite us](#) if you use the

## 3. Model selection and evaluation

### 3.1. Cross-validation: evaluating estimator performance

- 3.1.1. Computing cross-validated metrics

[https://scikit-learn.org/stable/model\\_selection.html](https://scikit-learn.org/stable/model_selection.html)

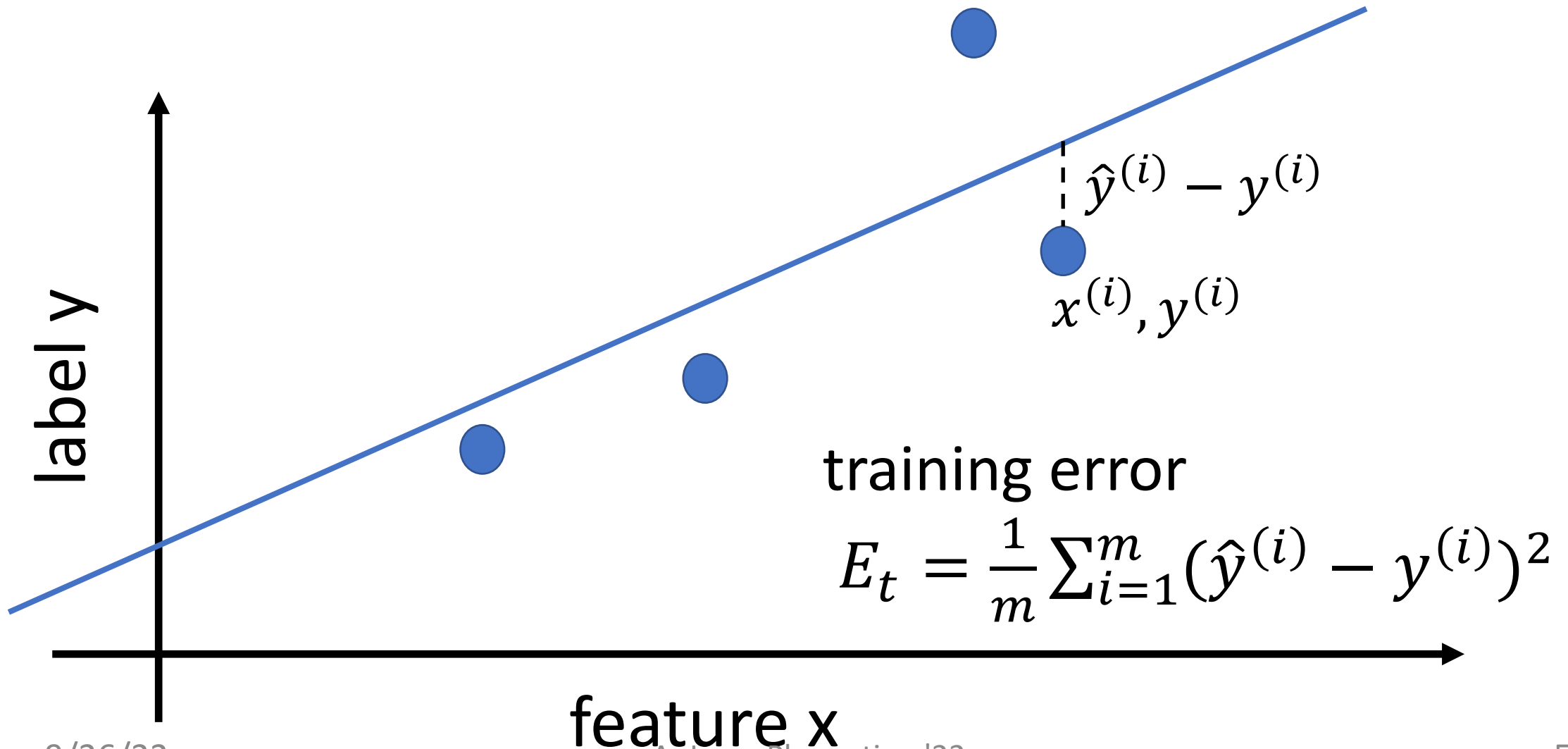
# Learning Goals

- large models (such as deep nets) are prone to overfitting
- train/val error are random variables under i.i.d. asspt.
- Know that there are pathological datasets where  $\text{val err} < \text{train err}$
- Know how to "read" val/train errors
- Know how to improve performance based on diagnosis
- Know some rule of thumb "sample size  $\gg$  effective number of tunable parameters"
- Know effective number of tunable parameter for lin.reg, and poly.reg

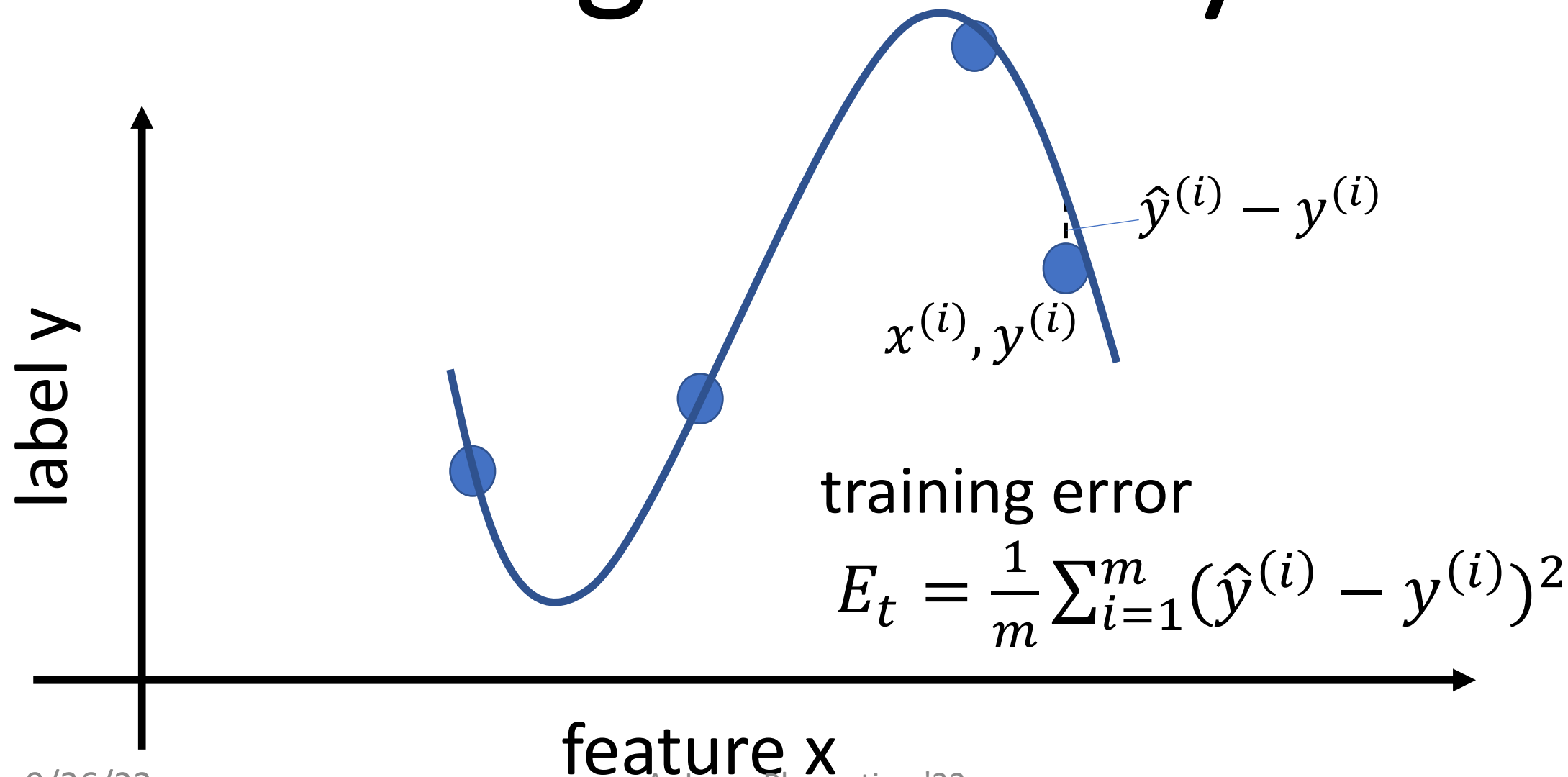
# What are three **main** **components** of machine **learning**?



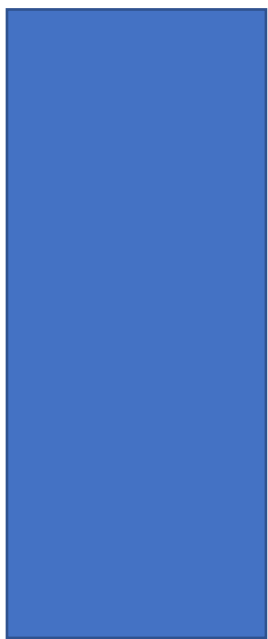
# Learn Linear Predictor



# Learn Degree 3 Polyn.



# Training Errors

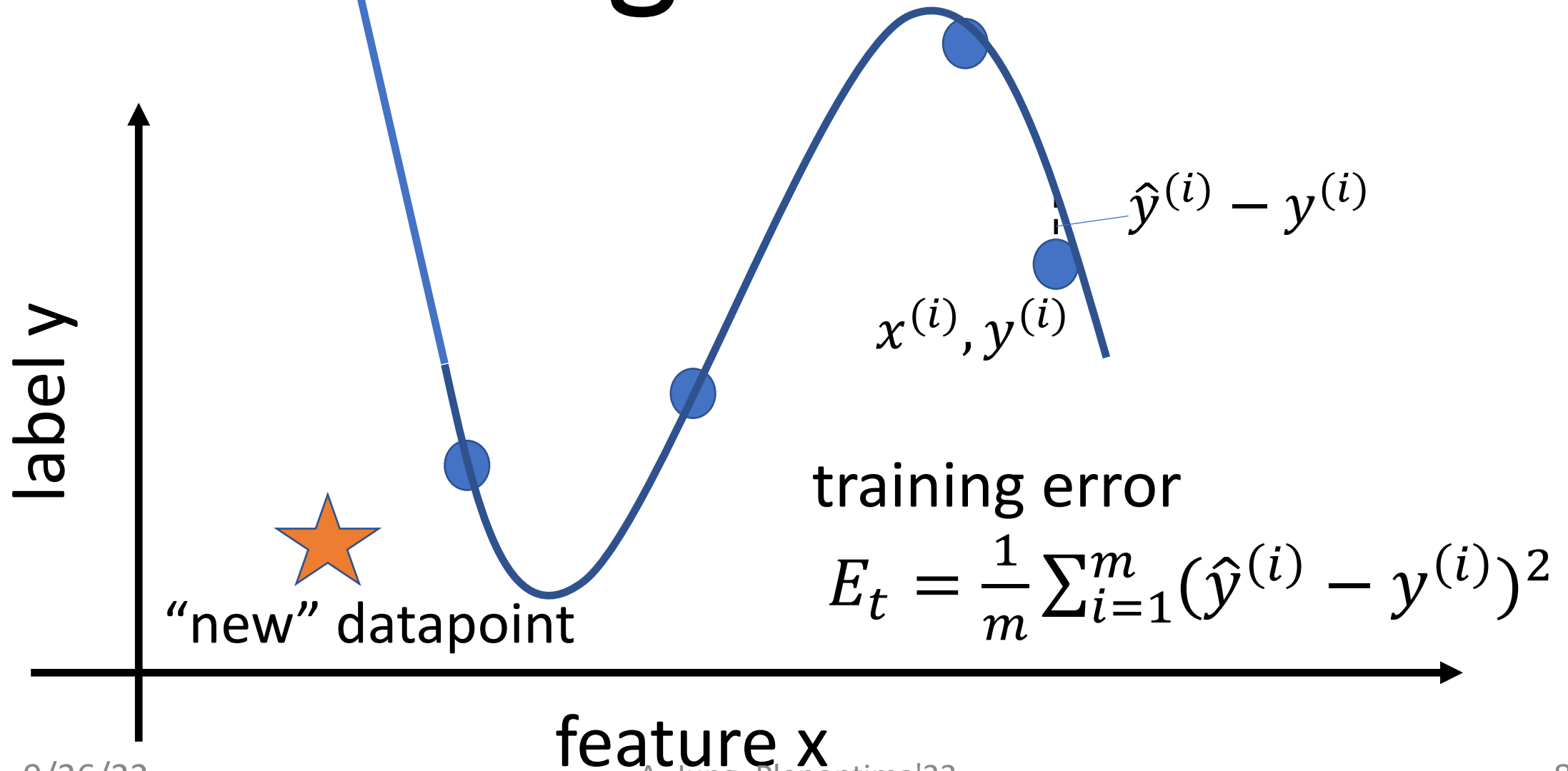


model 1  
linear predictors



model 2:  
degree 3 polyn.

# Overfitting





Small Training Error Does Not  
Imply Good Performance on  
New Data Points!

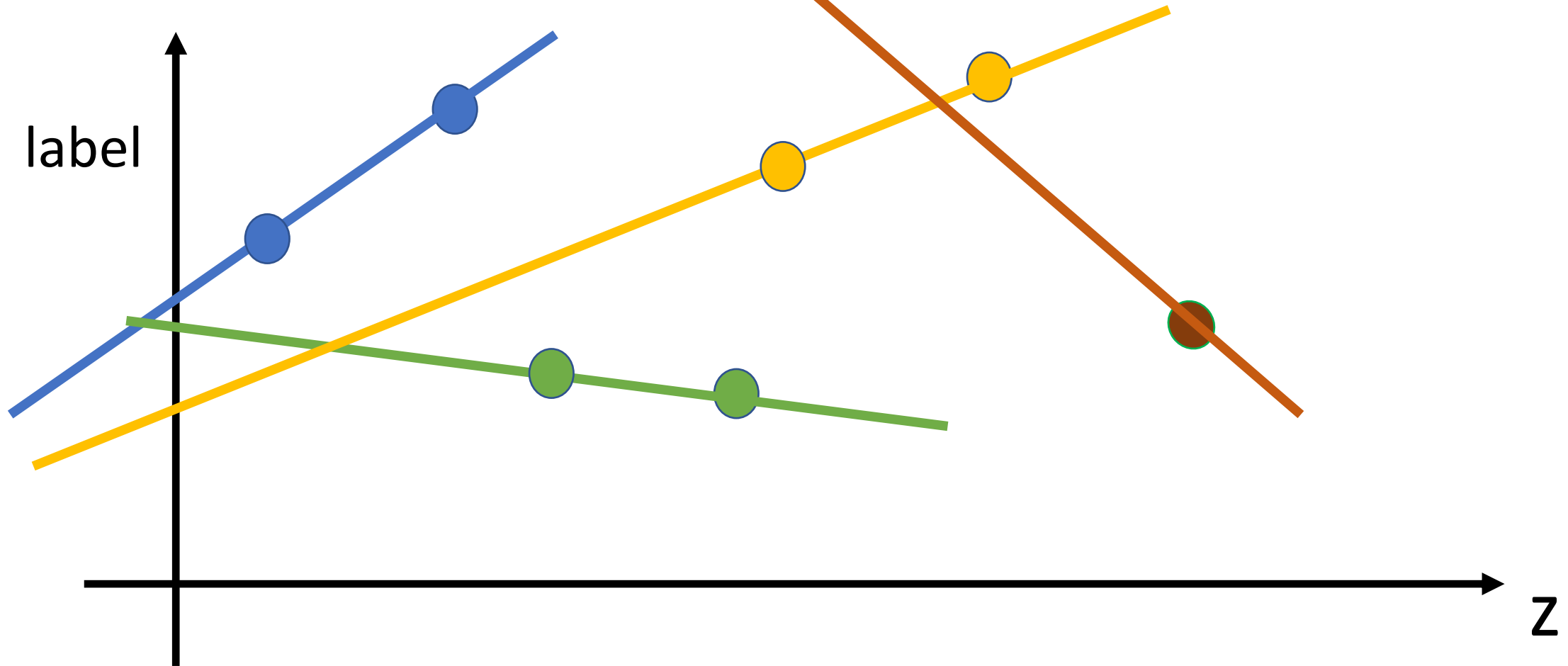
Small Training Error Merely  
Indicates That  
Optimization/Training  
Algorithm Works

# A Key Result

we can perfectly fit (almost) any  $m$  data points using polynomials of degree  $n-1$  as soon as

$$n \geq m$$

$m=2, n=2$



# Take Home Messages

- large models (e.g. deep nets) often overfit
- small training error does not mean much!
- COMPARE TRAINING WITH VALIDATION ERROR!
- use validation error for model selection/hyper.parm tuning

# Thank You !