Model Selection, Overfitting, Regularization

Explaining concepts with a polynomial fitting example

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

January 2021

UNIVERSITY OF CALGARY

Outline

- Learning Goals
- Experimental setup and model selection
- Overfitting and regularization
- Metrics
- Sumamry

Learning Goals

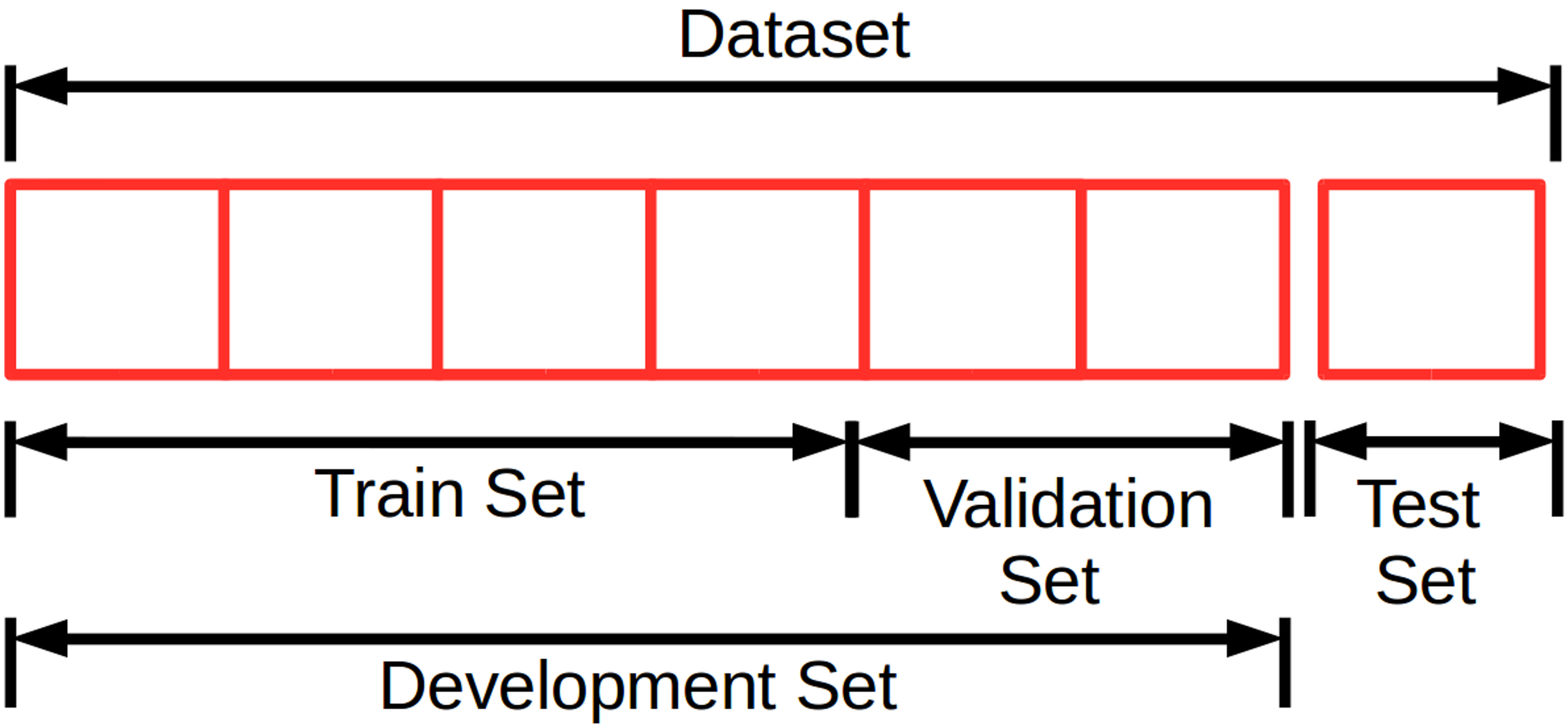
- Explain how to design your experiment
- Introduce how to select your model
- Introduce the concepts of *over-fitting*, *under-fitting*, and *model generalization*.
- Introduce the concept of *regularization* for reducing model *over-fitting*.

Hands-on Tutorial

- https://github.com/rmsouza01/ENEL645
    - **Tutorial 03**: Model selection, overfitting, regularization
- Based on the example presented in chapter 1 of the book: **Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.**

UNIVERSITY OF CALGARY

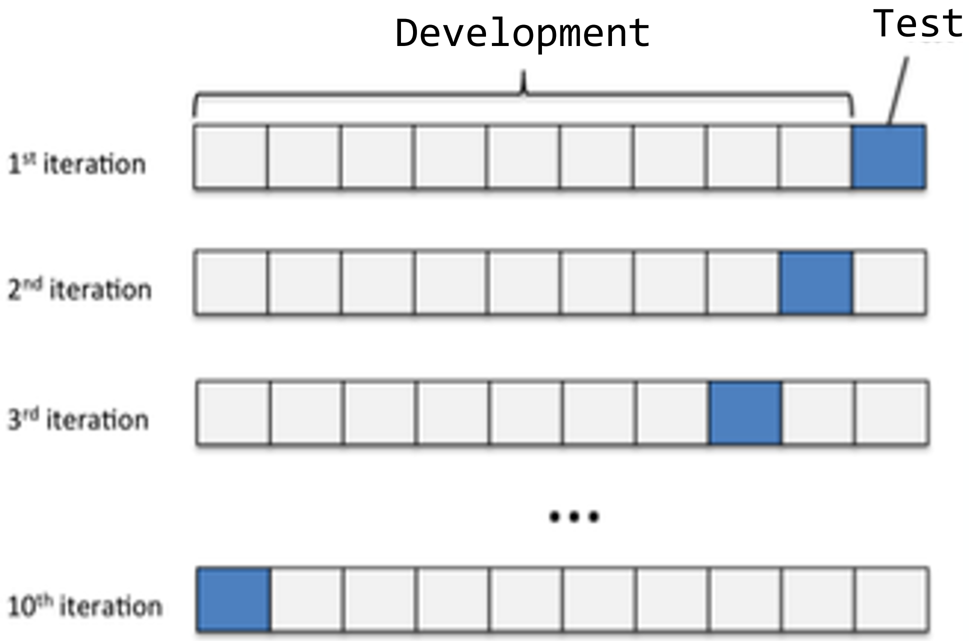
Experiment Design: Train, Validation and Test



- **Train set**: learn parameters of your models
- **Validation set**: model selection
- **Test set**: verify generalizability to unseen data

Experiment Design: k-fold cross validation

- Performs k iterations on the data
- Stratified k-fold: maintain the proportions of each class into folds (unbalance data)

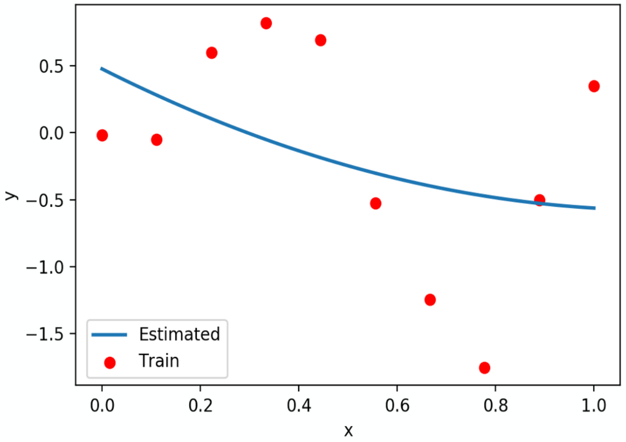


UNIVERSITY OF CALGARY

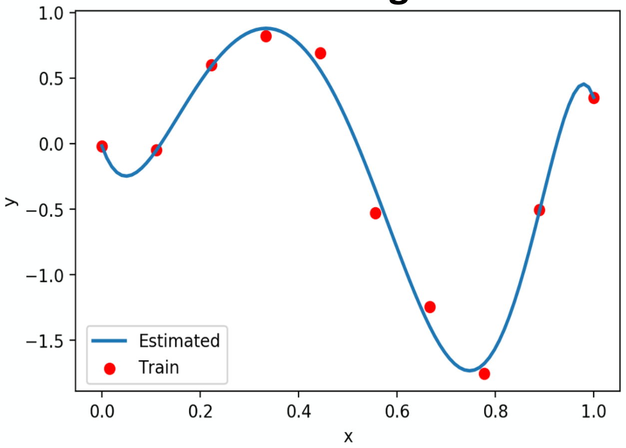
Under- and Over-fitting

- Under-fitting: too inflexible; captures no pattern
  - fitting a linear model to non-linear data
- Over-fitting: too flexible; fits to noise in the data
  - model is excessively complex (#features>>#samples or #parameters too high)
  - decision boundary does not generalize-> poor results for new samples

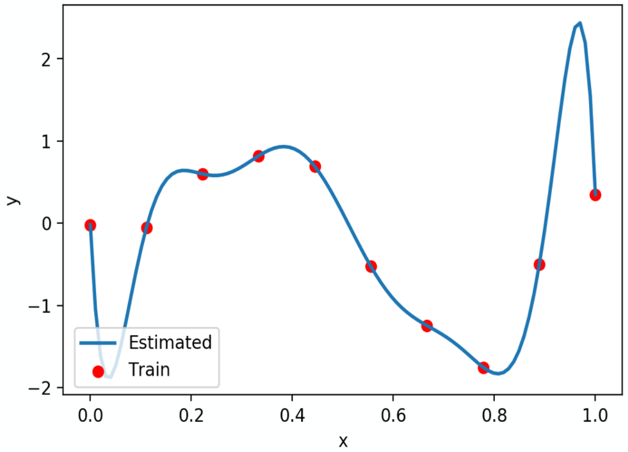
Under-fitting



Good fitting

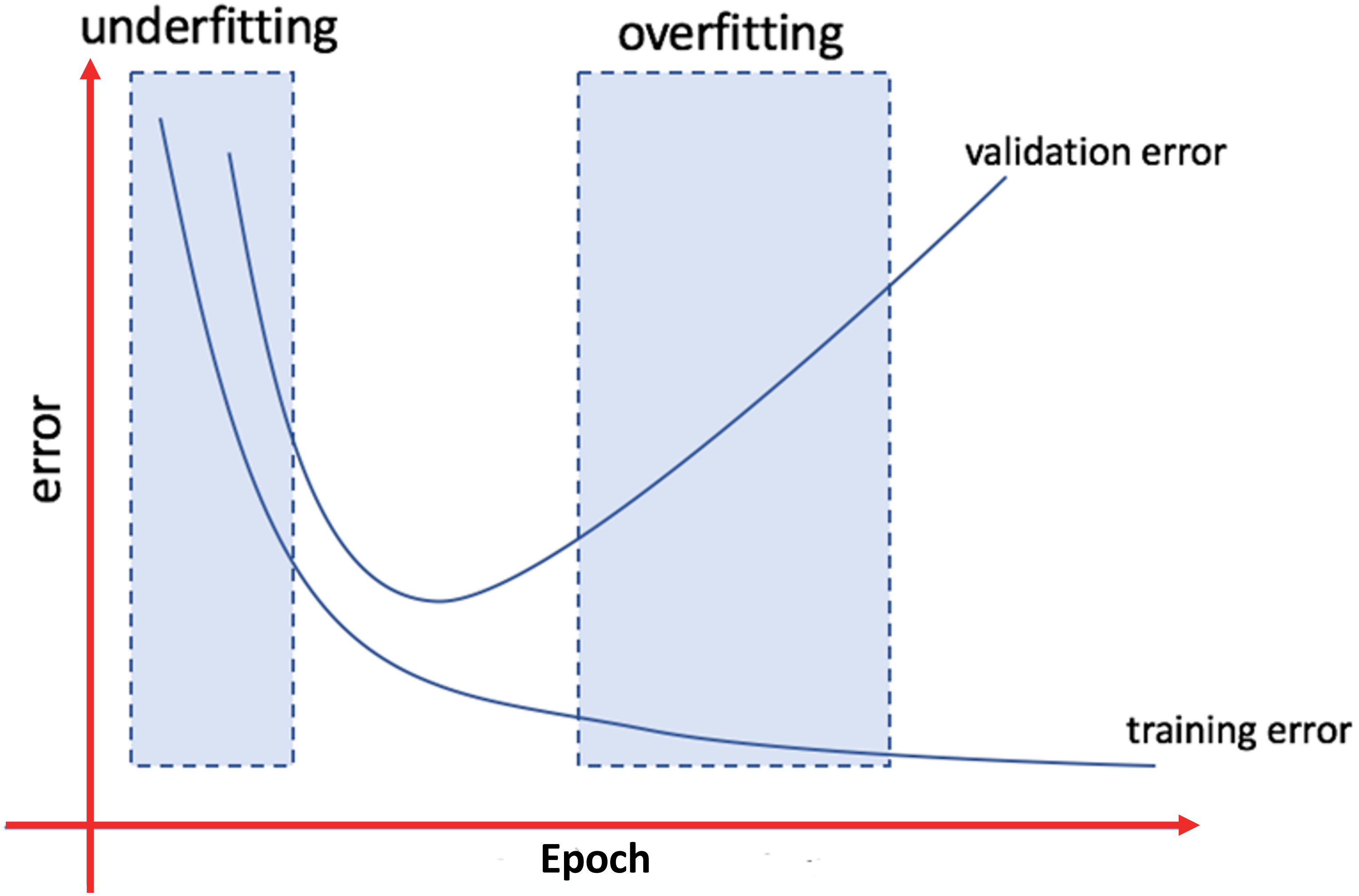


Over-fitting



UNIVERSITY OF CALGARY

Under- and Over-fitting



Techniques to Avoid Over-fitting

- More data
- Reduce model complexity (*i.e.*, number of trainable parameters)
- Regularization
- Dropout
- Data augmentation

Metrics - Classification

- Confusion matrix

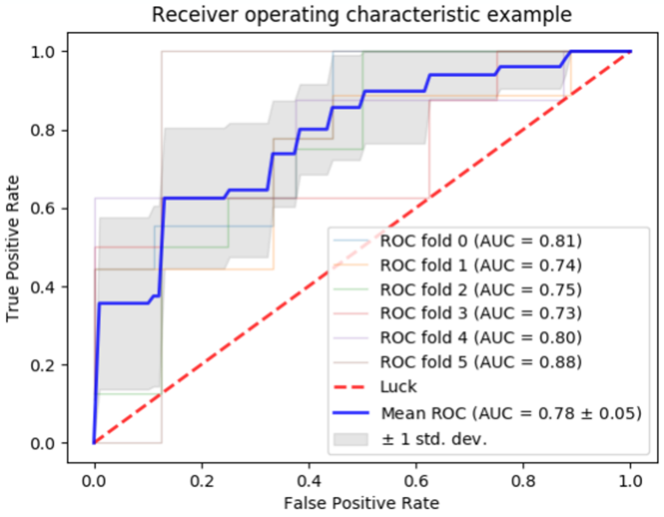
		Prediction outcome		
		positive	negative	
Actual value	positive	TP	FN	$TP + FN$
	negative	FP	TN	$FP + TN$
		$TP + FP$	$FN + TN$	

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

$$Sensitivity = TP / P$$

$$Specificity = TN / N$$

- Receiver operating characteristic (ROC)



UNIVERSITY OF CALGARY

Metrics - Regression

- Structural Similarity (SSIM)

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

- Normalized Root Mean Squared Error (NRMSE)

$$\operatorname{RMSD}(\hat{\theta}) = \sqrt{\operatorname{MSE}(\hat{\theta})} = \sqrt{\operatorname{E}((\hat{\theta} - \theta)^2)}.$$

$$\mathrm{NRMSD} = \frac{\mathrm{RMSD}}{y_{\max} - y_{\min}}$$

- Peak Signal to Noise Ratio (PSNR)

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \end{aligned}$$

Summary

- For large datasets, a single train/val/test split is often sufficient
- The validation set is used for model selection
- Overfitting makes your model less generalizable to new datasets
- Model overfitting can be mitigated by employing techniques, such as regularization

Thank you!