# Computational Communication Science 2
# Week 7 - Lecture
# »Rule-based Text Classification and an Introduction to Supervised Machine Learning«

Marthe Möller
Anne Kroon

a.m.moller@uva.nl, @marthemoller
a.c.kroon@uva.nl, @annekroon

May, 2022

Digital Society Minor, University of Amsterdam

# Today

What is SML?

The principles behind SML

SML step by step

# What is SML?

# What is SML?


Yu, J., Ma, X., & Han, T. (2016). Four-Dimensional Usability Investigation of Image CAPTCHA. *arXiv preprint arXiv:1612.01067*.

# What is SML?

Read more about this project in: Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2014). OverFeat: Integrated recognition, localization and detection using convolutional networks. *arXiv:1312.6229 [cs]*. Retrieved December 23, 2021, from http://arxiv.org/abs/1312.6229

# What is SML?

Machine Learning: A process whereby a machine learns how to predict a variable.

# What is SML?

Supervised Machine Learning (SML): "A form of machine learning, where we aim to predict a variable that, for a least part of our data is known."

"The goal of Supervised Machine Learning: estimate a model based on some data, and then use the model to predict the expected outcome for some new cases, for which we do not know the outcome yet."

Van Atteveldt, W., Trilling, D., & Calderon, C. A. (2022). *Computational analysis of communication*.

Wiley-Blackwell

# What is SML?

Machine Learning has a lot of similarities to regression analysis!

# The principles behind SML

# The principles behind SML

$$y = constant + b_1 * x_1 + b_2 * x_2$$

$x_1$ = bark? (0= no, 1 = yes)

$x_2$ = tail? (0 = no, 1 = yes)

$y$ = Is this a dog? ( 0 = definitely no, 1 = definitely yes)

# The principles behind SML

$$y = constant + b_1 * x_1 + b_2 * x_2$$

$$y = 0 + 0.8 * x_1 + 0.2 * x_2$$

$$y = 0 + 0.8 * 1 + 0.2 * 0$$

# The principles behind SML

$$y = 0 + 08 * 1 + 0.2 * 0$$

$$0.8 = 0 + 0.8 * 1 + 0.2 * 0$$

# The principles behind SML

$$0.8 = 0 + 0.8 * 1 + 0.2 * 0$$

Classification: a predictive modeling problem where a class label is predicted for a given example of input data.

# The principles behind SML

| Machine Learning Lingo | Statistics Lingo |
|---|---|
| Feature | Independent variable |
| Label | Dependent variable |
| Labeled dataset | Dataset with both independent and dependent variables |
| To train a model | To estimate |
| Classifier | Model to predict nominal outcomes |
| To annotate | To (manually) code |

Adapted from: Van Atteveldt, Trilling, & Arcilla (2021)

# The principles behind SML

Machine Learning: using a (regression) formula to predict a label.

Traditional usage of formulas in CS: to explain

Usage of formulas in ML: to predict

# Zooming out

We talked about:

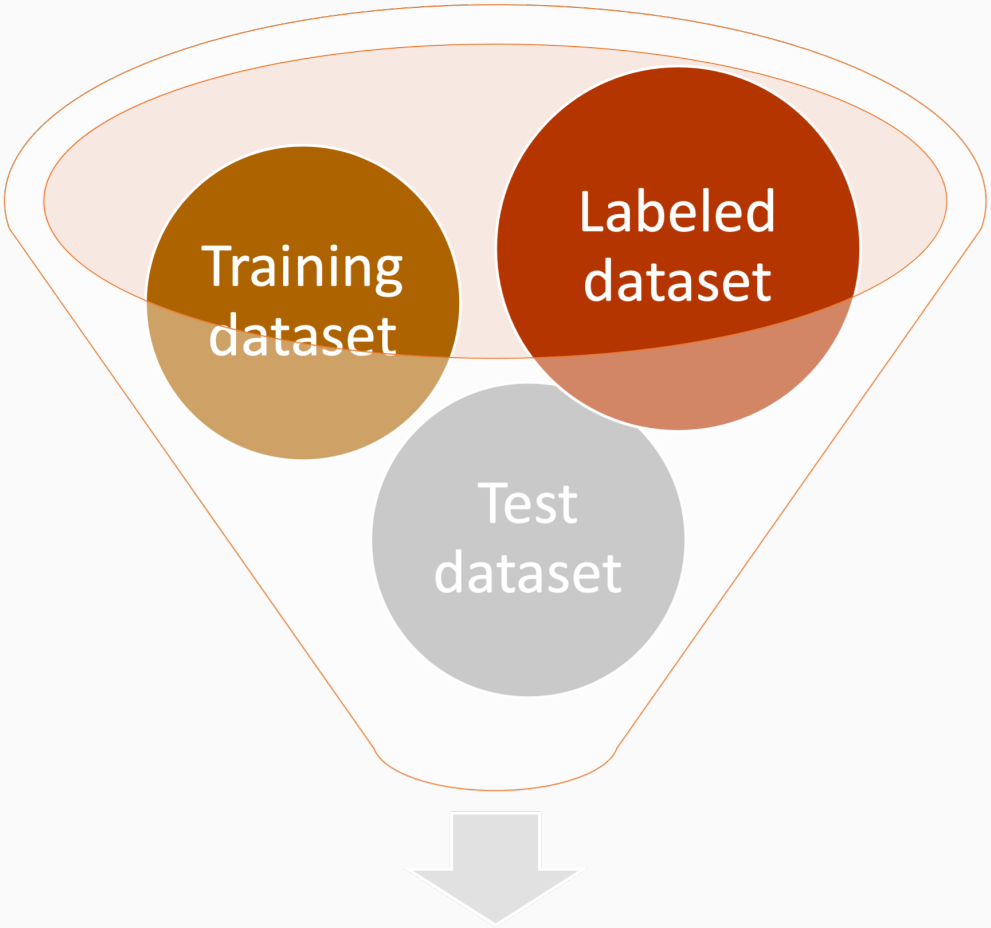
- The principles behind SML

Next, we will talk about:

- The steps of SML

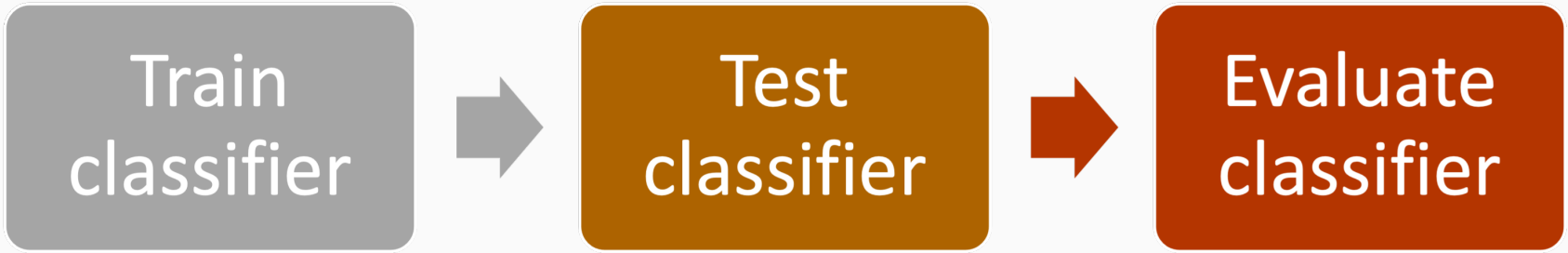
# SML step by step

# SML step by step

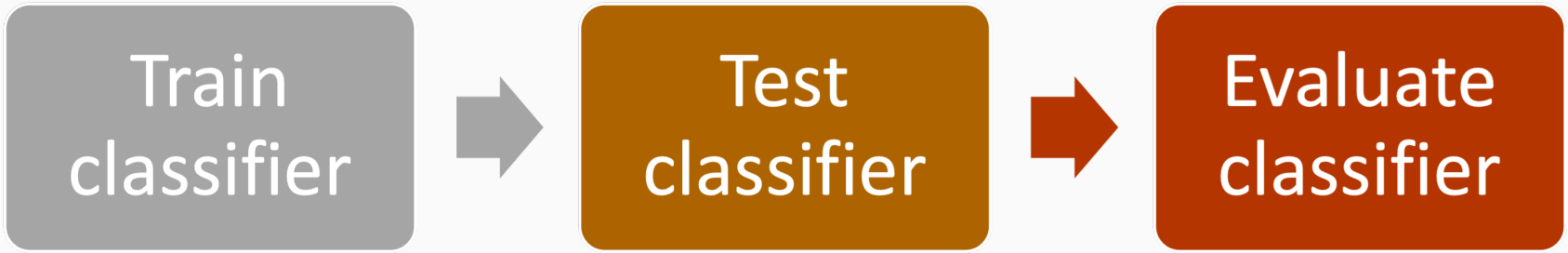



Machine Learning Process

# SML step by step



# SML step by step



# Zooming out

We talked about:

- The principles behind SML
- The steps of SML

Next, we will talk about:

- Some commonly used ML models