



**University of  
Zurich** <sup>UZH</sup>

Institute of Computational Linguistics

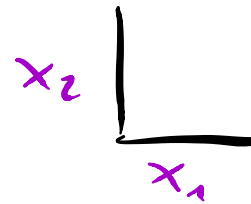
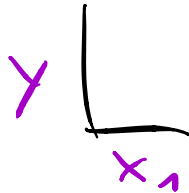
---

# Introduction to Machine Learning

Lesson 4: Cross Validation and Hyperparameter Search

Mathias Müller, Phillip Ströbel

True or false?



A pairplot shows the relationship between all features and the target variable.



Classification algorithms need training data with labels.

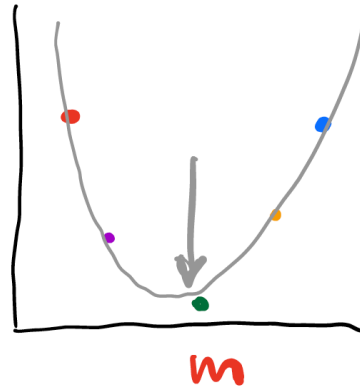
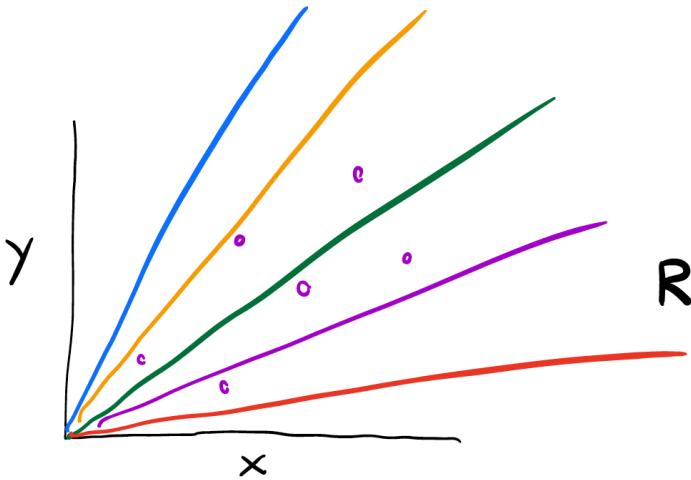
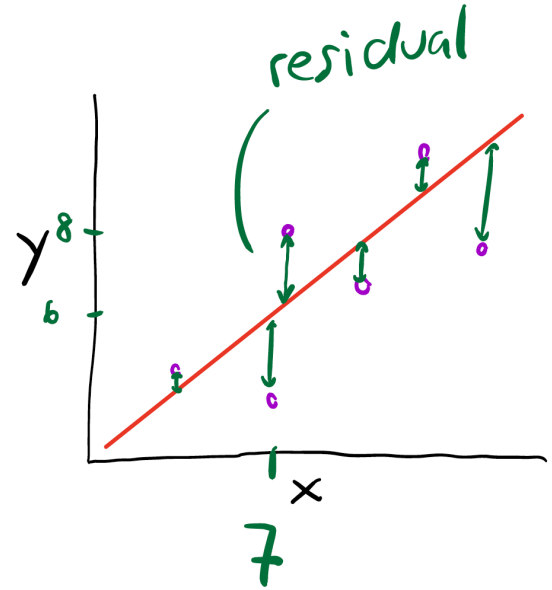
The k in K Nearest Neighbour is the ~~number of classes in the training data.~~

number of neighbors considered.

Wo waren wir

linear  
regression

---



## Topics of this lesson

- Cross Validation
- Hyperparameter search
- Automated Hyperparameter search



**University of  
Zurich** <sup>UZH</sup>

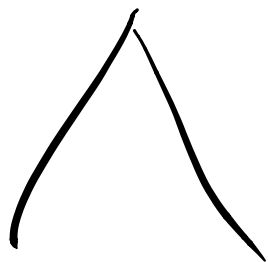
Institute of Computational Linguistics

# Cross-Validation

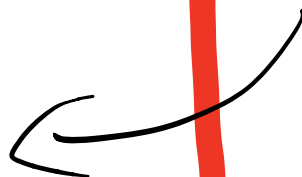
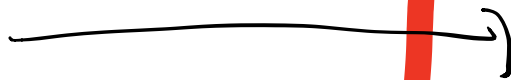


train set

test set



train dev



## Problem

- Even if we split data into train and test sets randomly,
  - all samples in the test set might be easy
  - all samples in the test set might be hard
- And both are unfair!

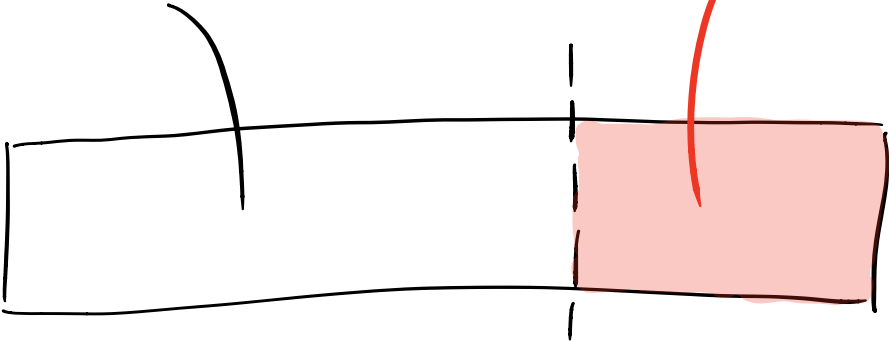




# Solution: Cross Validation

train set

test set



all data

# Solution: Cross Validation

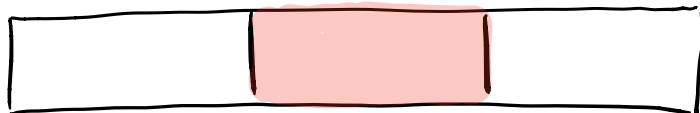
accuracy  
score

①



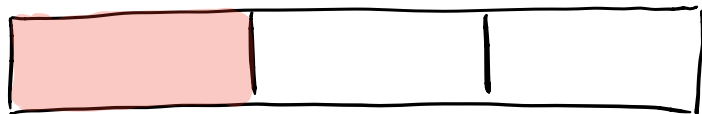
0.8

②



0.6

③

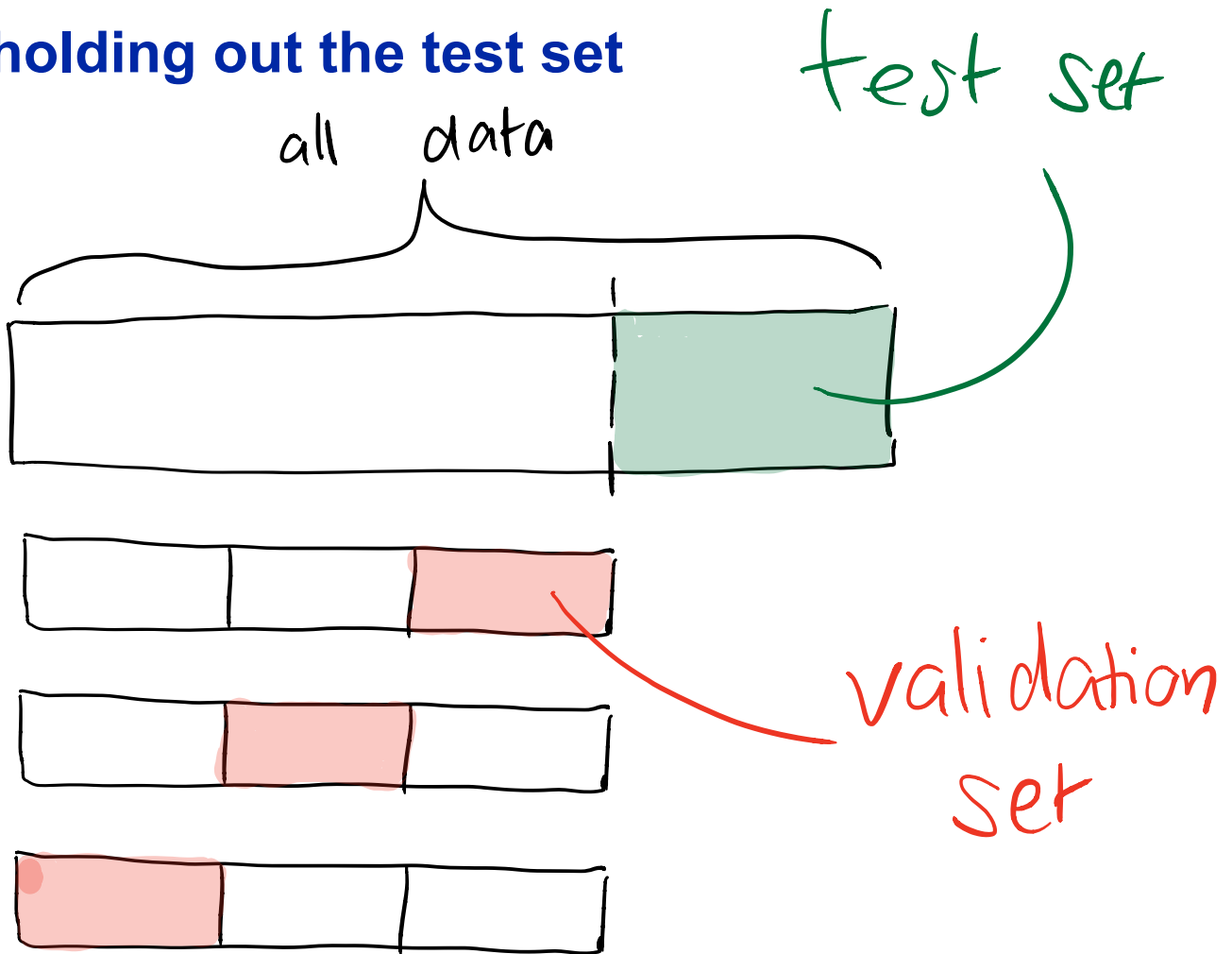


0.7

K-fold CV  
3-fold CV

0.7

# Truly holding out the test set





**University of  
Zurich**<sup>UZH</sup>

Institute of Computational Linguistics

# Hyperparameters

# Hyperparameters

# Parameter

- parameters whose value must be set by the user
- values might have an impact on performance

Estimator

KNN

MLP

Ridge Regression

Example Hyper-P.

$k=3$

$k=10$

learning rate 0.1

regularization strength

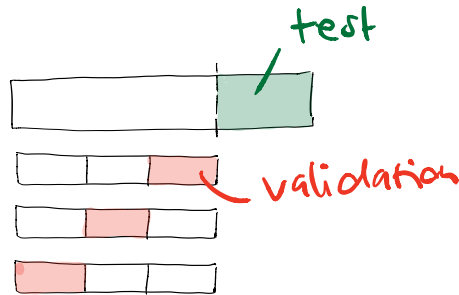
# Hyperparameter Search

- try out different values for hyperparameters
- compute cross validation score for each value

TS

'k' in KNN:

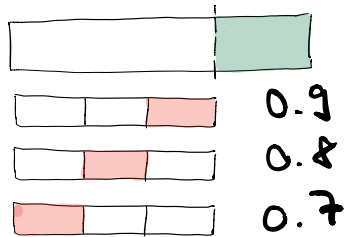
$k=3$



mean score

0.6 | ~~0.7~~

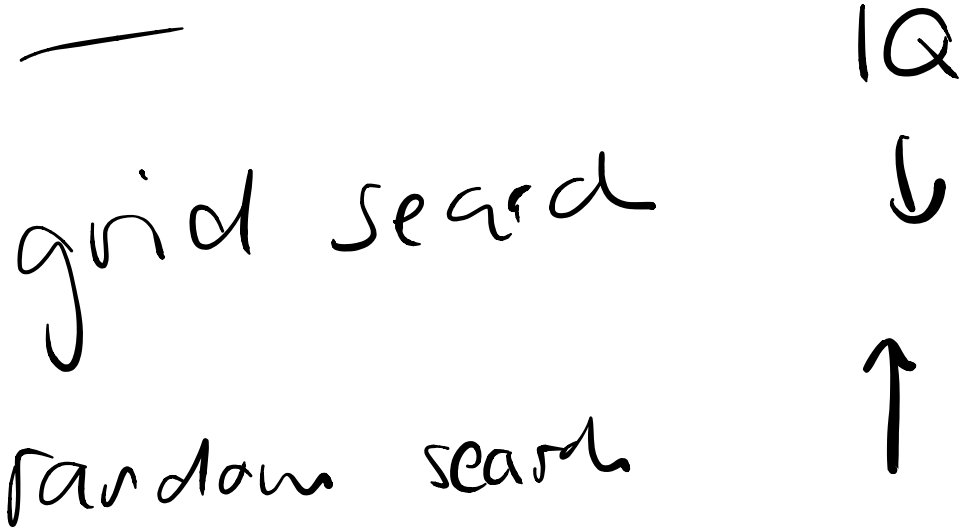
$k=10$



0.8 | 0.7

# Automated Hyperparameter Search

- well-known methods: grid search and random search





**University of  
Zurich** <sup>UZH</sup>

Institute of Computational Linguistics

# Grid Search and Random Search



## Grid Search

3-fold CV

k in KNN:

[1, 2, 3, 4, 5]

16

CV score

k=1	0.1
k=2	0.5
k=3	0.5
k=4	0.5
k=5	0.9

MCP

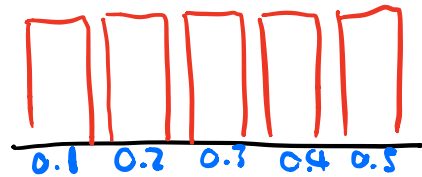
learning-rate (0.1, 0.01, 0.001, 0.0001)  
solver ("adam", "lbfgs", "sgd")  
hidden-layer-size (10, 20, 30, 40, 50)

$$4 \times 3 \times 5 \times 3 + 1 = \underline{181}$$

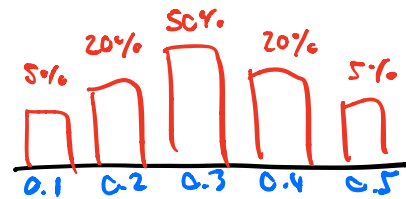
# Parameter Value Distributions

learning-rate = [0.1, 0.2, 0.3, 0.4, 0.5]

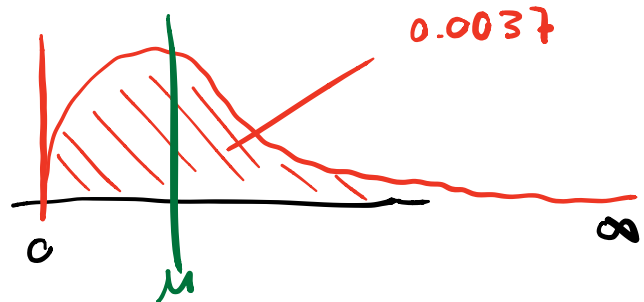
① discrete uniform



② discrete not uniform



③ continuous not uniform



# Random Search

# trials: 10

3-fold CV

CV score

①  $lr = 0.037$ ,  $ls = 10$ , solver = adam

0.6

②  $lr = 0.2$ ,  $ls = 20$ , solver = adam

0.5

⋮

⑩

$10 \times 3 + 1$

31

## Summary

- **Cross validation:** a technique to make sure your problem does not appear easier or harder than it really is
- **Hyperparameters:** some parameters must be set by the user, and they have an impact on performance
- **Hyperparameter search:** good values can be found automatically with grid search or random search