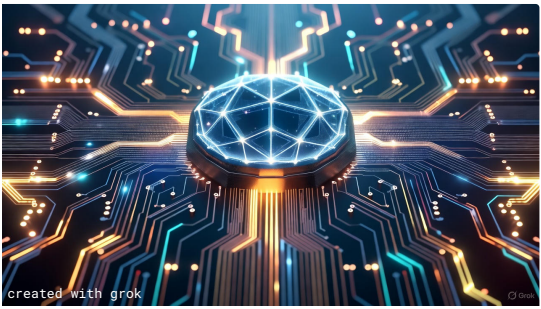
Advanced Machine Learning

Silvan Stadelmann - 28. Oktober 2025 - v0.0.1

github.com/silvasta/summary-aml



Contents

Representation

1 Learning objectives

Estimation of Dependences Based on Empirical Data

What is the learning problem?

$$y = f_\theta(x) + \eta \quad \text{with} \quad \nu \sim \mathsf{P}(\eta|0, \sigma^2)$$

2 Expected risk

- Conditional expected risk

- Total expected risk

3 Empirical risk

- Test and Train Data

Test data cannot be used before the final estimator has been selected!

Training error $\hat{R}(\hat{f}_n, \mathcal{Z}^{\text{train}})$ for Empirical Risk Minimizer (ERM) $\hat{f}_n$

4 Empirical test error and expected risk

Distinguish

5 Comparing algorithm performance on test data

6 Data

6.1 Feature space

- Measurement space $\mathcal{X}$

+ numerical $\mathcal{X} \subset \mathbb{R}^d$

+ boolean $\mathcal{X} = \mathbb{B}$

+ categorial $\mathcal{X} = \{1, ..., k\}$

Features are derived quantities or indirect observations which often significantly compress the information content of measurements.

Remark The selection of a specific feature space predetermines the metric to compare data; this choice is the first significant design decision in a machine learning system.

Taxonomy of Data

6.2 Example of Data

- monadic data

- dyadic data

- pairwise data

- polyadic data

7 Mathematical Spaces

- Topological spaces

- Metric space

- Euclidean vector spaces

- Probability Spaces

Regression

8 Linear Regression

- Statistical model

$$Y = X^{\mathsf{T}}\beta. \quad Y \in \mathbb{R}.\, X.\beta \in \mathbb{R}^{d+1}$$

- Residual Sum of Squares (RSS)

$$RSS(\beta) \;=\; (\mathbf{y} - \mathbf{X}\beta)^{\mathsf{T}}(\mathbf{y} - \mathbf{X}\beta)$$
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{y}$$

9 Gauss Markov Theorem

10 Bias/Variance Dilemma

- Tradeoff, split Error

- Identify error components

11 Bayesian Maximum A Posteriori (MAP) estimates

11.1 Ridge Regression

- Cost function

- Bayesian view

- Solution

Tikhonov regularization

11.2 LASSO

- Cost function

- Bayesian view

- Solution

11.3 Ridge vs. LASSO Estimation

12 Remarks on Shrinkage Methods

- Generalized Ridge Regression

Idea behind shrinkage When white noise is added to the data then all Fourier coefficients are increased by a constant on average. ⇒ Shrink all coefficients by the estimated noise amount to derive a robust predictor.

13 Model averaging is common practice

- Previous: Gaussian process motivated by Bayesian linear regression.

- Seldom: take MAP estimator in Bayesian setting.

- Bayesian approach: average models with different parameters (weighted according to prior).

- Cross validation: Take average over models trained on different folds.

- Winners of most Machine Learning competitions (e.g. on Kaggle): ensembles (weighted averages of models).

-

14 Combining Regressors - Bias

TODO: formula

-

15 Combining Regressors - Variance

TODO: formula

16 Ensemble Learning

The idea of classifier ensembles Boosting is an approach to machine learning based on the idea of creating a highly accurate prediction rule by combining many relatively weak and inaccurate rules.

- Computational advantage

- Statistical advantage

17 Induction Principles for Classifier Selection

I) Empirical Risk Minimization (ERM) Principle

II) Bayesian inference by model averaging

18 Motivation for Ensemble Methods

- Train several sufficiently diverse predictors

- Bagging

- Arcing

- Boosting

19 Weak Learners Used for Bagging or Boosting

Combining Classifiers

Bagging Classifiers

Classifier selection: First compare, then bag!

Bagging: The Mechanism

Decision Trees

Random Forests

The Idea of Boosting

AdaBoost

Data Reweighting

Boosted Classifier

Comparison of ensemble methods

20 Loss functions for classification

Support Vector Machines

Neural Networks

Transformer

Exercises

21 Problem 1 - Regression

- Linear Regression

- Ridge Regression

- Noisy Regression

E1.2.c - An Engineer's rule of thumb is to choose K as $\min \sqrt{n}, 10$

- Overfitting

- Cross Validation

- Generative vs. Discriminative Modeling