Machine Learning I

About this course

- *Introduction* to machine learning.
- 10 lectures covering a broad range of topics.
- It's a mathematical course (supplemented by practical exercises).
- The necessary math will be introduced as we go.

  For more, see Deisenroth et al., *Mathematics for Machine Learning* (free online).
- The focus is on introducing key concepts and developing intuitions.
- The slides and exercises are based on materials by Dmitry Kobak.
- Videos with similar content at `https://www.youtube.com/playlist?list=PL05umP7R6ij35ShKLDqccJSDntugY4FQT`

Formalities

- Lecture is co-taught by Sacha Sokoloski and Sebastian Damrich
- Lecture slides on Ilias before lecture
- 10 exercise sheets
  - Submission in groups of 2-3 (!) by 6pm on Mondays via ILIAS
  - Tutorials on Thursdays at 12:15 pm, room E.3.16 (Microbiology)
  - Pass 7 / 10 to qualify for exam
- 3 coding projects
  - Submission in groups of 2-3
  - Pass 2 / 3 to qualify for exam
- Examination
  - electronic midterm on 25.11. in Microbiology lecture hall
  - written final exam on 20.02., location TBD
  - final grade based on midterm (1/3) and exam (2/3)
- In case of issues, reach out (via ILIAS forum, email) before deadline!

Tentative course plan

Topic	Lecturer	Handout Date	Lecture Date	Tutorial Date	Tutorial Notes	Exercise (Due)	Practical (Due)
Intro	Sebastian	13-Oct	16-Oct	23-Oct		Ex. 1 (20-Oct)	—
Linear Regression	Sebastian	20-Oct	21-Oct	30-Oct		Ex. 2 (27-Oct)	—
Bias and Variance	Sebastian	27-Oct	28-Oct	06-Nov		Ex. 3 (03-Nov)	—
Regularization	Sebastian	03-Nov	04-Nov	13-Nov		Ex. 4 (10-Nov)	Pr. 1 (17-Nov)
Coding and LLMs	Sacha	—	11-Nov	20-Nov	Pr. 1 Review	—	—
Log Regression	Sacha	17-Nov	18-Nov	27-Nov		Ex. 5 (24-Nov)	—
Mid-term	Sebastian / Sacha	—	25-Nov	04-Dec		—	—
Neural Networks	Sacha	01-Dec	02-Dec	11-Dec		Ex. 6 (08-Dec)	—
Bagging and Boosting	Sacha	08-Dec	09-Dec	18-Dec		Ex. 7 (15-Dec)	Pr. 2 (07-Jan)
Clustering	Sacha	15-Dec	16-Dec	08-Jan		Ex. 8 (22-Dec)	—
—	—	—	—	15-Jan	Pr. 2 Review	—	—
PCA	Sebastian	12-Jan	13-Jan	22-Jan		Ex. 9 (19-Jan)	—
T-SNE	Sebastian	19-Jan	20-Jan	29-Jan		Ex. 10 (26-Jan)	Pr. 3 (02-Feb)
Advanced Methods	Sacha / Sebastian	26-Jan	27-Jan	05-Feb		Ex. 11 (02-Feb)	—
Backup / QA	Sacha / Sebastian	—	03-Feb	—		—	—
Exam	Sacha / Sebastian	—	20-Feb	—		—	—

Foundations · Supervised Learning · Unsupervised Learning · Additional Topics

- Use exercises and practicals as resources for learning
- Responsible use of LLMs is accepted
- Exams are similar to exercises, but with coding-related questions

What is machine learning?

ChatGPT

https://de.m.wikipedia.org/wiki/Datei:ChatGPT_logo.svg

DALLE2

https://openai.com/dall-e-2

Autonomous driving

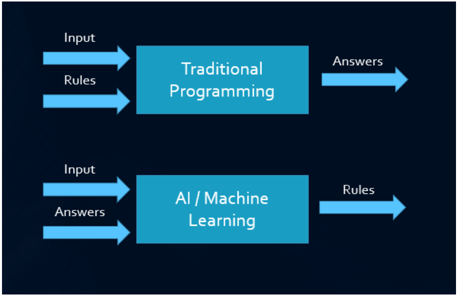
https://www.tesla.com/autopilot

AlphaStar

https://www.deepmind.com/blog/alphastar-mastering-the-real-time-strategy-game-starcraft-ii

What is machine learning?

"Machine learning (ML) is the study of computer algorithms that improve automatically through experience." (Wikipedia)



From the internet. Original source unknown.

What is machine learning?

“The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest.” (Murphy, *Machine Learning*)

“Machine learning is thus closely related to the fields of statistics and data mining, but differs slightly in terms of emphasis […]”

Machine learning vs. statistics

Both statistics and machine learning aim to “detect patterns in data” by building predictive models.

Statistics: use this as a tool to learn something about the world (*statistical inference*). Focus on simple, interpretable models. Develop theoretical analysis, work out statistical guarantees under some assumptions.

Machine learning: use this as a tool to actually make useful predictions. Focus on complicated, competitive models. Use large datasets. Be pragmatic. Relaxed attitude to statistical guarantees.

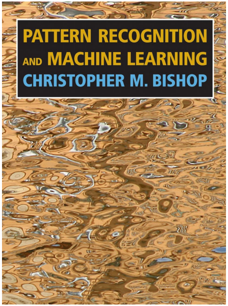
Breiman (2001) Statistical Modeling: The Two Cultures.

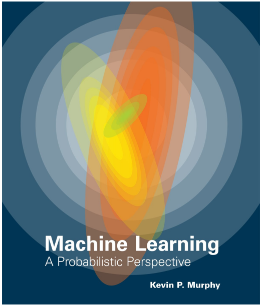
Machine learning vs. statistics: spectrum

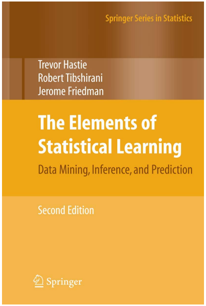
In practice, there is no boundary — it is a spectrum. It goes all the way from a one-sample t-test to ChatGPT.

https://xkcd.com/1838/

“Statistical learning”





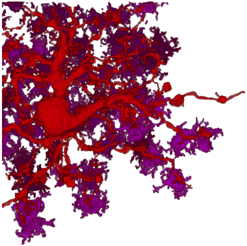


Types of machine learning problems

1. Supervised learning
   Example: distinguish cats from dogs using labelled pictures.
2. Unsupervised learning
   Example: Describe the distribution of cat and dog pictures without labels (maybe try and separate them?).
3. Reinforcement learning
   Example: play StarCraft II (successfully).

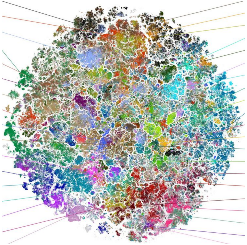
Yann LeCun: “Most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake.”

Machine learning in AG Berens



Neuronal Modeling

Safe Medical Diagnostics

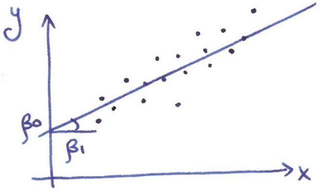


Representation learning & visualisation

We are looking for rotation project, master thesis, and PhD students!

Simple linear regression

Simple linear regression



Supervised learning problem. Regression (not classification) problem.

Training data: $\{(x_i, y_i)\}_{i=1}^{n}$.

Model: $\hat{y} = f(x) = \beta_0 + \beta_1 x$.

Two coefficients: *intercept* and *slope*. We want to *fit* the model to the data.

Loss function

To fit the model means to find β_0 and β_1 so that $f(x_i) \approx y_i$.

Loss function (aka cost function):

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}\big(y_i - f(x_i)\big)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.$$

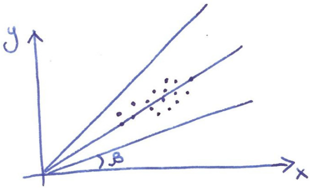
Mean squared error (MSE). Why MSE?

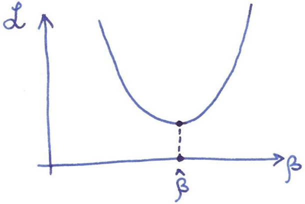
Ordinary least squares (OLS).

Baby linear regression

Consider slope-only model: $f(x) = \beta x$.

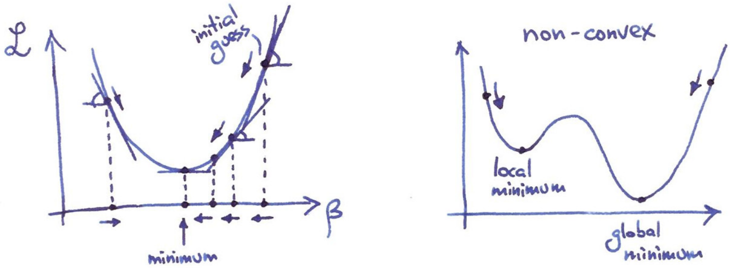
The loss: $\mathcal{L}(\beta) = \frac{1}{n}\sum_i (y_i - \beta x_i)^2$.





EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Baby gradient descent



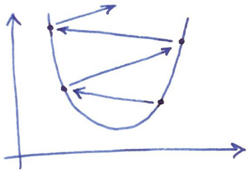
Update rule:

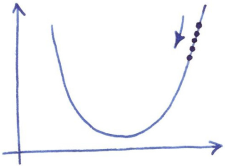
$$\beta \leftarrow \beta - \eta \frac{d\mathcal{L}(\beta)}{d\beta}.$$

Here η is the learning rate.

Learning rate

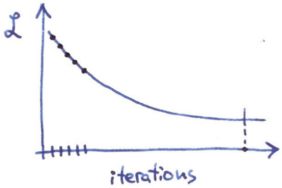
$$\beta \leftarrow \beta - \eta \frac{d\mathcal{L}(\beta)}{d\beta}$$





Too large η — divergence. Too small η — slow convergence.

Stopping criterion



$$\beta \leftarrow \beta - \eta \frac{d\mathcal{L}(\beta)}{d\beta}$$

Baby gradient descent cont.

We need to compute the derivative of the loss:

$$\mathcal{L}(\beta) = \frac{1}{n}\sum_i (y_i - \beta x_i)^2.$$

We get:

$$\mathcal{L}'(\beta) = \frac{1}{n}\sum_i 2(y_i - \beta x_i)(-x_i) = -\frac{2}{n}\sum_i x_i(y_i - \beta x_i).$$

Baby analytical solution

$$\mathcal{L}'(\beta) = -\frac{2}{n}\sum_i x_i(y_i - \beta x_i).$$

At the minimum:

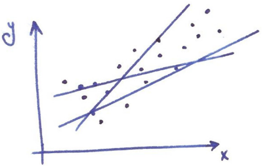
$$\sum_i x_i y_i - \hat{\beta}\sum_i x_i^2 = 0.$$

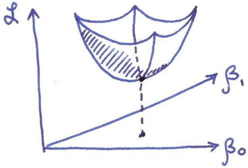
We obtain:

$$\hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}.$$

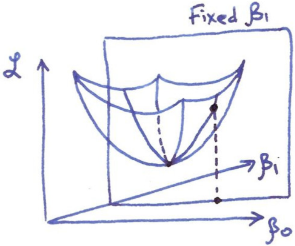
Back to simple linear regression

$$\mathcal{L}(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$





Introducing *partial derivatives*



$$\beta_0 \leftarrow \beta_0 - \eta \frac{\partial \mathcal{L}}{\partial \beta_0}$$

$$\beta_1 \leftarrow \beta_1 - \eta \frac{\partial \mathcal{L}}{\partial \beta_1}$$

Introducing *gradient*

Update rules for each parameter:

$$\beta_0 \leftarrow \beta_0 - \eta \frac{\partial \mathcal{L}}{\partial \beta_0}$$
$$\beta_1 \leftarrow \beta_1 - \eta \frac{\partial \mathcal{L}}{\partial \beta_1}$$

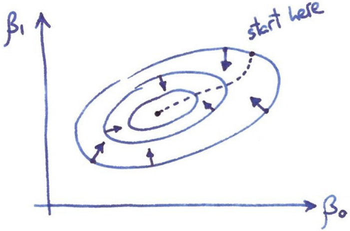
In vector form:

$$\vec{\beta} \leftarrow \vec{\beta} - \eta \nabla \mathcal{L}.$$

Gradient descent

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$\vec{\beta} \leftarrow \vec{\beta} - \eta \nabla \mathcal{L}$$



Computing the gradient

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

We need partial derivatives:

$$\frac{\partial \mathcal{L}}{\partial \beta_0} = -\frac{2}{n}\sum(y_i - \beta_0 - \beta_1 x_i)$$
$$\frac{\partial \mathcal{L}}{\partial \beta_1} = -\frac{2}{n}\sum(y_i - \beta_0 - \beta_1 x_i)x_i$$

Exercise: derive the analytical solution for $\hat{\beta}_0$ and $\hat{\beta}_1$.