

Introduction to Machine Learning

Bernd Bischl, Fabian Scheipl, Daniel Schalk

https://github.com/compstat-lmu/lecture_i2ml

The course is organized as a digital lecture for ML beginners for statistics Bachelors and Data Science ESG student, which should be as self-contained and enable self-study as much as possible:

- Slides with lecture videos
- Interactive tutorials: Currently online quizzes and some very simple online coding exercises
- Complemented by a week-long inverted-classroom block course (code demos, slide discussion, supervised exercises)
- **Produce lecture like a piece of software / library: Literate programming, Github, collaborative, modular chunks, issues, pull requests, etc.**

- Risk minimization
- Linear models, KNN, Naive Bayes
- Performance evaluation, cross-validation, ROC
- Decision trees, random forests
- Tuning

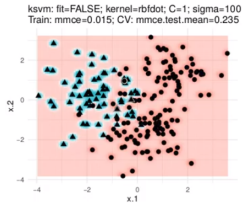
Prerequisites

The course is targeted at ML beginners with a basic, university level, education in maths

- Simple linear algebra
- Simple calculus
- Simple probability theory
- Some stats knowledge, you should now what mean, variance, bias, etc., is (Linear) Modelling from a stats perspective can help, but is not required, we sometimes compare to that
- Working knowledge of R

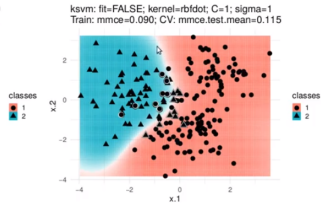
OVERFITTING

Overfitting learner

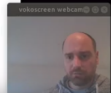


Better training set performance
(seen examples)

Non-overfitting learner



Better test set performance
(unseen examples)



Concept - Interactive Tutorials (Quiz)

Introduction to Machine Learning - Day 2

Introduction to Classification

Logistic Regression

Approaches to Classification

The Two Cultures of Statistical

Start Over

✓ Exercises

(7) Quiz

✓ Which statements are true?

- ☒ Classification is a supervised learning task. ✓
- ☐ The decision boundary is independent of the used model.
- ☒ Binary classification uses two discriminant functions.
- ☐ Linear classifiers can just learn linear decision boundaries.
- ☐ For the discriminant approach we must have a loss function for minimization. ✓
- ☒ The generative and discriminant approach are basically the same.
- ☐ The generative approach is a probabilistic approach. ✓

Incorrect.

✓ Which statements are false?

- ☒ The discriminant approach assumes a data generating process in which the features have different distributions conditional on the class of the target variable. ✓
- ☒ The generative approach attempts to minimize a loss function. ✓
- ☒ Linear discriminant analysis is a discriminant approach, while quadratic discriminant analysis is not. ✓
- ☐ The discriminant approach tries to model the discriminant score function directly.
- ☐ Logistic regression is a discriminant approach

Correct!

Next Topic

Concept - Interactive Tutorials (Examples)

Introduction to Machine Learning - Day 2

Introduction to Classification

Logistic Regression

Approaches to Classification

The Two Cultures of Statistical

Start Over

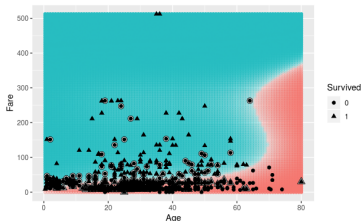
(P) Training a logistic regression with non-linear decision boundaries

The next demonstration shows how to include the features **Age** and **Fare** as polynomials and the effect on the decision boundary. As mentioned in the video, it is possible to transform a linear classifier into a non-linear classifier by just mapping features into a higher dimensional feature space (feature map):

Code [Start Over](#)

[Run Code](#)

```
1 library(ggplot2)
2
3 # Change degree and threshold here:
4 degree = 3
5 threshold = 0.5
6
```



[Previous Topic](#)

[Next Topic](#)

Outlook

- Cut lectures / videos down to smaller chunks
- Collaborative development with other lecturers / sites
- More code demos
- Better auto-correction of programming exercises
- Advertise material to other LMU departments / programs