LEUPHANA
UNIVERSITÄT LÜNEBURG

INTRODUCTION TO MACHINE LEARNING

01. Introduction
Winter 2020/2021


Who am I?

Burkhardt Funk

— Professor of Information Systems, Institute of Information Systems, Leuphana University of Lüneburg
— Program Coordinator Major Business Information Systems (College)
— Member of the Research Center for Digital Transformation
— Several administrative functions (e.g. PhD & examination board, VP until 2016)

Short CV

— Professor for IS, since 2003, visiting scholar at UVa and Stanford
— 20 years of entrepreneurial experience (founded 10+ companies), consulting, and supervisory board → happy to discuss your entrepreneurial activities/ ideas
— PhD in computational physics (Wuppertal), studied physics & computer science in Kiel, Würzburg, Stony Brook/NY

Research focus

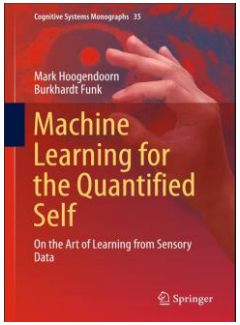
Decision support systems, understanding user behavior, and machine learning applications

Research contexts

- E-Commerce
- E-Health
- Digital Transformation

Methods

- Machine learning
- Bayesian statistics



Books

GET.ON Gesundheits Training.Online

Digi-Exist

E-COMPARED

HBI | CARE

SEM-A²

DATAx

Projects

- What was your primary study program?
- What are you currently doing?
- How much of a data scientist are you already?
- What brought you here?
- A fun fact/ hobby?

Introduce yourself

Today's objectives

- Overview and introduction of the course
- Administrative and organizational topics
- Get an intuition of what machine learning means
- Introduction to a dataset that will be used throughout the course

What is this course about?

Introduction to Machine Learning

=

Conceptual Foundation

+

Techniques

Agenda

- **Introduction**
- Learning problem & linear classification
- Linear models: regression & logistic regression
- Non-linear transformation, overfitting & regularization
- Support Vector Machines and kernel learning
- Neural Networks: shallow [and deep]
- Theoretical foundation of supervised learning
- Unsupervised learning

Course work and exam

During the course

- Discuss mathematical / theoretical concepts (use the chance to ask and jointly reflect the concepts)
- Implement tasks in your favorite programming language (Python, R, MATLAB)

Before and after the course

- Read provided material, text books, video lectures
- Solve problem sets and be able to present solutions in class

Exam

- Write scientific paper (setting: project or paper)
- Present topic in class (Feb. 5th 2021)
- 4 problem sets

Schedule

	Date	Time	Topic	Room
1	16.10.2020	14:00-18:30	Learning problem & linear classification	40.606
2	30.10.2020	14:00-16:30	Linear models: regression & logistic regression	zoom
3	06.11.2020	14:00-16:30	Non-linear transformation, overfitting & regularization	zoom
4	20.11.2020	14:00-16:30	Support Vector Machines and kernel learning	zoom
5	27.11.2020	14:00-16:30	Neural Networks: shallow [and deep]	zoom
6	11.12.2020	14:00-16:30	Theoretical foundation of supervised learning	zoom
7	15.01.2021	14:00-16:30	Unsupervised learning	zoom
8	05.02.2021	14:00-18:30	Presentations	to be done

Resources – selected textbooks

- **Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H. T. (2012). *Learning from data* (Vol. 4). New York, NY, USA:: AMLBook.**
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics.
- S. Shalev-Shwartz & S. Ben-David (2014). Understanding Machine Learning. Cambridge
- Tom Mitchell (1997) Machine Learning. McGraw-Hill
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Kevin P. Murphy (2012) Machine Learning – A Probabilistic Perspective. MIT Press
- ***Additional material will be provided via mystudy***

Other freely accessible online resources

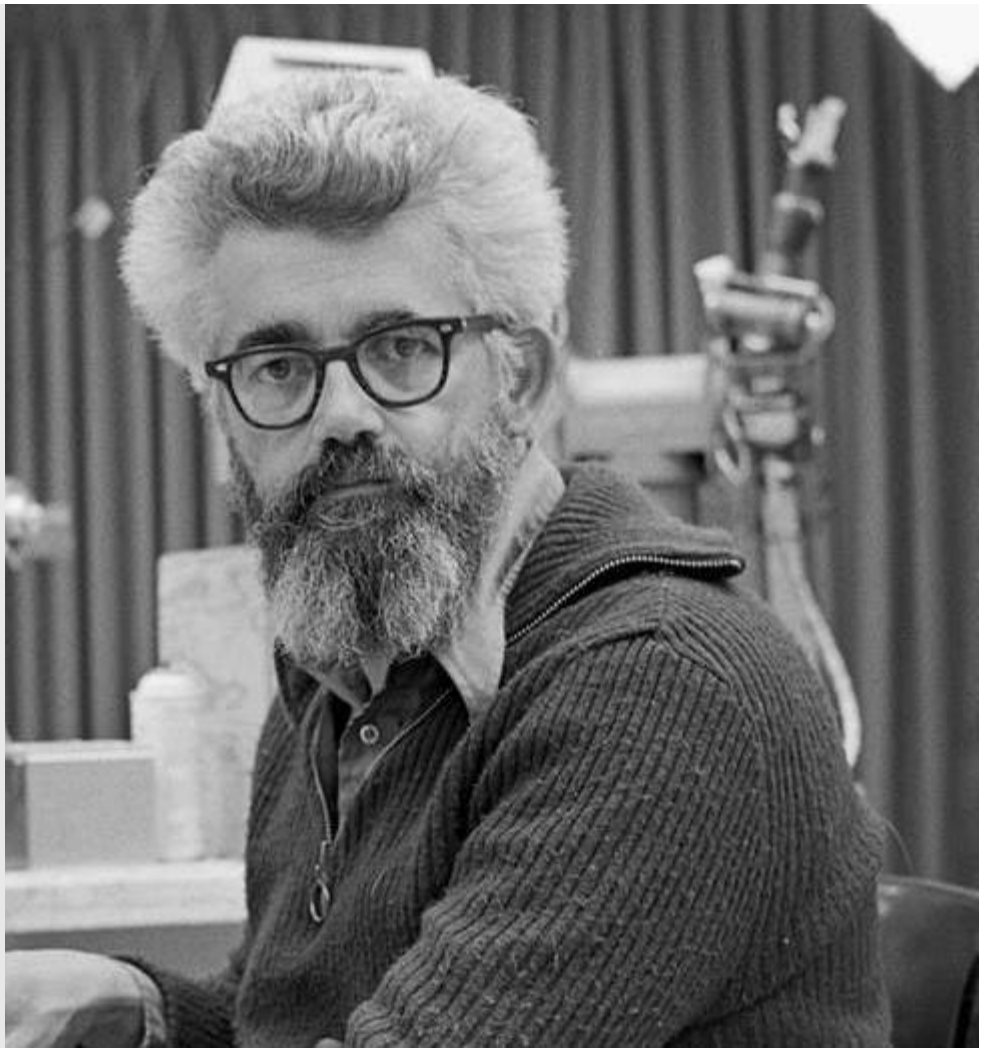
- Yaser Abu-Mostafa – Learning from Data. (https://www.youtube.com/watch?v=mbyG85GZ0PI&list=PLD63A284B7615313A)
- Andrew Ng – Machine Learning. (https://www.youtube.com/watch?v=PPLop4L2eGk&list=PLLssT5z_DsK-h9vYZkQkYNWcItqhlRJLN)
- Nando de Freitas – Machine Learning. (https://www.youtube.com/watch?v=w2OtwL5T1ow&list=PLE6Wd9FR--EdyJ5lbFl8UuGjecvVw66F6&index=1)
- Victor Lavrenko – online lectures (https://www.youtube.com/user/victorlavrenko)
- Towards Data Science blog on Medium (https://towardsdatascience.com/machine-learning/home)

Artificial Intelligence

Artificial Intelligence (AI) tries to make computers and things smart

AI = “human intelligence exhibited by machines”

Examples: rule-based and expert systems, knowledge representation & reasoning



John McCarthy

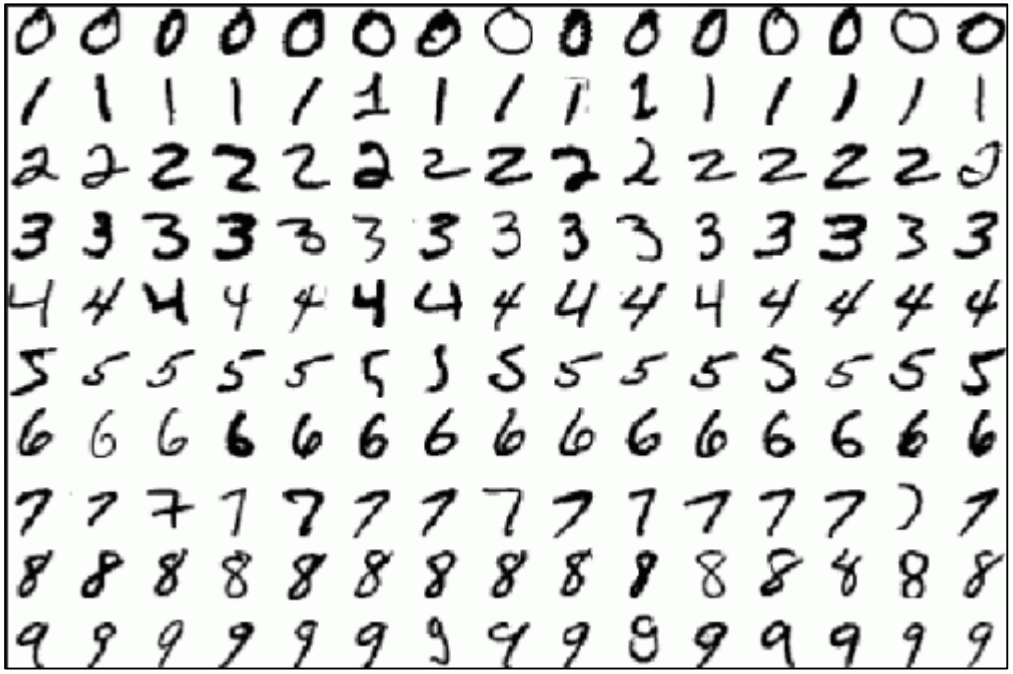
Machine learning is a subset of AI

"You can think of machine learning and artificial intelligence as a set of Russian dolls nested within each other. Machine learning is a subset of AI, which is an umbrella term for any computer program that does something smart."

Source: skymind.ai

Some interesting Machine Learning examples

- **MNIST dataset**; contains 60,000 handwritten digits with labels (28x28 pixel images)
- Algorithms try to recognize the digits by looking at the pixel details
- Many implementations for recognition exist; the best algorithms reach an error rate of below 1%



Source: Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, november 1998

Time for a quiz – MNIST dataset

1. What makes it hard for a computer to learn the task of recognizing handwritten digits?

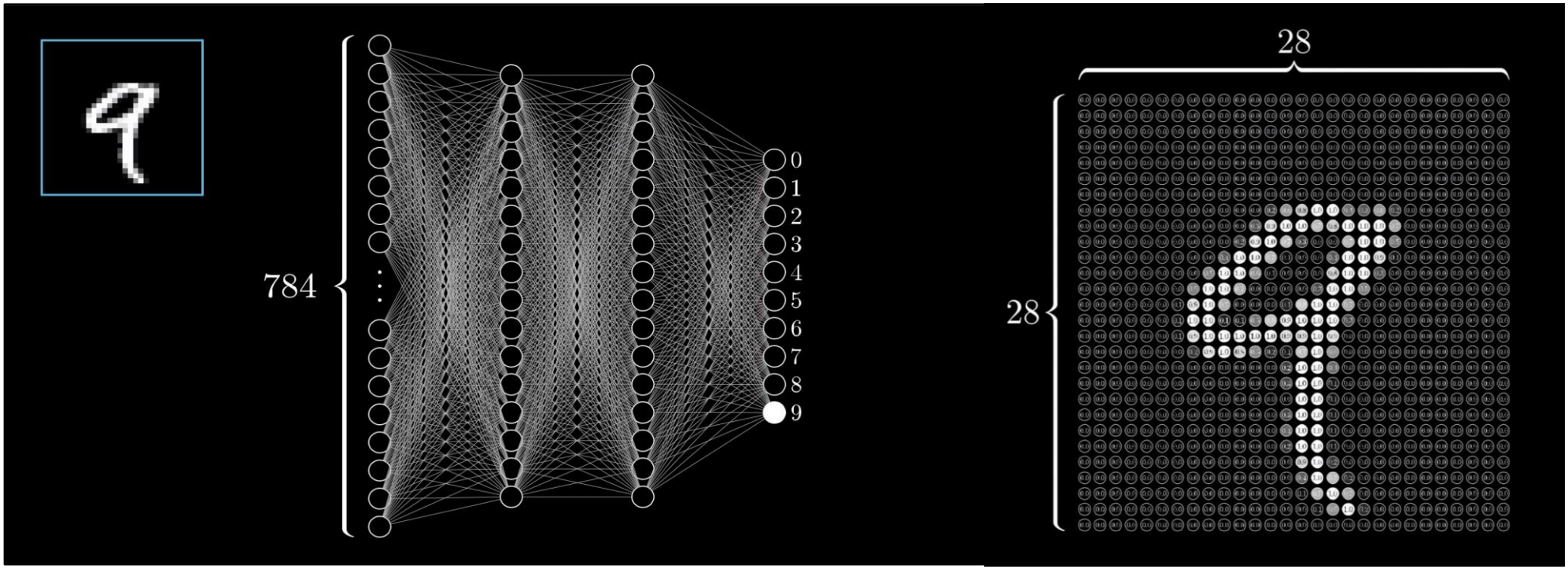
2. What **features** (properties of the images) would you recommend for learning to label handwritten digits?

MNIST dataset details

— 28 x 28 = 784 pixel grid for each digit example

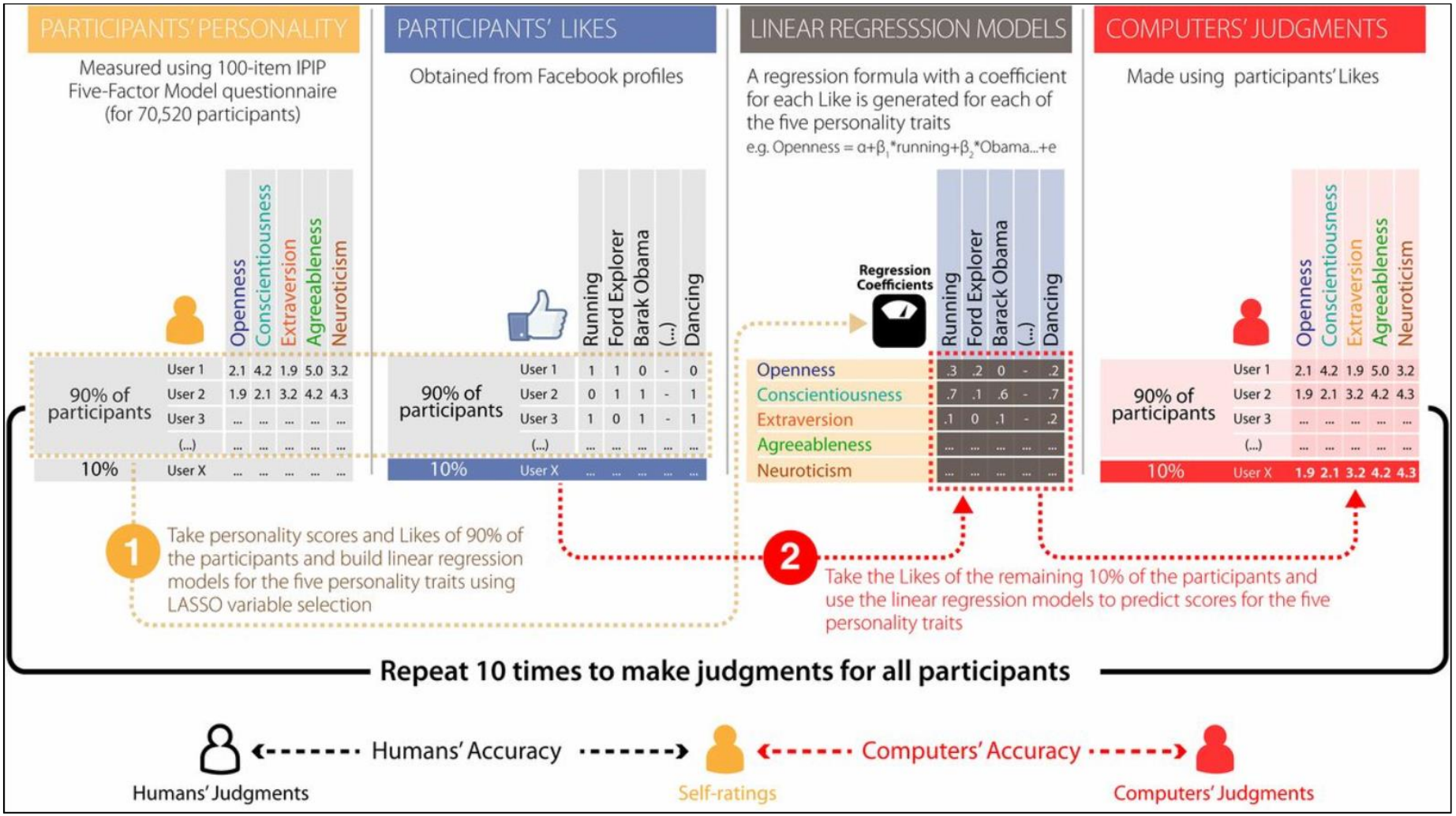
— Each pixel can take a value between 0 (black pixel) and 1 (white pixel)

— Here, a Neural Network tries to predict the unseen example of the digit "9"



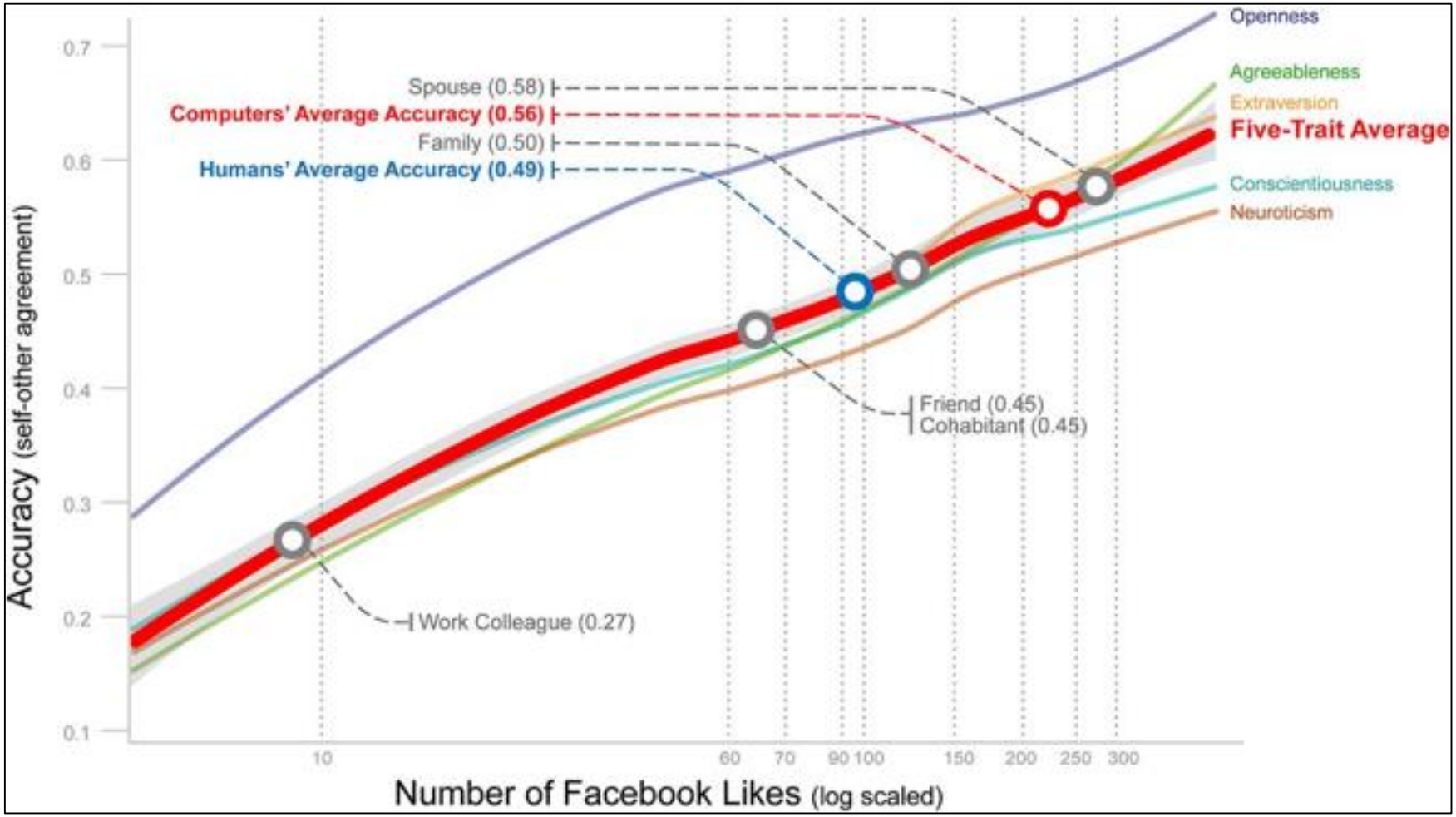
Source: 3Blue1Brown

Psychology: Facebook likes explain personality traits



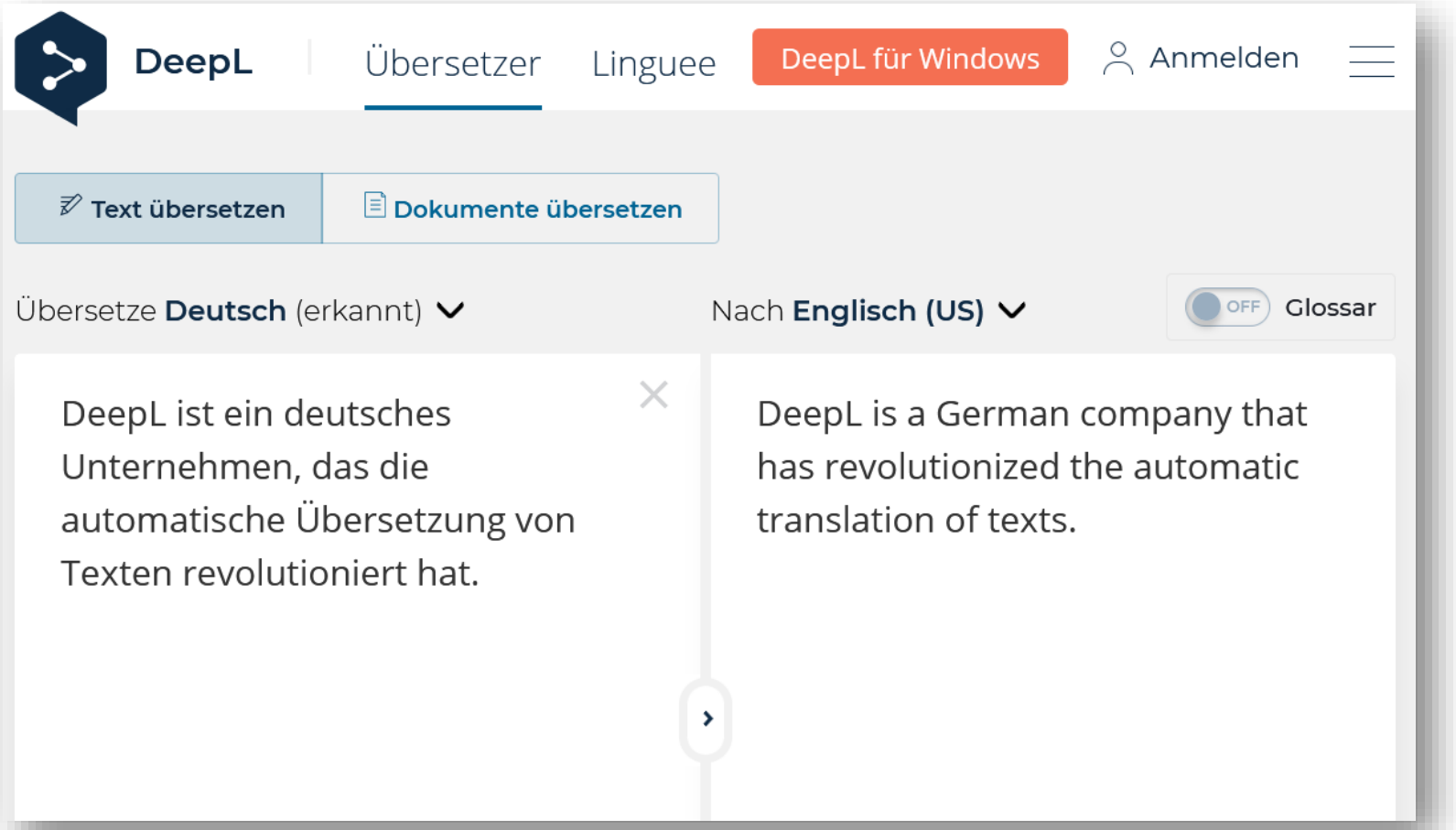
Youyou et al. PNAS 2015;112:1036-1040

How many likes are needed to beat a friend's/partner's prediction

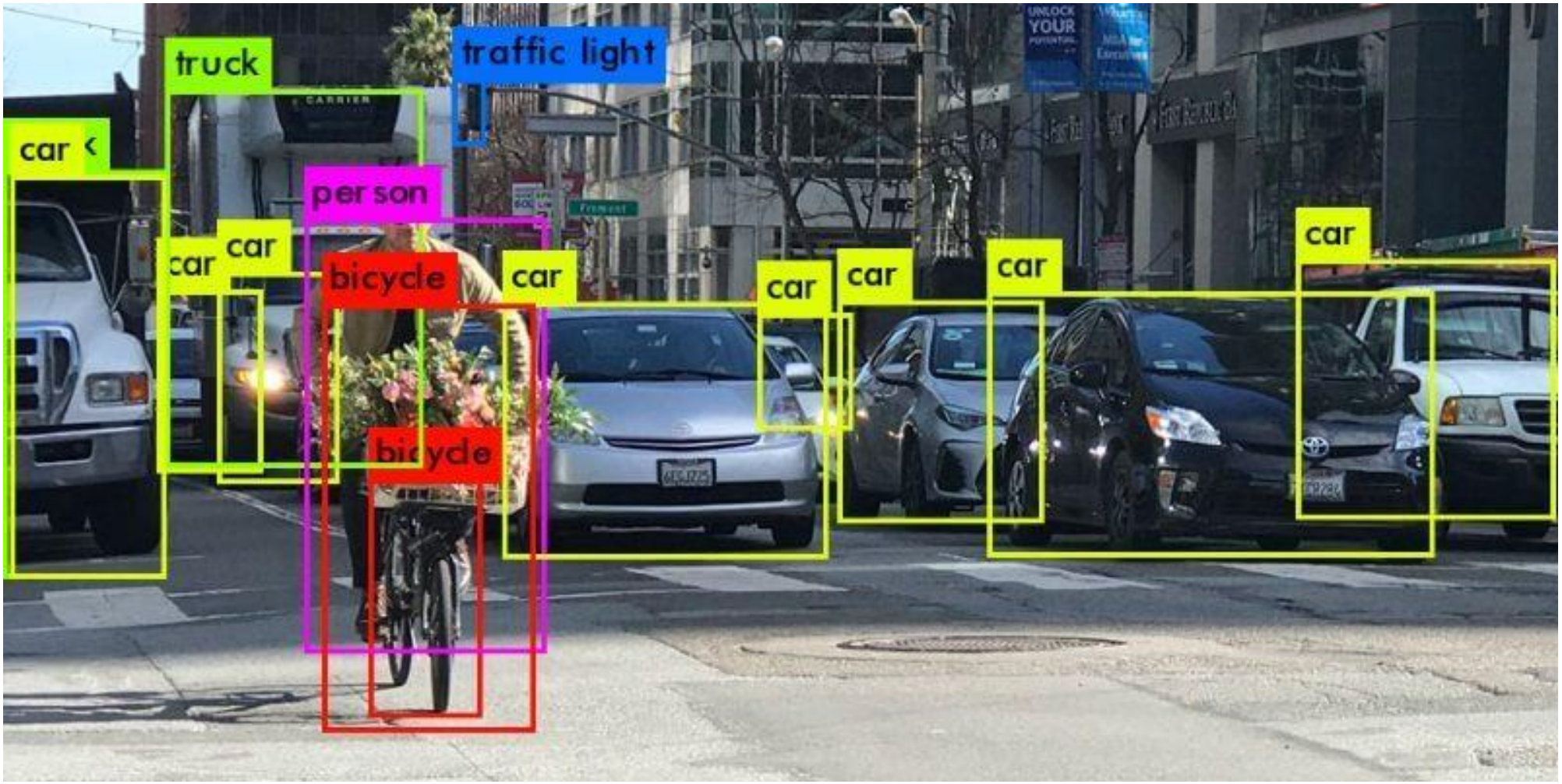


Source: Youyou et al. PNAS 2015;112:1036-1040

Machine translation

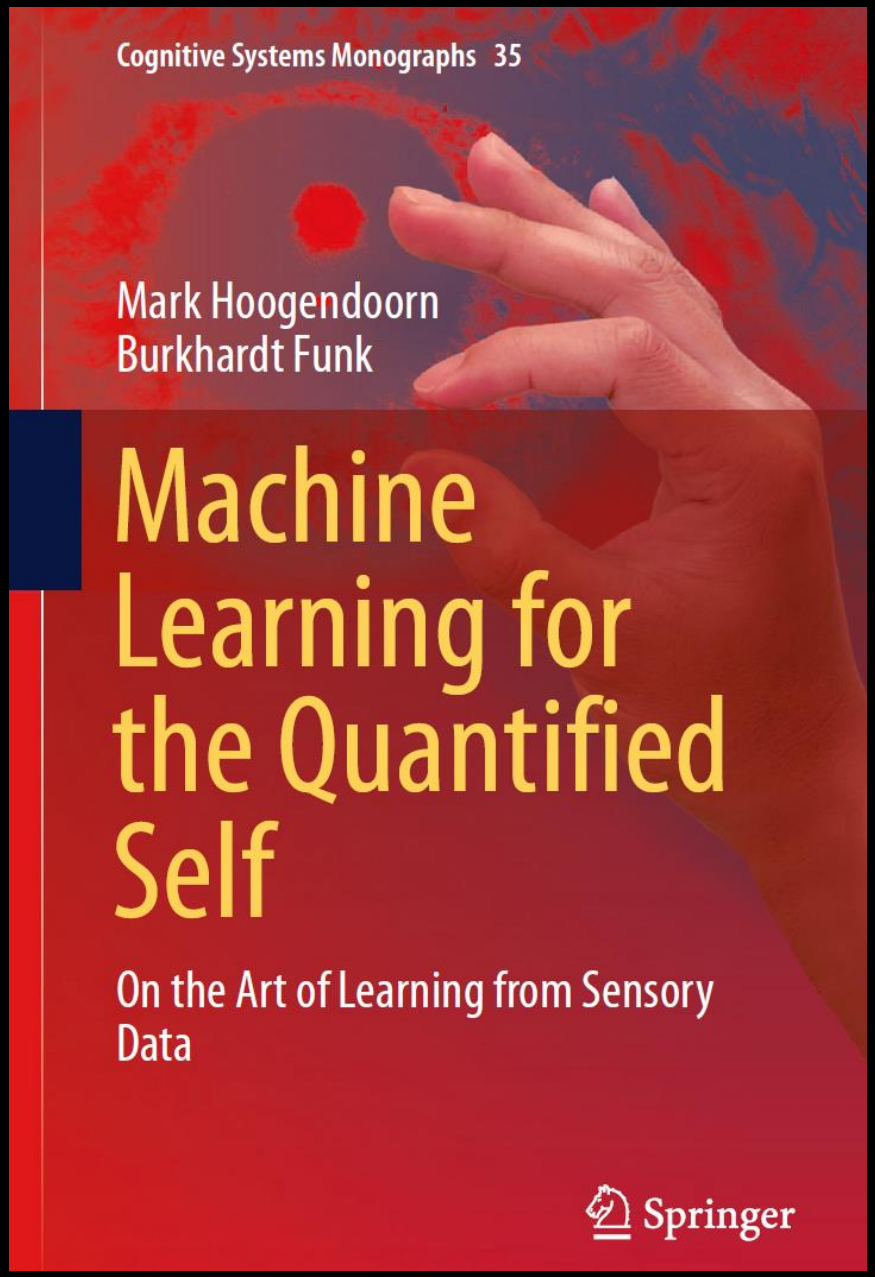


Real-time image recognition by traffic cameras



Source: Azati.ai

Like to play around with sensor data?



Everybody dance now

https://www.youtube.com/watch?v=PCBTZh41Ris

Chan, C., Ginosar, S., Zhou, T., & Efros, A. A. (2019). Everybody dance now. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5933-5942).

Generic machine learning process

An example dataset – Rainfall prediction

- We can train ML models to predict the occurrence, probability of occurrence, and amount of rainfall on the following day, based on present day information
- **Features:** 22; **Note:** for our analyses, we only consider continuous features (16)
- Features include pressure, temperature, evaporation, sunshine, wind speed, etc.
- **Targets:** binary target for rainfall and continuous target for amount of rainfall
- **Note:** when fitting an ML model, remember to use either the binary or continuous target. Failure to remove the other target may result in leakage of future information in your models
- **Samples:** 142,193; **Note:** for simplicity, remove NA values before fitting the models

A glimpse of the dataset

Date	Location	Evaporation	Sunshine	WindGust...	Humidity9...	Pressure9...	Cloud9am	Temp9am	RISK_MM	RainTomor...
2008-12-29	Albury	NA	NA	46	49	1004.8	NA	21.6	1.2	Yes
2009-01-22	Albury	NA	NA	98	60	1005.3	4	26.1	6.4	Yes
2009-02-12	Albury	NA	NA	46	58	1017	2	17	3	Yes
2009-03-10	Albury	NA	NA	50	51	1019.5	NA	20.1	1.2	Yes
2009-03-12	Albury	NA	NA	37	52	1019.5	NA	22.2	5.8	Yes
2009-03-13	Albury	NA	NA	31	82	1017.4	8	19	3	Yes
2009-03-14	Albury	NA	NA	69	82	1012.7	NA	19.9	11.6	Yes
2009-03-25	Albury	NA	NA	30	69	1017.4	8	18.3	1.8	Yes
2009-04-02	Albury	NA	NA	28	59	1022.6	NA	18.4	8.6	Yes

- Choose between the two target variables while training models (continuous or binary)
- Download the dataset: https://www.kaggle.com/jsphyg/weather-dataset-rattle-package
- **Note:** clean the dataset before fitting the models, in order to make it machine readable. Data scientists are also data janitors, e.g. Converting the target to 0 or 1

Source: Kaggle

Let‘s have a look at the code

```python
# -*- coding: utf-8 -*-
"""
Created on Thu Oct  1 12:19:07 2020
Machine learning - lecture_01
Analysing the Australia Rain dataset to predict rainfall for the following day
"""

# loading required libraries
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split, cross_validate
from sklearn.metrics import roc_auc_score, roc_curve, confusion_matrix
import matplotlib.pyplot as plt
from sklearn import preprocessing

# importing the data https://www.kaggle.com/jsphyg/weather-dataset-rattle-package
weather = pd.read_csv("weatherAUS.csv")

# removing the target variable for the amount of rainfall (additional target variable)
weather = weather.drop(columns=['RISK_MM'])
```

Confusion Matrix – a simple evaluation metric for classification

Confusion Matrix		Actual values	
		Rain (1)	No Rain (0)
Predicted values	Rain (1)	True Positive (TP)	False Positive (FP)
	No Rain (0)	False Negative (FN)	True Negative (TN)

$$\textit{Recall} = \frac{TP}{TP + FN} \qquad \textit{Precision} = \frac{TP}{TP + FP} \qquad \textit{F - measure} = \frac{2*Recall*Precision}{Recall + Precision}$$

- **Recall:** how many of the actual values for the positive class (rain) are correctly predicted
- **Precision:** how many of the predicted values for the positive class (rain) are correct
- **F1 measure:** Harmonic mean of recall and precision; to strike a balance between the two

Source: Sarang Narkhede. Towards Data Science: Understanding Confusion Matrix

Receiver Operating Characteristic Curve (ROC)

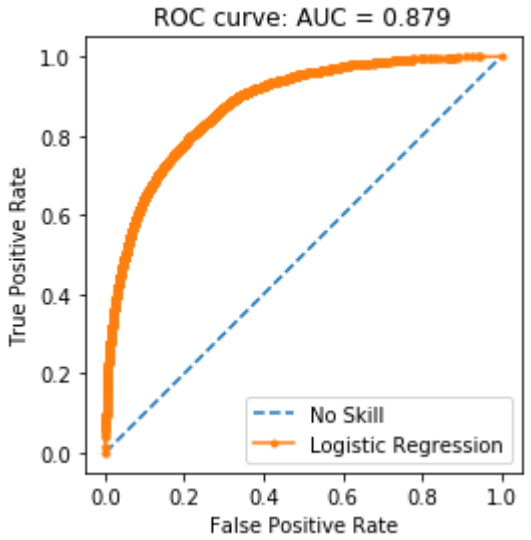
- The ROC curve gives an idea about the performance of a classification model at various thresholds (eg. of a threshold: probability of rain > 0.5 to classify it as rain)
- The Area Under the Curve (AUC) of the ROC curve gives an idea about how efficient the model is at distinguishing between the two classes
- Threshold can be selected based on how many false positives you're willing to accept and the AUC can be used to compare model performance

$$\text{TPR /Recall / Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{FPR} = 1 - \text{Specificity} = \frac{FP}{TN + FP}$$

AUC ~ 1 => perfect predictions
AUC = 0.5 => no capacity to distinguish between classes
AUC ~ 0 => opposite predictions



Source: Sarang Narkhede. Towards Data Science: Understanding AUC-ROC curve

First task

— Use the code for the rain fall case as a template and analyze the heart disease data

https://myshare.leuphana.de/?t=c18d76caf7fefec89aed4d3e51ad336b

— Any challenges/ insights?

Any questions/ thoughts?