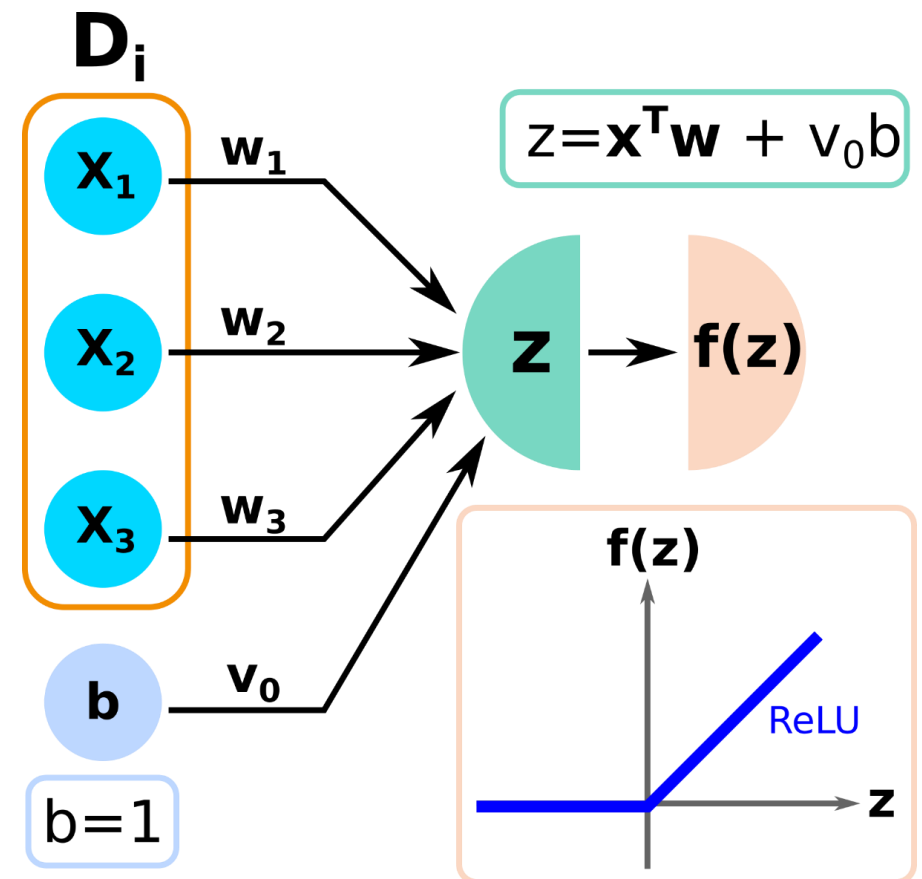# Short Overview on Machine Learning

Dr. Stefan Zahn

What is Artificial Intelligence and Machine Learning?

◤ AI enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.

◤ Machine learning is a subfield of artificial intelligence that gives computers the ability to learn without explicitly being programmed.

   ◤ Write a code where y = 3x + 4 is used to predict y based on x.

      ➢ No machine learning since it is limited to a single correlation.

   ◤ Write a code which get the relation of y = 3x + 4 based on a linear regression and use this property subsequently to predict y based on x.

      ➢ Machine learning since it can be used on any 1D-data to find a linear correlation

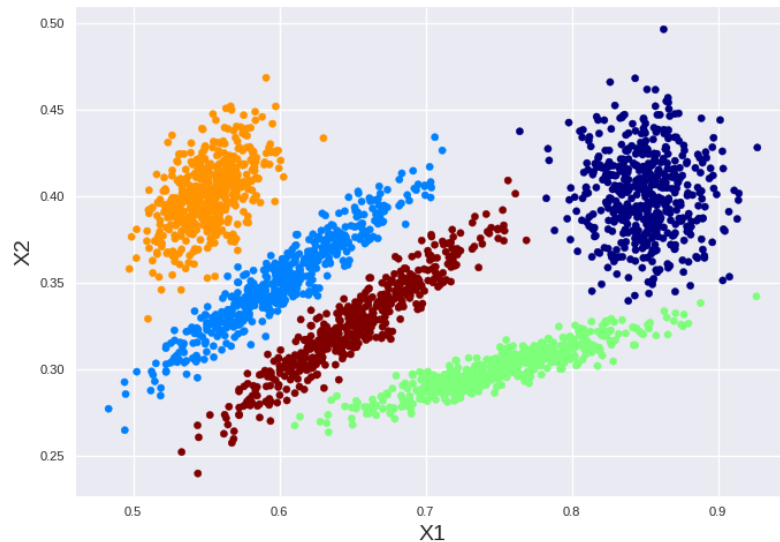◤ Deep learning refers to usage of neuronal networks architectures with many layers.

Artificial Intelligence

Machine Learning

Deep Learning
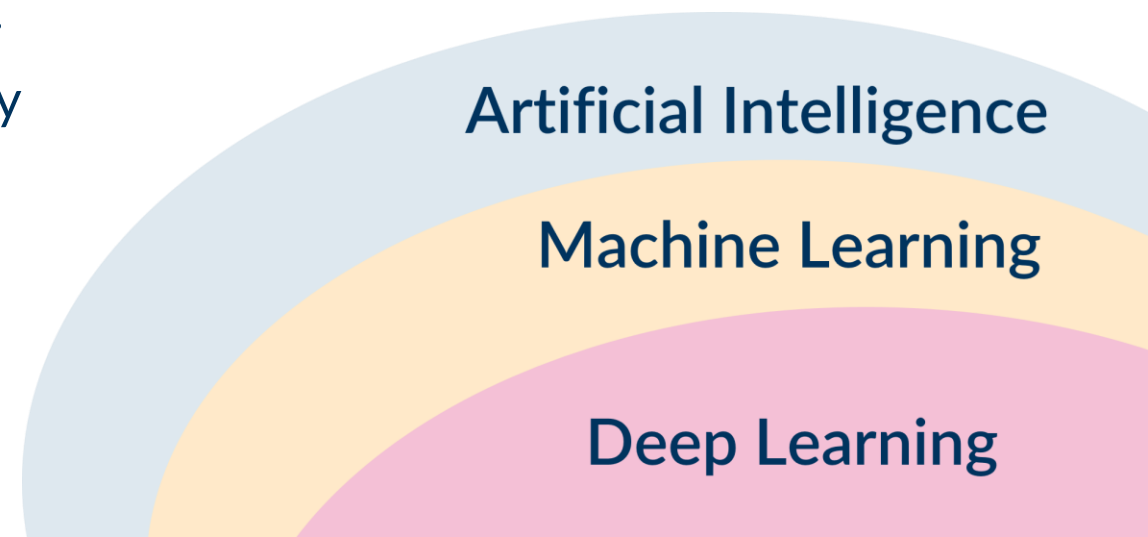
Types of Machine Learning

Machine Learning Types

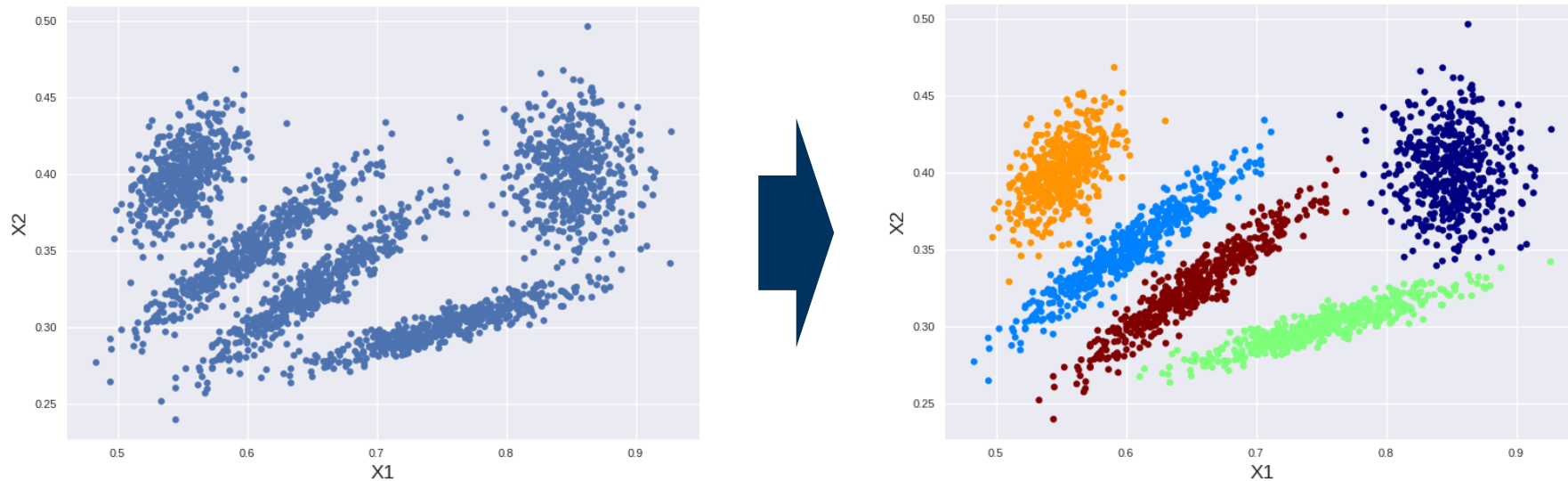◢ Supervised: A models trained on labeled data to predict labels for new input

X1 = 0.85

X2 = 0.40

it's the dark
blue group

Machine Learning Types

▰ Supervised: A models trained on labeled data to predict labels for new input

▰ Unsupervised: Model is trained on unlabeled data to find patterns, relations etc.
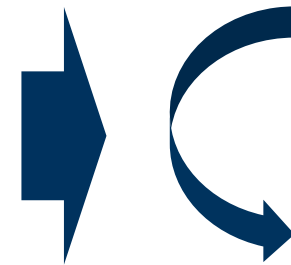
Machine Learning Types

- Supervised: A models trained on labeled data to predict labels for new input

- Unsupervised: Model is trained on unlabeled data to find patterns, relations etc.

- Semi-Supervised: A small subset of data is labeled. Agent assigns pseudo-labels to the rest of the data and trains a supervised algorithm on data with labels of highest confidence.
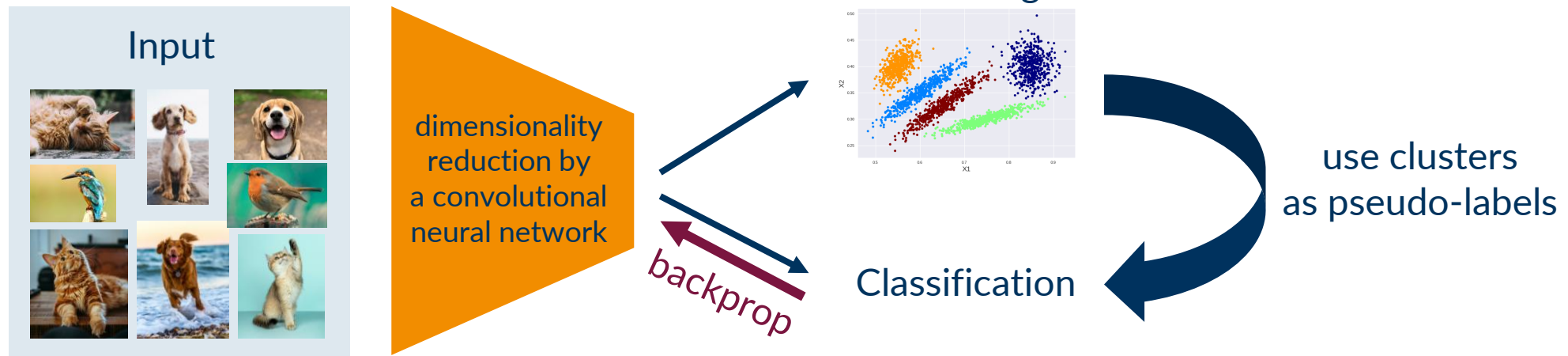
X1 = 3.0

X2 = 2.1

it's the dark blue group

Machine Learning Types

◢ Supervised: A models trained on labeled data to predict labels for new input

◢ Unsupervised: Model is trained on unlabeled data to find patterns, relations etc.

◢ Semi-Supervised: A small subset of data is labeled. Agent assigns pseudo-labels to the rest of the data and trains a supervised algorithm on data with labels of highest confidence.

◢ Self-supervised: The algorithm generate reasonable labels based on the analysis of data and trains a supervised model subsequently.


Concept of Deep Clustering

Machine Learning Types

▰ Supervised: A models trained on labeled data to predict labels for new input

▰ Unsupervised: Model is trained on unlabeled data to find patterns, relations etc.

▰ Semi-Supervised: A small subset of data is labeled. Agent assigns pseudo-labels to the rest of the data and trains a supervised algorithm on data with labels of highest confidence.

▰ Self-supervised: The algorithm generate reasonable labels based on the analysis of data and trains a supervised model subsequently.

▰ Reinforcement Learning: Optimize a policy based on trial and error

  ➢ Analyse data
  ➢ Select action based on a policy
  ➢ Do action
  ➢ Get reward or penalty based on the outcome
  ➢ Update policy
  ➢ Iterate until optimal policy is found

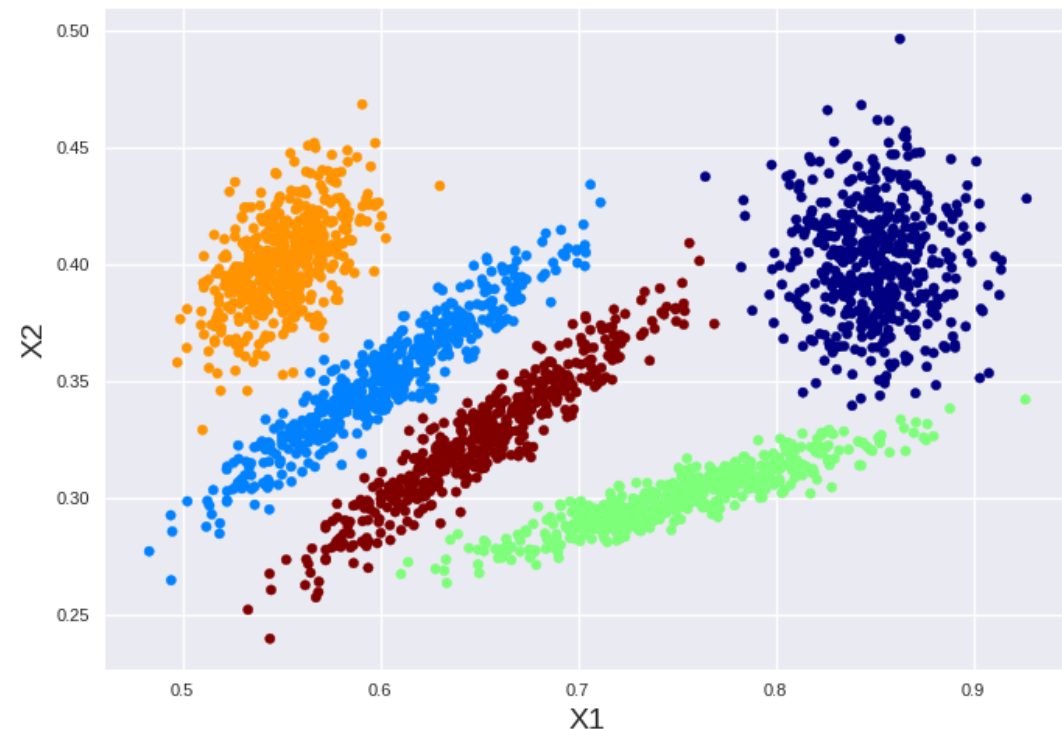# Machine Learning Techniques introduced during the Course

IOM

Regression
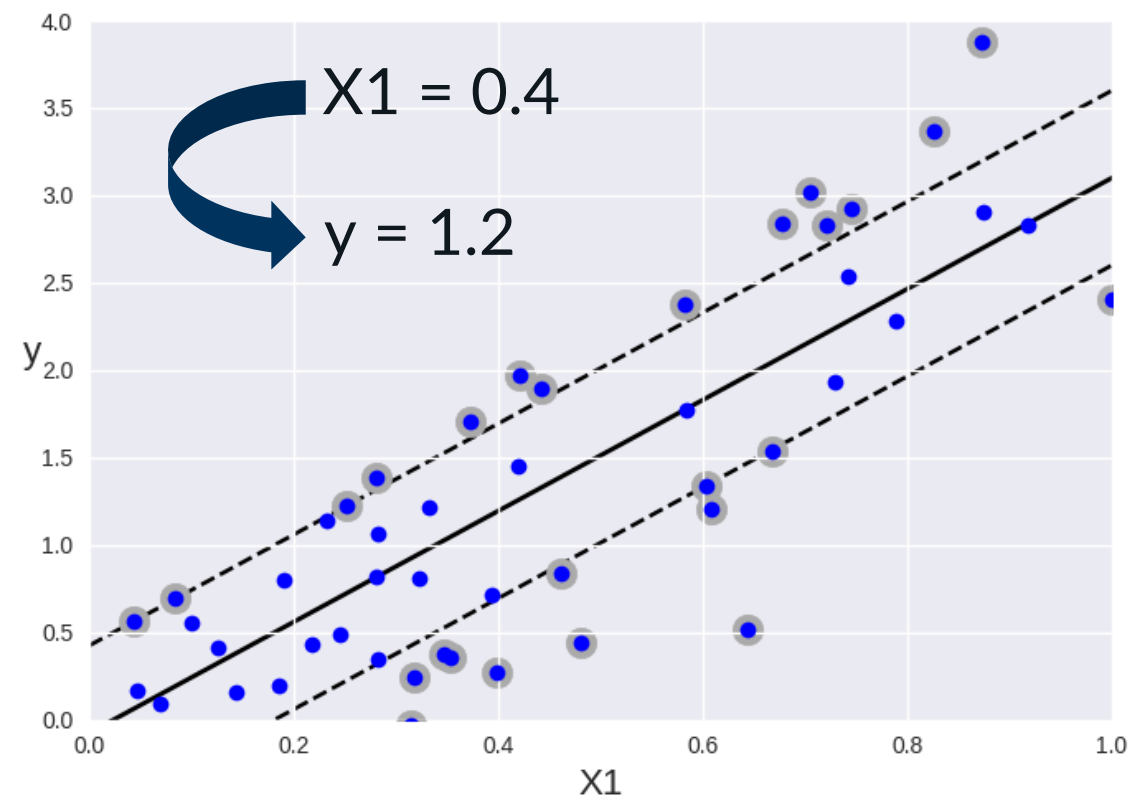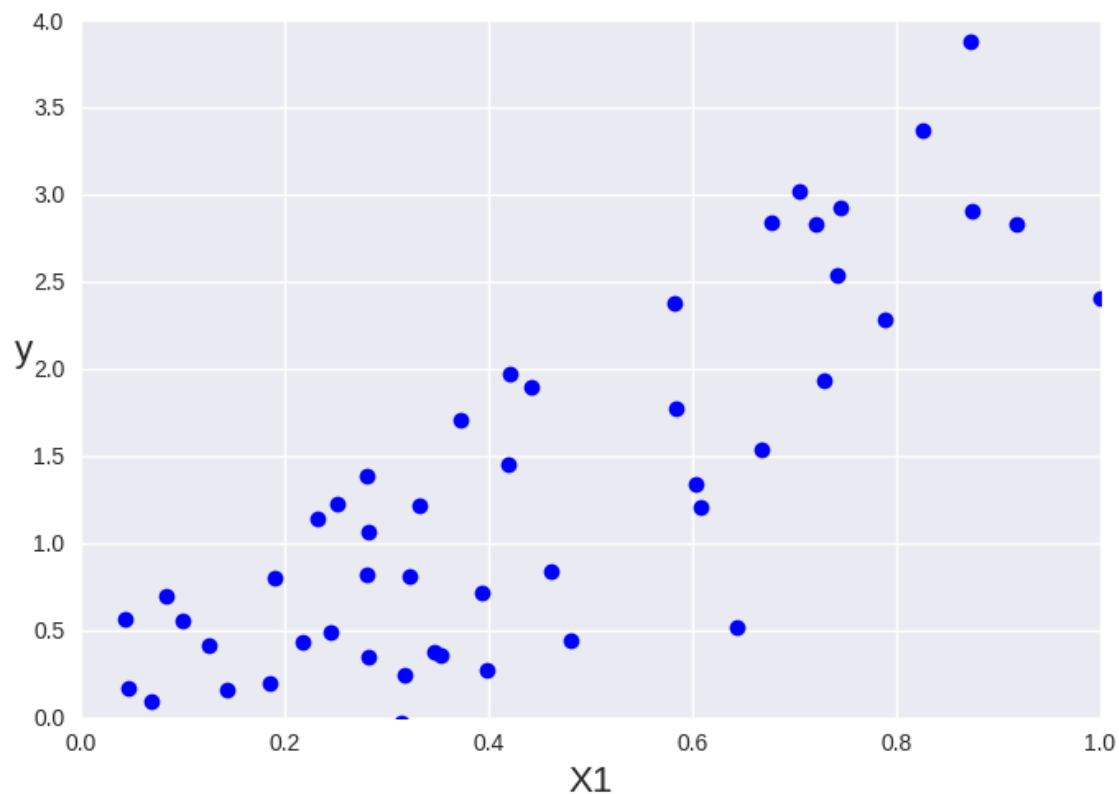
- Find a correlation between input and output (supervised technique)
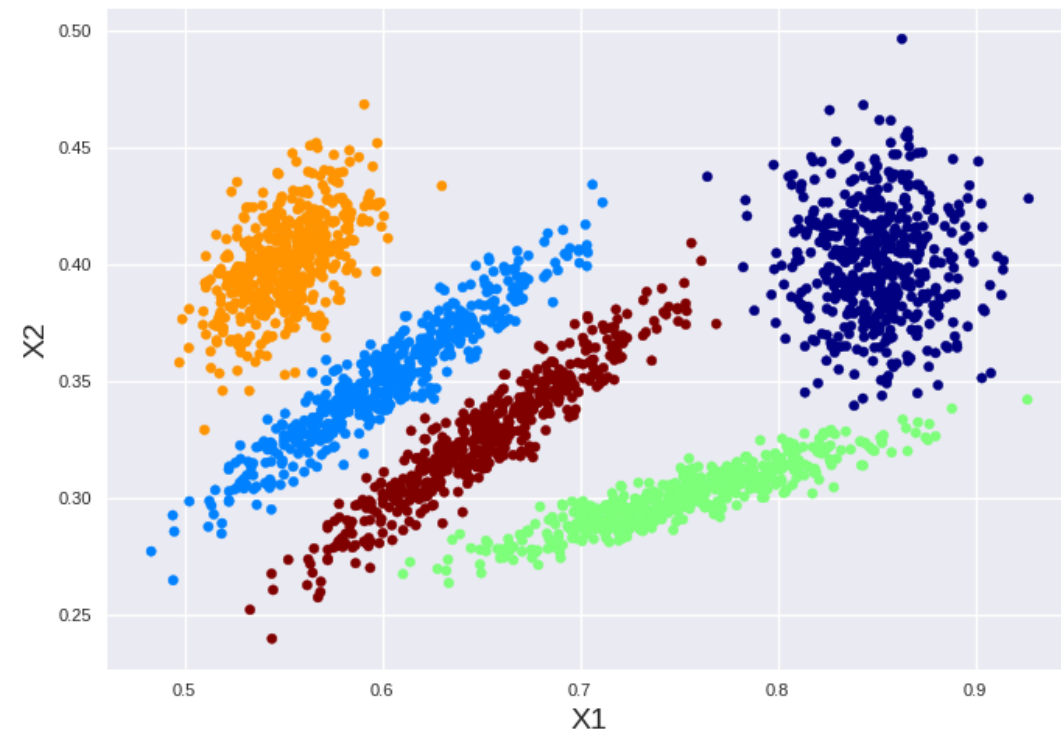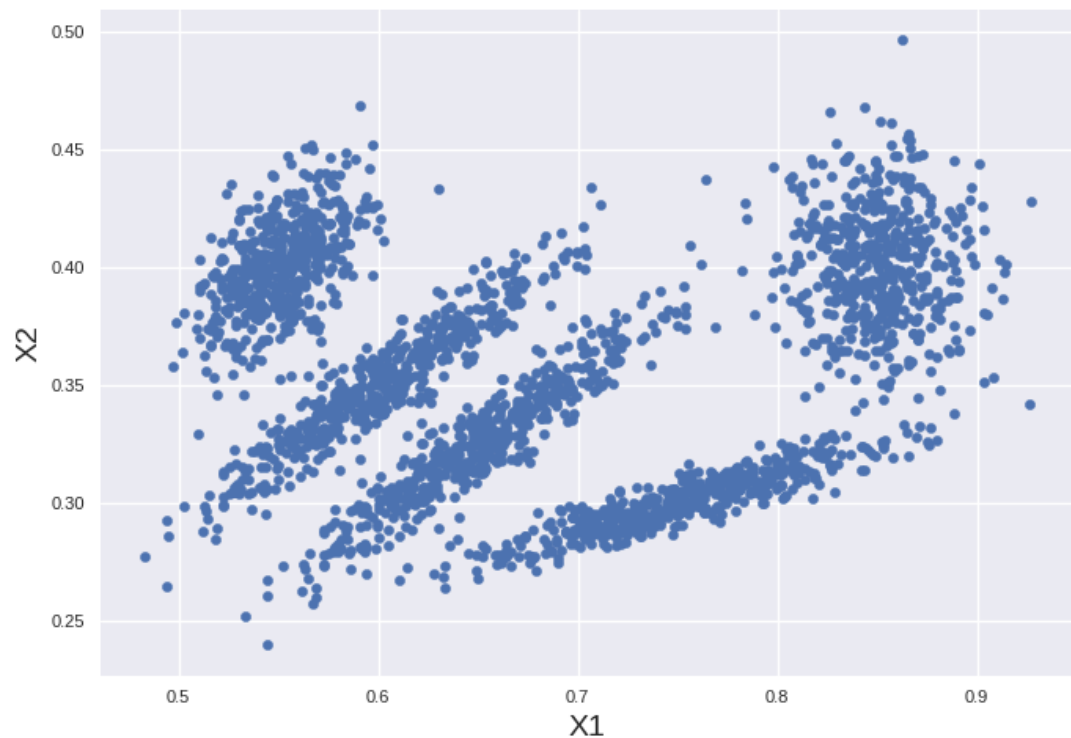- Focus on Monday

Classification

X1 = 0.85

X2 = 0.40

it's the dark blue group

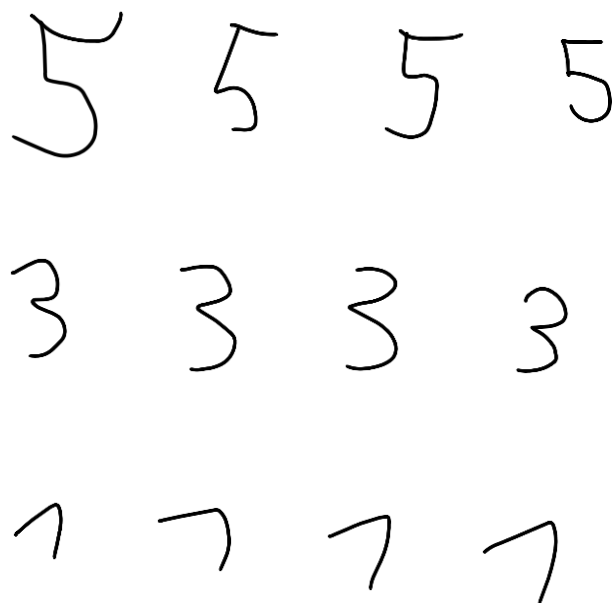▰ Find a correlation of input to a group (supervised technique)

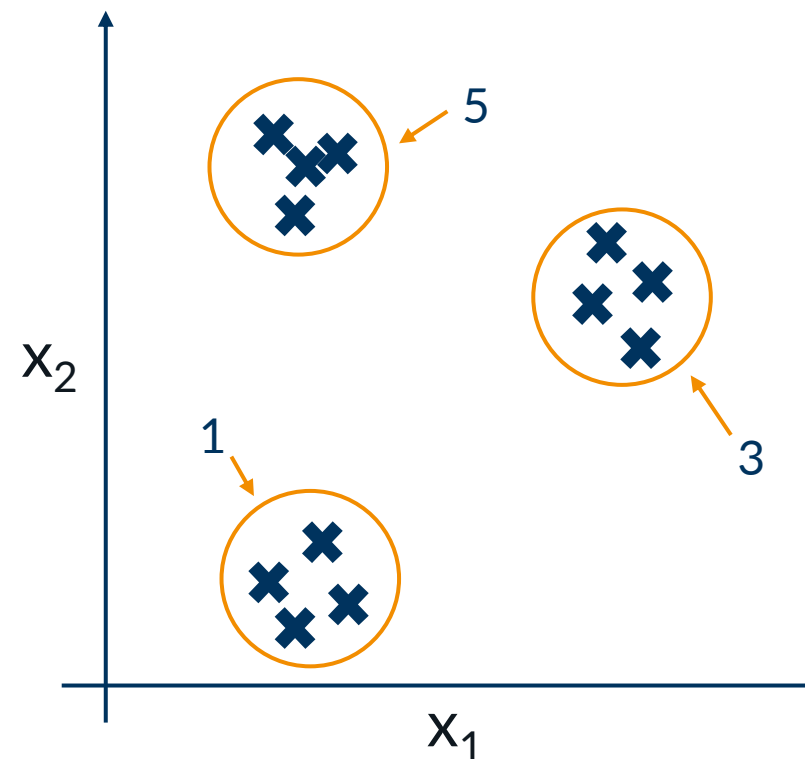▰ Focus on Tuesday

Clustering

- Find groups in a data set (unsupervised technique)
- Focus on Tuesday

Dimensionality Reduction

figures of numbers
(images are stored as a matrix)

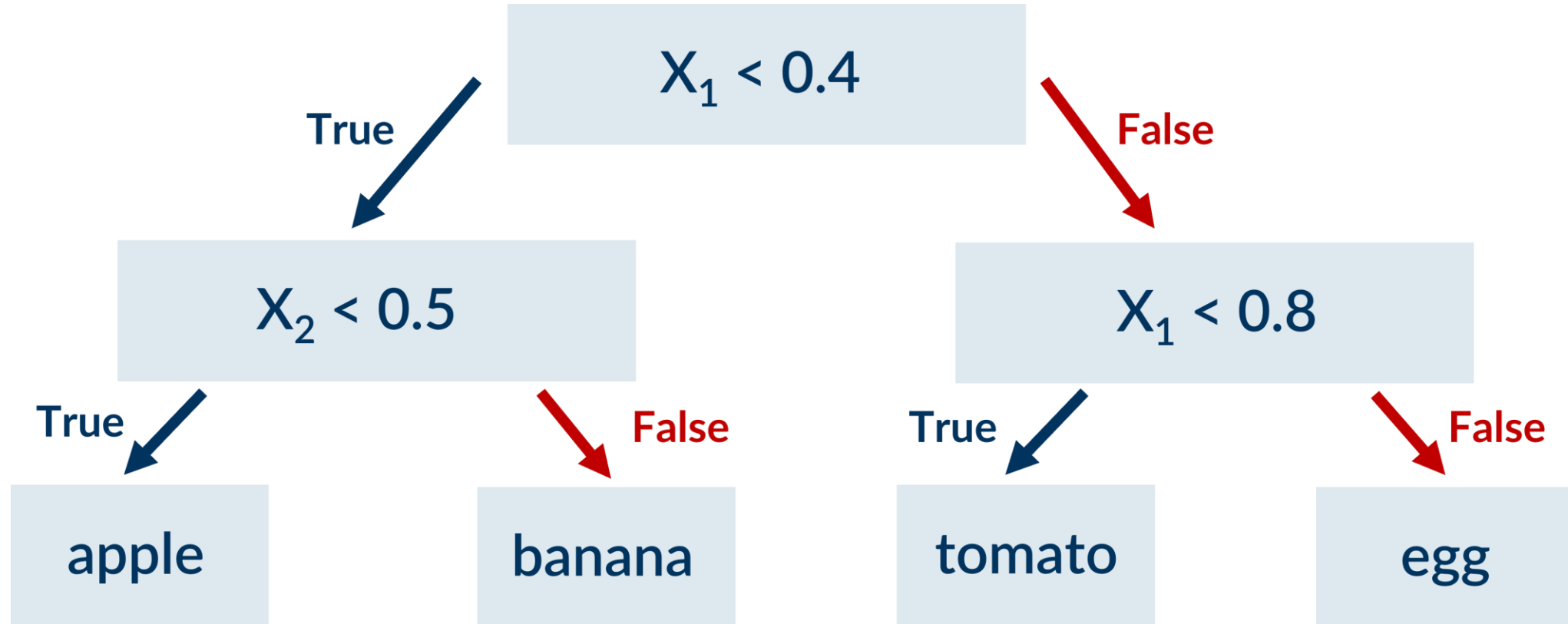### 2-dimensional matrix

x_2

x_1

5

3

1

- Find a lower dimensional representation for the data (unsupervised technique)

- Focus on Wednesday (session on data visualization)

4 Types of Machine Learning Models
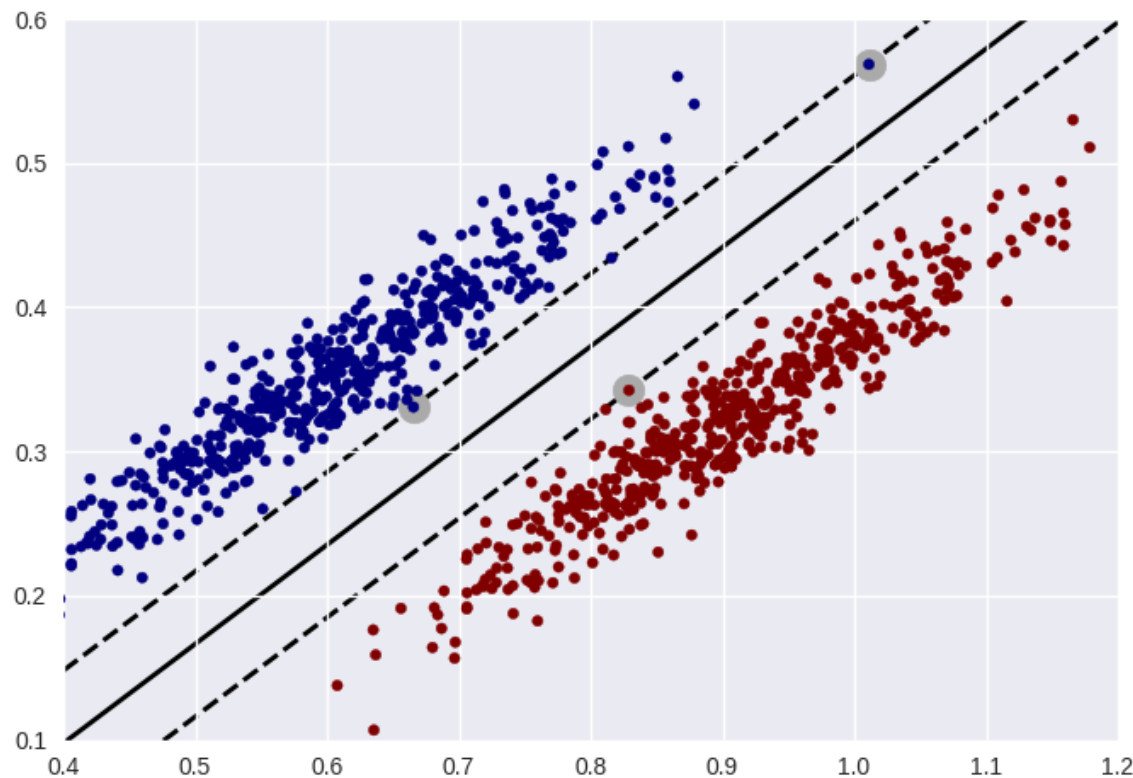# for Regression or Classification

ML models: Decision Trees and Random Forests

$X_1 < 0.4$

True

False

$X_2 < 0.5$

$X_1 < 0.8$

True

False
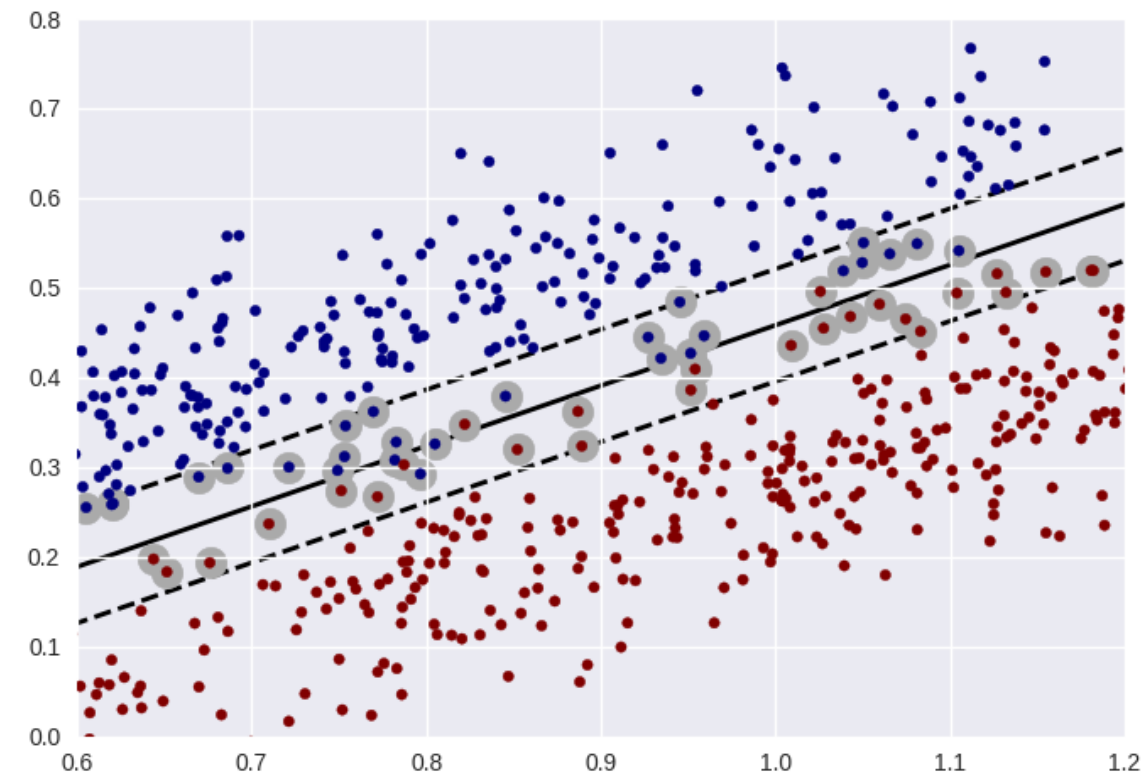
True

False

apple

banana

tomato

egg

- Decision trees are very sensitive to training data

- Random forests are an ensemble of decision trees to be less prone to training data

IOM

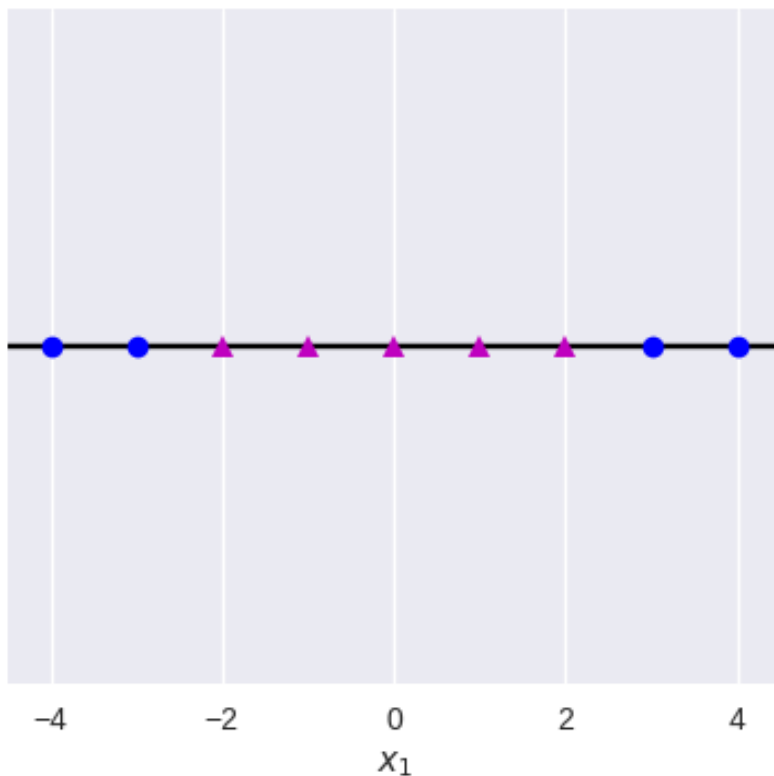# ML models: Support Vector Machines

hard margin SVM

soft margin SVM

▰ Basic concept for classification: Put the widest possible linear street between two classes

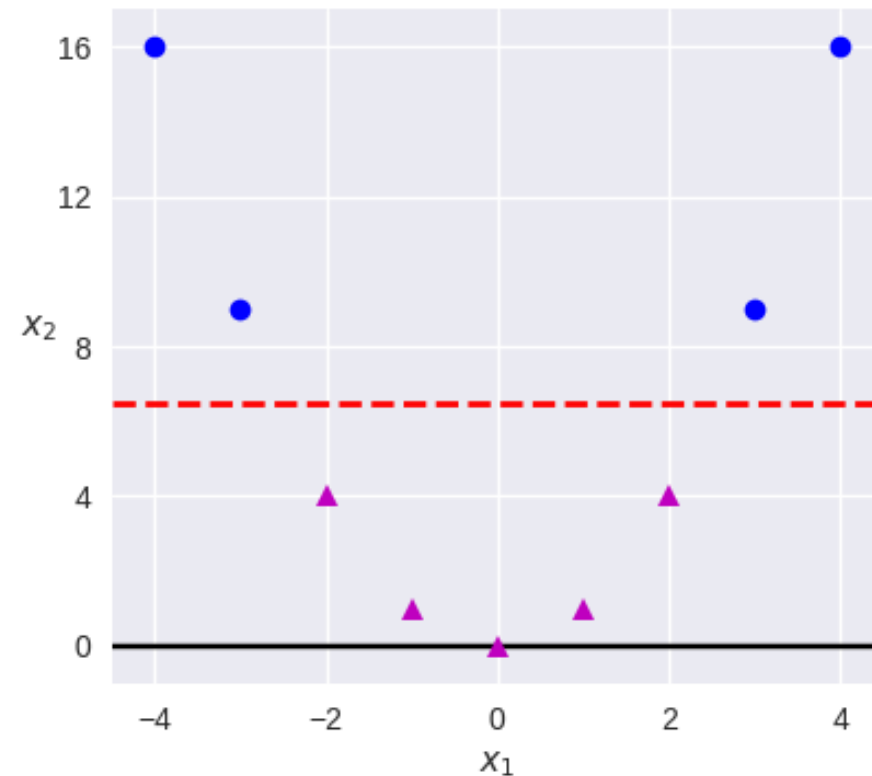▰ Soft margin allows margin violation controlled by hyperparameters

lecture on SVC: https://tinyurl.com/SVC-math

ML models: Support Vector Machines

original data

transformed data

kernel

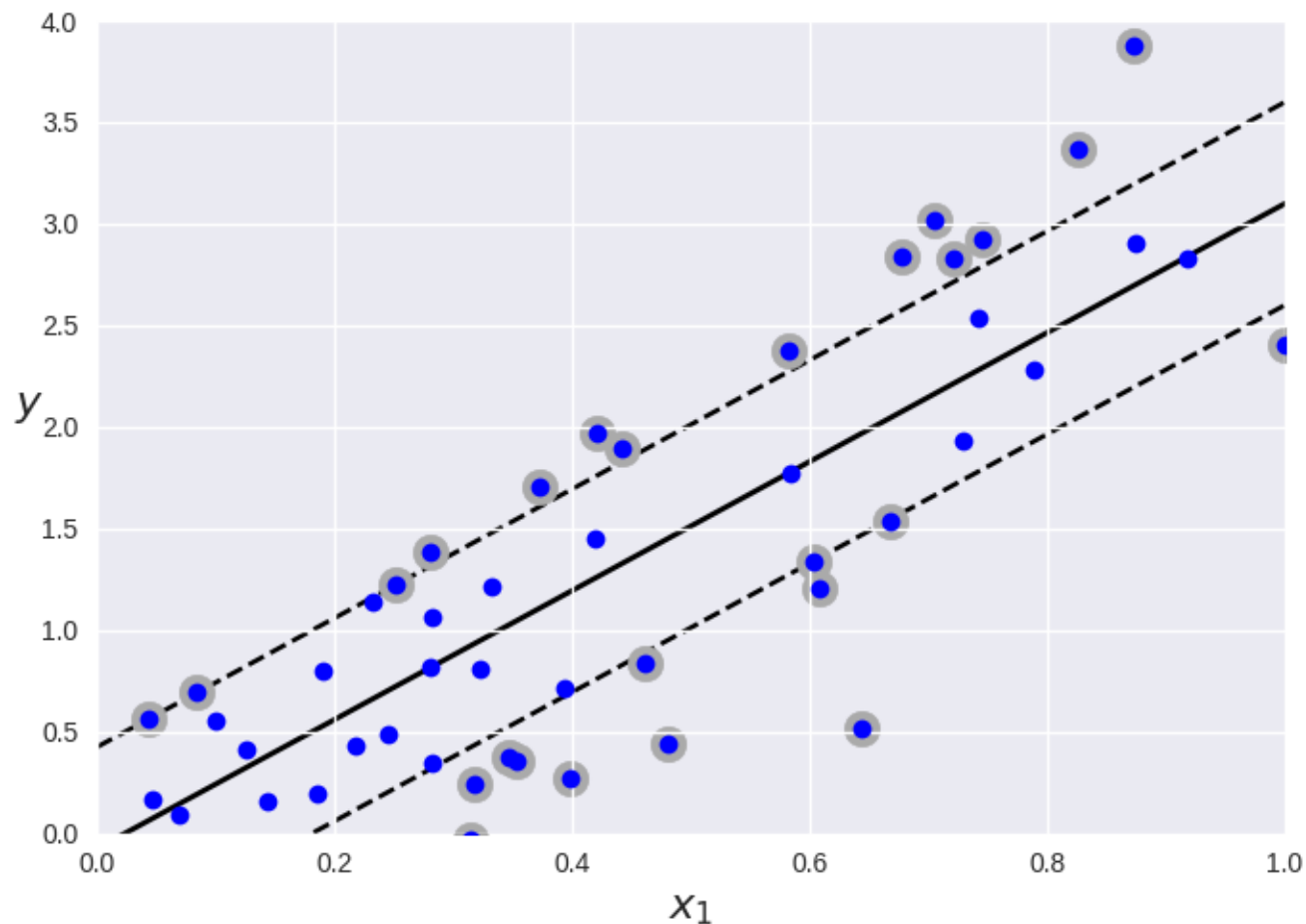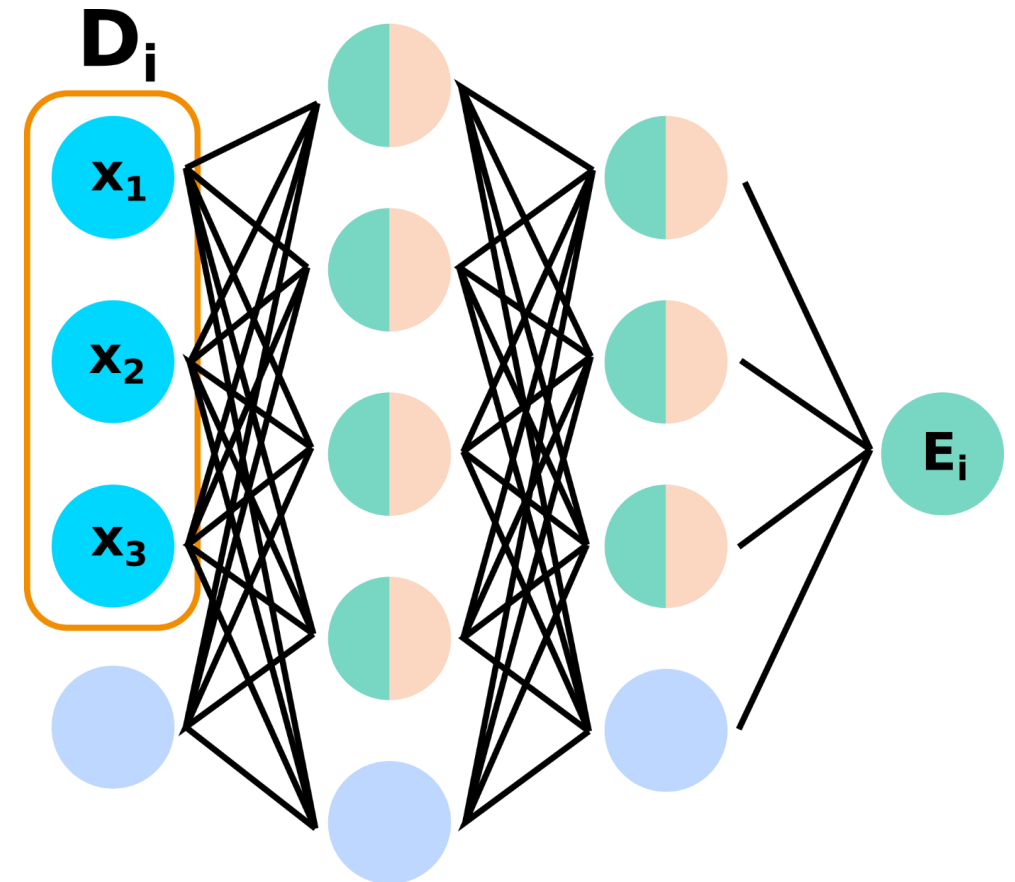"transforms" from 1D to
2D problem which is
linear separable

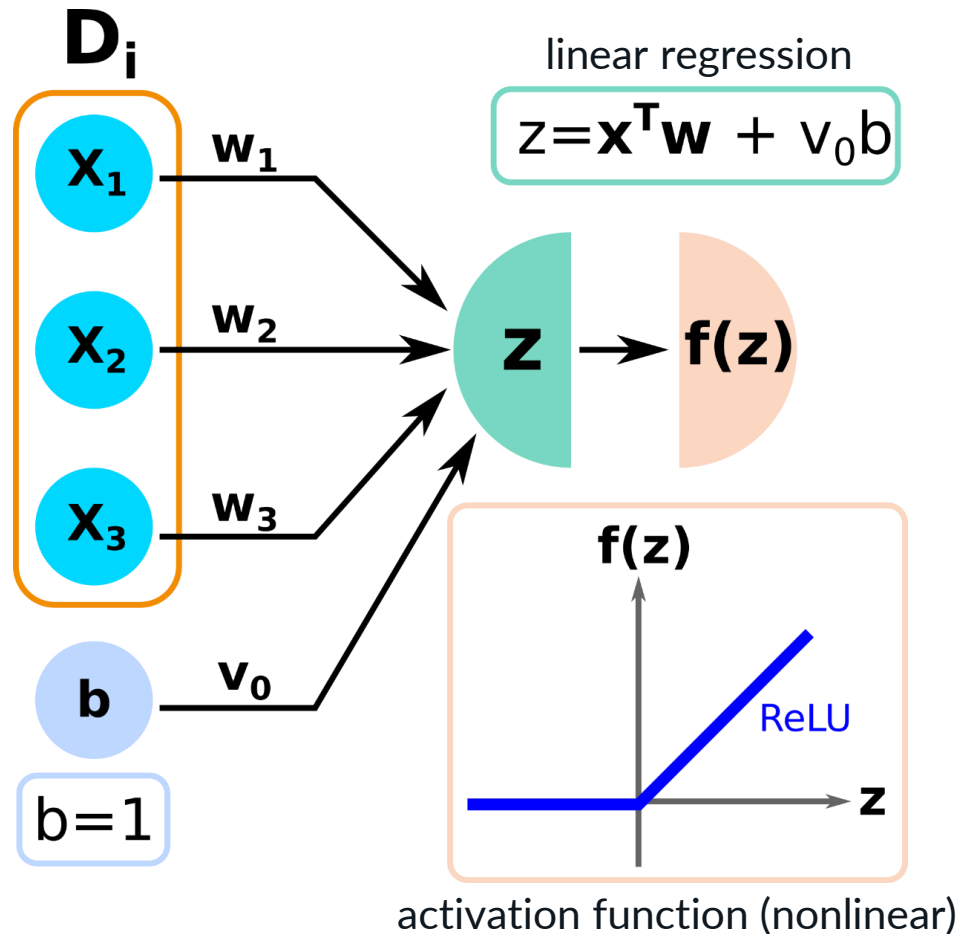◢ Kernel trick allows to apply SVM to nonlinear problems

◢ Basic concept: Input features are "transformed" before classification is done

lecture on SVC: https://tinyurl.com/SVC-math

Support Vector Machines

◢ Basic concept for regression: Put as much as possible data points on a linear street

◢ Hyperparameter control wide of the street (good for noisy data)

lecture on SVC: https://tinyurl.com/SVC-math

ML models: Neuronal Networks

linear regression

$$z = \mathbf{x}^T\mathbf{w} + v_0 b$$

$b=1$

activation function (nonlinear)

- Basic concept: linear regression where the output is passed to a nonlinear function
- Deep neuronal networks have multiple hidden layers

Central Steps of a Machine Learning Project

◢ Look at the big picture

◢ Get the data

◢ Create a test dataset (only for supervised techniques)

◢ Explore and visualize the data

◢ Prepare the data for the machine learning algorithm

◢ Select a model and train it

◢ Fine-tune the model

◢ Present your model and make it ready for later usage

➢ We go through these steps for a regression model to predict the energy of a hydrogen bond.

acceptor atom

donor atom
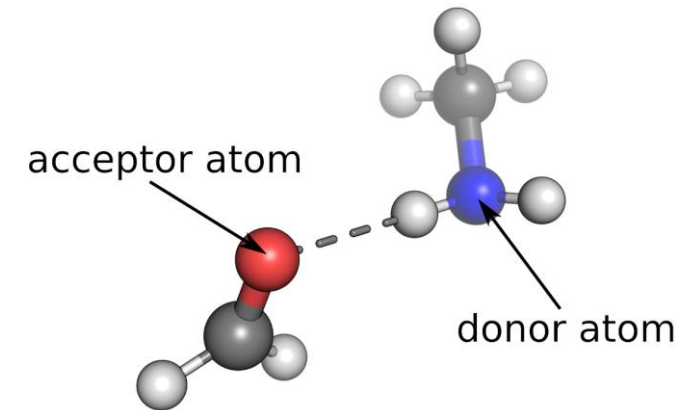
What is scikit-learn?

◢ Python-based machine learning library for:
  ◢ Regression
  ◢ Classification
  ◢ Clustering
  ◢ Dimensionality Reduction
  ◢ Tools for data preprocessing and hyperparameter tuning

◢ Large number of models suited for small and medium sized data sets

◢ Not optimized for deep learning approaches

◢ Allows to setup a machine learning model and optimize hyperparameters with few lines of Python code