ECS 171: Machine Learning

Summer 2023
Edwin Solares
easolares@ucdavis.edu
Linear Classification & SVM Intro

Group Project

Project Requirements:
Form a group of up to 6
Determine Project Goals, submit for approval
Find a Dataset (UCI ML datasets, kaggle, google datasets)
Data Exploration
- Read up on what the data means

Preprocess Data
- Imputation
- Normalization vs standardization
- Data Encoding
- Data transformation

Group Project

Project Requirements (cont…):
Preliminary Data Visualization
- PairPlot
- Correlation Matrix

Choose a model
- Supervised vs Unsupervised?
- Maybe both?
- feature engineering all the way to building prediction models, and active learning.

Evaluation and testing
Digest and dissect results
Writeup

Group Project

Write up Requirements:

Github

- Create documentation via Readme
- Background
  - What is the story behind the data, previous work
- Introduction
  - What are the motivations behind the project
  - What are the objectives of the project
  - What is the broader impact?
- Methodology
  - Explain how you preprocessed, choosing of model, implementation, how did you deal with over/under fitting?
- Data visualization and meaning. Describe the plot
  - What is the message behind each visualization

Group Project

Write up Requirements:

Github

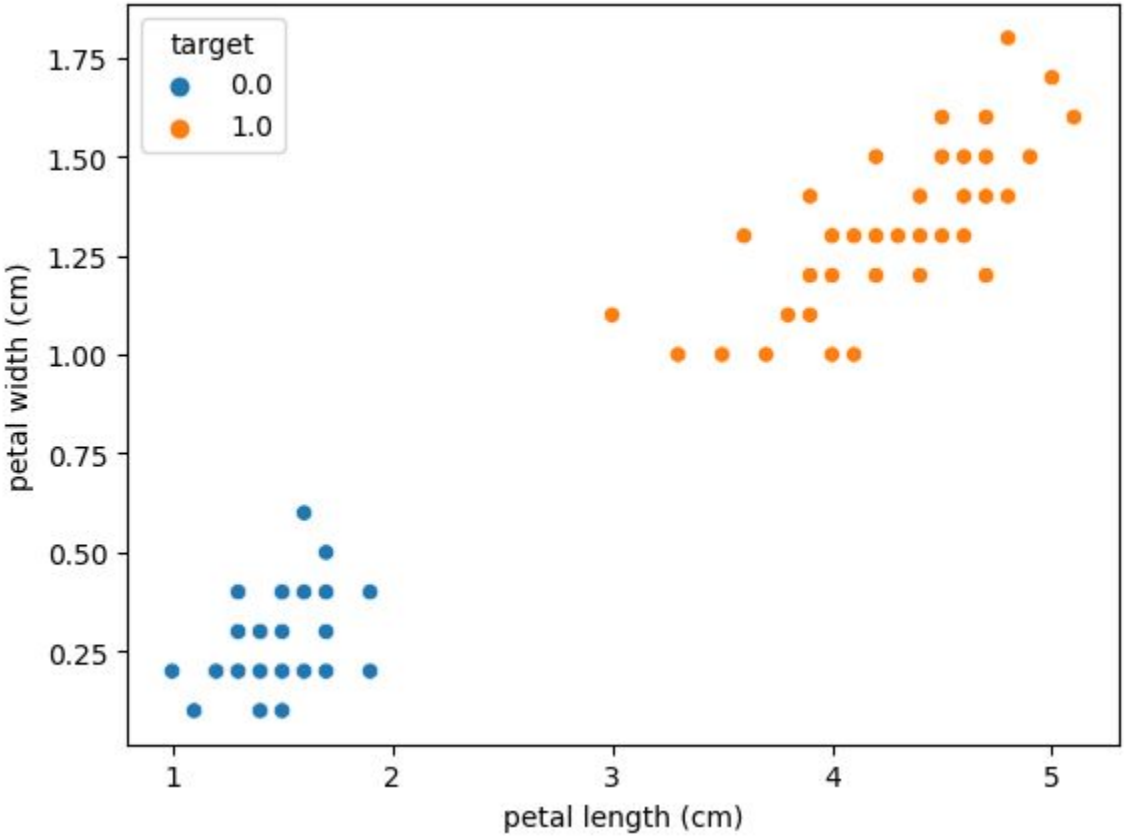
- Conclusions
- Possible future follow ups/investigations

What is Classification

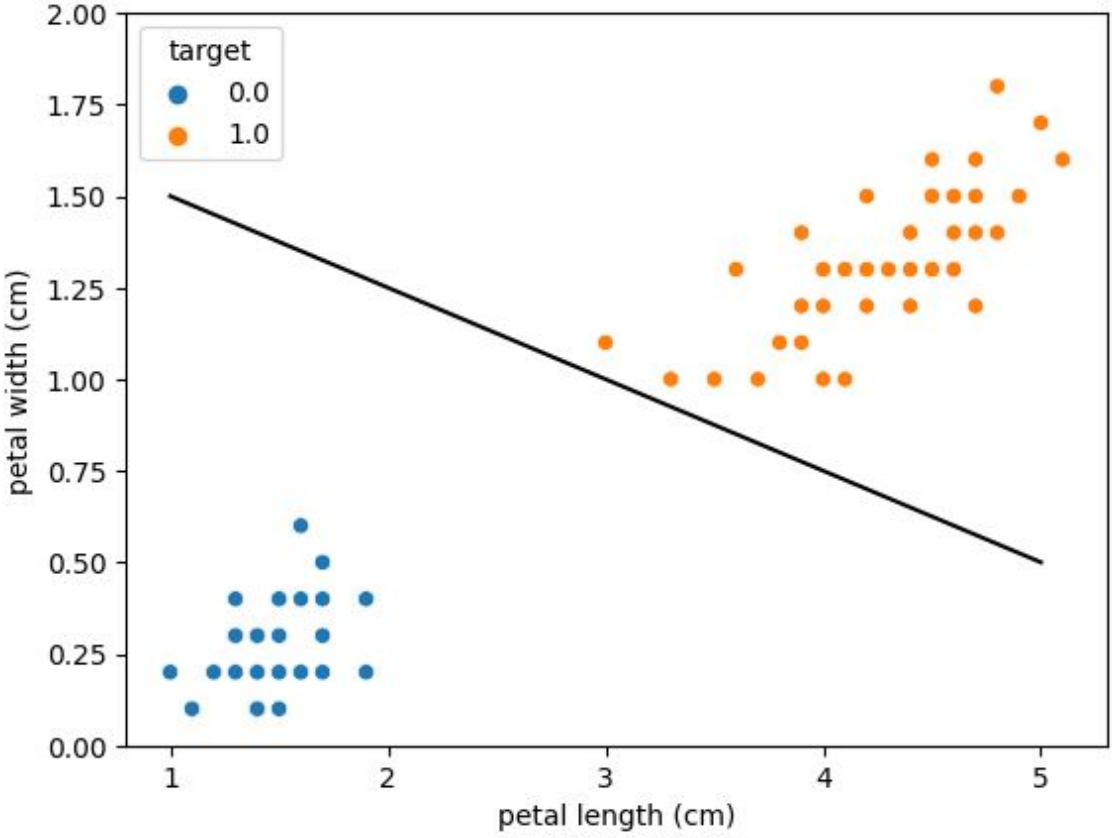
Given a dataset with labels

- Can we create a function (linear, polynomial, other) to split data
- Can we use a discriminating function to predict classes of new observations
- Examples:
  - Spam classification, malware classification, species classification

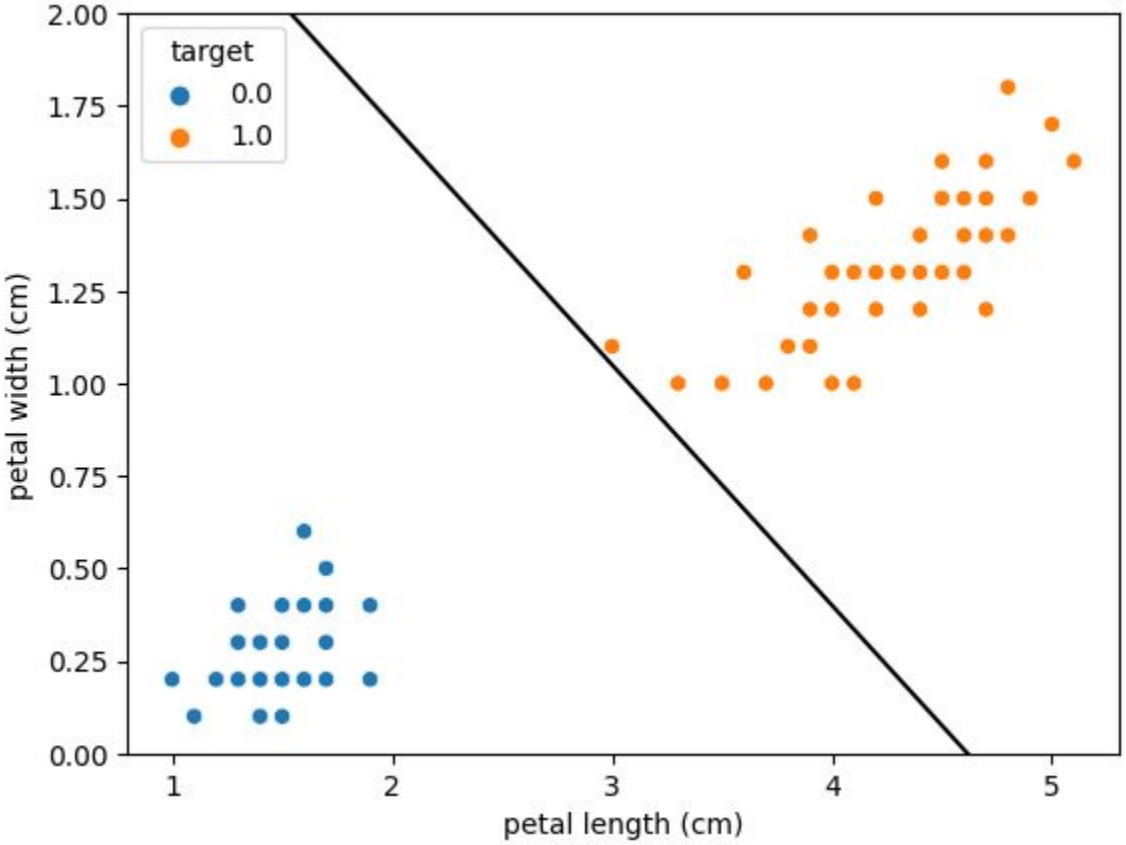
What is Classification



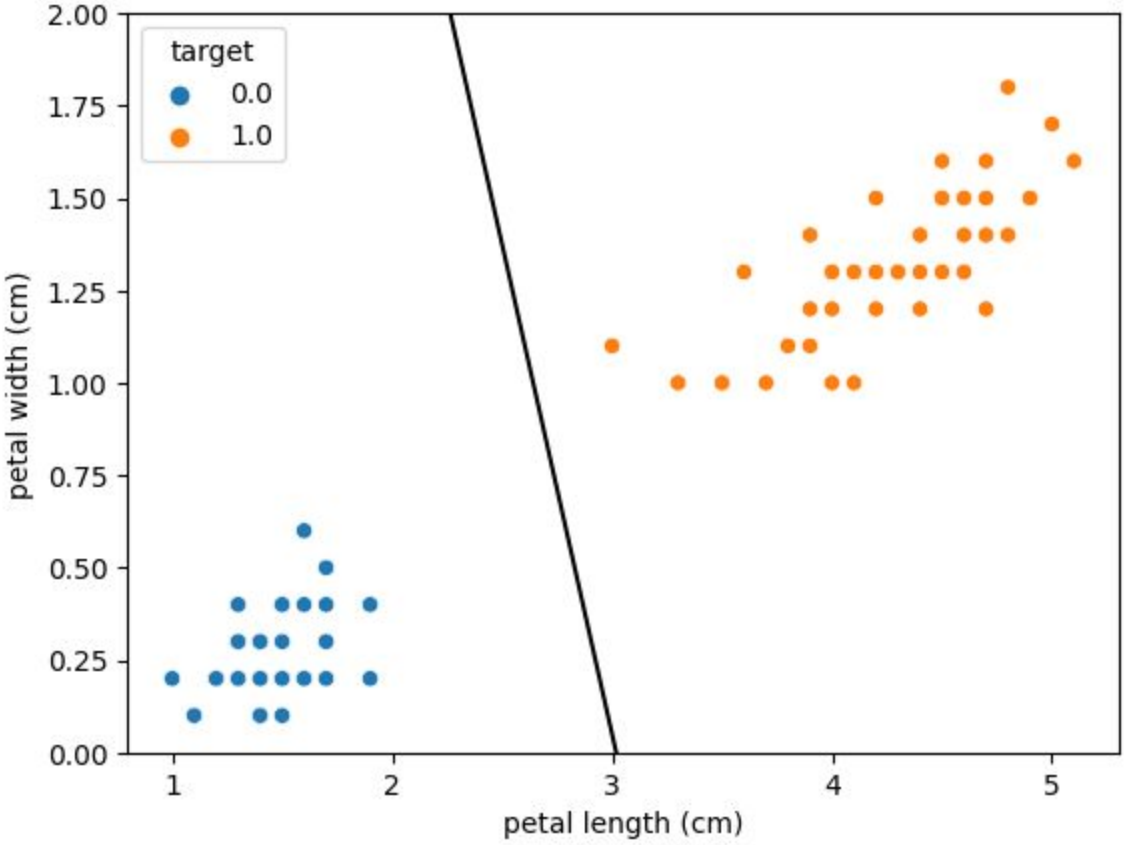
Using a line to define a boundary



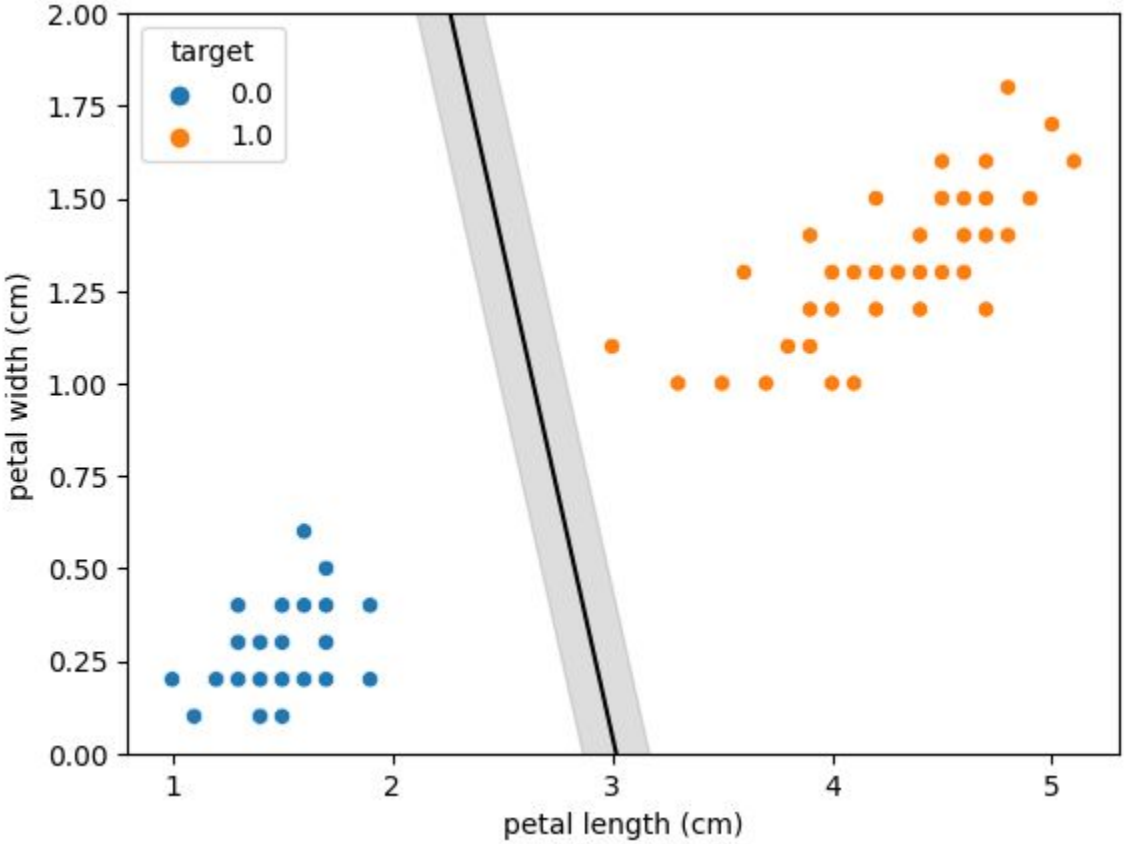
Which is a good fit?



Which is a good fit?

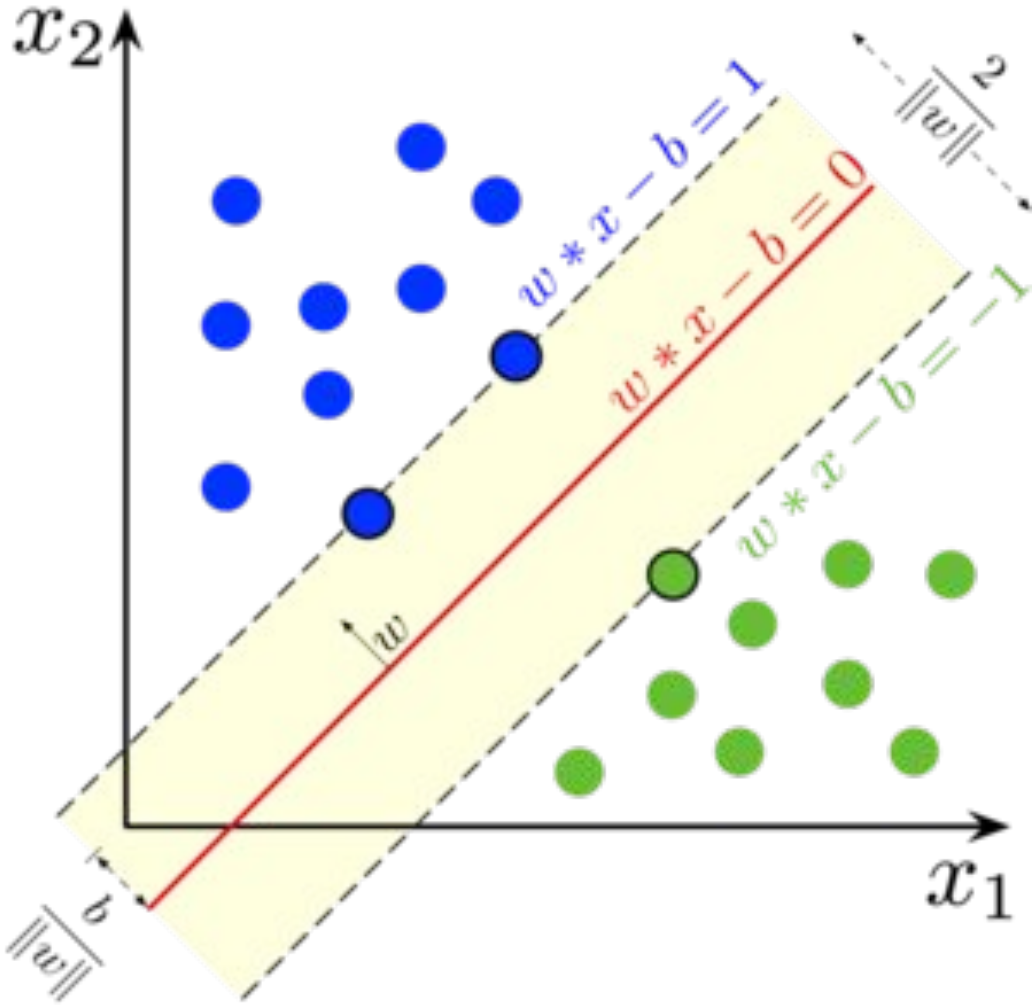


Adding a Margin

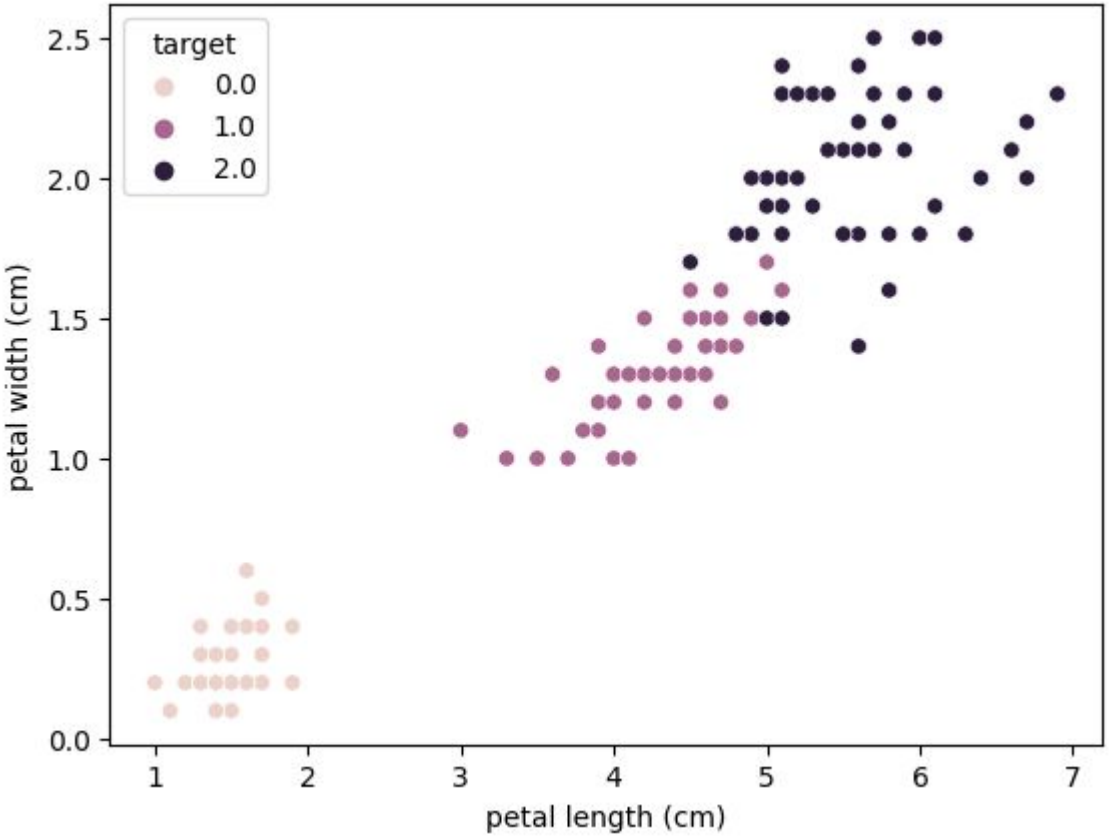


What is Classification

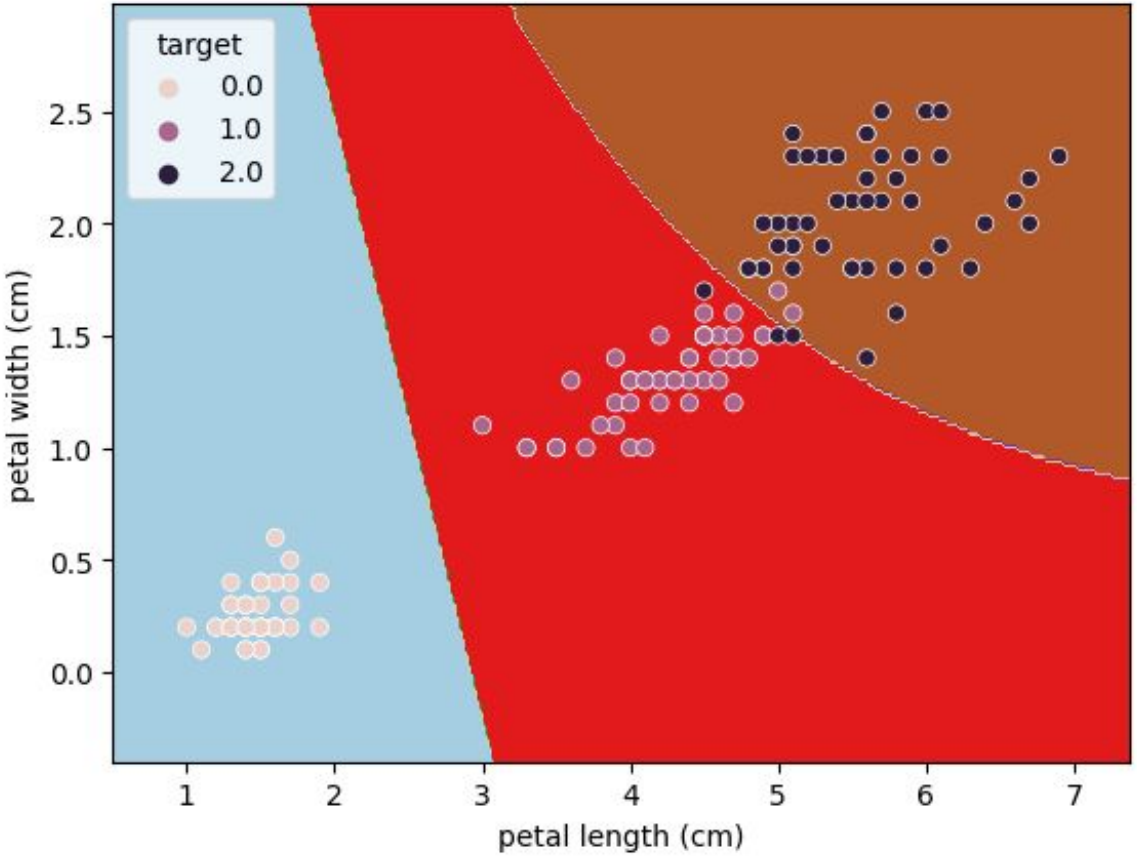
Wikipedia



Complex Data



Using an SVM on Complex Data

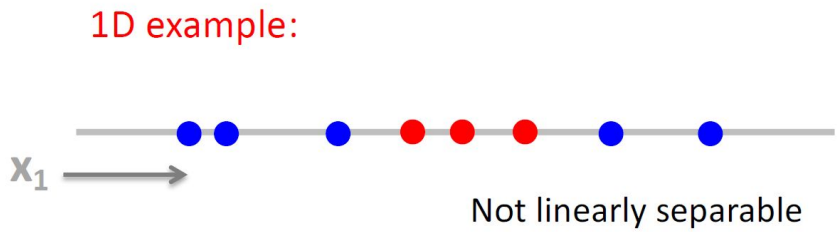


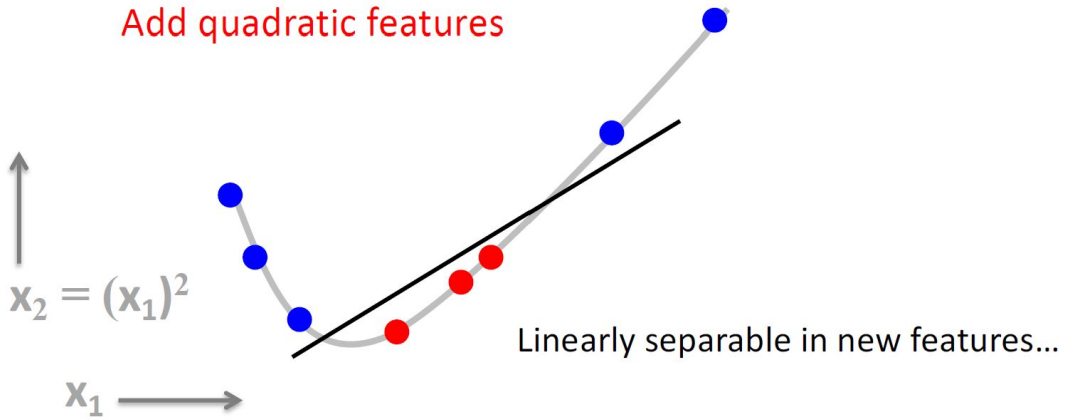
Support Vector Machine Defined

Machine learning algorithm for classification of data

- When used for classification finds a line with max w (margin width) between two classes
- Powerful due ability to transform data into hyperplanes for better classification
- Can be linear, polynomial or other function

Adding features for better classification



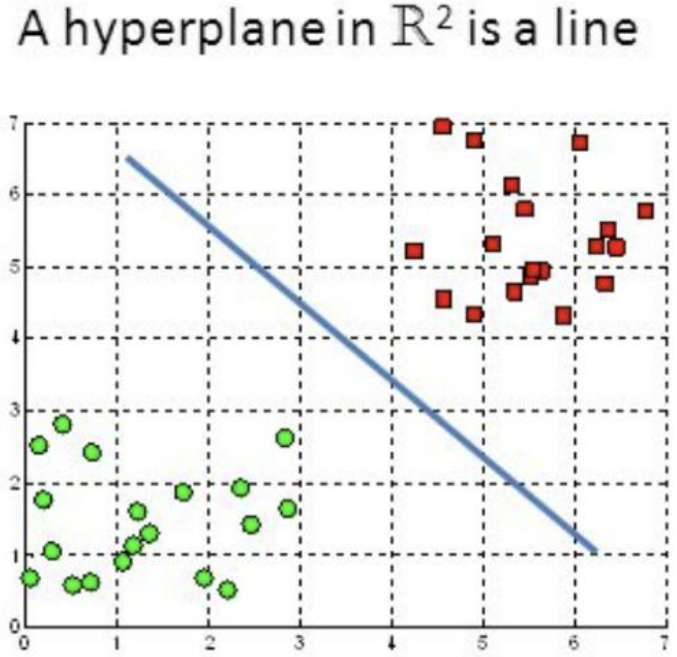


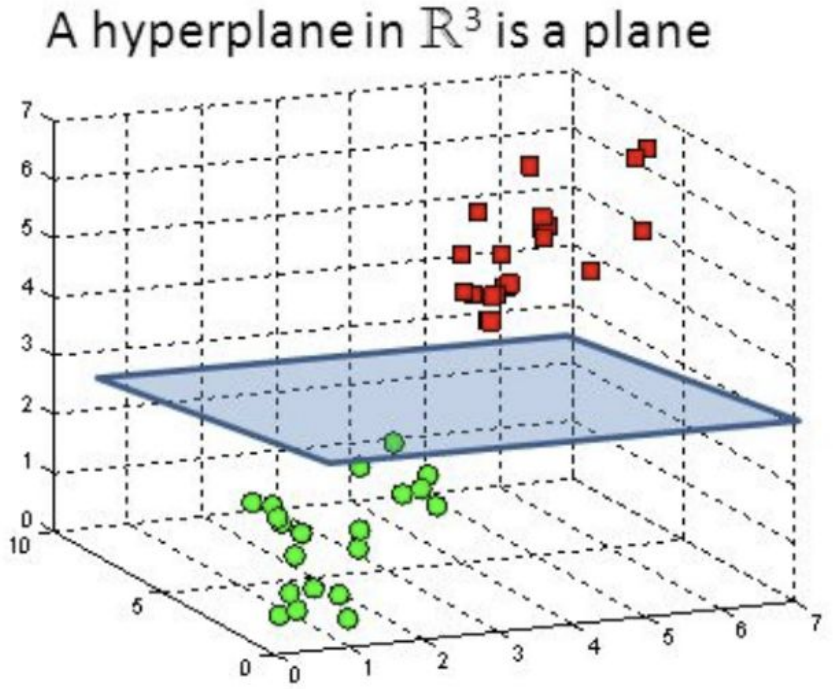
Kernel

Dr. Ihler

Kernel + Plane

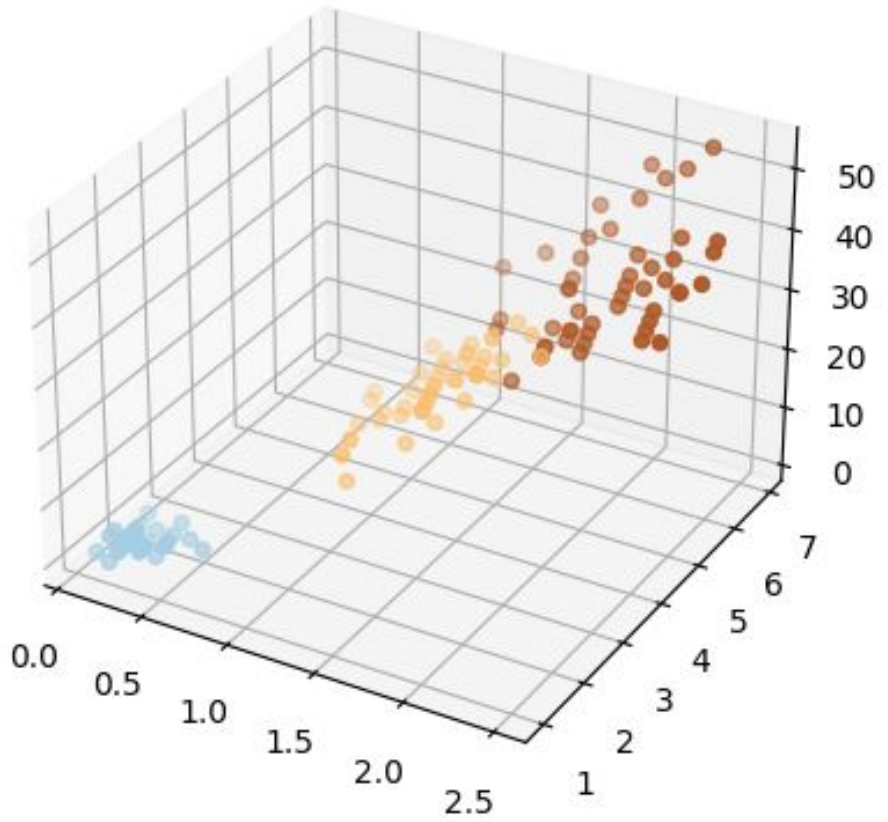
Towardsdatascience.com





Use a kernel to increase dimensionality

Here we square our values



Jupyter Notebooks Time!

Jupyter Notebooks Time!

https://colab.research.google.com