SVMs and Hyperparameters

TM Quest

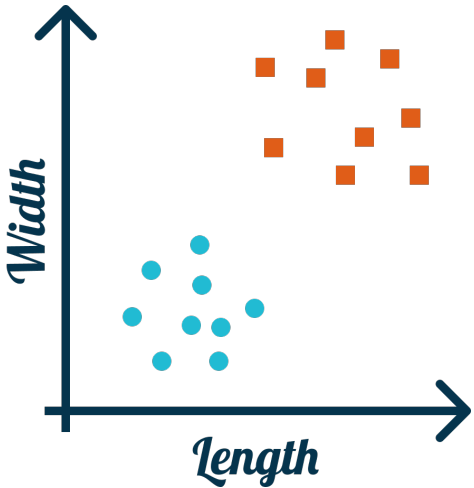
Overview

What Will we Learn in This Module?

- What is a support vector machine (SVM)?
  - How to implement SVM.
  - What are soft and hard margins?
- What are hyperparameters?
  - What do the different hyperparameters do in a SVM?
  - How to do hyperparameter search.

Support Vector Machine

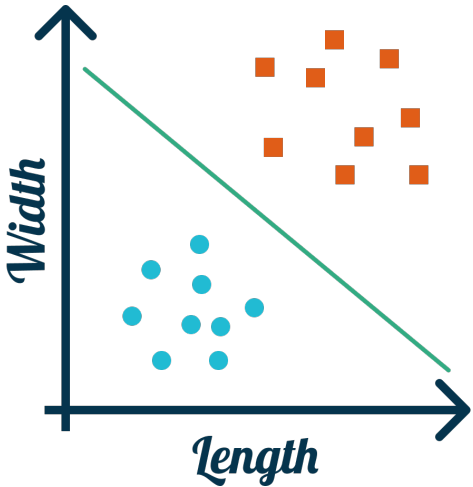
Motivation



Goal

Make a model separating the two classes ■ and ● based on the two features width and length.

Support Vector Machine



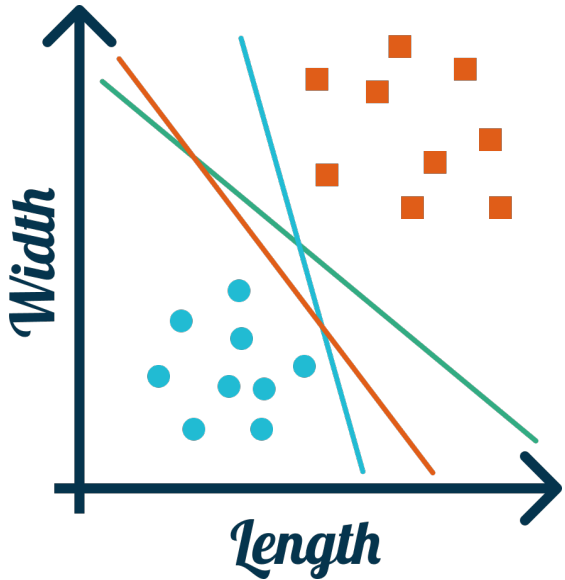
Goal

Make a model separating the two classes ■ and ● based on the two features width and length.

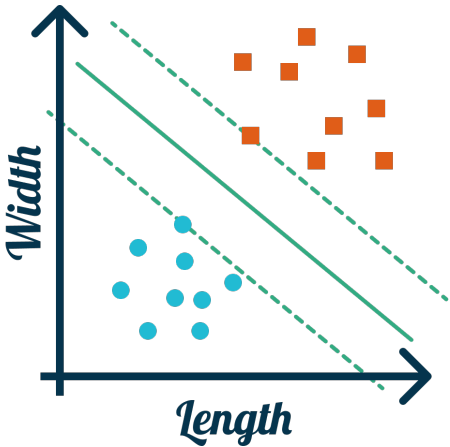
Idea

Separating the classes with a line.

What is the Best Line/hyperplane?



What is the Best Line?



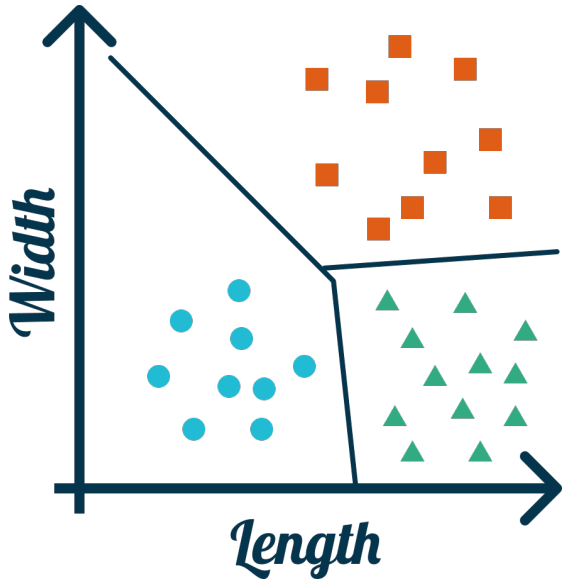
Best Line

The best line is maximizing the ”road” between the two categories.

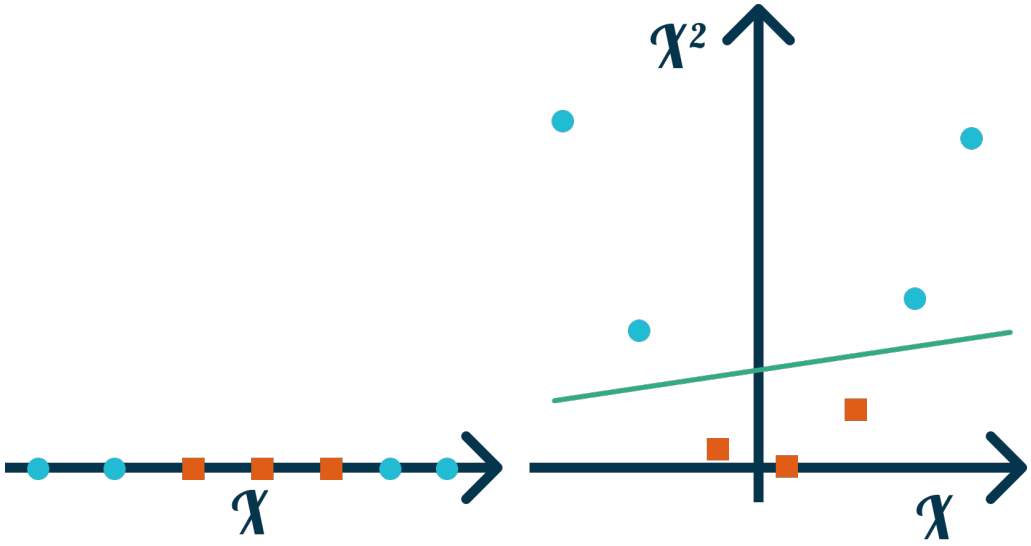
Decision Boundary The line/surface — where we change decisions.

Decision Margin The area within - - used to train the model.

Several Categories

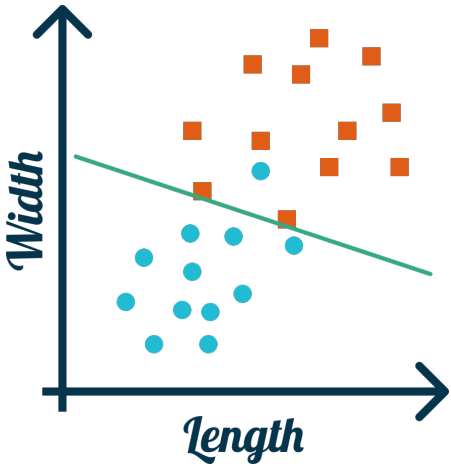


Adding Features



Hyperparameters

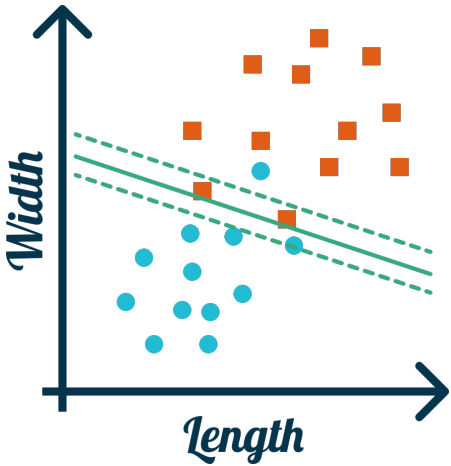
What if there are no separating lines/hyperplanes?



Motivation

- Can't always separate the categories.
- Need to penalize the mistakes.
- We can choose how to penalize mistakes through hyperparameters.

Margins



Soft and hard margins

Hard Margin The model tries to separate the categories without allowing for missclassification within the margins.

Soft Margin The model allows for some missclassifications, but penalizes misclassification.

Hyperparameters

Definition

Hyperparameters are parameters saying how the model should train and are set by the user before the training step.

C Hyperparameter in SVM

Adds a penalty for each missclassification.

- Small C means small penalty, and typically larger margin.
- Big C means large penalty, and typically smaller margin.
- Default value C=1.

Degree Hyperparameter in SVM

Adds polynomial features up to a specified degree. Default value 1.

Hyperparameter Search

Grid Search

Running through the model with different hyperparameters in a grid pattern to find the best hyperparameters.

Example

n \ C	C=1	C=5	C=10
n=1	1, 1	5, 1	10, 1
n=2	1, 2	5, 2	10, 2

- Let C and n be the constants in SVM.
- Can take long time:
  - Cross validation with 5 folds.
  - Three hyperparameters with 5 values each.
  - Then we train $5 * 5 * 5 * 5 = 625$ times!