

EMBEDDED VISION DESIGN 3

SUPERVISED ML

JEROEN VEEN



HAN_ UNIVERSITY
OF APPLIED SCIENCES

QUIZ TIME

- Individual, multiple-choice questions
- Online: <http://www.socrative.com> room **1PTGB6PY**
- Open book quiz, so books and slides can be consulted
- **HAN student number**, so NOT your name, nickname or anything else.
- Quiz starts exactly at class hour and takes 10 minutes.
- Be on time and have your equipment prepared.

CONTENTS

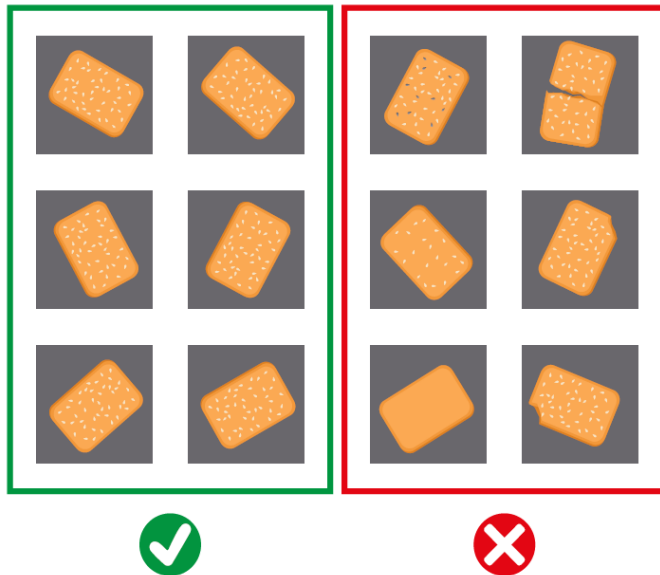
- Classification
- k-nearest neighbors
- Support vector machines
- Decision trees
- Random forests
- Bagging and boosting
- Try it out

INTRODUCTION

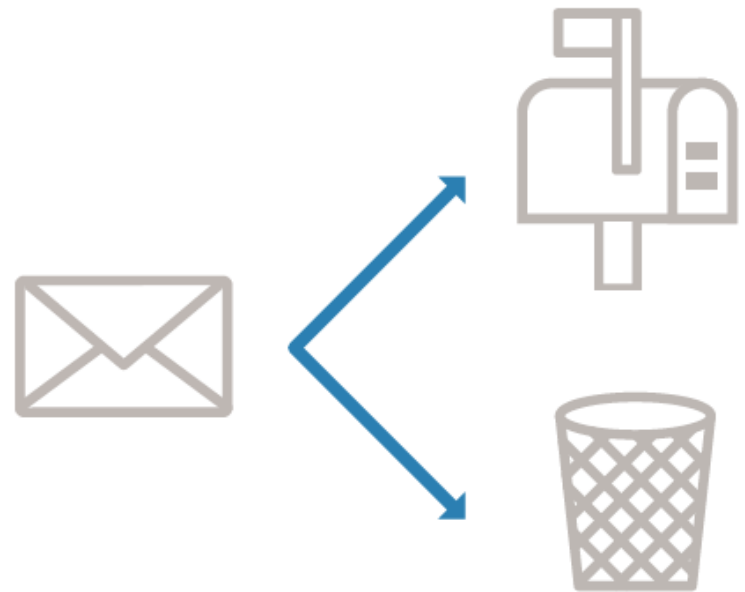
- A supervised learning algorithm takes a known set of input data and known responses to the data (labels) and trains a model to generate reasonable predictions of the labels for new input data.
- Use supervised learning if you have existing data for the labels you are trying to predict.
- All supervised learning techniques are a form of either classification (discrete output) or regression (continuous output).

BINARY CLASSIFICATION

- Sample falls in either of 2 classes



Source: Basler, Artificial Intelligence in Image Processing



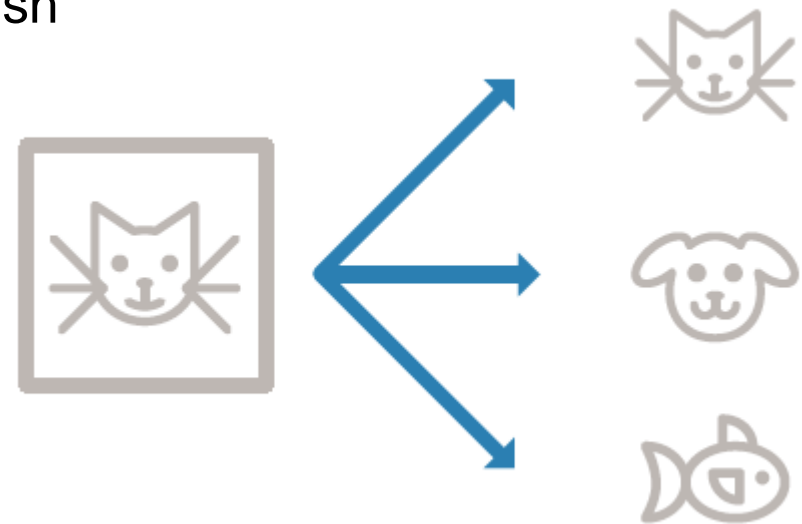
Source: Mathworks, Applying Supervised Learning

MULTI-CLASS CLASSIFICATION

- Sample falls in either of 3 or more classes
- E.g. classify image as cat, dog, fish
- E.g. digit recognition



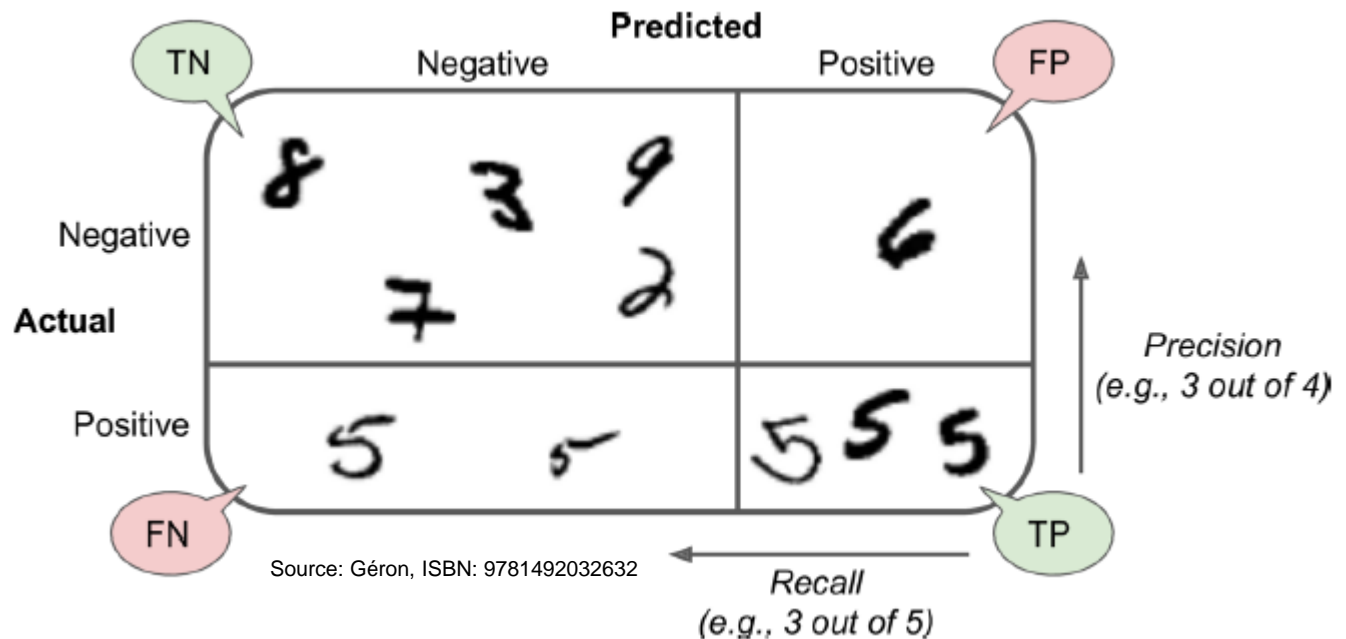
Source: MNIST dataset



Source: Mathworks, Applying Supervised Learning

CLASSIFICATION PERFORMANCE

- Confusion matrix



- More on performance in another lecture

MULTI-LABEL CLASSIFICATION

- Multiple labels may be assigned to each sample
- Generalization of multi-class classification

Three Type of Classification Tasks

YAHOO!
JAPAN

Binary Classification



- Spam
- Not spam

Multiclass Classification



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

Multi-label Classification



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

Source: <https://www.microsoft.com/en-us/research/uploads/prod/2017/12/40250.jpg>

MULTI-OUTPUT CLASSIFICATION

- Disjoint label combinations.
- Both the number of properties and the number of classes per property is greater than 2



K NEAREST NEIGHBOR (KNN)

- <https://youtu.be/AoeEHqVSNOw>

KNN

- The simplest classifier
- Assume feature vectors near each other are similar
- Categorizes objects based on the classes of their nearest neighbors
- No training required
- Intuitive
- Benchmark

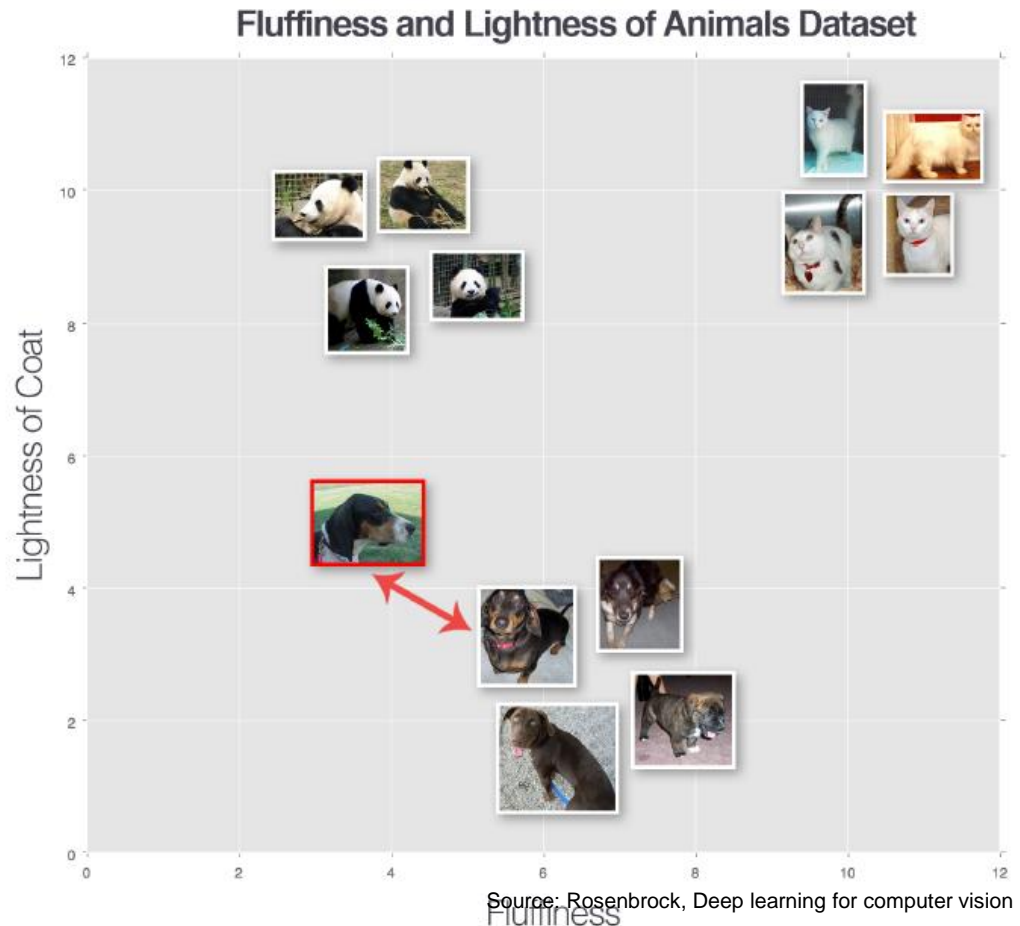


Source: Mathworks, Applying Supervised Learning

“Tell me who your neighbors are, and I’ll tell you who you are”

MAKING PREDICTIONS

- Comparing to every example is very slow
- More suited for low-dimensional feature spaces (which images are not)

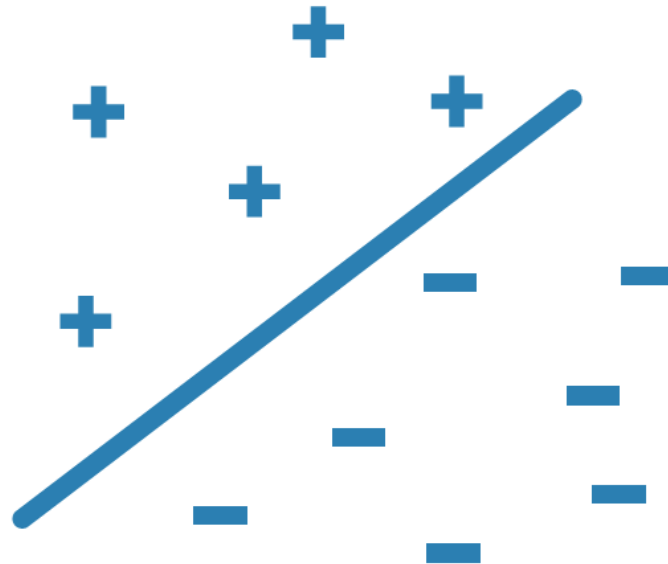


SUPPORT VECTOR MACHINE (SVM)

- Powerful and versatile ML model, capable of performing linear or nonlinear classification, regression, and even outlier detection.
- Simple and easy to interpret.
- Binary classifier, i.e. for data that has exactly two classes.
- For multi-class data, reduction to several binary problems needed.

LINEAR SEPARATION

- Finding the linear decision boundary (hyperplane) that separates all data points of one class from those of the other class.



Source: Mathworks, Applying Supervised Learning

SKITTLE SORTING ON ARDUINO

- Color sensor gives 1 RGB pixel
- SVM color classification

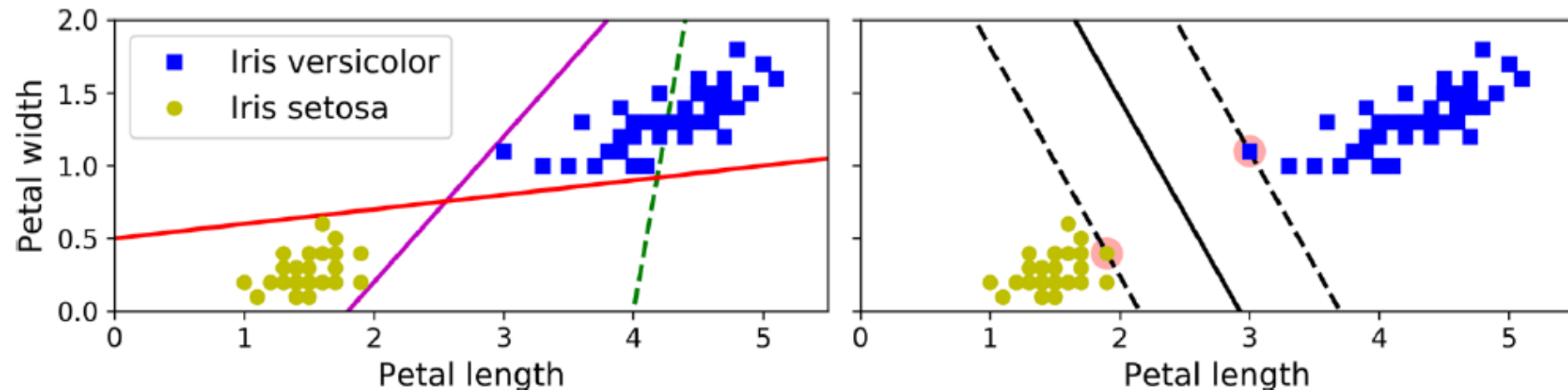
See
<https://eloquentarduino.github.io/2020/03/how-to-train-a-color-classification-machine-learning-classifier-directly-on-your-arduino-board/>



Source:
<https://create.arduino.cc/projecthub/user421848217/how-to-make-color-sorting-machine-8278c9>

LARGE MARGIN CLASSIFICATION

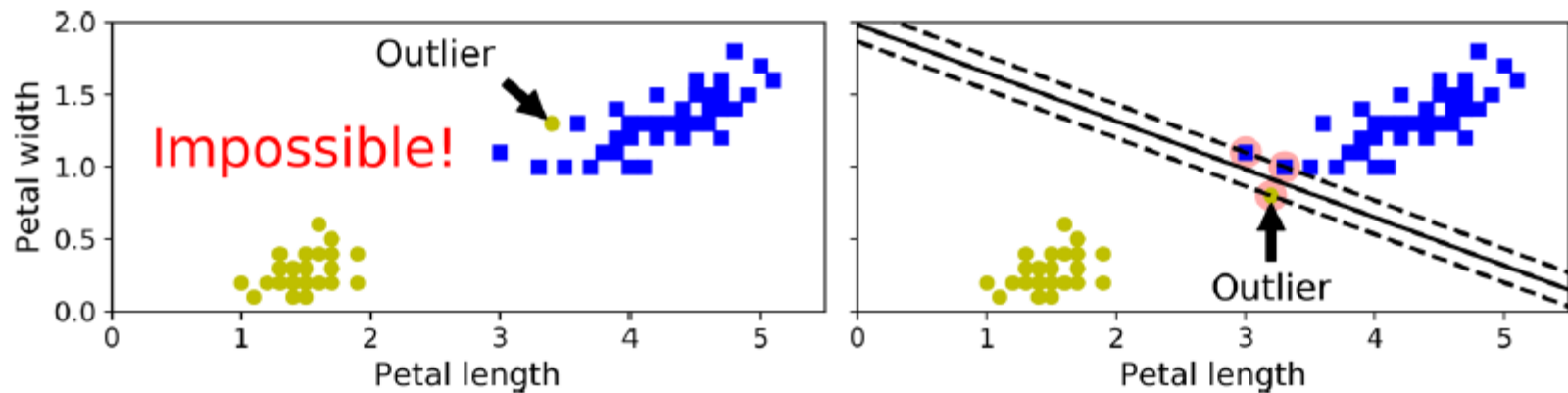
- Decision boundary of an SVM classifier not only separates the two classes but also stays as far away from the closest training instances as possible.



Source: Géron, ISBN: 9781492032632

SOFT MARGIN CLASSIFICATION

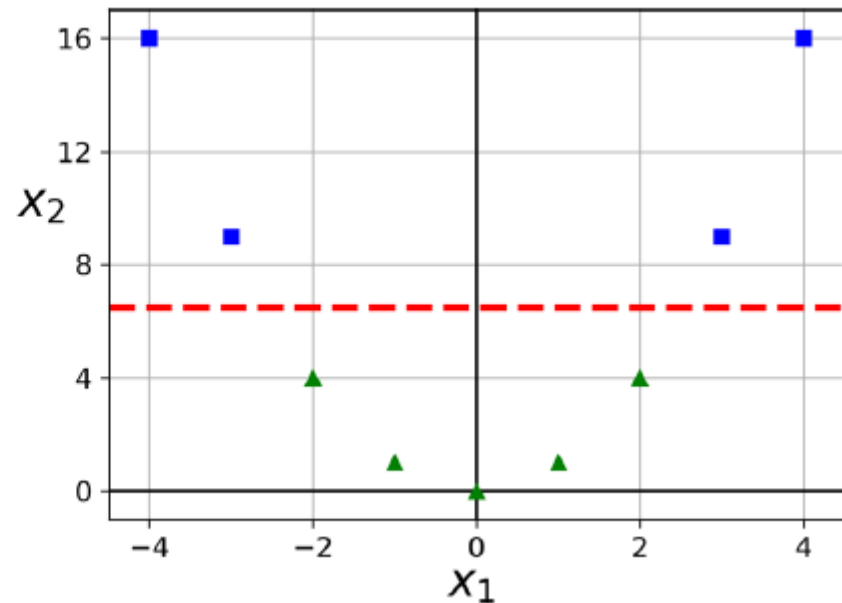
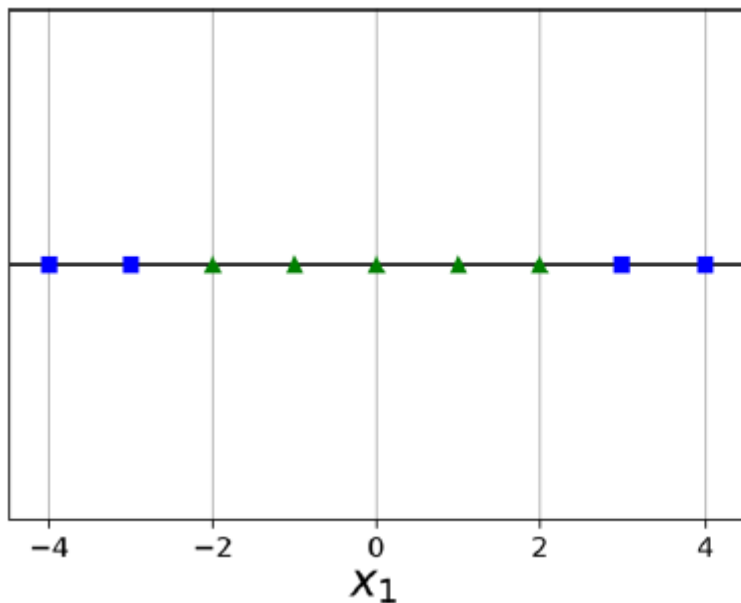
- More flexible model



Source: Géron, ISBN: 9781492032632

- Balance between keeping the street as large as possible and limiting the margin violations

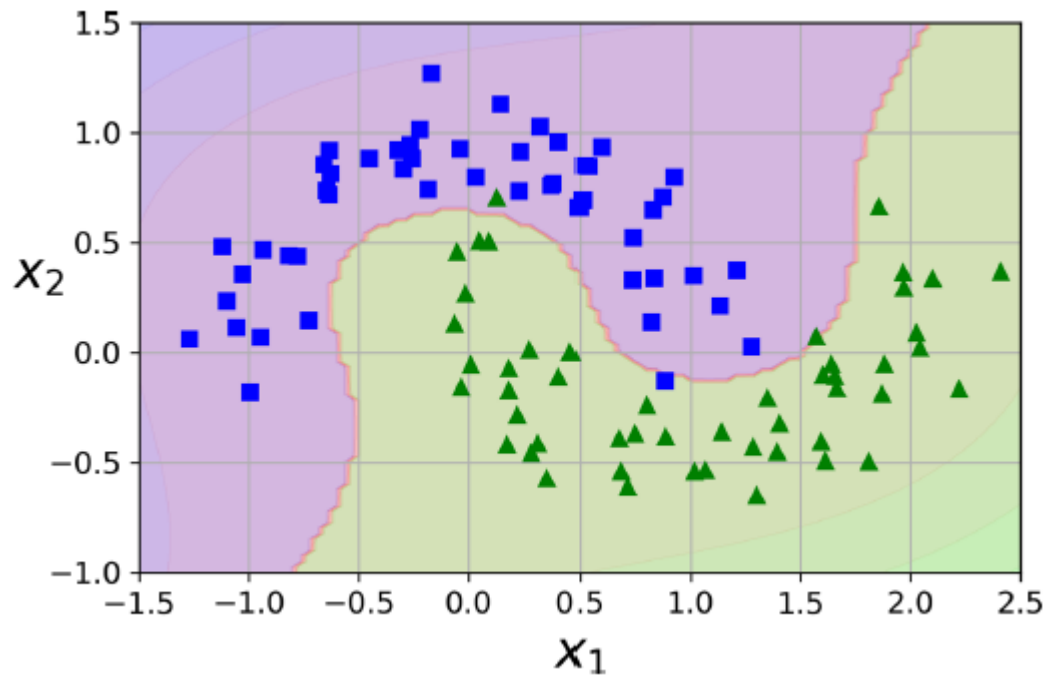
ADDING FEATURES TO MAKE A DATASET LINEARLY SEPARABLE



Source: Géron, ISBN: 9781492032632

NONLINEAR SVM CLASSIFICATION

- E.g. polynomial kernel



Source: Géron, ISBN: 9781492032632

THE KERNEL TRICK

- Represent data through a set of pairwise similarity comparisons between the original data and references
- A similarity function measures how much each sample resembles a particular landmark.

Kernel Definition

- A function that takes as its inputs vectors in the original space and returns the dot product of the vectors in the feature space is called a *kernel function*
- More formally, if we have data $\mathbf{x}, \mathbf{z} \in X$ and a map $\phi: X \rightarrow \mathbb{R}^N$ then

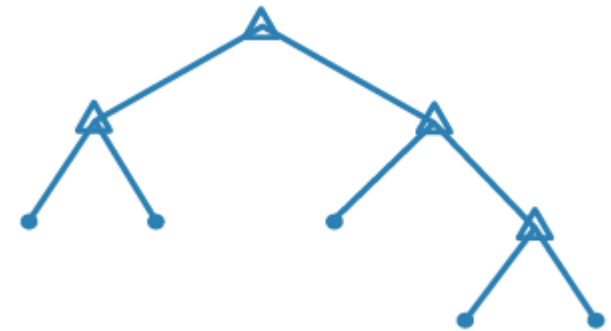
$$k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$$

is a kernel function

Source: <https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f>

DECISION TREES

- Predict responses to data by following the decisions in the tree from the down to a leaf node.
- Easy to interpret
- Fast to fit
- Minimize memory usage

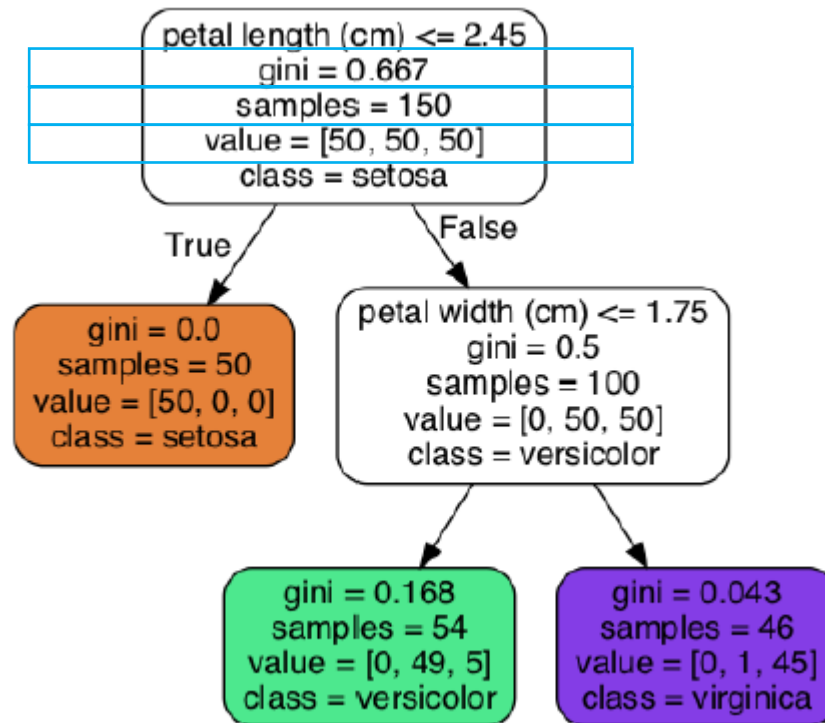


Source: Mathworks, Applying Supervised Learning

See: <https://youtu.be/tNa99PG8hR8>

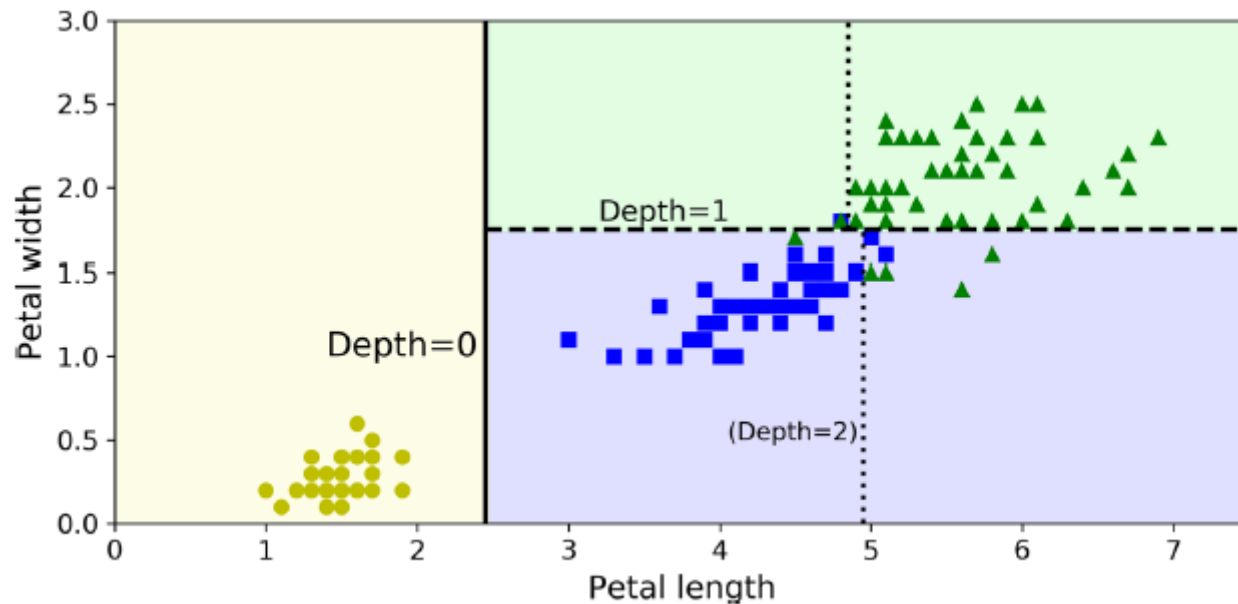
MAKING PREDICTIONS

- E.g. Iris Decision Tree



DECISION TREE BOUNDARIES

- white box models



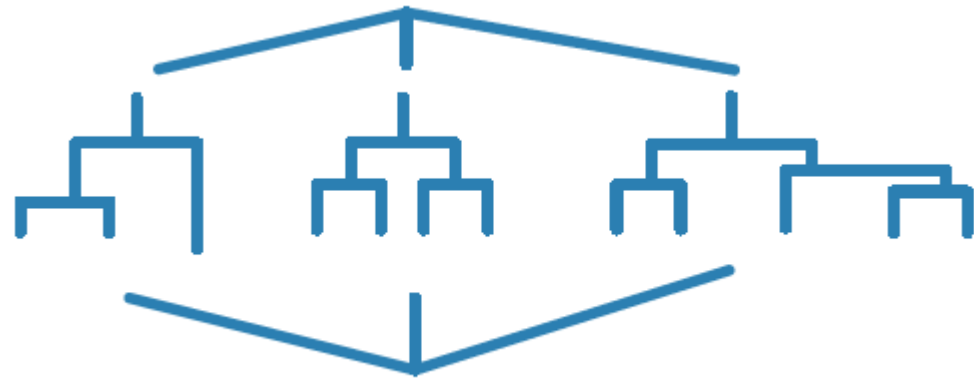
Source: Géron, ISBN: 9781492032632

CHOOSING THE BEST CLASSIFIER?

- There is no “best”
- No free lunch theorem:
Averaged over all possible types of data distributions, all classifiers perform the same.... (see [http://en.wikipedia.org/wiki/ No_free_lunch_theorem](http://en.wikipedia.org/wiki/No_free_lunch_theorem))
- However, given certain data and set of constraints, there may be a best classifier. Thus, when faced with real data it's a good idea to try-out many
- In a future lesson we will discuss more supervised learning methods

ENSEMBLE LEARNING

- Wisdom of the crowd
- Group of predictors
- Random forest

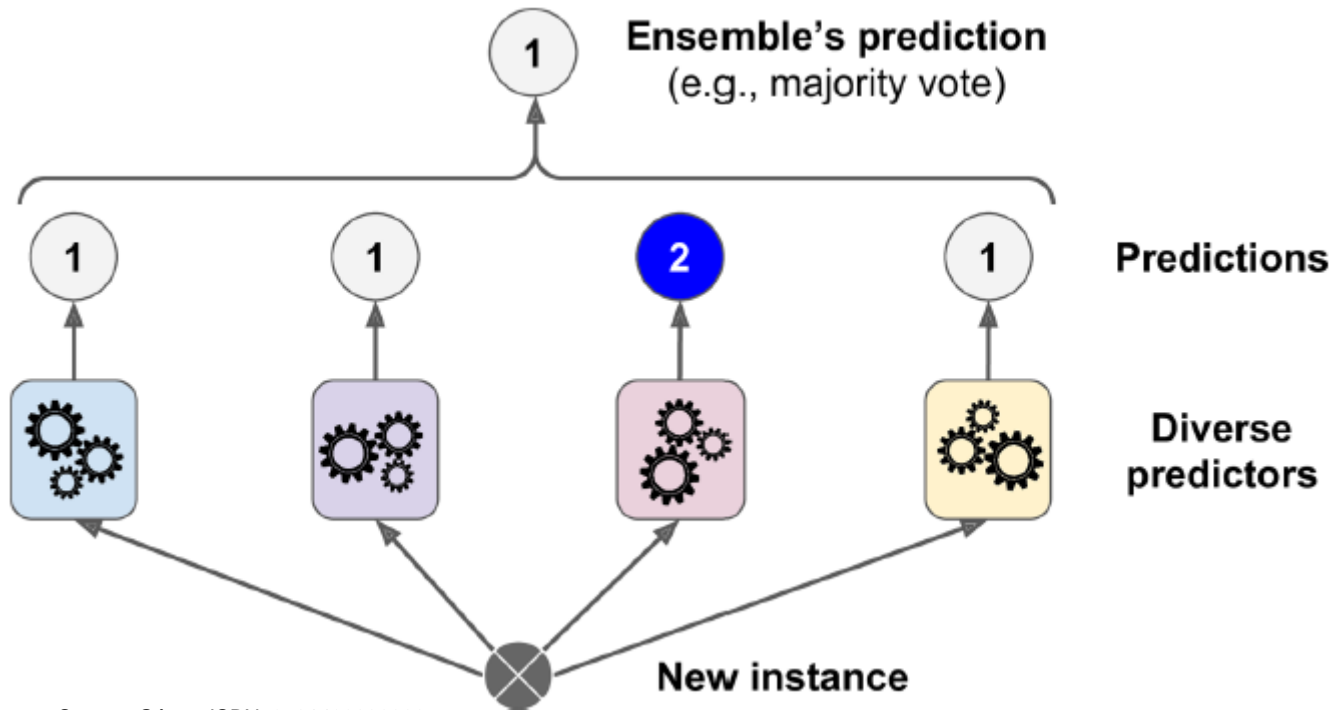


Source: Mathworks, Applying Supervised Learning

- Several “weaker” decision trees are combined into a “stronger” ensemble

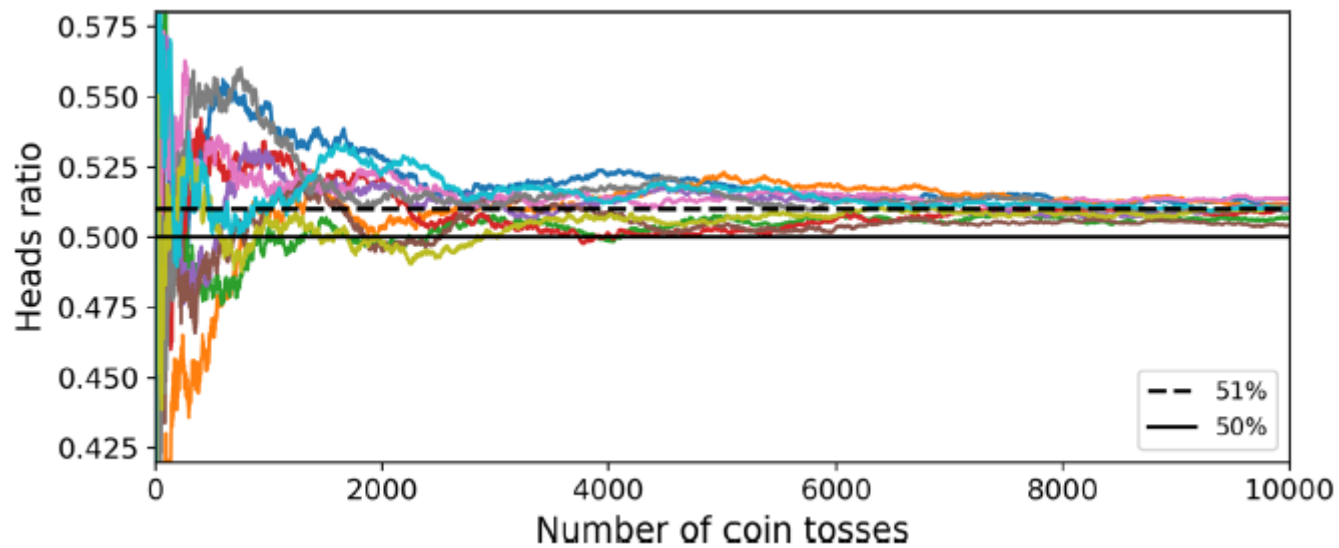
HARD VOTING CLASSIFIER

- Majority-vote can be strong given sufficient diversity



Source: Géron, ISBN: 9781492032632

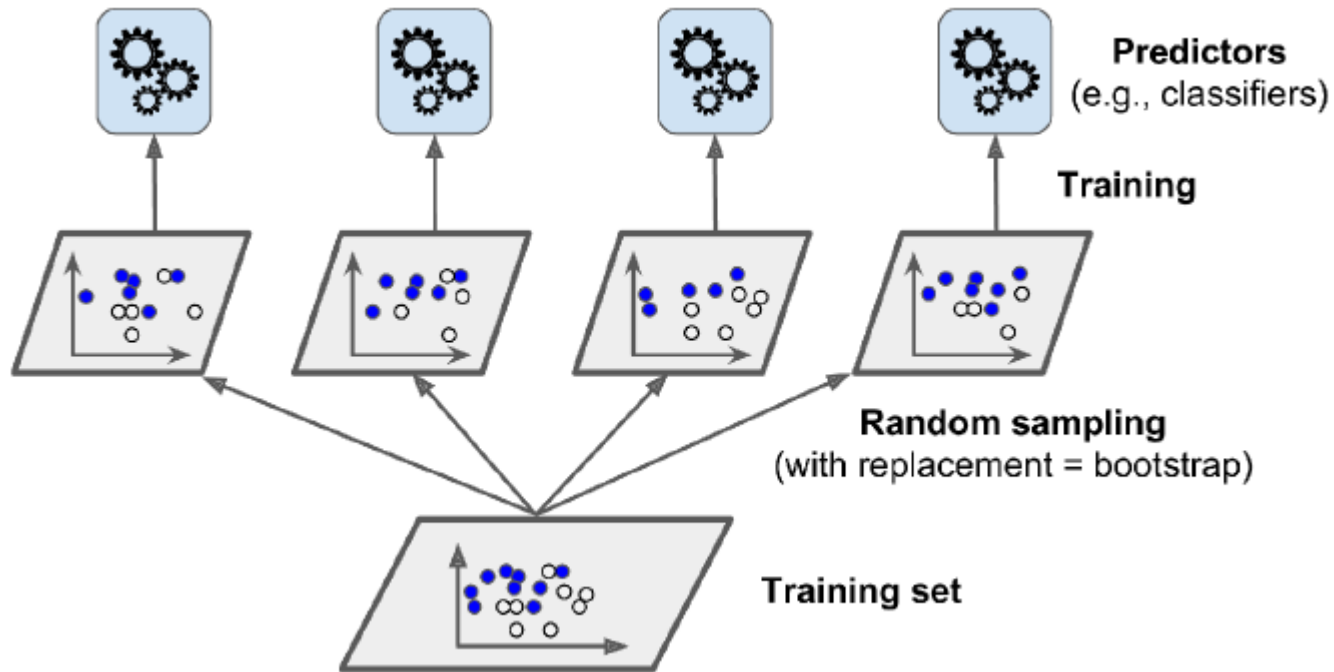
LAW OF LARGE NUMBERS



Source: G ron, ISBN: 9781492032632

BAGGING PREDICTORS

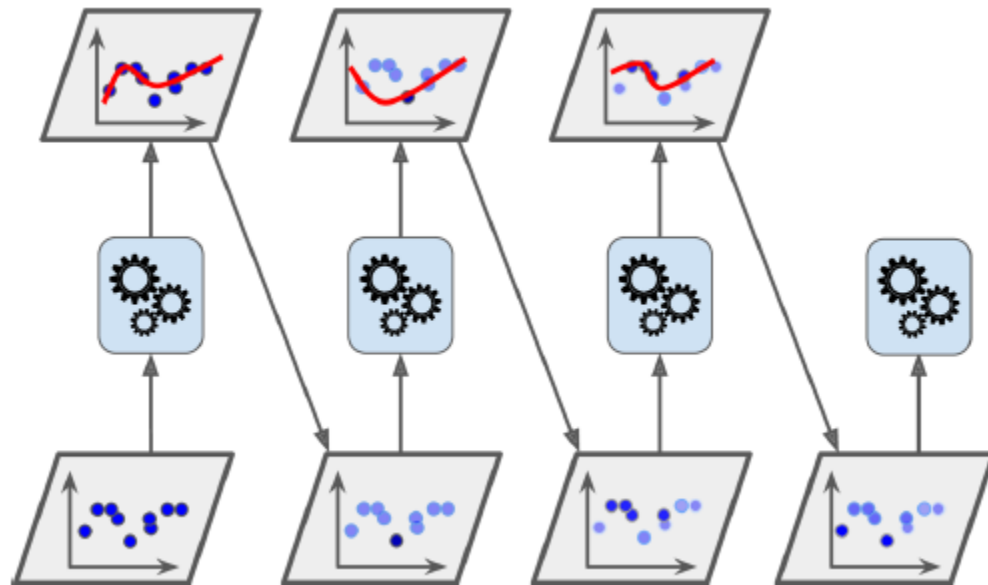
- Trees are trained independently on bootstrapped data



Source: Géron, ISBN: 9781492032632

BOOSTING

- Sequentially adding predictors to an ensemble, each one correcting its predecessor



Source: Géron, ISBN: 9781492032632

TRY IT OUT

- A First Application Classifying Iris Species with KNN from Introduction to Machine Learning with Python (11p)

<https://scikit-learn.org/stable/install.html>



PITFALLS

- Assuming that generalization is possible
- **Inductive bias**
the set of assumptions a learner uses to predict results given inputs it has not yet encountered.
- Induce an approximate function
- Class imbalance

NEXT TIME

- Hands-on:
 - Sci-kit learn
 - Training a classifier
 - Tuning and thinking about performance
- Theory:
 - Regression
 - How do classification and regression relate?