

Spatial Ecology

Introduction to Machine Learning

Antonio Fonseca

Agenda

1) Intro to machine learning

- Defining learning
- Supervised vs Unsupervised learning
- The framework of learning algorithms

2) Example of Supervised learning

- Support Vector Machine (SVM)
- Optimization of SVM
- Extension of SVM to regression (SVR)

What is machine learning?

Machine learning is the process of identifying patterns in data.

Two kinds of machine learning

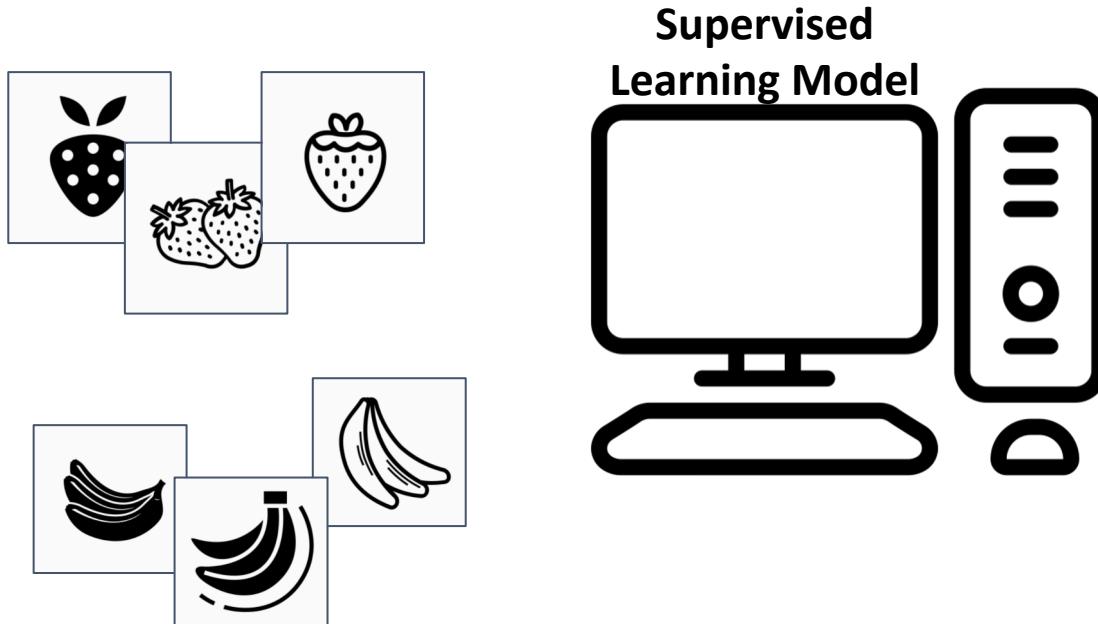
Supervised learning

- Have a bunch of labelled data,
want to label new data

Two kinds of machine learning

Supervised learning

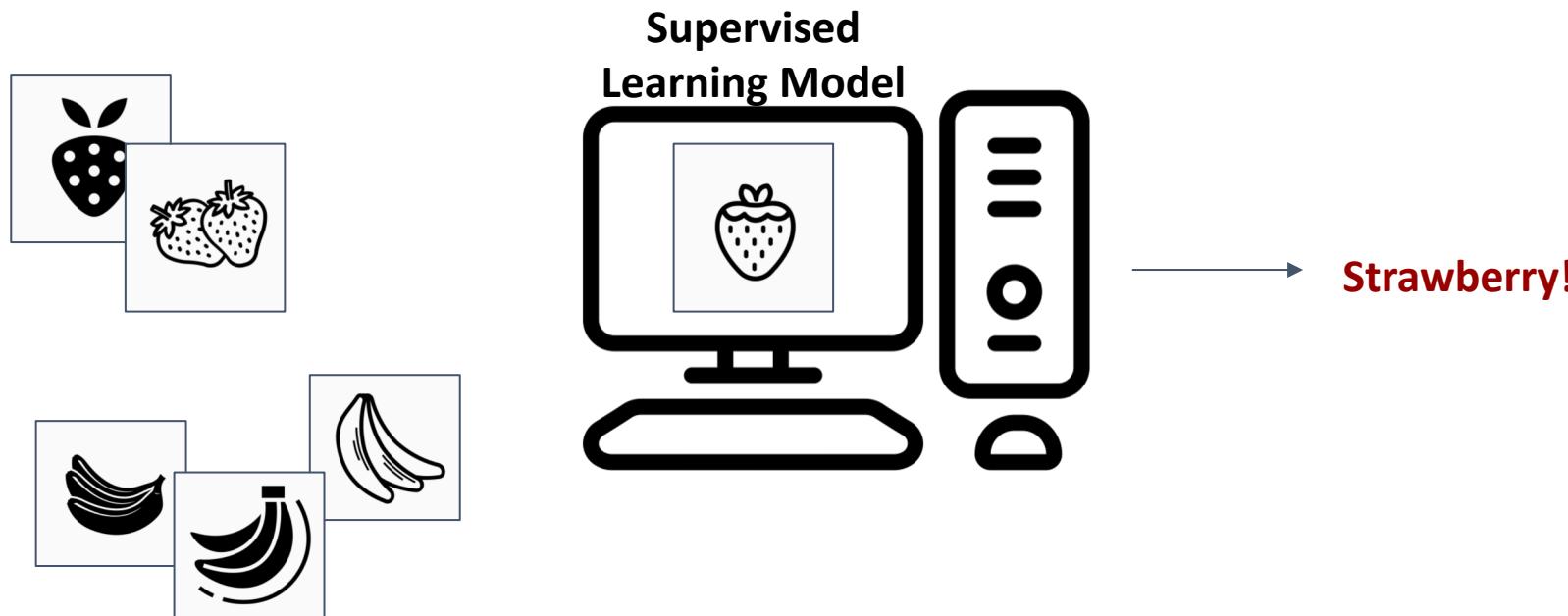
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

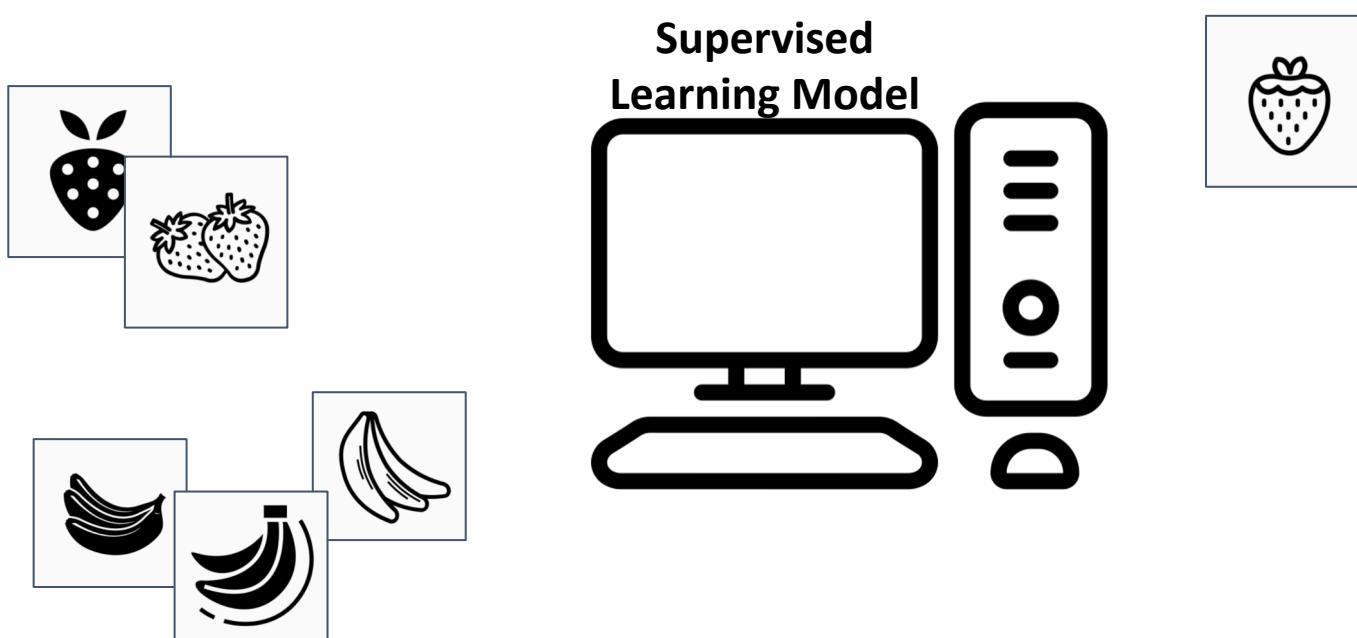
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

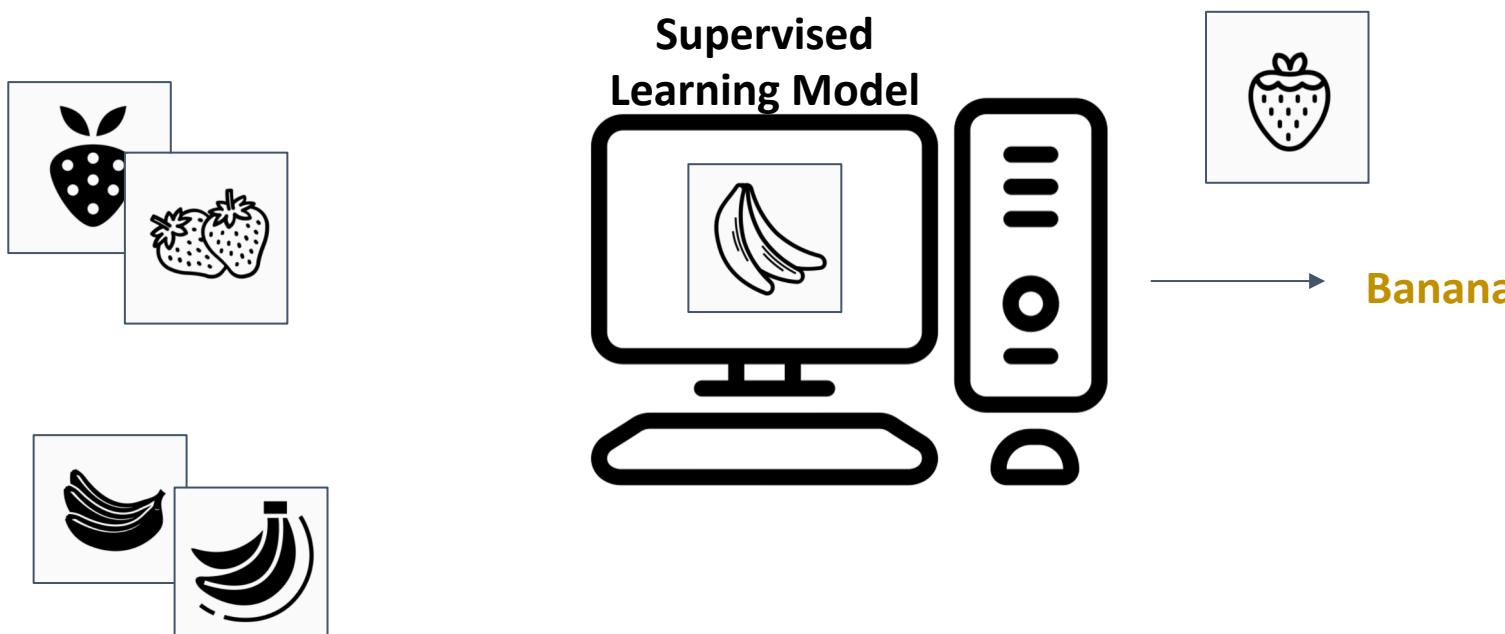
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

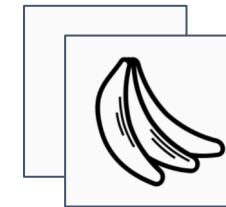
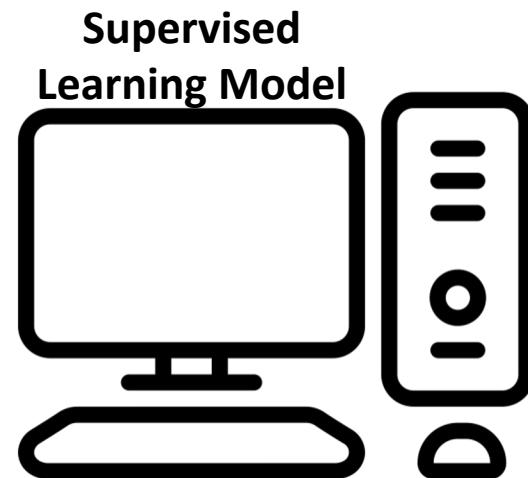
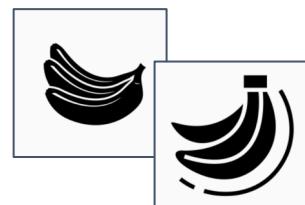
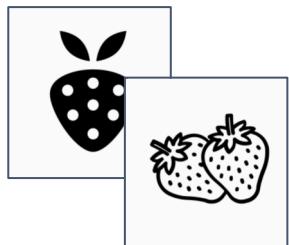
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

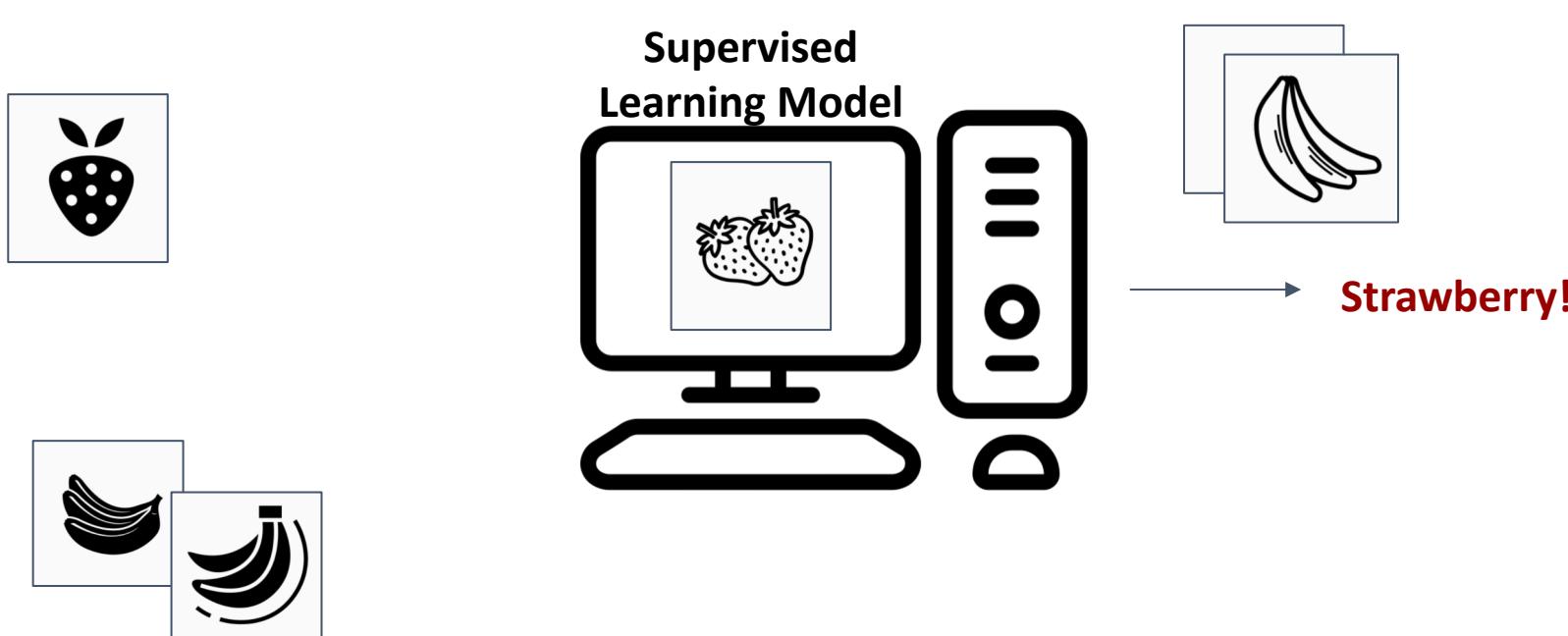
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

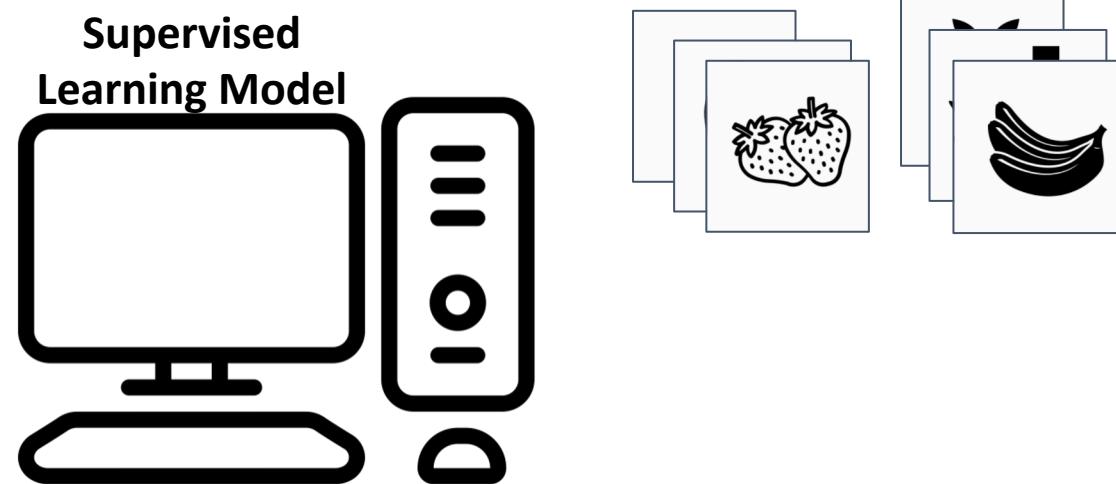
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

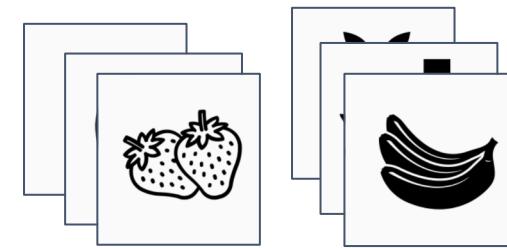
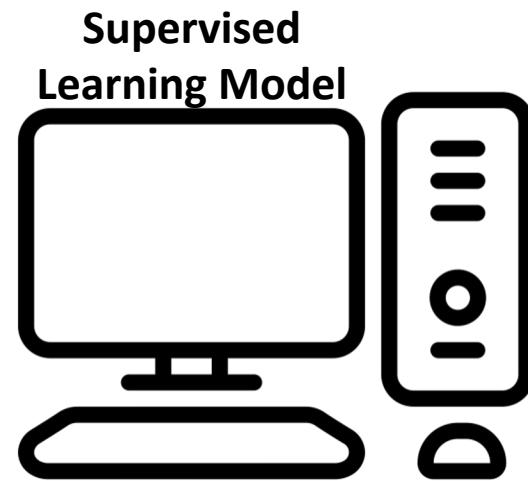
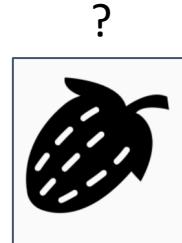
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

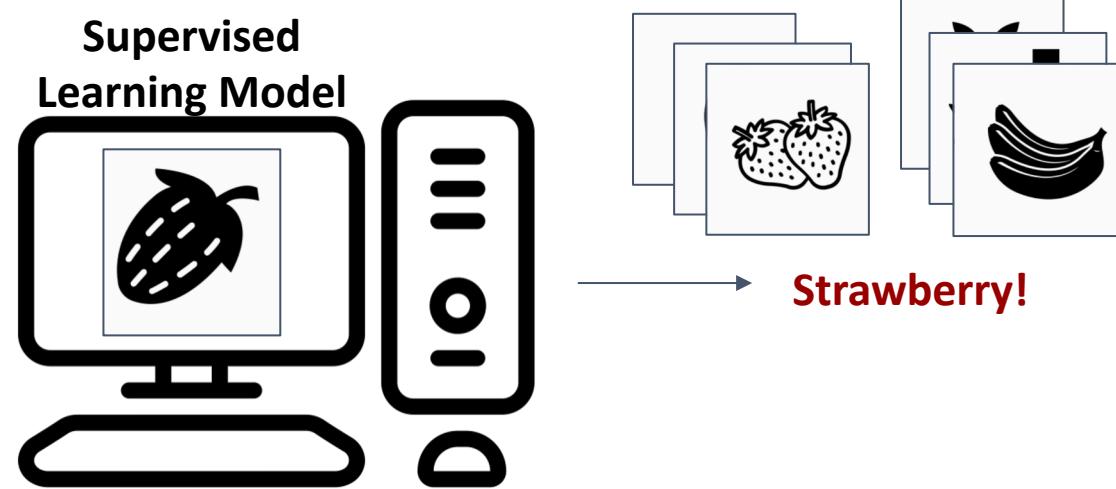
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

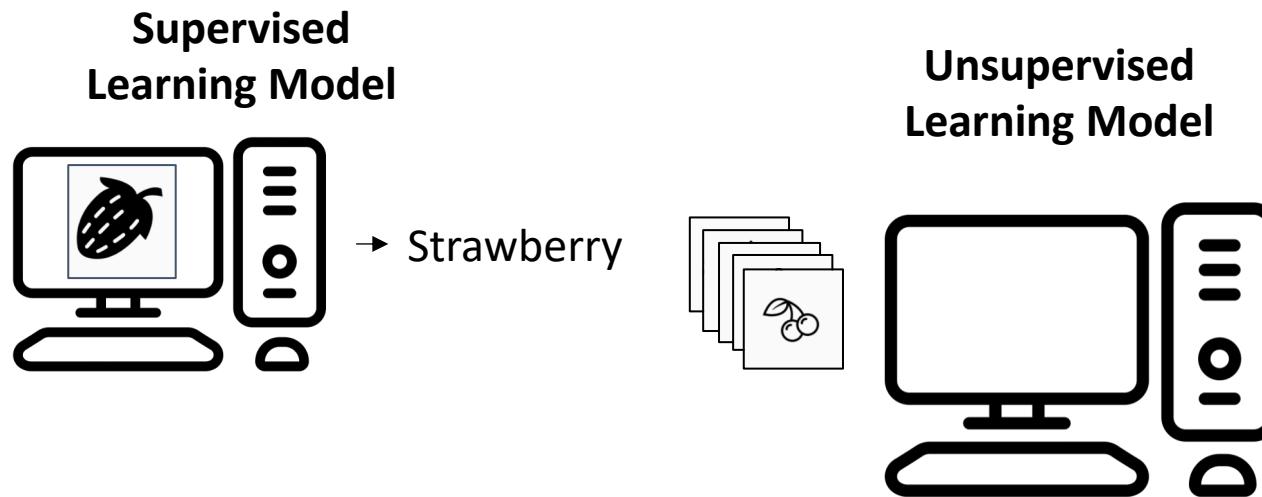
- Have a bunch of labelled data,
want to label new data



Two kinds of machine learning

Supervised learning

- Have a bunch of labelled data, want to label new data



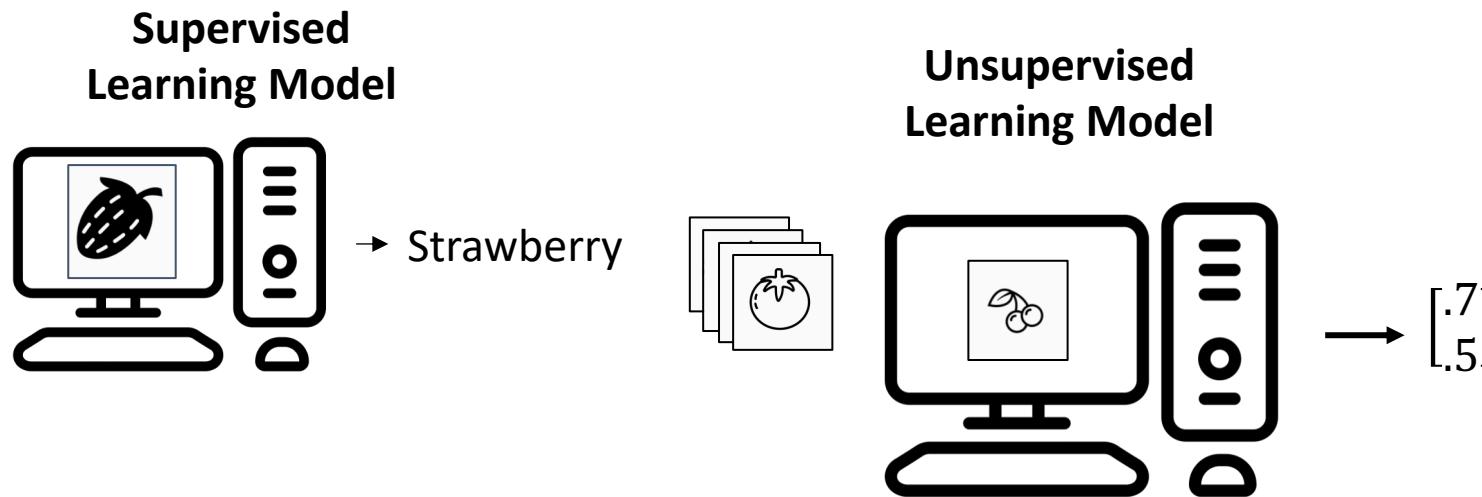
Unsupervised learning

- Have a bunch of unlabeled data, want to organize it

Two kinds of machine learning

Supervised learning

- Have a bunch of labelled data, want to label new data



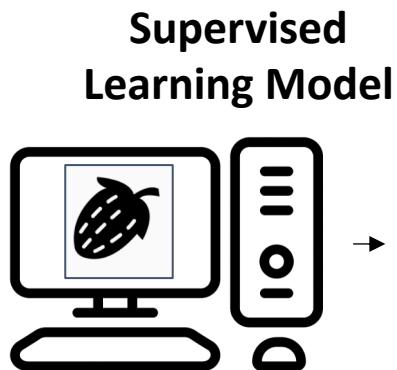
Unsupervised learning

- Have a bunch of unlabeled data, want to organize it

Two kinds of machine learning

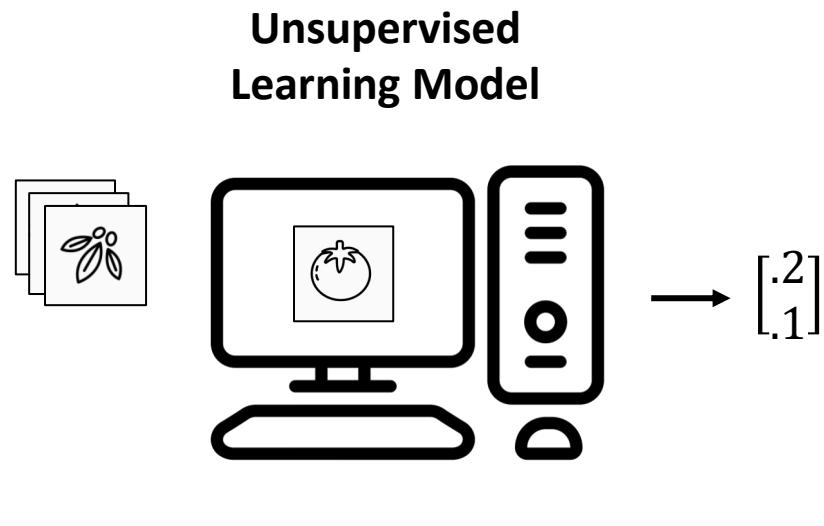
Supervised learning

- Have a bunch of labelled data, want to label new data



Unsupervised learning

- Have a bunch of unlabeled data, want to organize it

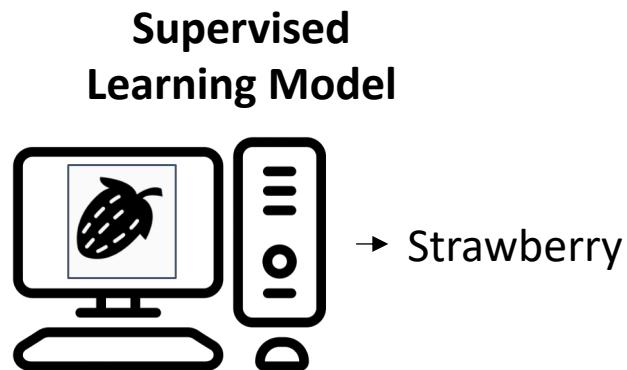


$$\begin{bmatrix} \text{apple} \\ \text{orange} \end{bmatrix} \begin{bmatrix} .7 \\ .5 \end{bmatrix}$$

Two kinds of machine learning

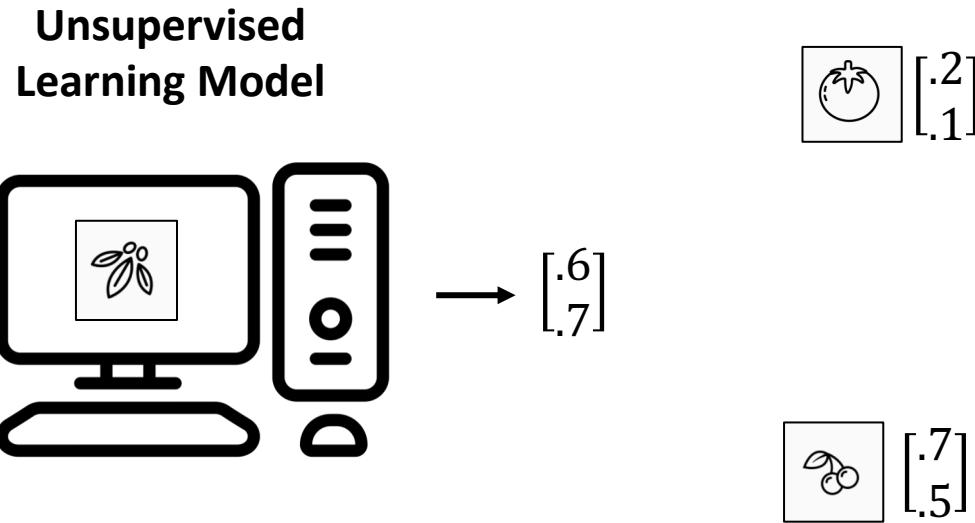
Supervised learning

- Have a bunch of labelled data, want to label new data



Unsupervised learning

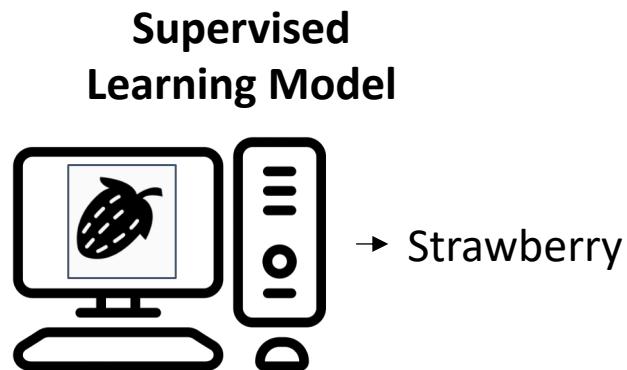
- Have a bunch of unlabeled data, want to organize it



Two kinds of machine learning

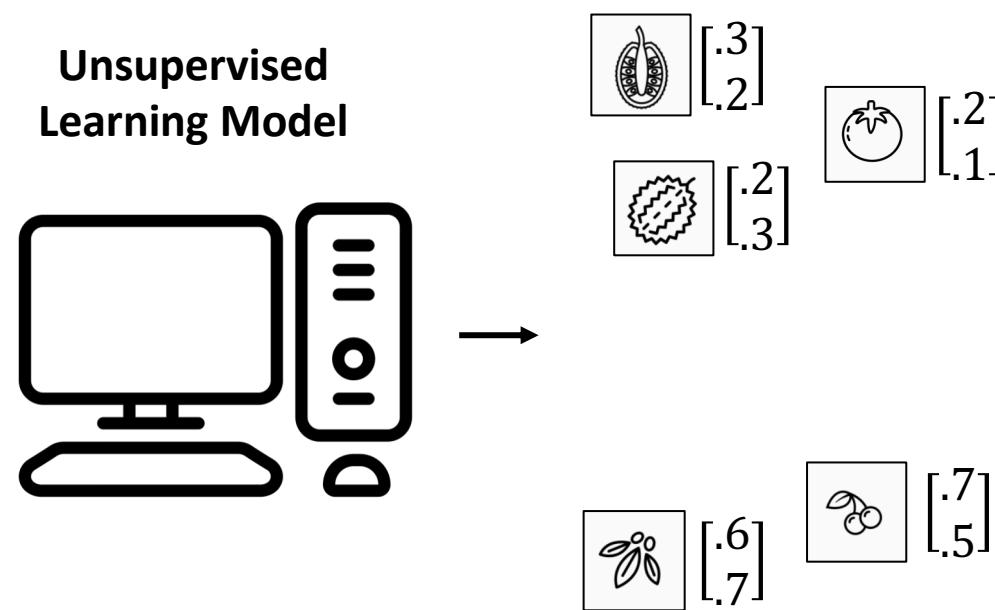
Supervised learning

- Have a bunch of labelled data, want to label new data



Unsupervised learning

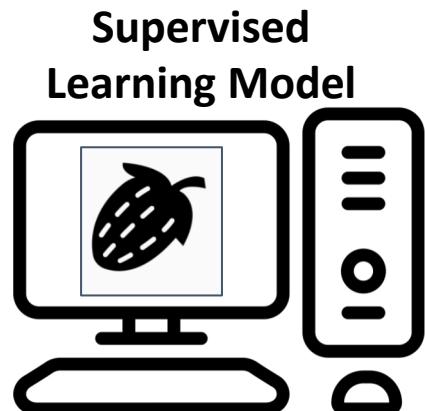
- Have a bunch of unlabeled data, want to organize it



Two kinds of machine learning

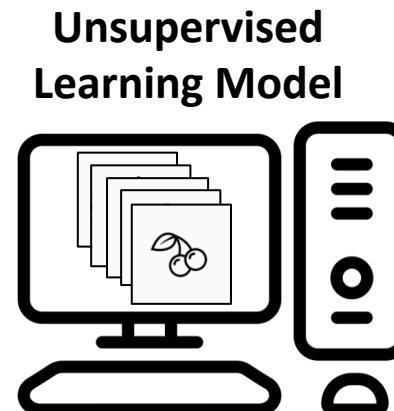
Supervised learning

- Have a bunch of labelled data, want to label new data
- Learn a function $f(X) \rightarrow Y$ where all values of Y are known for some samples of X



Unsupervised learning

- Have a bunch of unlabeled data, want to organize it
- Learn an embedding $f(X) \rightarrow Y, X \in \mathbb{R}^n, Y \in \mathbb{R}^m, n \gg m$
- Lower dimensional, easier to interpret (e.g. as clusters)

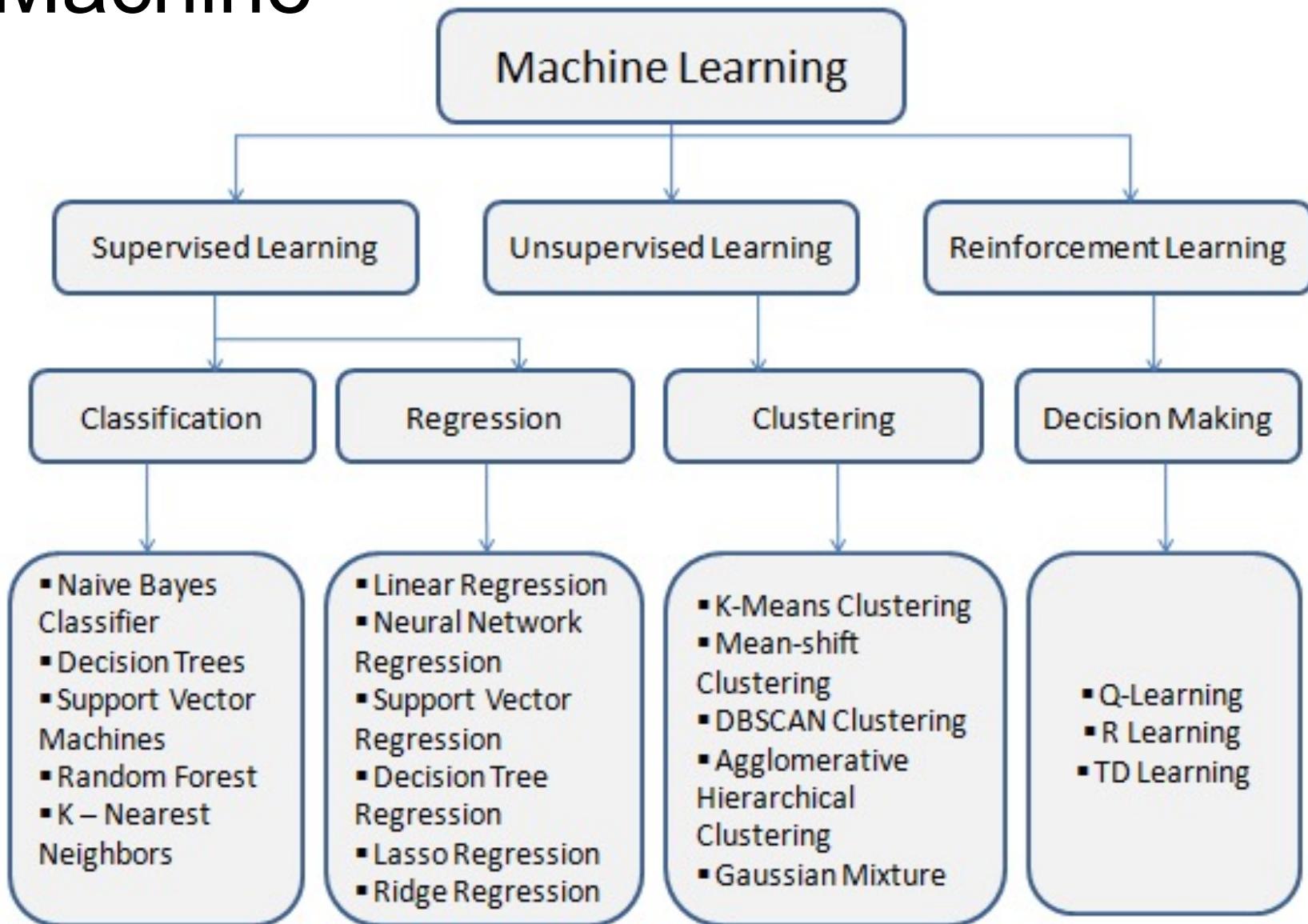


Learning algorithms

“A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P** , if its performance at tasks in **T** , as measured by **P** , improves with experience **E**.”

Tasks (T)	Performance (P)	Experience (E)
Transcription		
Machine Translation	Accuracy rate	Supervised Learning
Classification		
Anomaly detection		
Synthesis and sampling		Unsupervised Learning
:		
Regression	Adjusted R ² RMSE/MSE/MAE	Reinforcement Learning

Types of Machine Learning



Putting these frameworks in perspective

■ "Pure" Reinforcement Learning (cherry)

- ▶ The machine predicts a scalar reward given once in a while.
- ▶ **A few bits for some samples**

■ Supervised Learning (icing)

- ▶ The machine predicts a category or a few numbers for each input
- ▶ Predicting human-supplied data
- ▶ **10→10,000 bits per sample**

■ Unsupervised/Predictive Learning (cake)

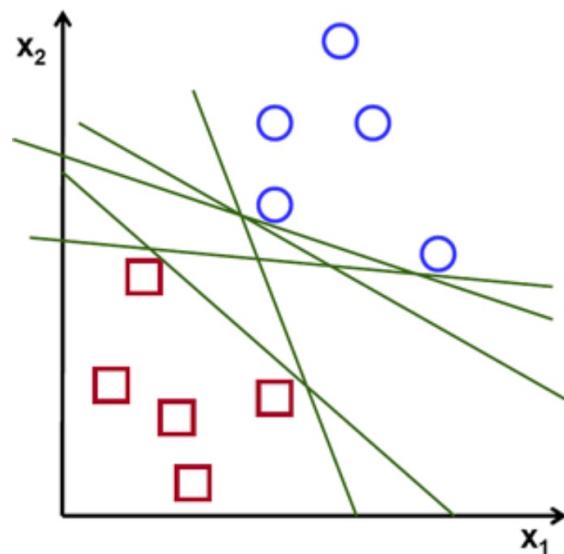
- ▶ The machine predicts any part of its input for any observed part.
- ▶ Predicts future frames in videos
- ▶ **Millions of bits per sample**

■ (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)



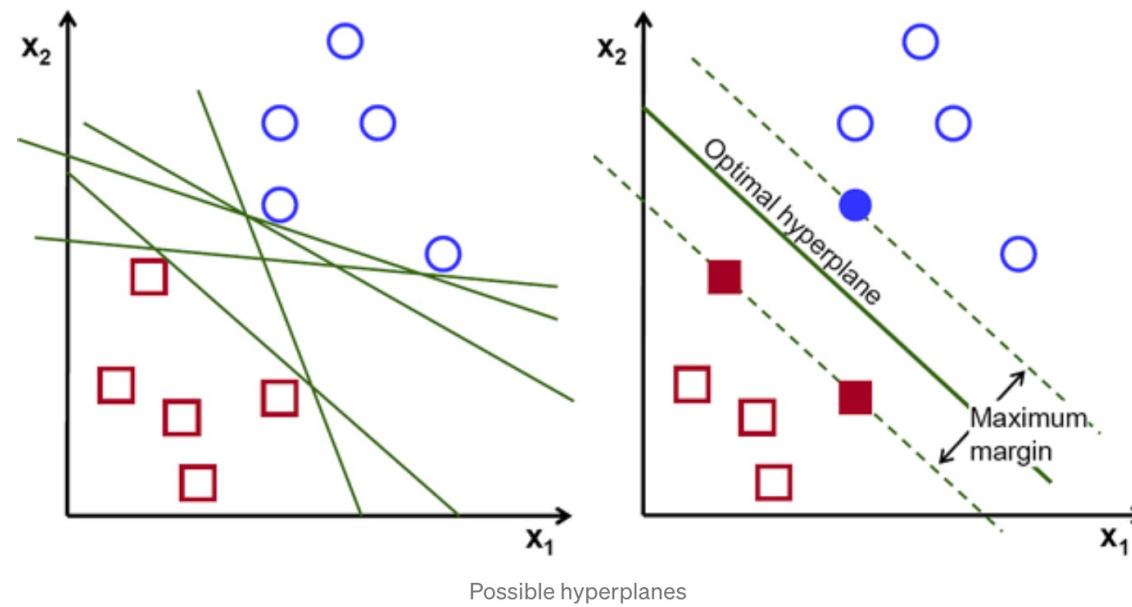
Support Vector Machine

Find a hyperplane in an N-dimensional space that distinctly classifies the data points.



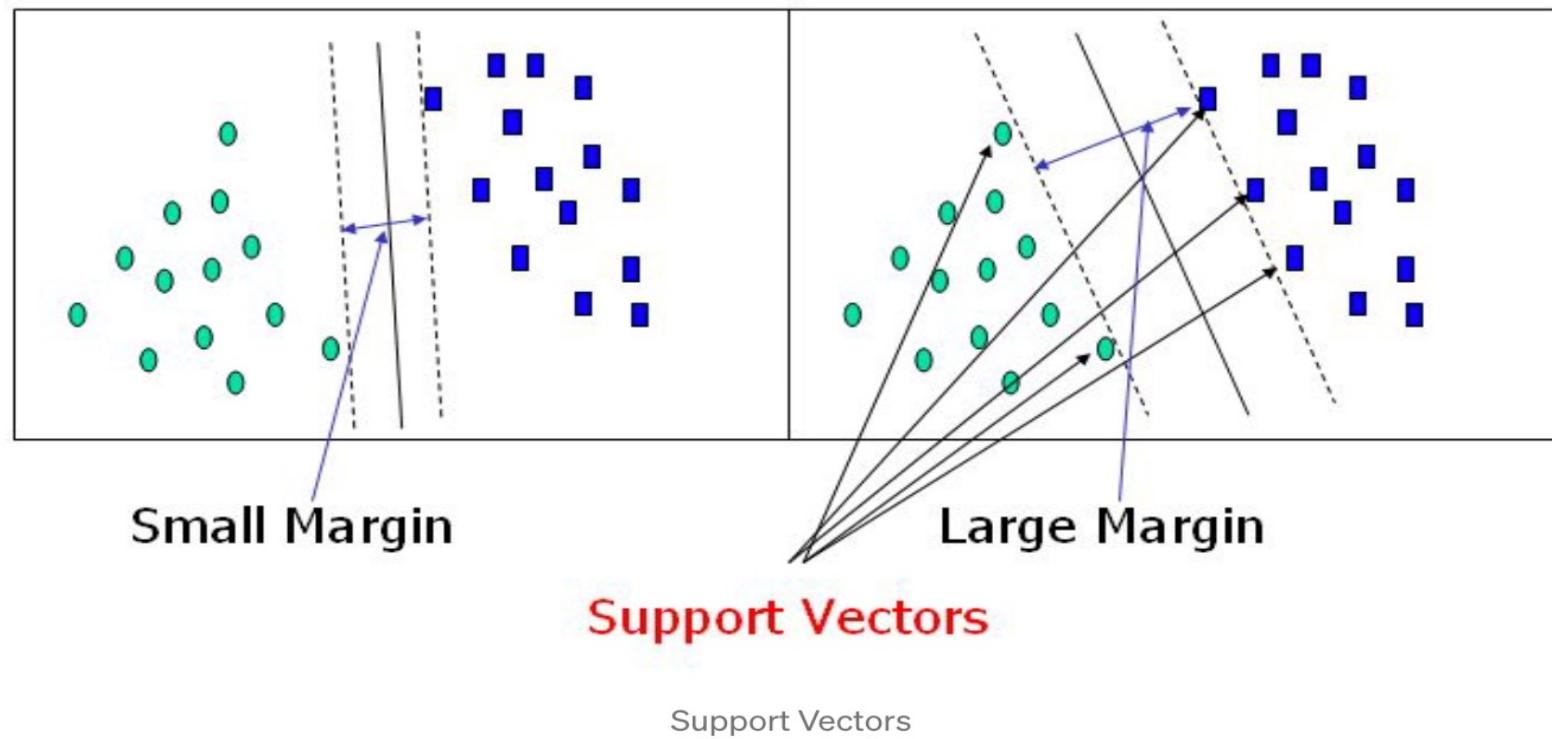
Support Vector Machine

Find a hyperplane in an N-dimensional space that distinctly classifies the data points.



Support Vector Machine

Maximize the margin of the classifier.



SVM Optimization

Hinge loss function

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases}$$

Loss function for the SVM

$$\min_w \lambda \|w\|^2 + \sum_{i=1}^n (1 - y_i \langle x_i, w \rangle)_+$$

Gradients

$$\frac{\delta}{\delta w_k} \lambda \|w\|^2 = 2\lambda w_k$$

$$\frac{\delta}{\delta w_k} (1 - y_i \langle x_i, w \rangle)_+ = \begin{cases} 0, & \text{if } y_i \langle x_i, w \rangle \geq 1 \\ -y_i x_{ik}, & \text{else} \end{cases}$$

Updating the weights:

No misclassification

$$w = w - \alpha \cdot (2\lambda w)$$

Misclassification

$$w = w + \alpha \cdot (y_i \cdot x_i - 2\lambda w)$$

Support Vector Machine for Regression

- The best fit line is the hyperplane that has the maximum number of points.
- The fit time complexity of SVR is more than quadratic with the number of samples
- SVR scales poorly with number of samples (eg, >10k samples)
 - For large datasets, **Linear SVR** or **SGD Regressor**