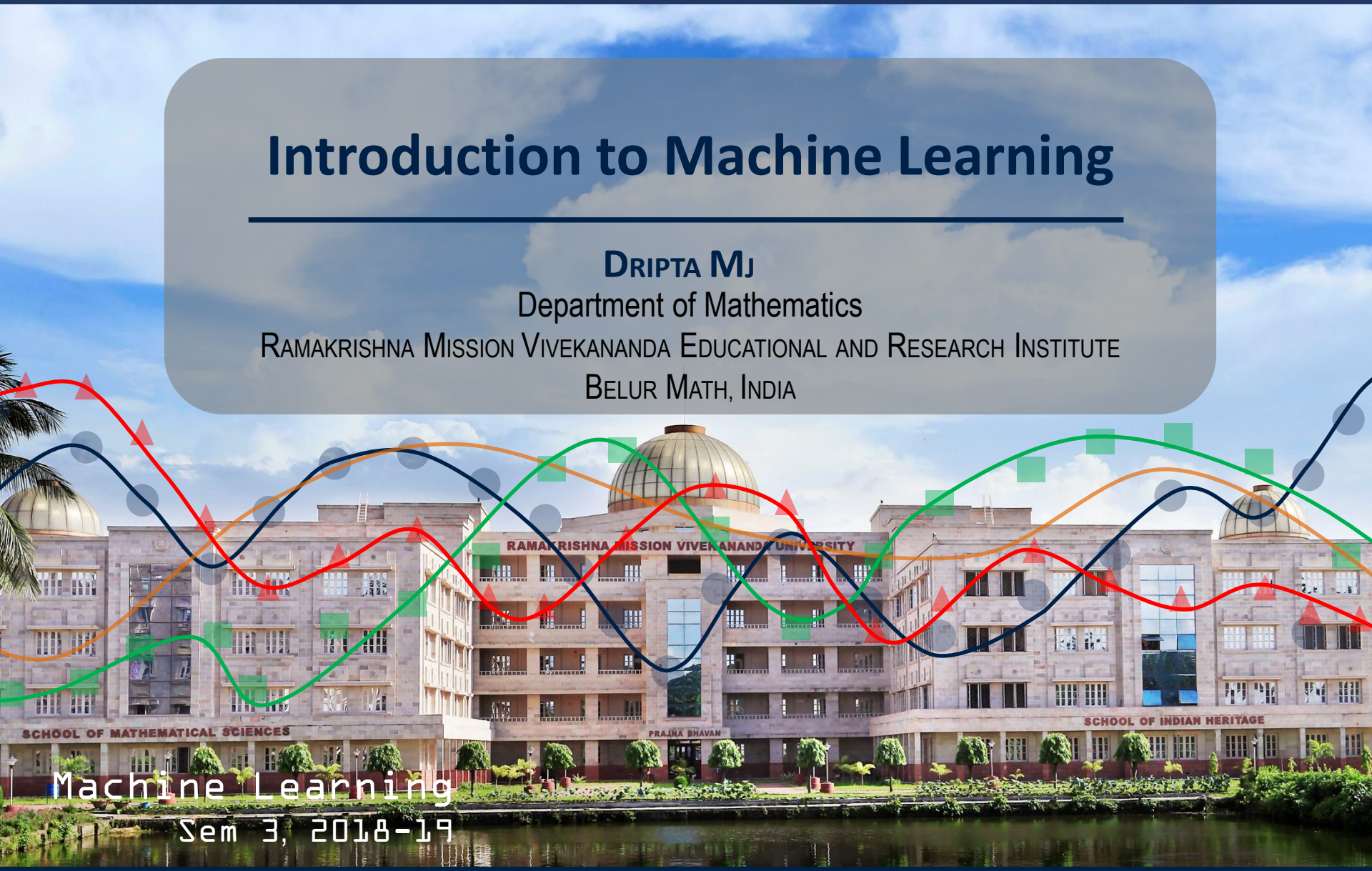
Introduction to Machine Learning

Dripta Mj
Department of Mathematics
Ramakrishna Mission Vivekananda Educational and Research Institute
Belur Math, India



Machine Learning
Sem 3, 2018-19

Why?

The world is data rich!

Astronomy

Social Networks

Healthcare

Banking

Genomics

Weather measurements

Dogs and Cats …..





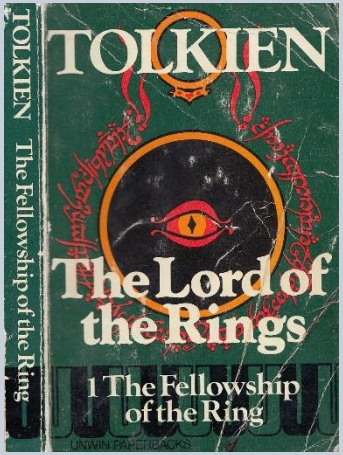




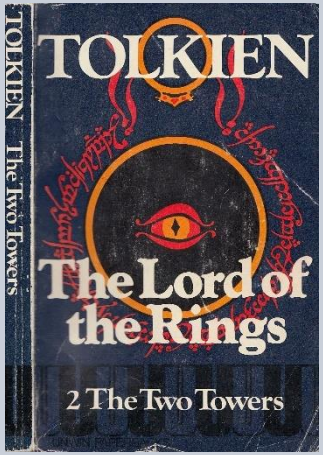

?

?

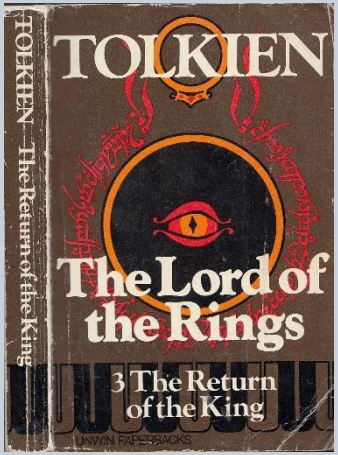
Product recommendation

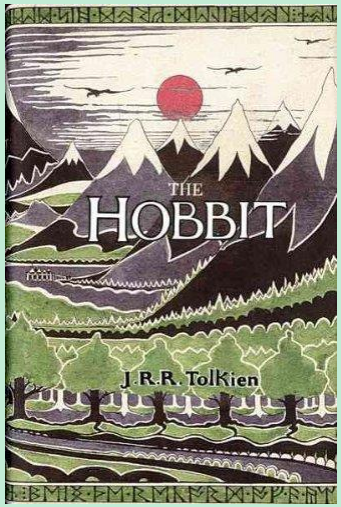


+



+





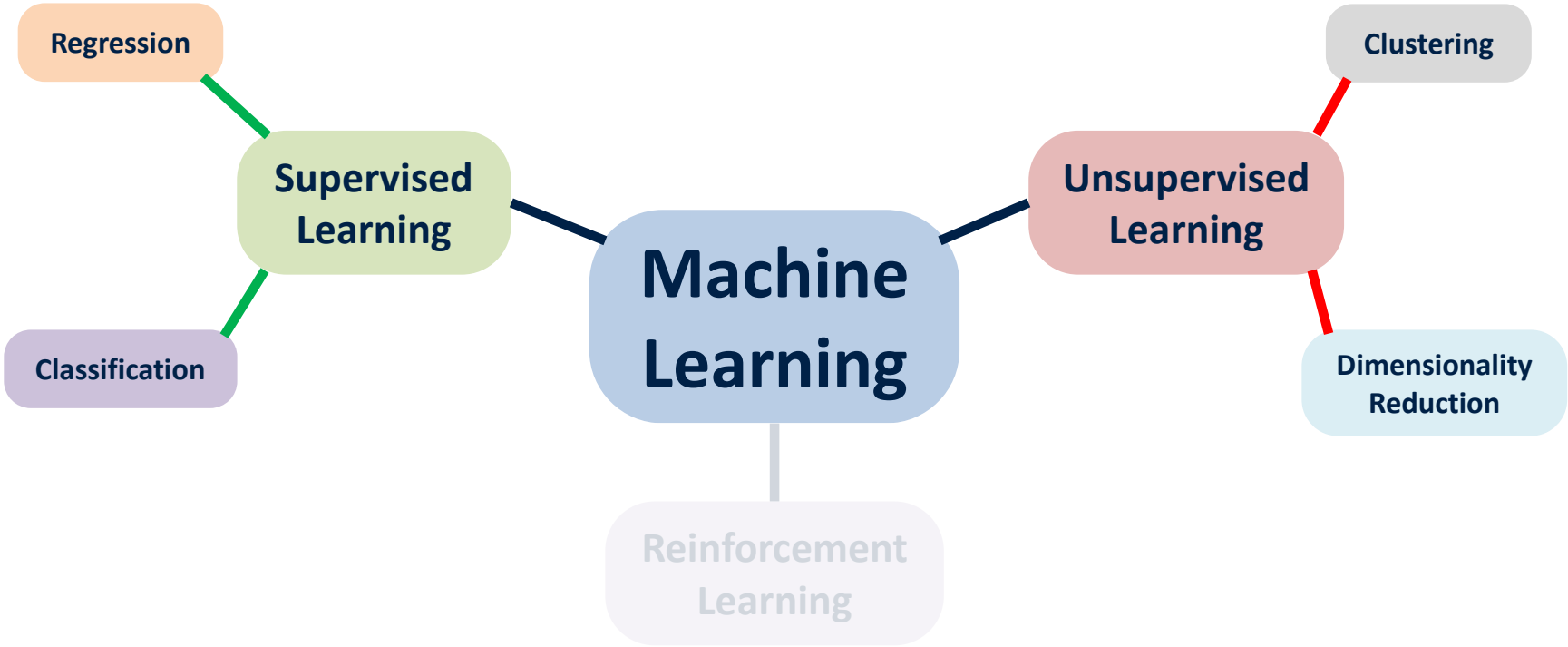
Images from *amazon.com*

ML depends on

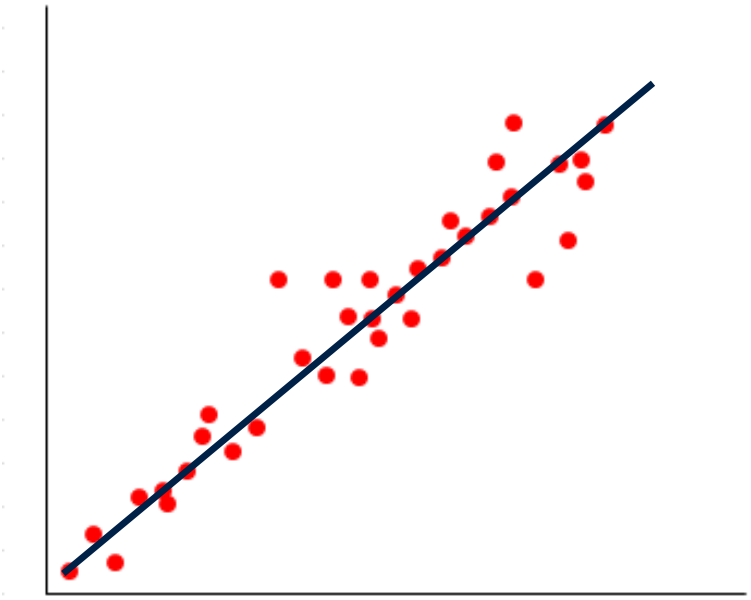
- Statistics: Probability theory, Sampling .....
- Mathematics: Linear Algebra, Multivariate Calculus,....
- Computer Science: Data structures, Programming .....
- Some domain knowledge.

Machine Learning

Definition: "*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*"

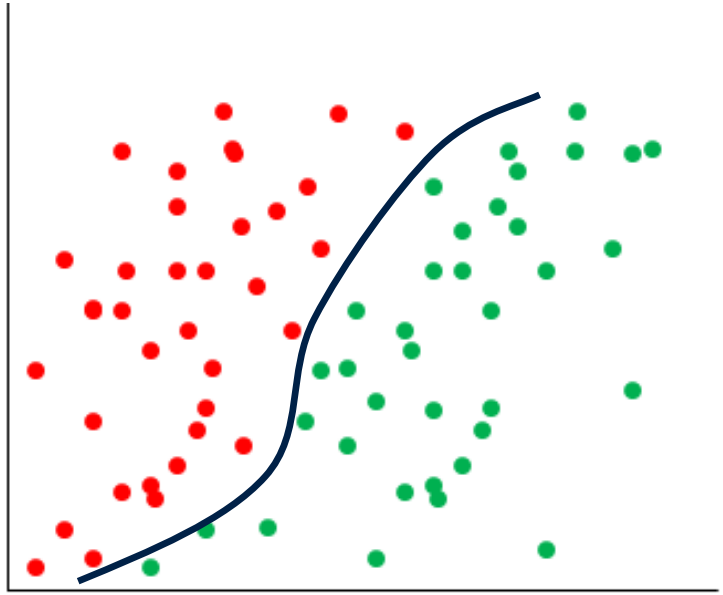


Regression



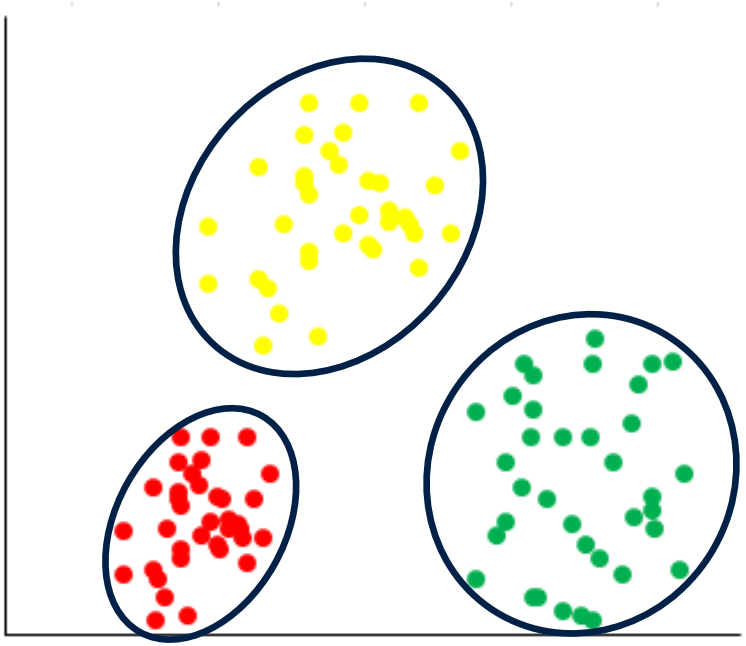
Supervised

Classification



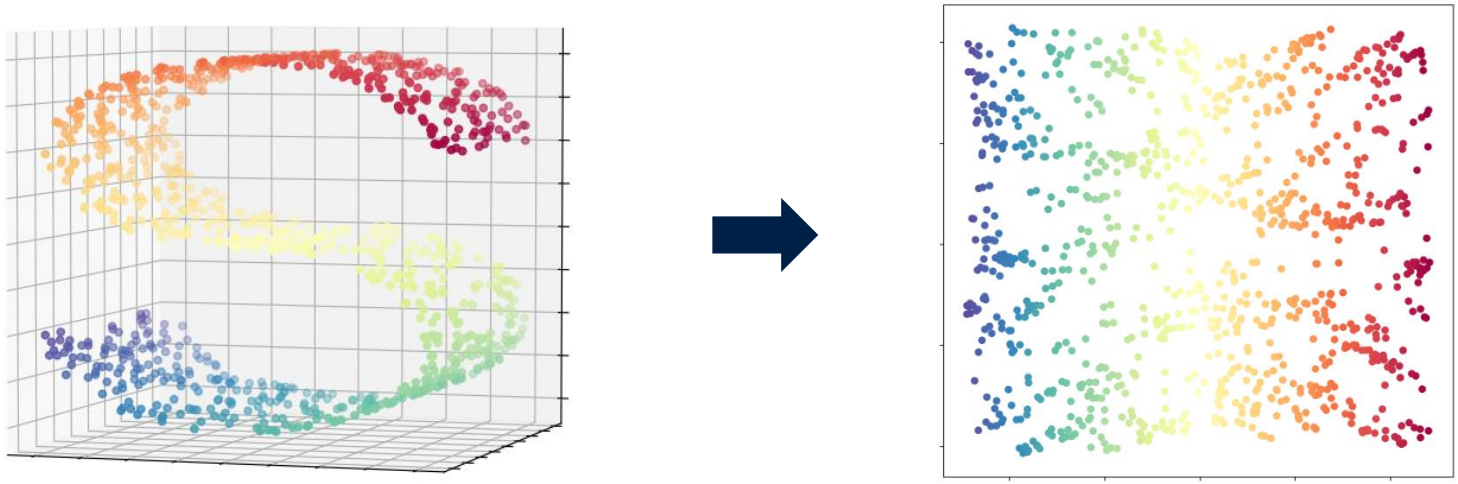
Supervised

Clustering



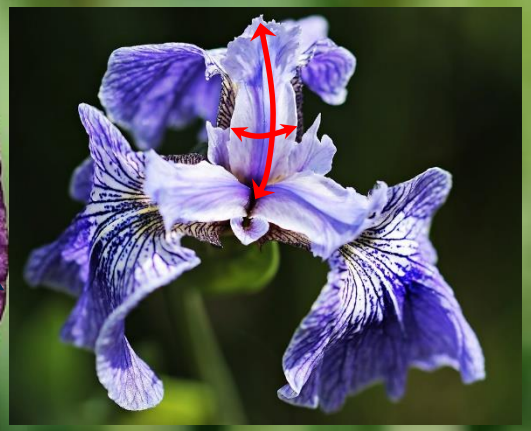
Unsupervised

Dimensionality reduction



Features

- Attributes used to represent input data.
- Features of *Iris* species:
  - Sepal Length
  - Sepal Width
  - Petal Length
  - Petal Width



Iris dataset

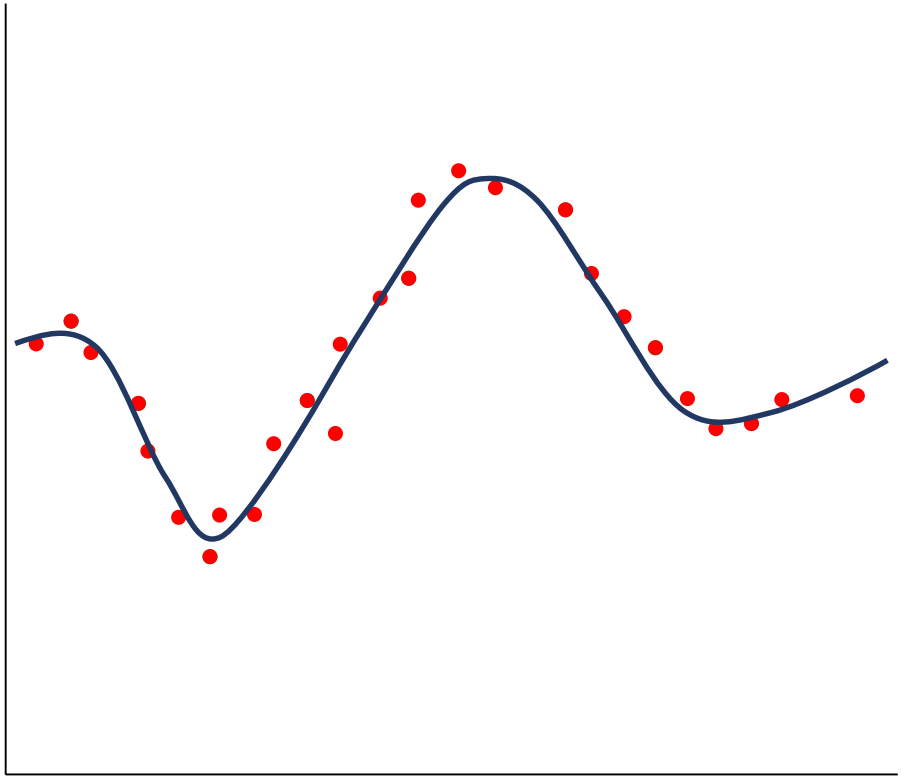
INPUTS

Sepal Length (cm)	Sepal Width (cm)	Petal Length (cm)	Petal Width (cm)
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5	3.6	1.4	0.2
5.4	3.9	1.7	0.4
4.6	3.4	1.4	0.3
5	3.4	1.5	0.2
4.4	2.9	1.4	0.2
.	.	.	.
.	.	.	.

OUTPUTS

Species	
Iris Setosa	0
Iris Virginica	1
Iris Versicolor	2

Training and Test data

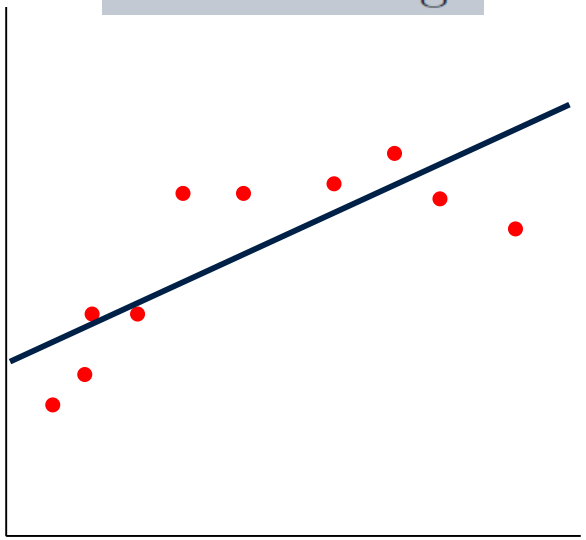


- Training data: Used for training the ML algorithm.
- Test data: Used for assessing the performance of the ML algorithm.

Model selection

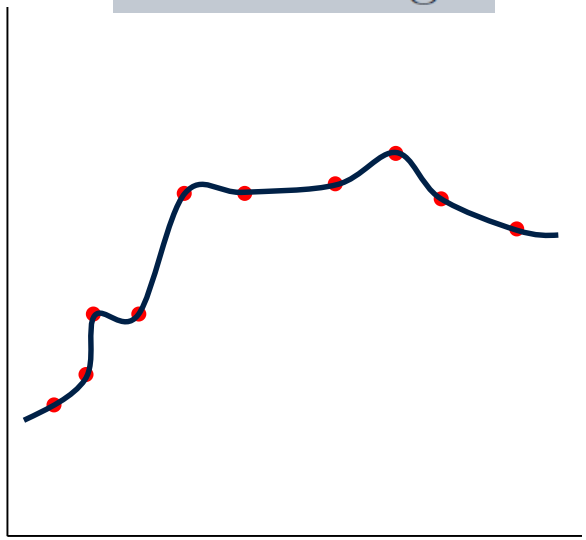
- Inductive bias of the ML algorithm.
- Hypothesis class (of functions) $\mathbb{H}$.

Underfitting



Complexity of $\mathbb{H}$ is low.

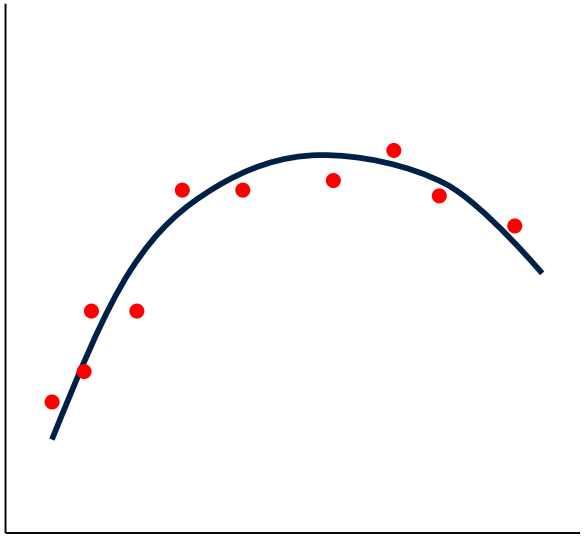
Overfitting



Complexity of $\mathbb{H}$ is high.

- Very complex hypothesis could lead to overfitting.
- Model selection $\rightarrow$ choosing the right $\mathbb{H}$.

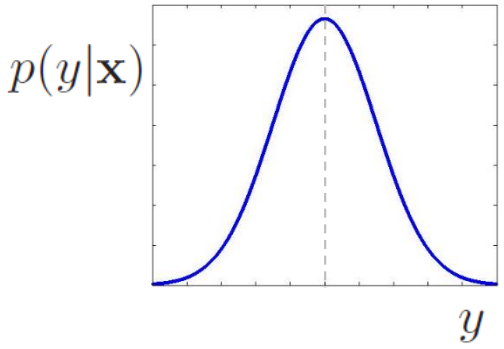
Generalization



- Larger class of functions $\rightarrow$ more complexity of the hypothesis class $\mathcal{C}(\mathbb{H})$.
- Objective: Good prediction at unobserved locations $\rightarrow$ good **generalization**.

Probabilistic modelling

- Many cases of supervised learning need estimation of the distribution $p(y|\mathbf{x})$ over possible outputs y for input $\mathbf{x}$.



- Expected value of the output is the mean of the distribution.
- Gives an estimate of the uncertainty of predictions.
- Two major types of probabilistic modelling approaches:
  - Discriminative modelling: The conditional distribution $p(y|\mathbf{x})$ is estimated directly. The distribution $p(\mathbf{x})$ is not modelled. For example, using $p(y|\mathbf{x}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x}, \sigma^2)$ to model regression problem.
  - Generative modelling: The conditional distribution $p(y|\mathbf{x})$ is estimated using the joint distribution $p(y, \mathbf{x})$ and the distribution $p(\mathbf{x})$ as $p(y|\mathbf{x}, \boldsymbol{\theta})$ $= p(y, \mathbf{x}|\boldsymbol{\theta})/p(\mathbf{x}|\boldsymbol{\theta})$. These type of approaches model both y and $\mathbf{x}$.