

Mastering Machine Learning with Python
(27th Aug 2024 - 18th Oct 2024)

iHUB DivyaSampark, IIT Roorkee

Support Vector Machines

Ritvij Bharat Private Limited

Theory and Intuition - Kernel Trick and Math

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Let's briefly talk about general mathematics of SVM and how it is related to the Scikit-Learn class calls.

Feel free to consider this an “optional” lecture.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Hyperplanes Defined

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

x_2

x_1

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$

x2



x1

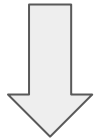
Led by : Shreyas Shukla

Mastering Machine Learning with Python

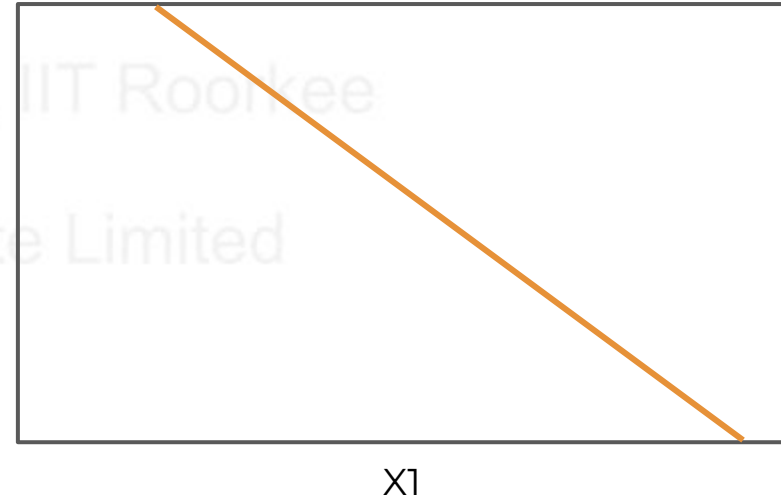
(27th Aug 2024 - 18th Oct 2024)

Hyperplanes Defined

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$



$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0$$



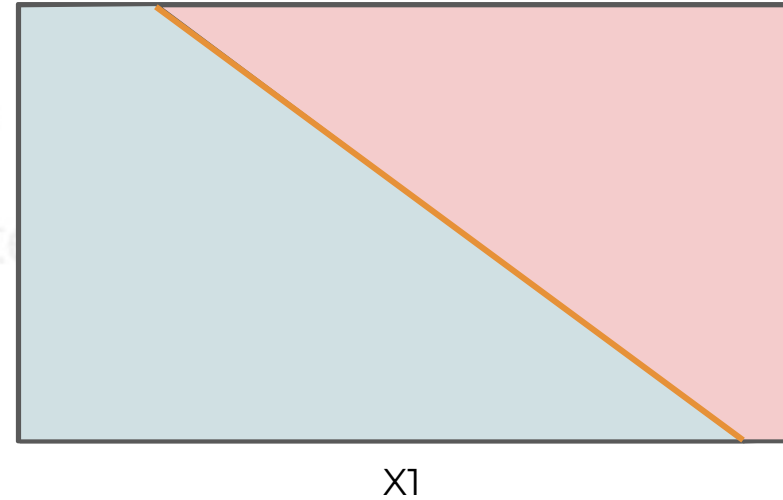
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Separating Hyperplanes

iHUB DivyaSampark,
and
Ritvij Bharat Private



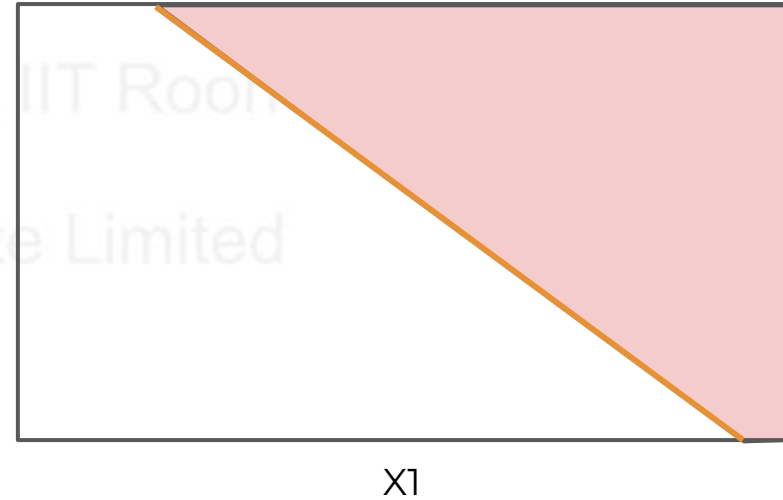
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p > 0$$



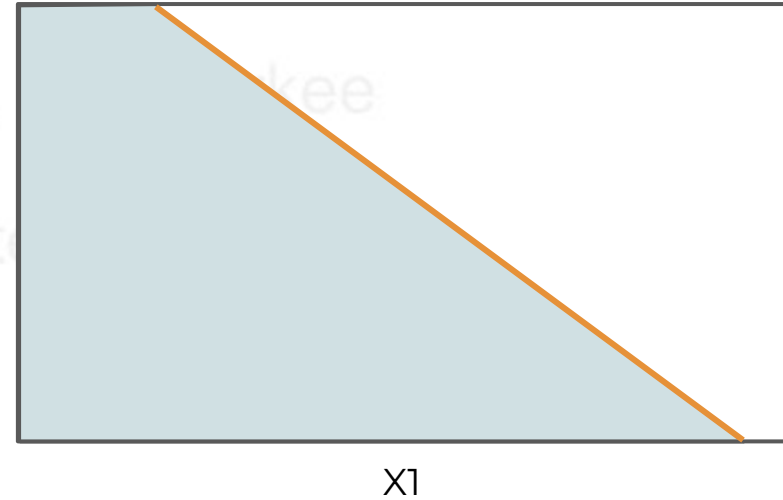
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

iHUB DivyaSampark,
and
Ritvii Bharat Private

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p < 0$$



Led by : Shreyas Shukla

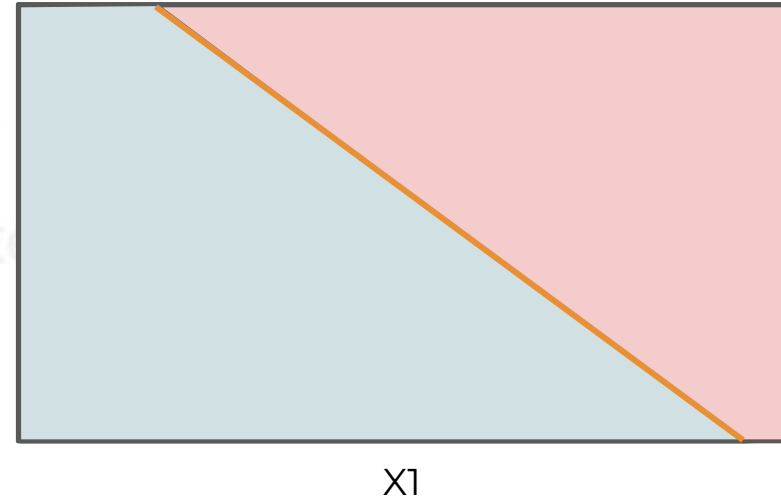
Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Separating Hyperplanes

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p > 0$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p < 0$$



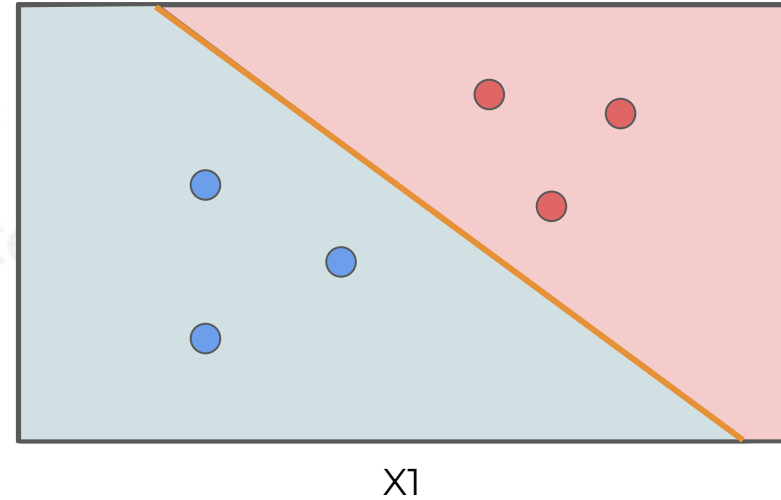
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Data Points

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$



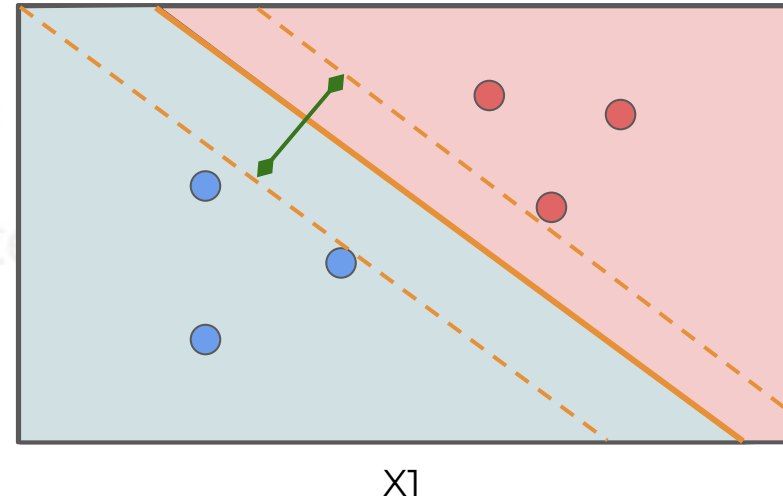
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Max Margin Classifier

$$\underset{\beta_0, \beta_1, \dots, \beta_p, M}{\text{maximize}} \quad M$$



Led by : Shreyas Shukla

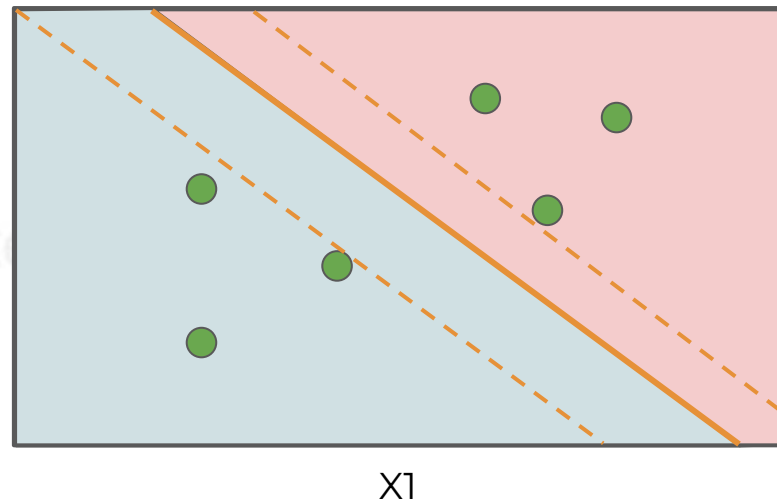
Max Margin Classifier

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

$$\text{maximize } M$$
$$\beta_0, \beta_1, \dots, \beta_p, M$$

$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M \quad \forall i = 1, \dots, n.$$



Mastering Machine Learning with Python

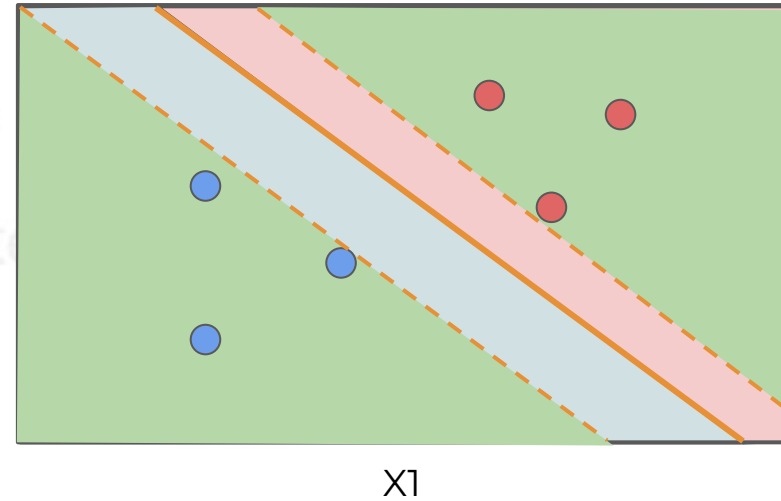
(27th Aug 2024 - 18th Oct 2024)

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

$$\text{maximize } M$$
$$\beta_0, \beta_1, \dots, \beta_p, M$$

$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M \quad \forall i = 1, \dots, n.$$



Mastering Machine Learning with Python

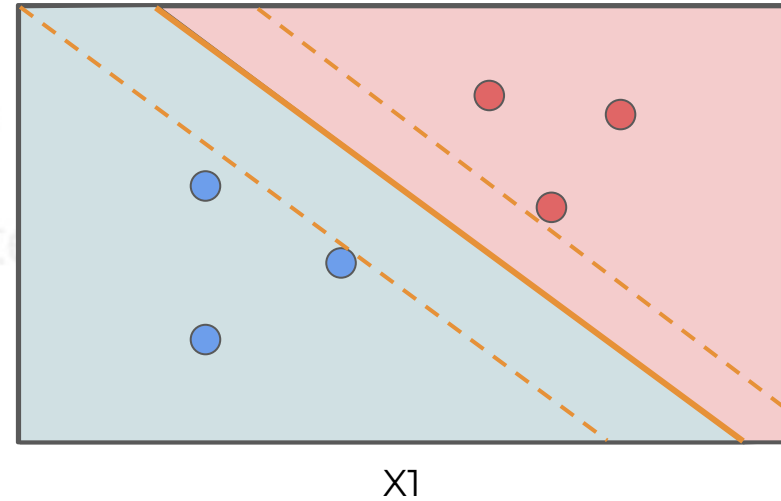
(27th Aug 2024 - 18th Oct 2024)

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

$$\text{maximize } M$$
$$\beta_0, \beta_1, \dots, \beta_p, M$$

$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M \quad \forall i = 1, \dots, n.$$



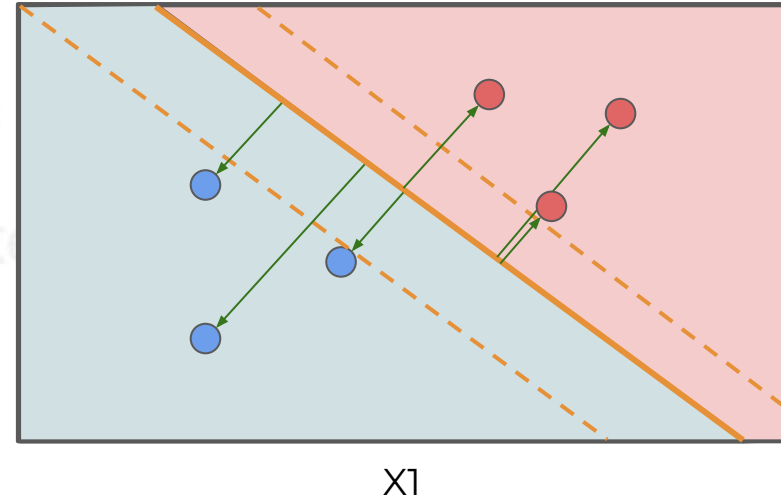
Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

iHUB DivyaSampark,
and
Ritvij Bharat Private

subject to $\sum_{j=1}^p \beta_j^2 = 1$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$$



Shukla

Mastering Machine Learning with Python

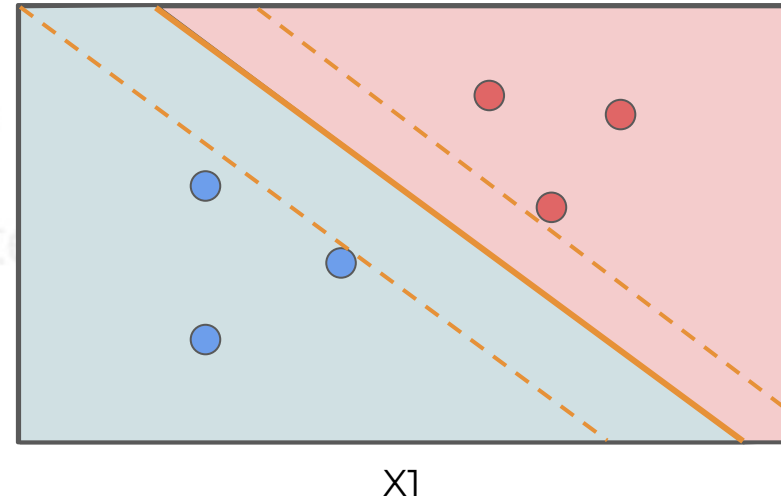
(27th Aug 2024 - 18th Oct 2024)

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

$$\text{maximize } M$$
$$\beta_0, \beta_1, \dots, \beta_p, M$$

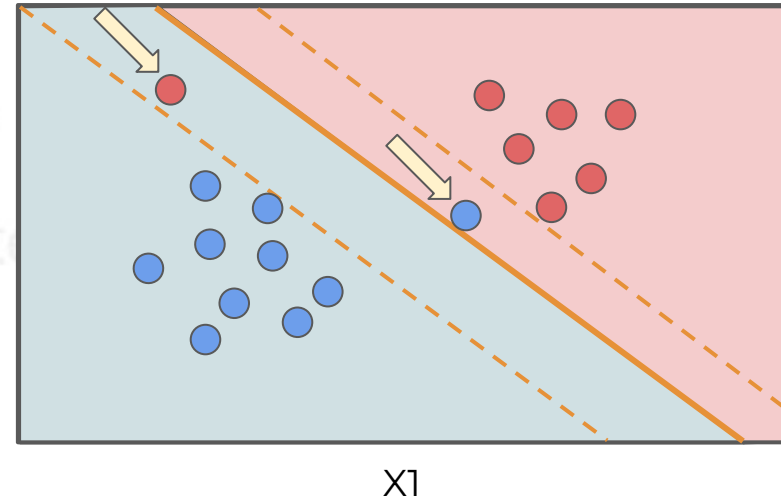
$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M \quad \forall i = 1, \dots, n.$$



Support Vector Classifier

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$



Led by : Shreyas Shukla

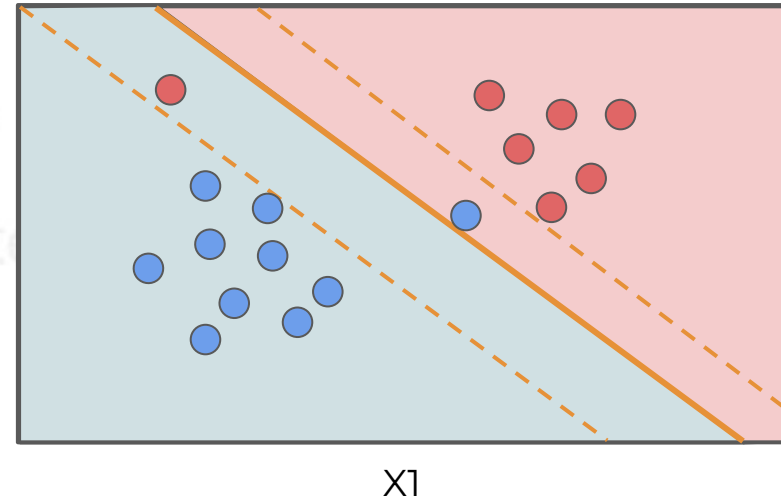
Support Vector Classifier

$$\underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} \quad M$$

$$\text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i)$$



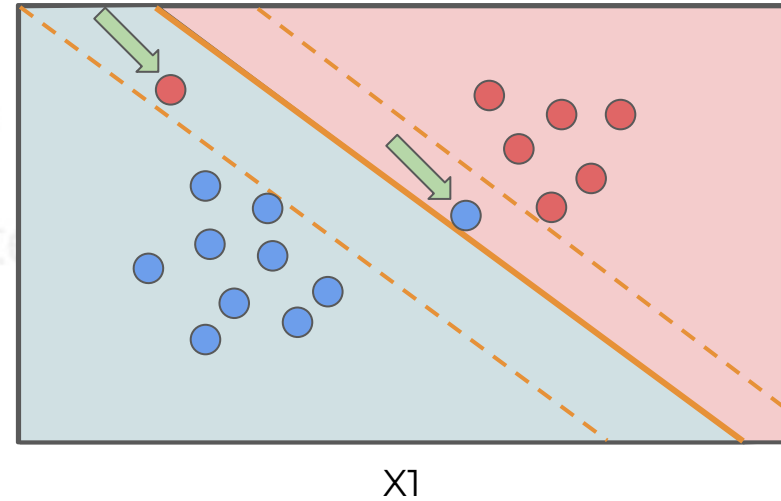
Support Vector Classifier

$$\underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} \quad M$$

$$\text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i)$$



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

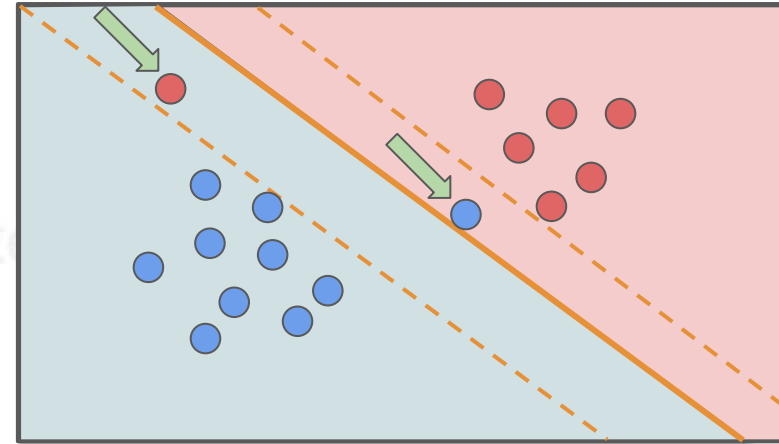
C : float, default=1.0

Regularization parameter. The strength of the regularization is inversely proportional to C. Must be strictly positive. The penalty is a squared L2 penalty.

$$\epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i)$$

x2



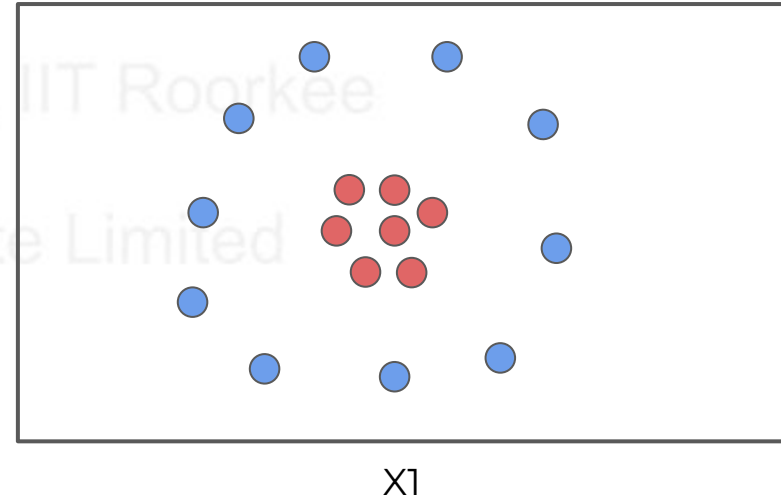
x1

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Support Vector Machines

$$X_1, X_2, \dots, X_p,$$



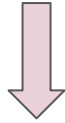
Led by : Shreyas Shukla

Mastering Machine Learning with Python

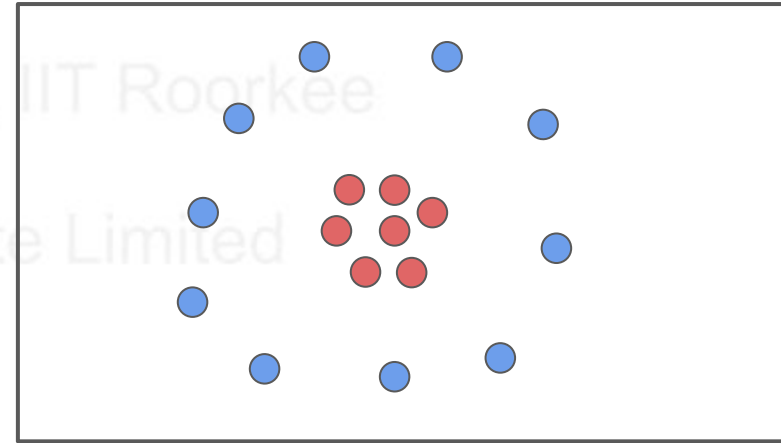
(27th Aug 2024 - 18th Oct 2024)

Support Vector Machines

$$X_1, X_2, \dots, X_p,$$



$$X_1, X_1^2, X_2, X_2^2, \dots, X_p, X_p^2$$



Led by : Shreyas Shukla

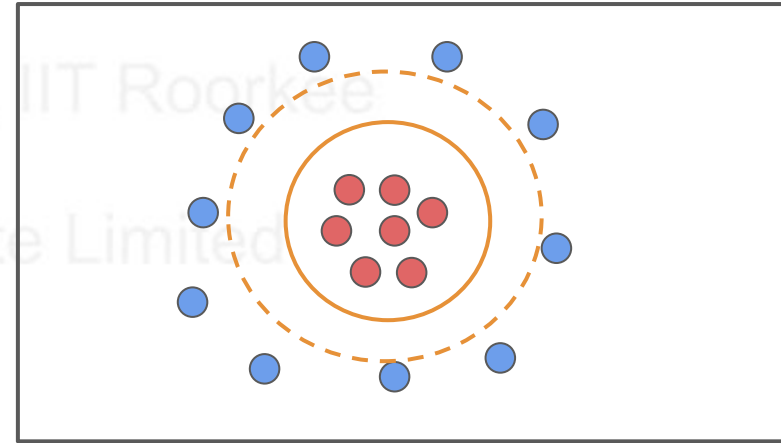
Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Support Vector Machines

$$\underset{\beta_0, \beta_{11}, \beta_{12}, \dots, \beta_{p1}, \beta_{p2}, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} \quad M$$

X2



X1

Led by : Shreyas Shukla

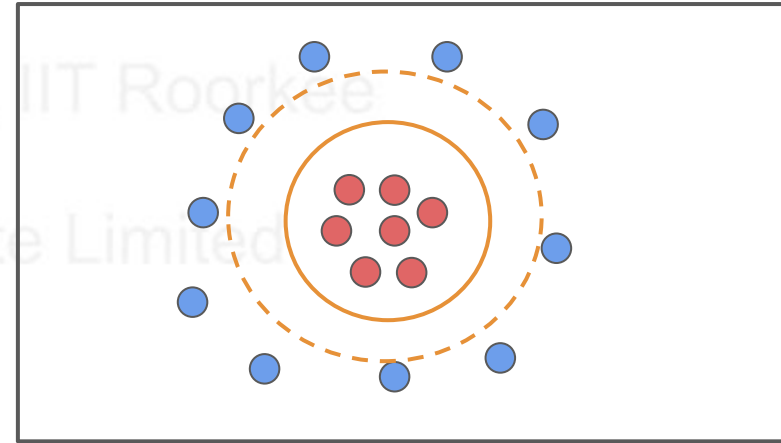
Support Vector Machines

$$X_1, X_1^2, X_2, X_2^2, \dots, X_p, X_p^2$$

$$\underset{\beta_0, \beta_{11}, \beta_{12}, \dots, \beta_{p1}, \beta_{p2}, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} \quad M \quad \text{X2}$$

$$\text{subject to } y_i \left(\beta_0 + \sum_{j=1}^p \beta_{j1} x_{ij} + \sum_{j=1}^p \beta_{j2} x_{ij}^2 \right) \geq M(1 - \epsilon_i)$$

$$\sum_{i=1}^n \epsilon_i \leq C, \quad \epsilon_i \geq 0, \quad \sum_{j=1}^p \sum_{k=1}^2 \beta_{jk}^2 = 1.$$



X1

eyas Shukla

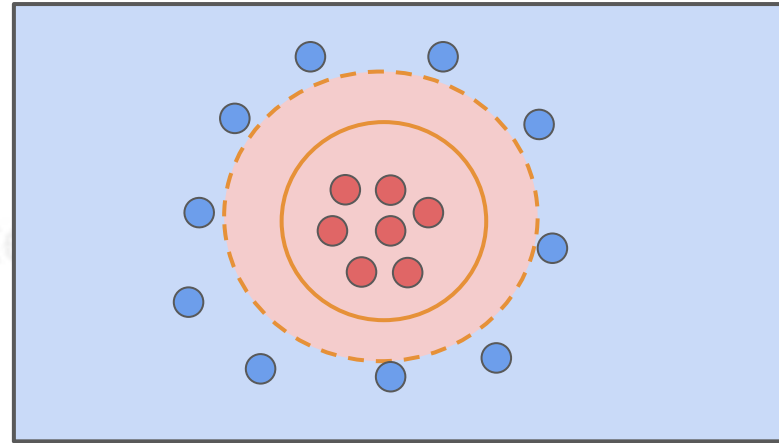
Support Vector Machines

$$X_1, X_1^2, X_2, X_2^2, \dots, X_p, X_p^2$$

$$\underset{\beta_0, \beta_{11}, \beta_{12}, \dots, \beta_{p1}, \beta_{p2}, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} \quad M \quad \text{X2}$$

$$\text{subject to } y_i \left(\beta_0 + \sum_{j=1}^p \beta_{j1} x_{ij} + \sum_{j=1}^p \beta_{j2} x_{ij}^2 \right) \geq M(1 - \epsilon_i)$$

$$\sum_{i=1}^n \epsilon_i \leq C, \quad \epsilon_i \geq 0, \quad \sum_{j=1}^p \sum_{k=1}^2 \beta_{jk}^2 = 1.$$



X1

eyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

How to deal with very large feature space? As polynomial order grows, the number of computations necessary to solve for margins also grows!

We use **Kernel trick** which makes use of the **inner product** of vectors, also known as the **dot product**.

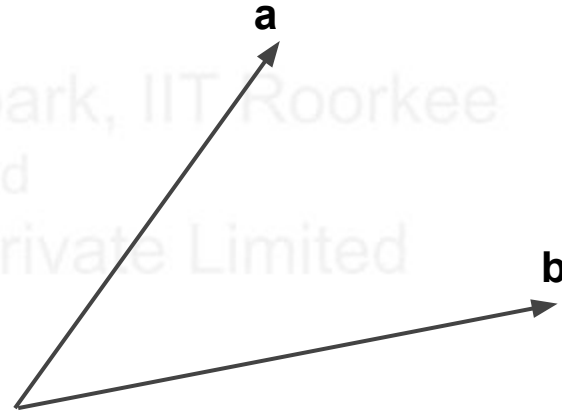
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Dot Product

$$\langle a, b \rangle = \sum_{i=1}^r a_i b_i$$



Led by : Shreyas Shukla

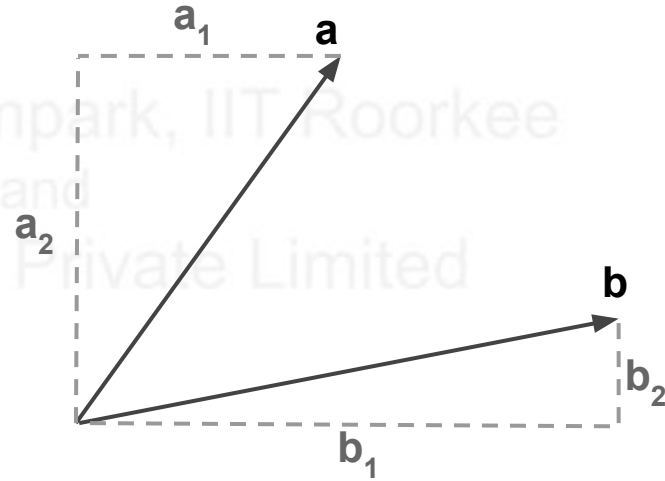
Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Dot Product

$$\langle a, b \rangle = \sum_{i=1}^r a_i b_i$$

$$a \cdot b = a_1 b_1 + a_2 b_2$$



Led by : Shreyas Shukla

Mastering Machine Learning with Python

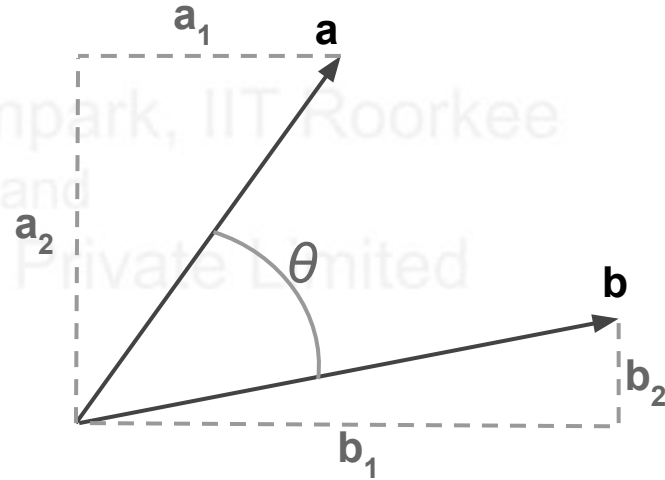
(27th Aug 2024 - 18th Oct 2024)

Dot Product

$$\langle a, b \rangle = \sum_{i=1}^r a_i b_i$$

$$a \cdot b = a_1 b_1 + a_2 b_2$$

$$a \cdot b = |a||b|\cos(\theta)$$



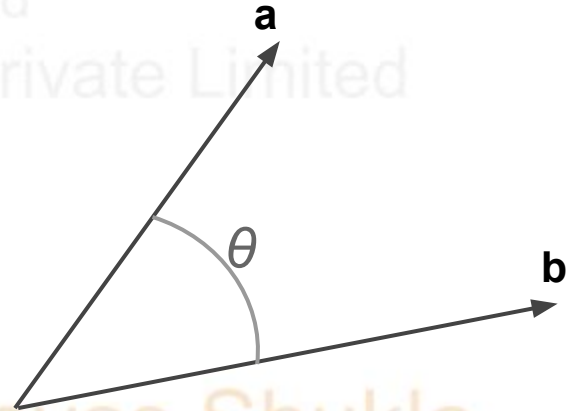
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Notice how the dot product can be thought of as a similarity between the vectors.

$$a \cdot b = |a||b|\cos(\theta)$$



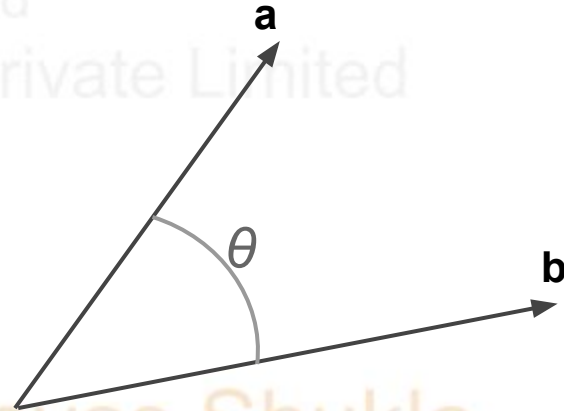
Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- $\cos(0^\circ) = 1$
- $\cos(90^\circ) = 0$
- $\cos(180^\circ) = -1$

$$a \cdot b = |a||b|\cos(\theta)$$



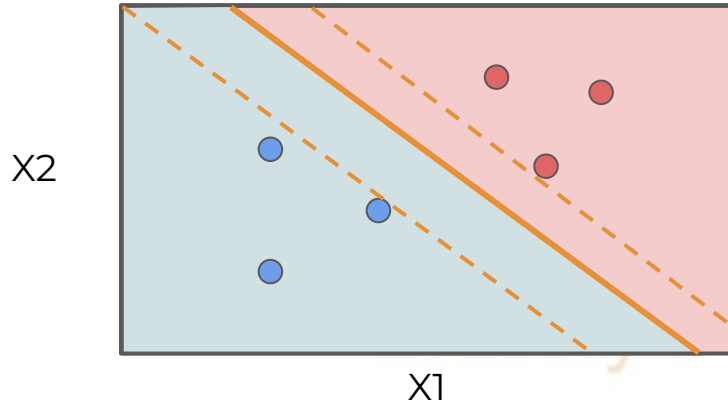
Led by : Shreyas Shukla

(27th Aug 2024 - 18th Oct 2024)

Let's discuss Kernel Trick

Linear Support Vector Classifier rewritten:

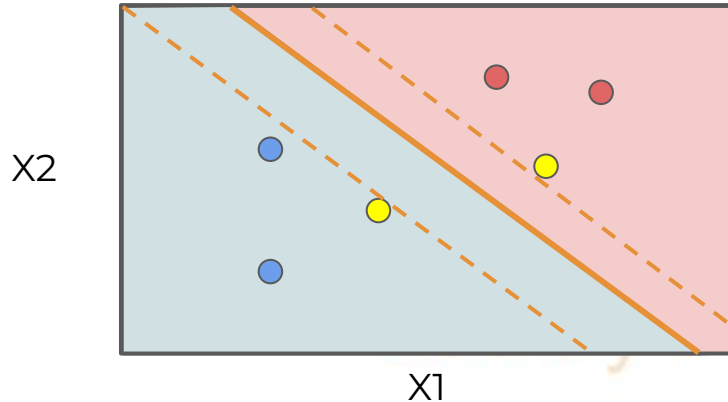
$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i \langle x, x_i \rangle$$



Calculating the inner products of all pairs of training observations

Linear Support Vector Classifier rewritten:

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i \langle x, x_i \rangle$$



Only non-zero for the support vectors.

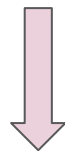
Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Linear Support Vector Classifier rewritten:

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i \langle x, x_i \rangle$$



$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle$$

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Linear Support Vector Classifier rewritten:

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle$$

\mathcal{S} collection of indices of these support points

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Kernel Function

$$K(x_i, x_{i'}) = \sum_{j=1}^p x_{ij} x_{i'j}$$

A kernel is a function that quantifies the similarity of two observations.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Kernel Function

$$K(x_i, x_{i'}) = \sum_{j=1}^p x_{ij} x_{i'j} \quad f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Kernel Function

$$K(x_i, x_{i'}) = \sum_{j=1}^p x_{ij} x_{i'j} \quad f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle$$

The diagram illustrates the relationship between the kernel function, the model function, and the inner product. Red boxes highlight the summation in the kernel function, the inner product term in the model function, and the inner product definition. Red arrows point from the definition of the inner product to both the kernel function and the model function.

$$\langle a, b \rangle = \sum_{i=1}^r a_i b_i$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Kernel Function

$$K(x_i, x_{i'}) = \sum_{j=1}^p x_{ij} x_{i'j} \quad f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i K(x, x_i)$$


The diagram illustrates the relationship between the kernel function $K(x_i, x_{i'})$ and the inner product $\langle a, b \rangle = \sum_{i=1}^r a_i b_i$. A red arrow points from the kernel function in the first equation to the inner product, and another red arrow points from the inner product to the kernel function in the second equation.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Polynomial Kernel

$$K(x_i, x_{i'}) = \left(1 + \sum_{j=1}^p x_{ij} x_{i'j}\right)^d \quad f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle$$
$$\langle a, b \rangle = \sum_{i=1}^r a_i b_i$$


Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Radial Basis Kernel

$$K(x_i, x_{i'}) = \exp\left(-\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2\right) \quad f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle$$
$$\langle a, b \rangle = \sum_{i=1}^r a_i b_i$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

The use of **kernels** as a replacement is known as the **kernel trick**.

Kernels allow us to avoid computations in the enlarged feature space, by only needing to perform computations for each distinct pair of training points

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Intuitively we've already seen inner products act as a measurement of similarity between vectors.

The use of kernels can be thought of as a measure of similarity between the original feature space and the enlarged feature space.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Tree Based Methods

Led by : Shreyas Shukla

Tree Based Methods

Three main methods:

- Decision Trees
- Random Forests
- Boosted Trees

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Decision Trees

Theory and Intuition

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

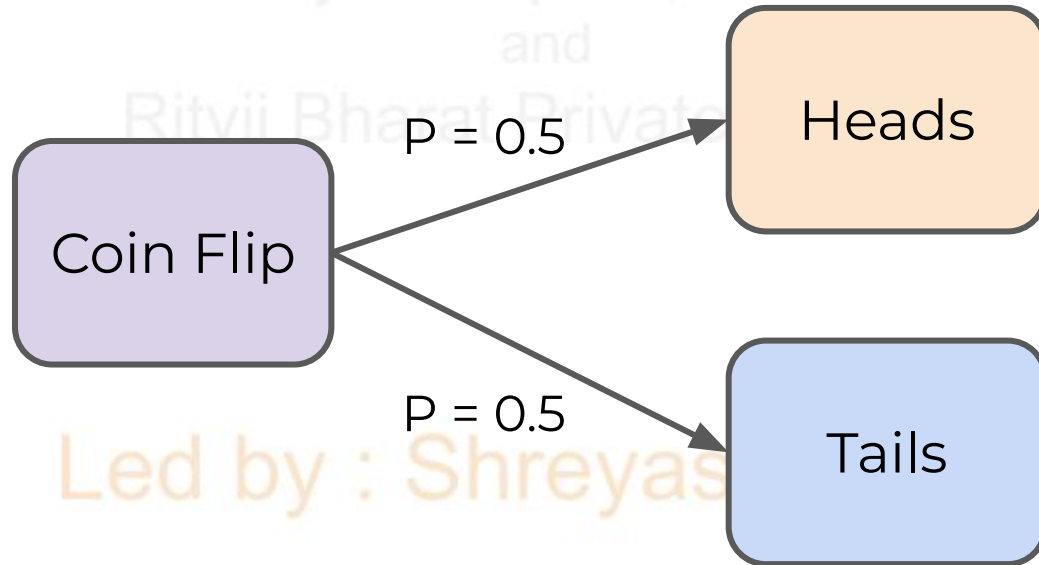
- While the use of basic decision trees for modeling choices and outcomes have been around for a very long time, statistical decision trees are a more recent development.
- **Note the difference here!**

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

The general term “decision tree” can refer to a flowchart mapping out outcomes.



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Decision Tree Learning refers to the statistical modeling that uses a form of decision trees, where node splits are decided based on an information metric.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Decision trees methods is basically the ability to split data based on information from features.

We need a mathematical definition of information and the ability to measure it.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

The ability to measure and define information will become more important as we learn the mathematics of tree based methods.

Let's talk about the development of decision trees.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

1963: First publication of regression tree algorithm by Morgan and Sonquist

1963: Morgan and Sonquist created piecewise-constant model with splits.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

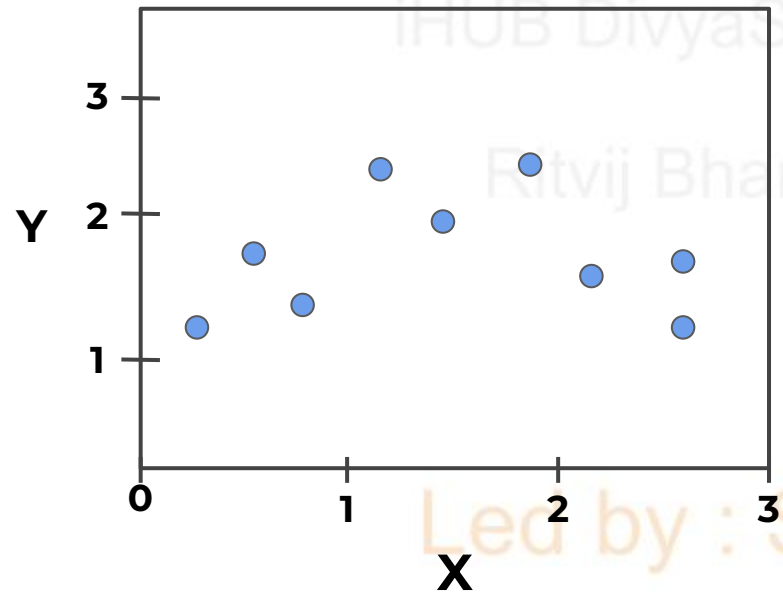
1963: Piecewise-constant regression tree



Led by : Shreyas Shukla

Mastering Machine Learning with Python

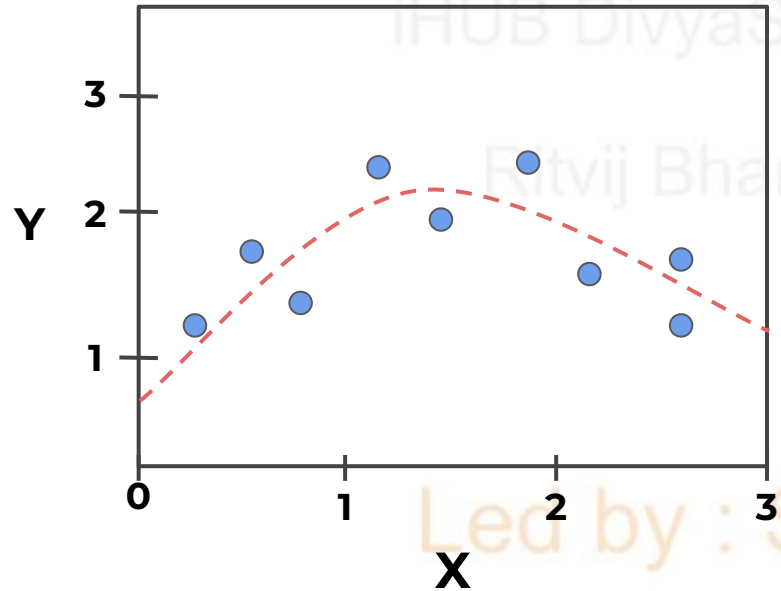
(27th Aug 2024 - 18th Oct 2024)



Led by : Shreyas Shukla

Mastering Machine Learning with Python

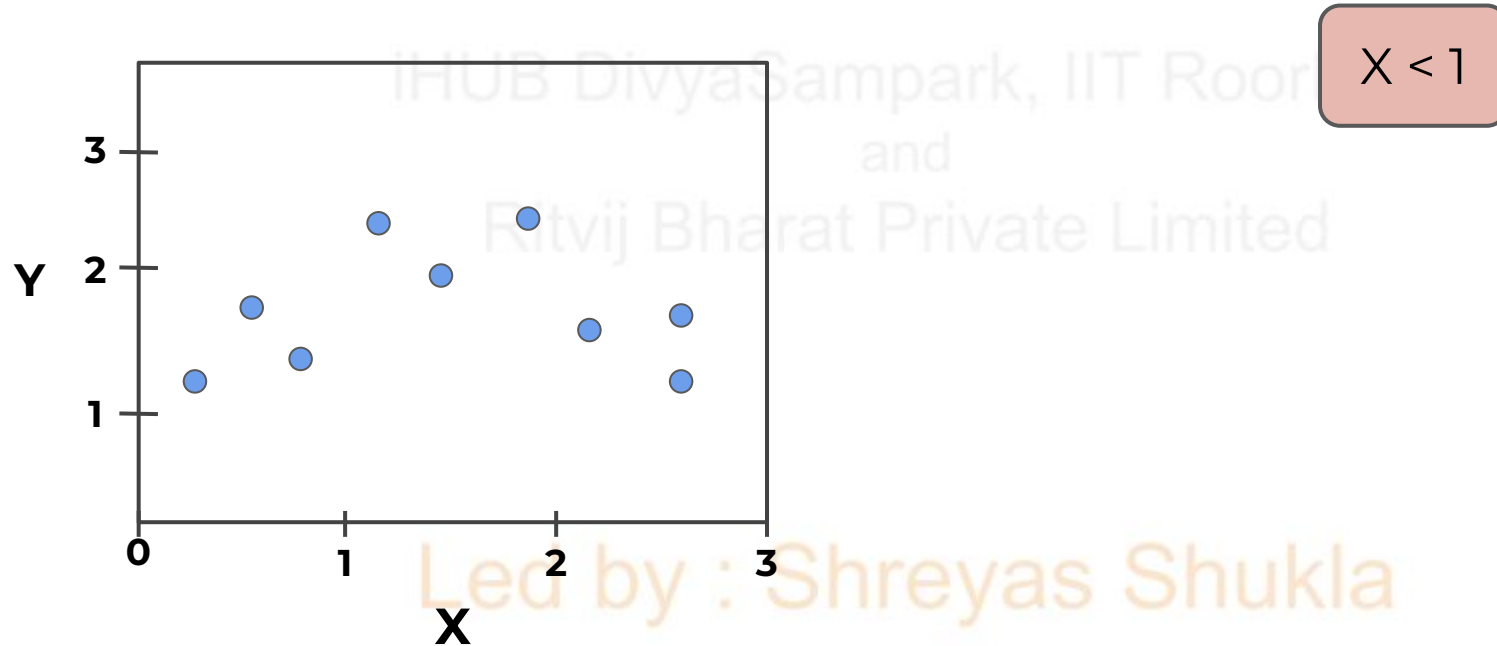
(27th Aug 2024 - 18th Oct 2024)



Led by : Shreyas Shukla

Mastering Machine Learning with Python

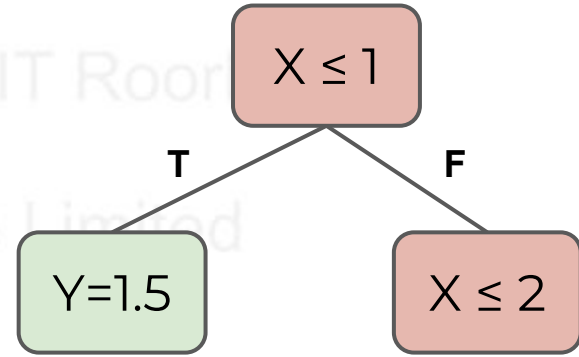
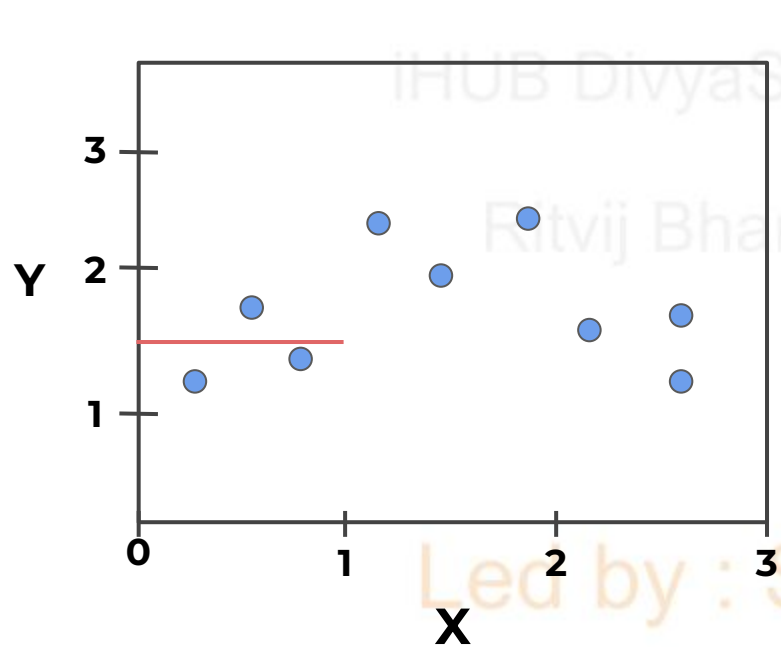
(27th Aug 2024 - 18th Oct 2024)



Led by : Shreyas Shukla

Mastering Machine Learning with Python

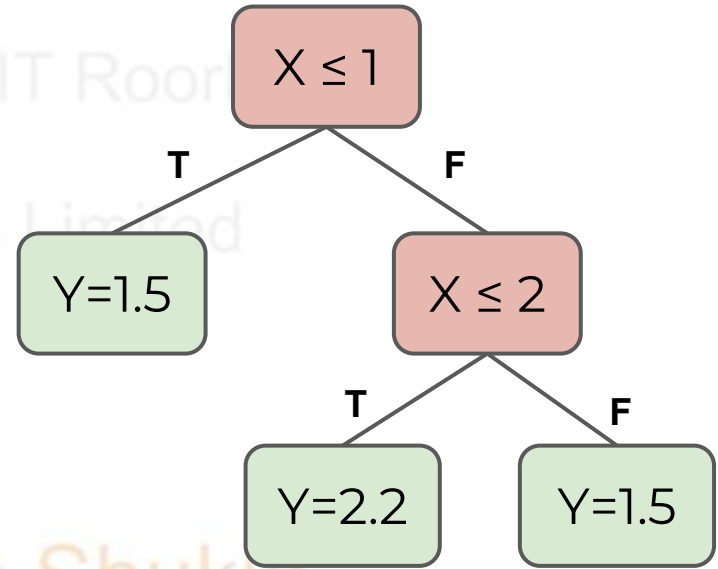
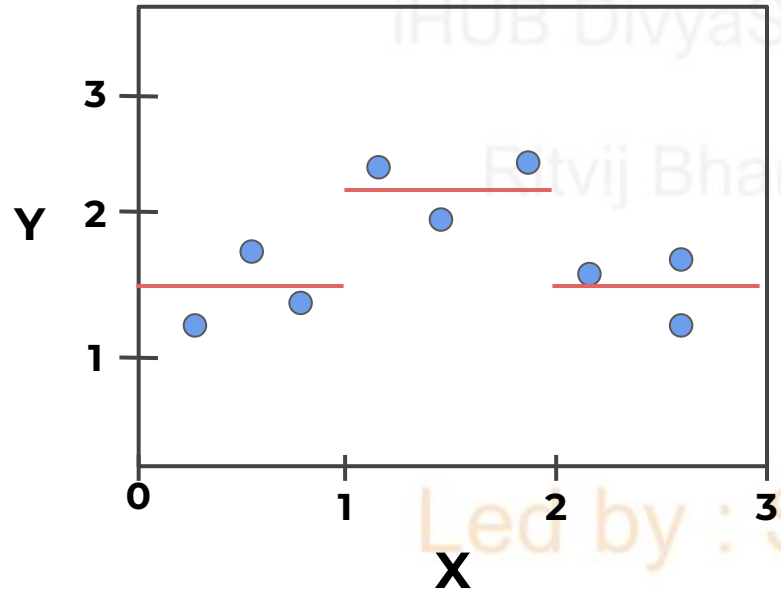
(27th Aug 2024 - 18th Oct 2024)



Led by : Shreyas Shukla

Mastering Machine Learning with Python

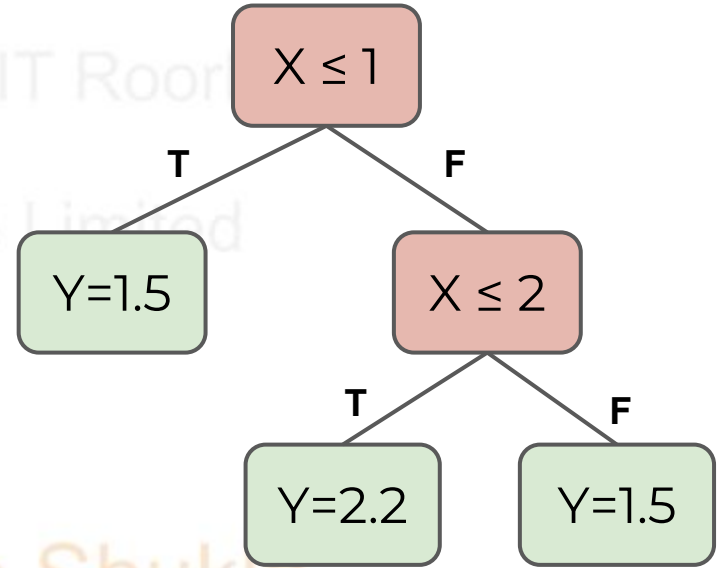
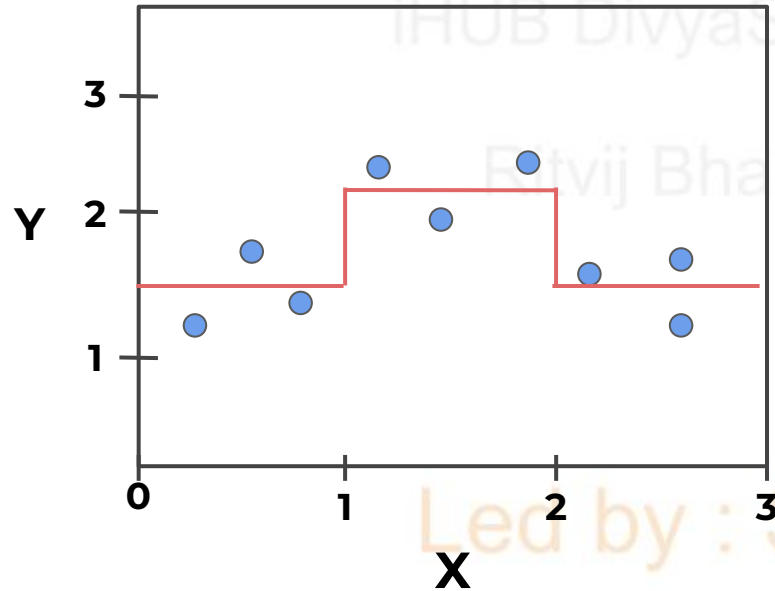
(27th Aug 2024 - 18th Oct 2024)



Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)



Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Node impurity

iHUB DivyaSampark, IIT Roorkee
and
Ritvii Bharat Private Limited

$$\phi(t) = \sum_{i \in t} (y_i - \bar{y})^2$$

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Decision Trees

Decision Tree Basics

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Let us understand some terminology about the decision tree components.

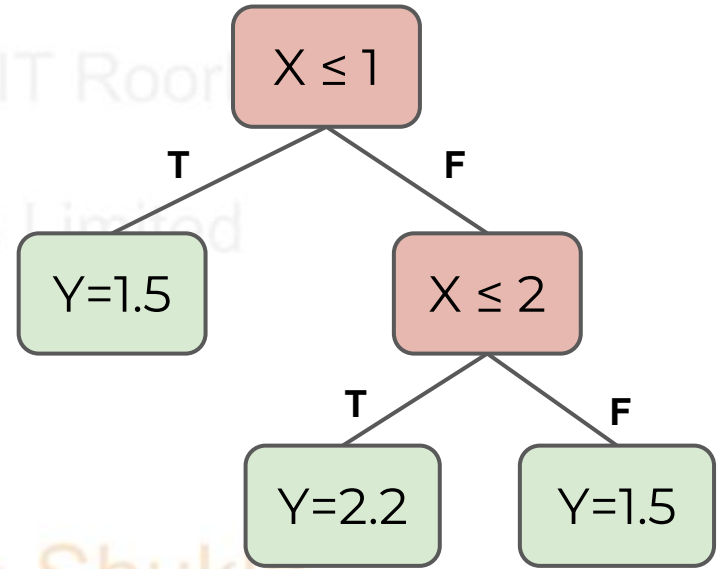
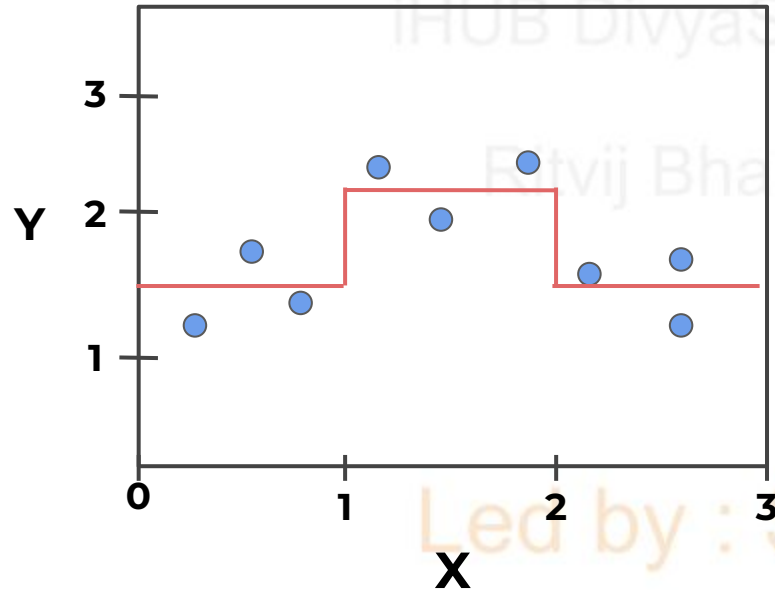
IHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

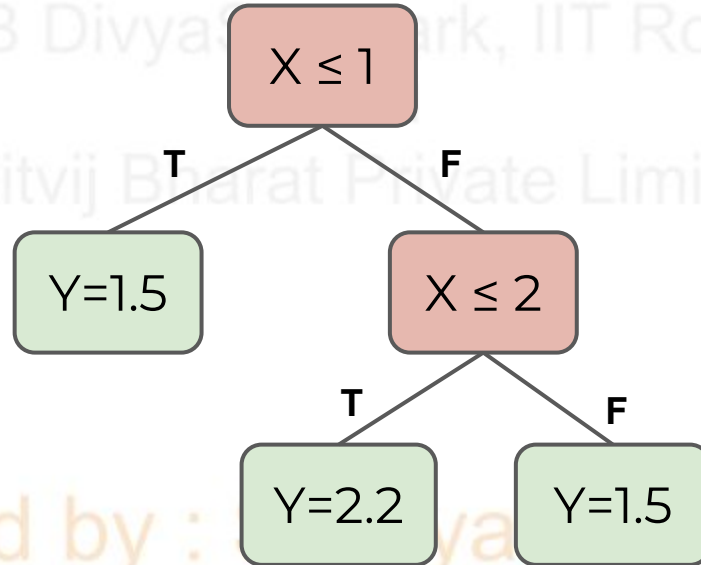
Recall our simple regression tree:



Led by : Shreyas Shukla

Mastering Machine Learning with Python

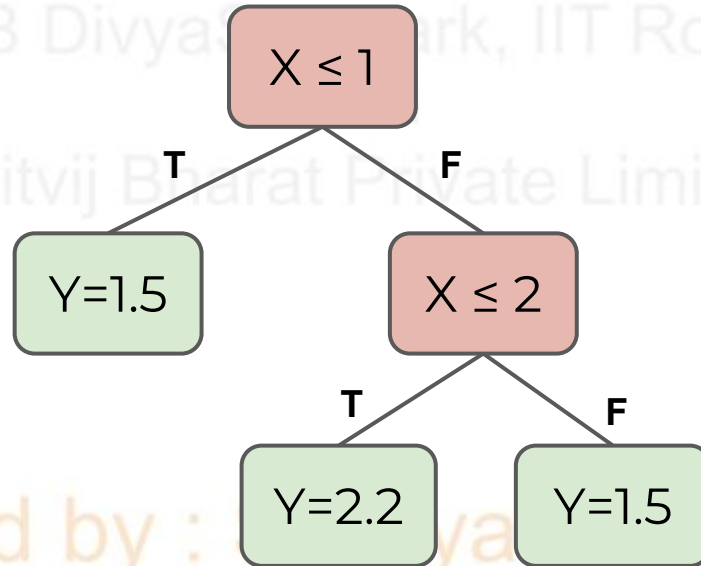
(27th Aug 2024 - 18th Oct 2024)



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

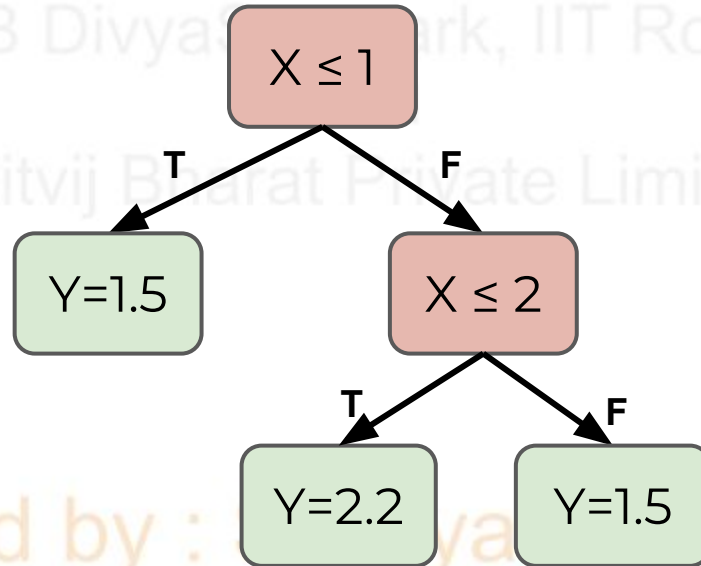
Splitting



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

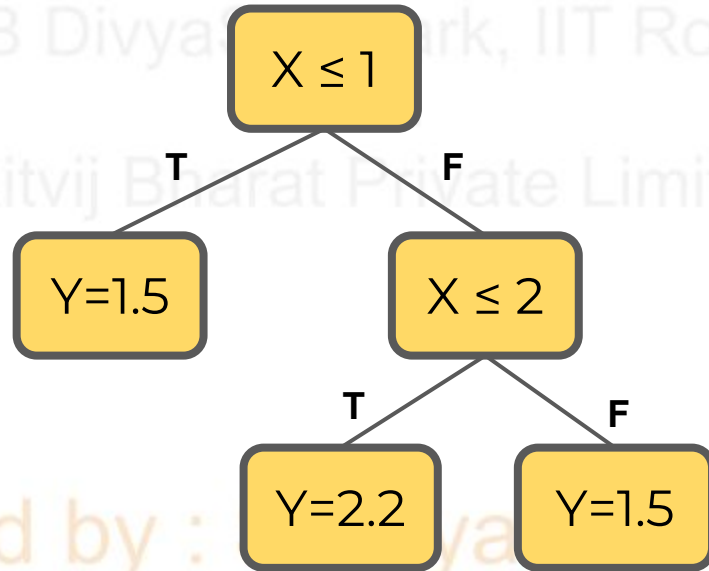
Splitting



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

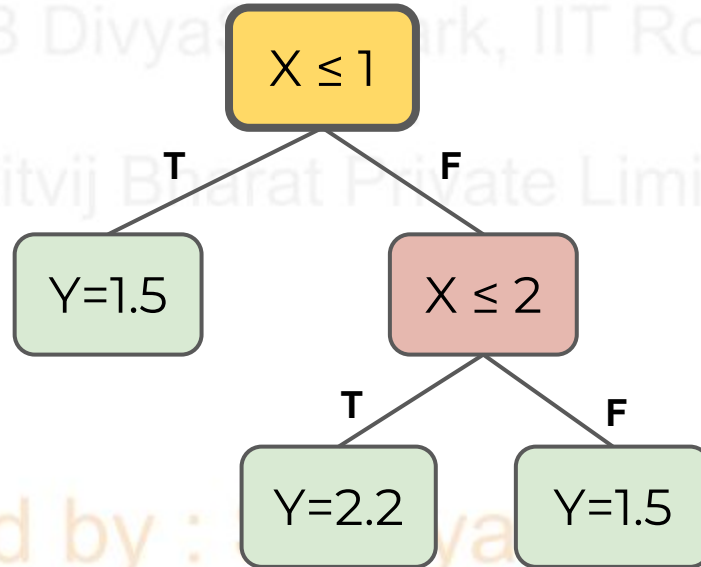
Nodes:



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

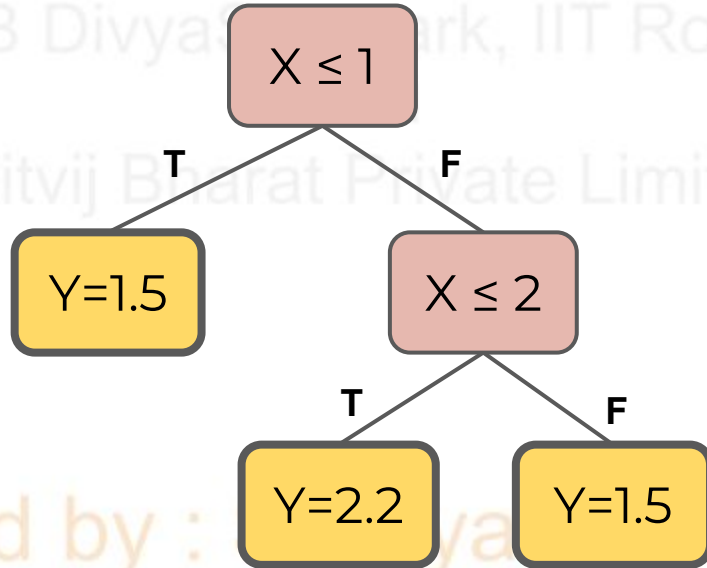
Root Node:



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

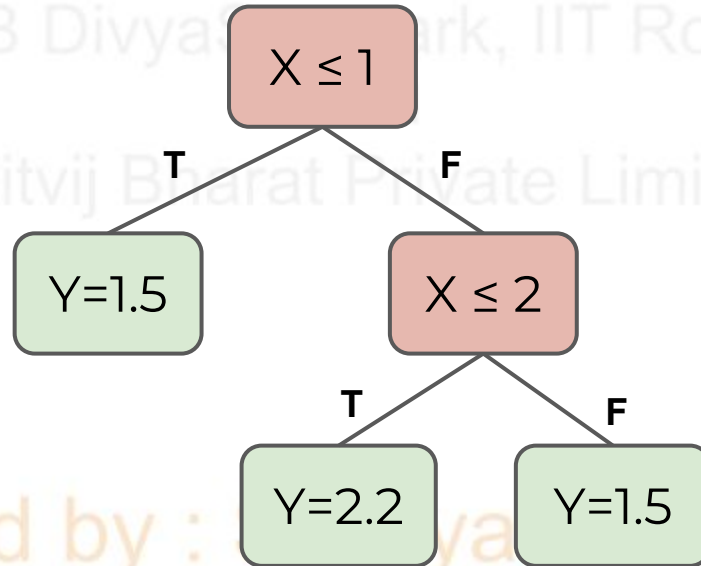
Leaf (Terminal) Nodes:



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

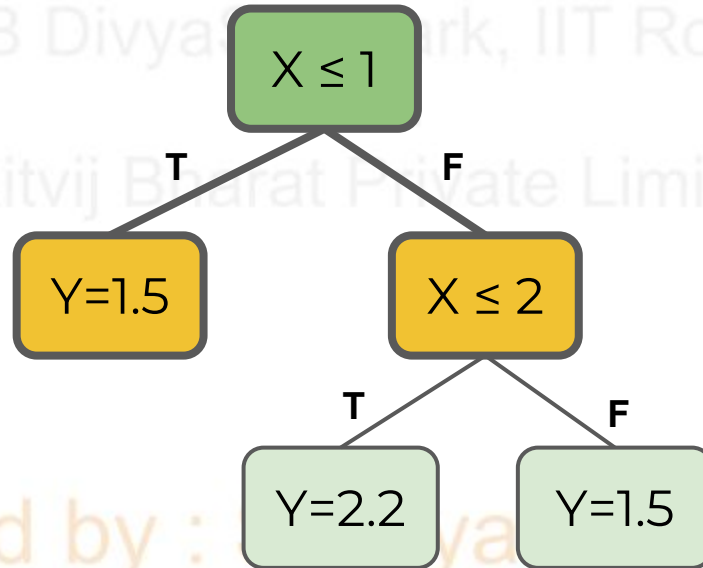
Parent and Children Nodes:



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

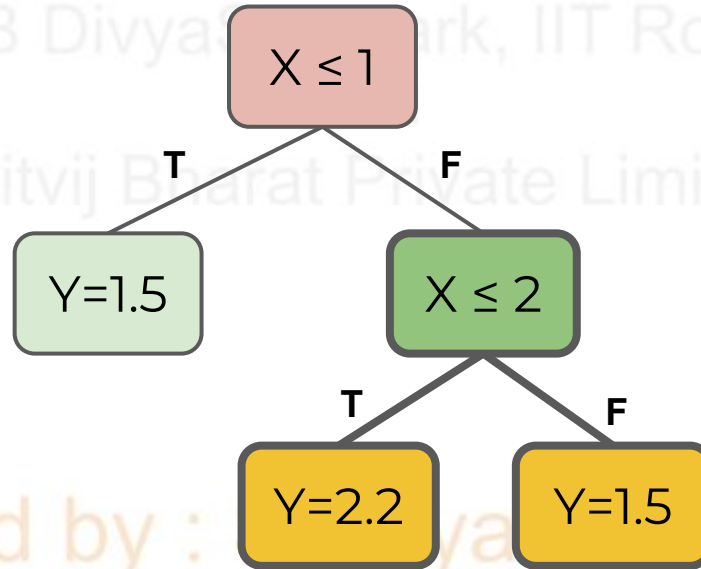
Parent and Children Nodes:



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

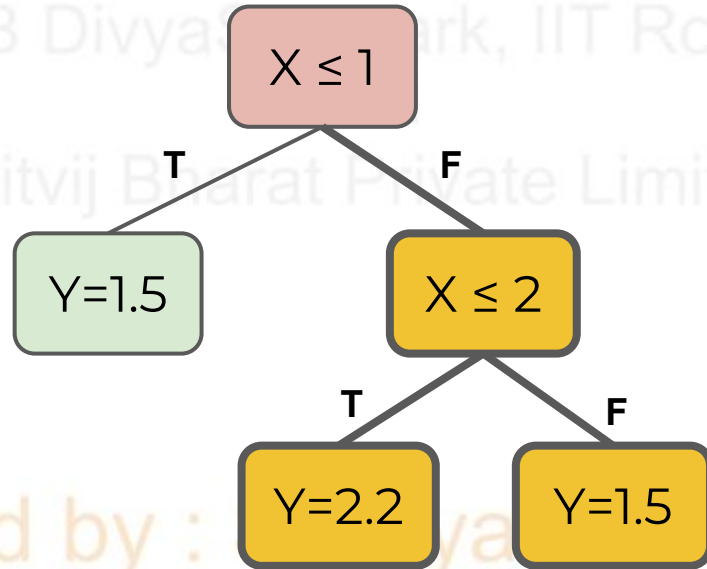
Parent and Children Nodes:



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

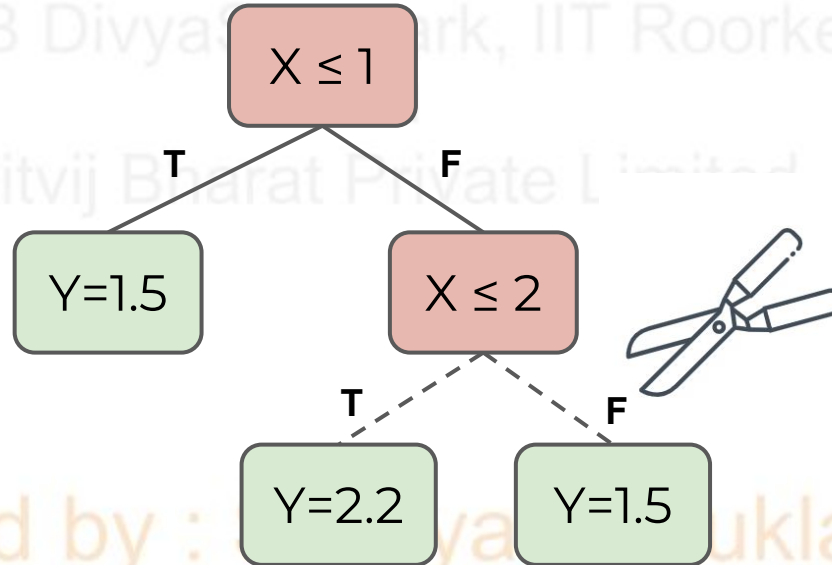
Tree Branches (Sub Trees):



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

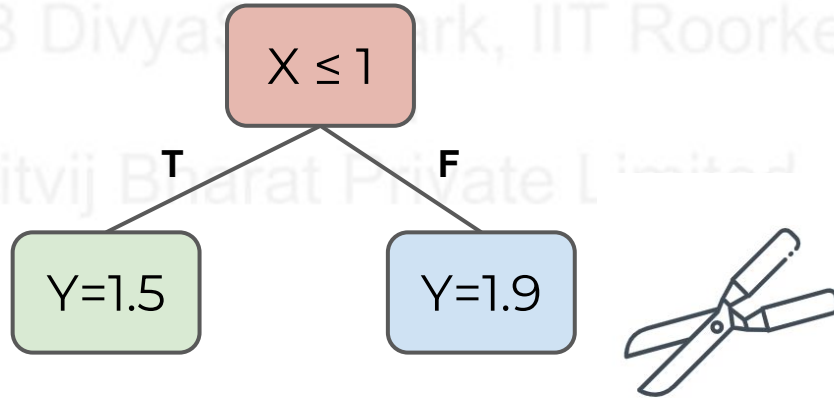
Pruning:



Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Pruning:



Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Let's begin constructing a tree!

