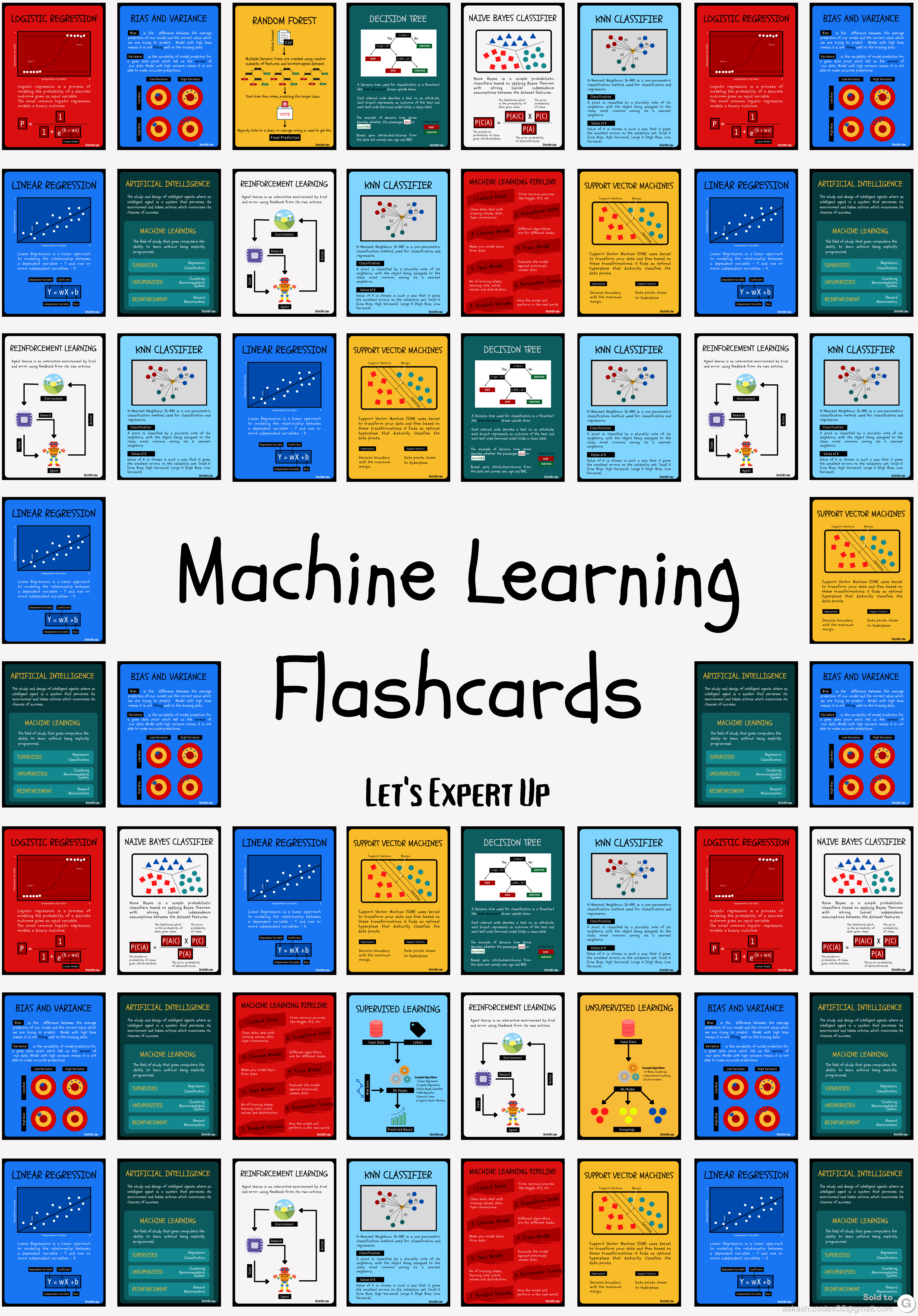


Machine Learning Flashcards

Let's Expert Up

ARTIFICIAL INTELLIGENCE

The study and design of intelligent agents where an intelligent agent is a system that perceives its environment and takes actions which maximizes its chances of success.

MACHINE LEARNING

The field of study that gives computers the ability to learn without being explicitly programmed.

SUPERVISED

Regression
Classification

UNSUPERVISED

Clustering
Recommendation System

REINFORCEMENT

Reward Maximization

MACHINE LEARNING PIPELINE

1. Collect Data

From various sources like Kaggle, UCI, etc

2. Transform Data

Clean data, deal with missing values, data type conversions

3. Choose Model

Different algorithms are for different tasks

4. Train Model

Make you model learn from data

5. Test Model

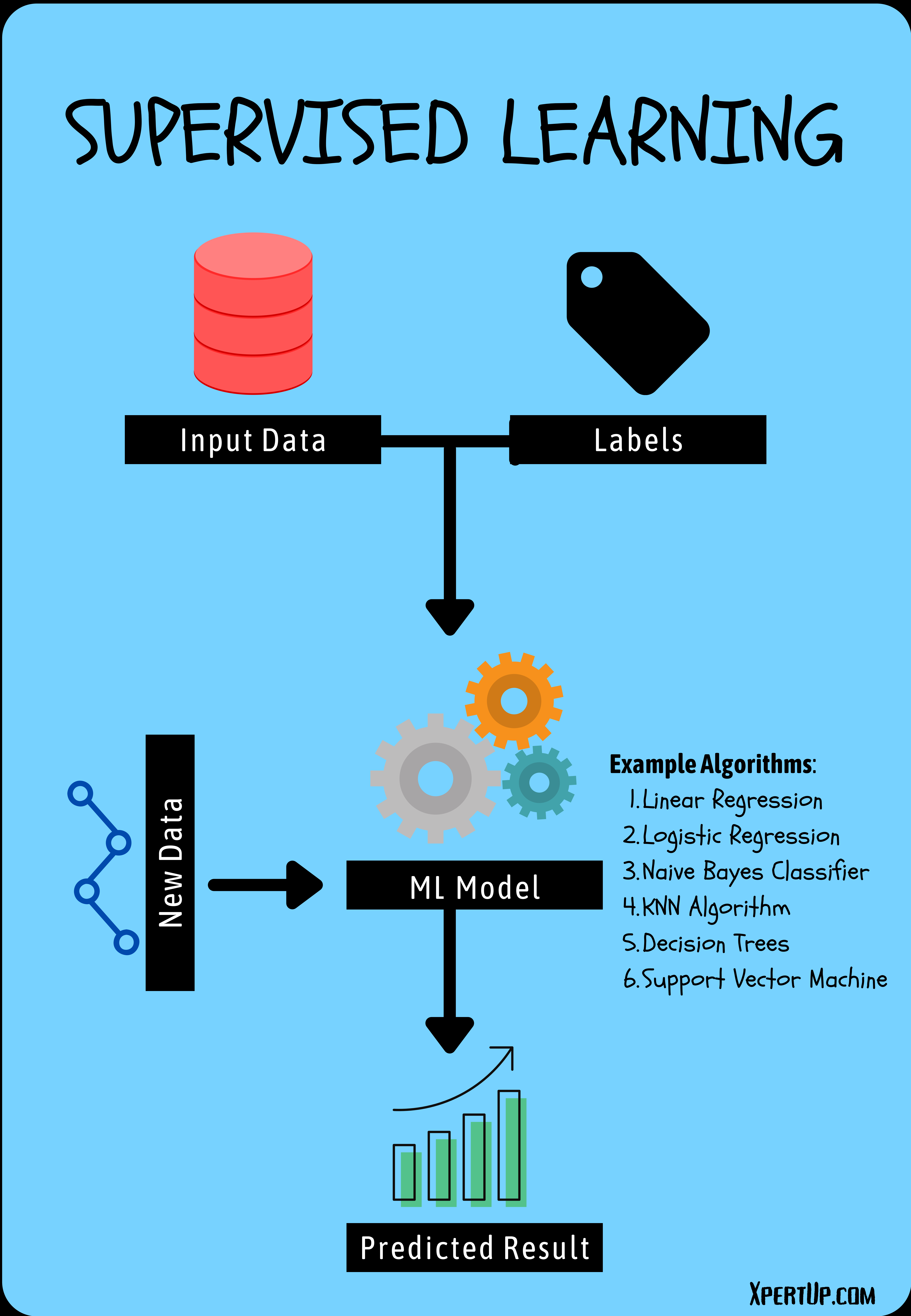
Evaluate the model against previously unseen data

6. Parameter Tuning

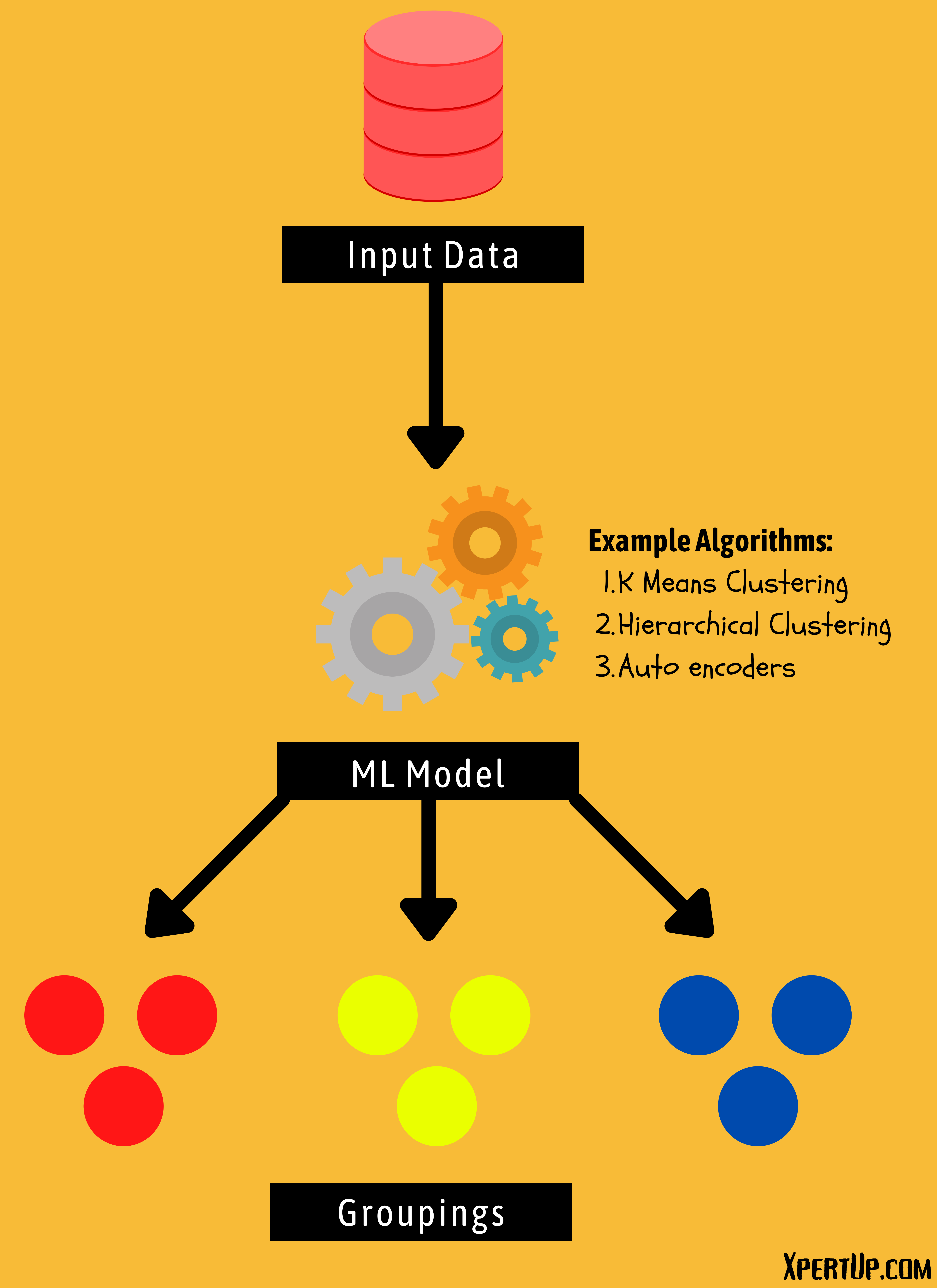
No of training steps, learning rate, initial values and distribution

7. Predict Values

How the model will perform in the real world

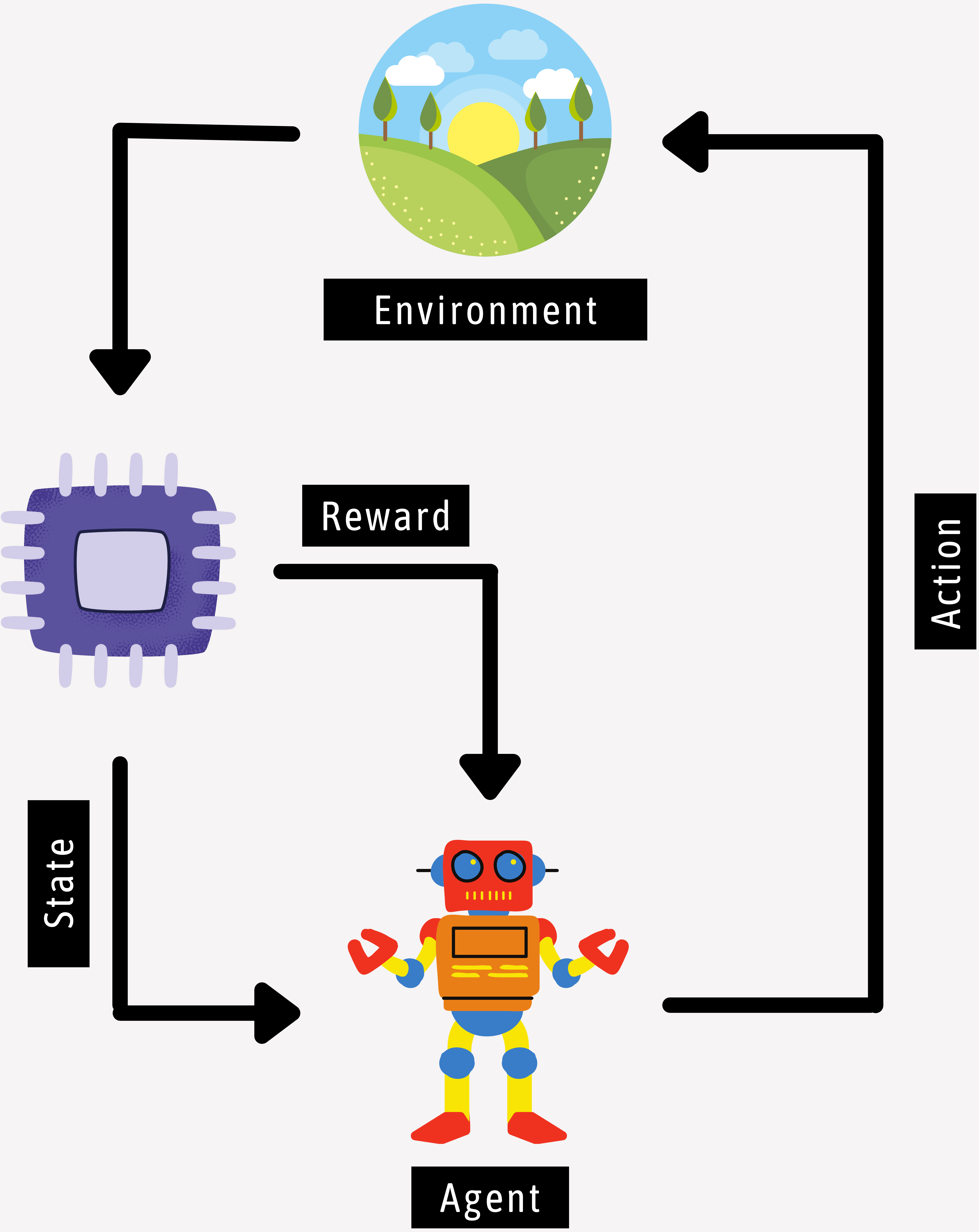

SUPERVISED LEARNING
Input Data
Labels
New Data
ML Model
Example Algorithms:
1.Linear Regression
2.Logistic Regression
3.Naive Bayes Classifier
4.KNN Algorithm
5.Decision Trees
6.Support Vector Machine
Predicted Result
XPERTUP.COM

UNSUPERVISED LEARNING



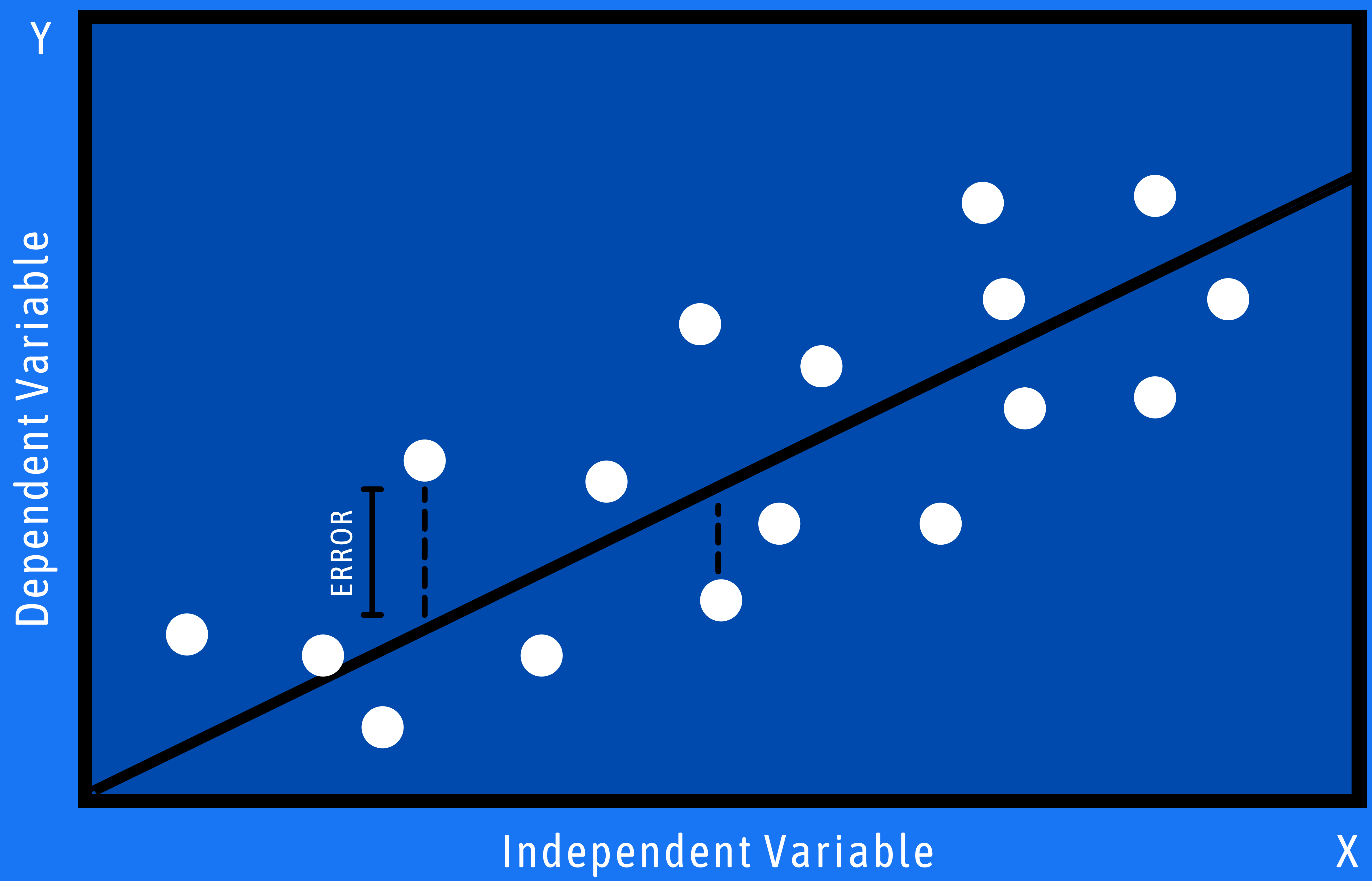
REINFORCEMENT LEARNING

Agent learns in an interactive environment by trial and error using feedback from its own actions.

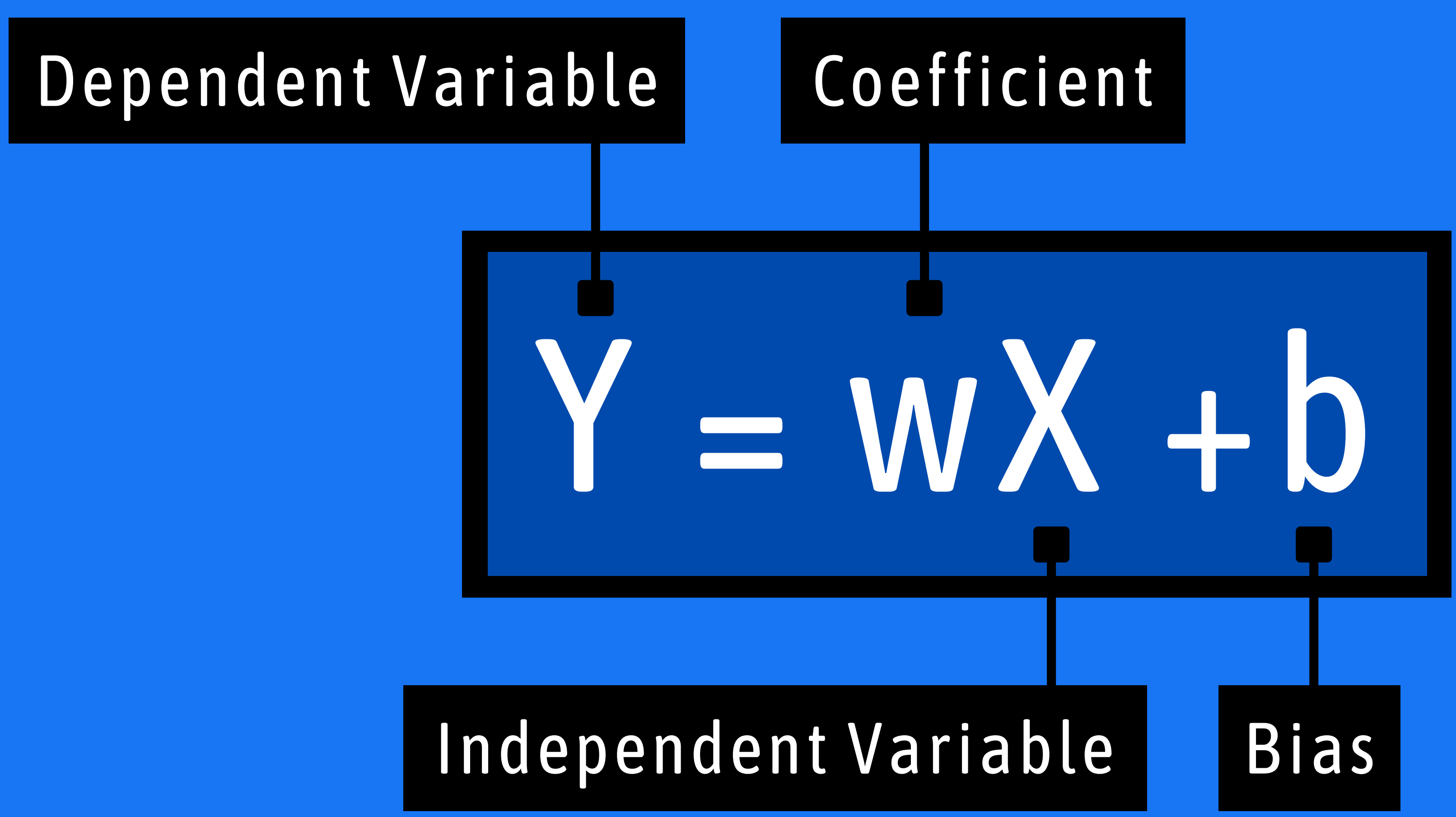


XPERTUP.COM

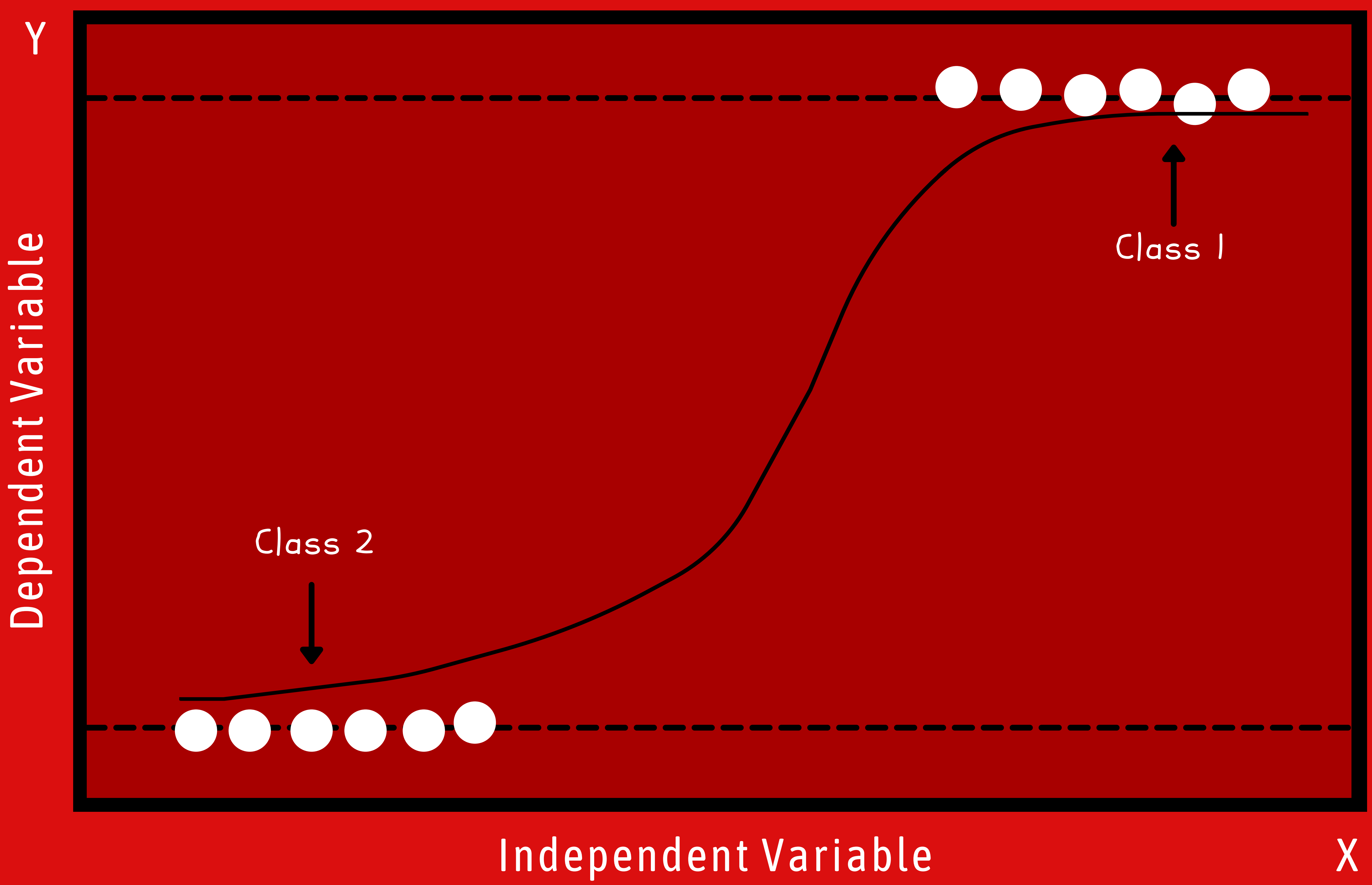
LINEAR REGRESSION



Linear Regression is a linear approach to modeling the relationship between a dependent variable - Y and one or more independent variables - X.



LOGISTIC REGRESSION



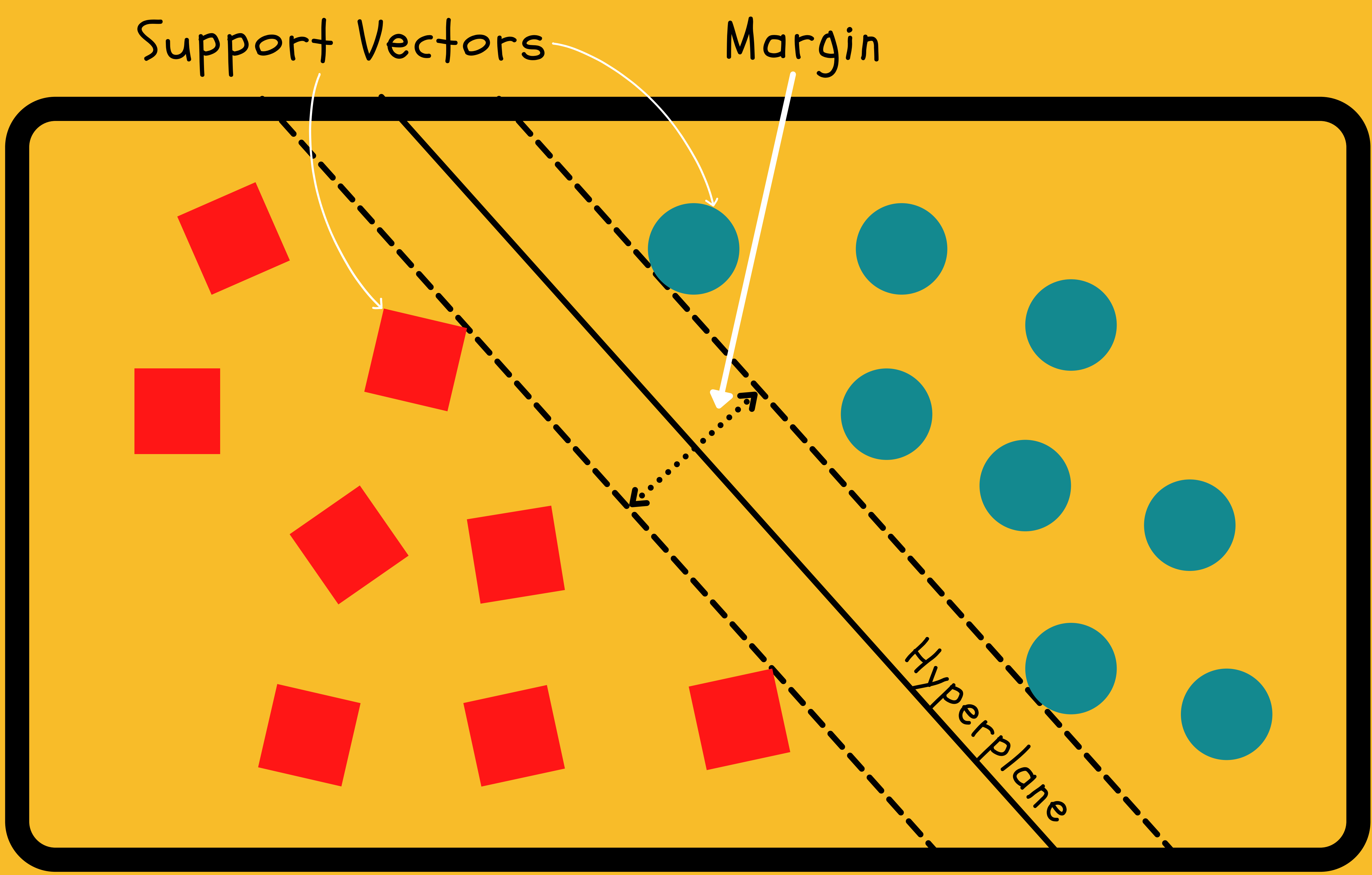
Logistic regression is a process of modeling the probability of a discrete outcome given an input variable.
The most common logistic regression models a binary outcome.

$$P = \frac{1}{1 + e^{-(b + wx)}}$$

Linear Model

XPERTUP.COM

SUPPORT VECTOR MACHINES



Support Vector Machine (SVM) uses kernel to transform your data and then based on these transformations it finds an optimal hyperplane that distinctly classfies the data points.

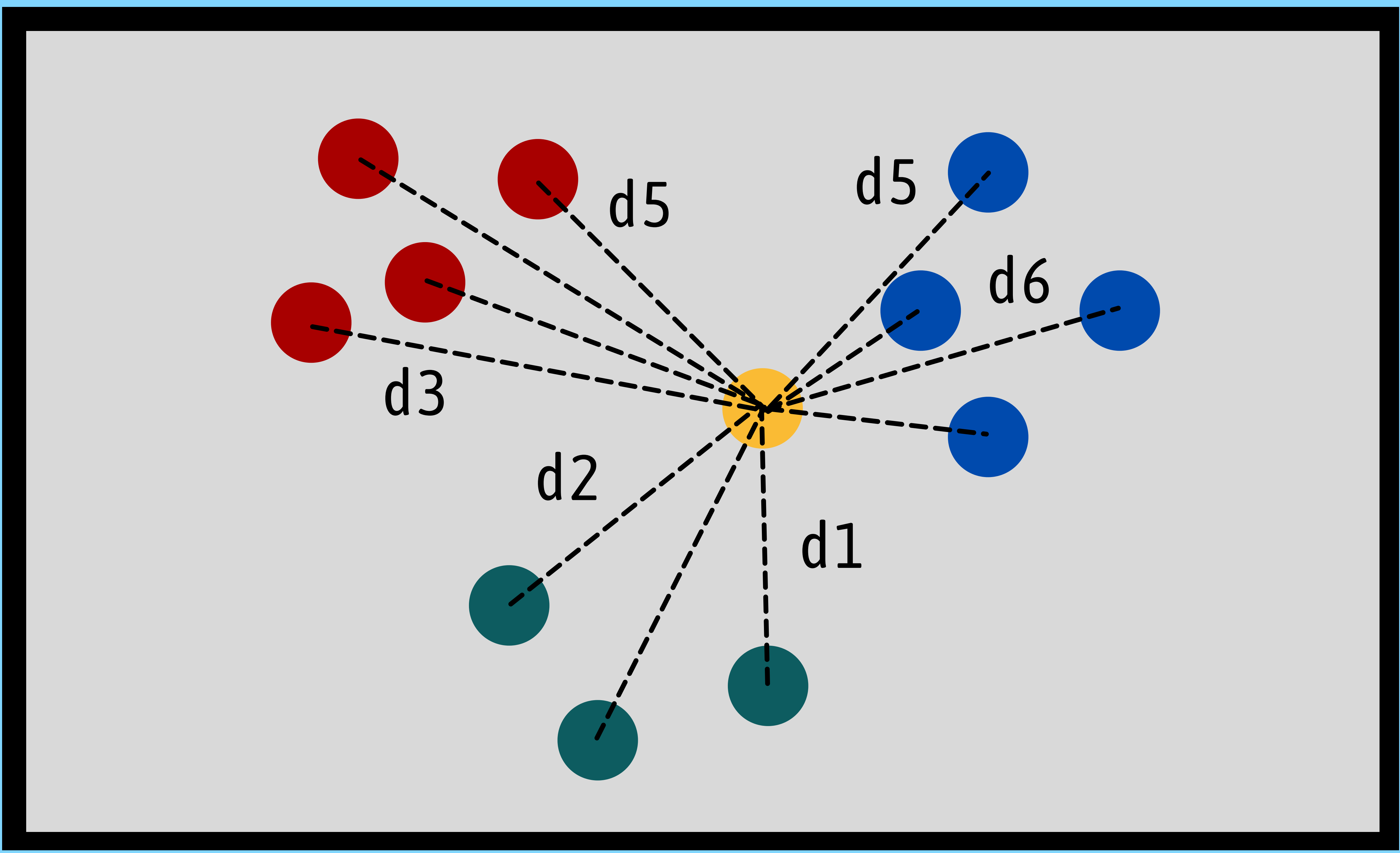
Hyperplane

Decision boundary with the maximum margin.

Support Vectors

Data points closer to Hyderplane.

KNN CLASSIFIER



K-Nearest Neighbour (k-NN) is a non-parametric classification method used for classification and regression.

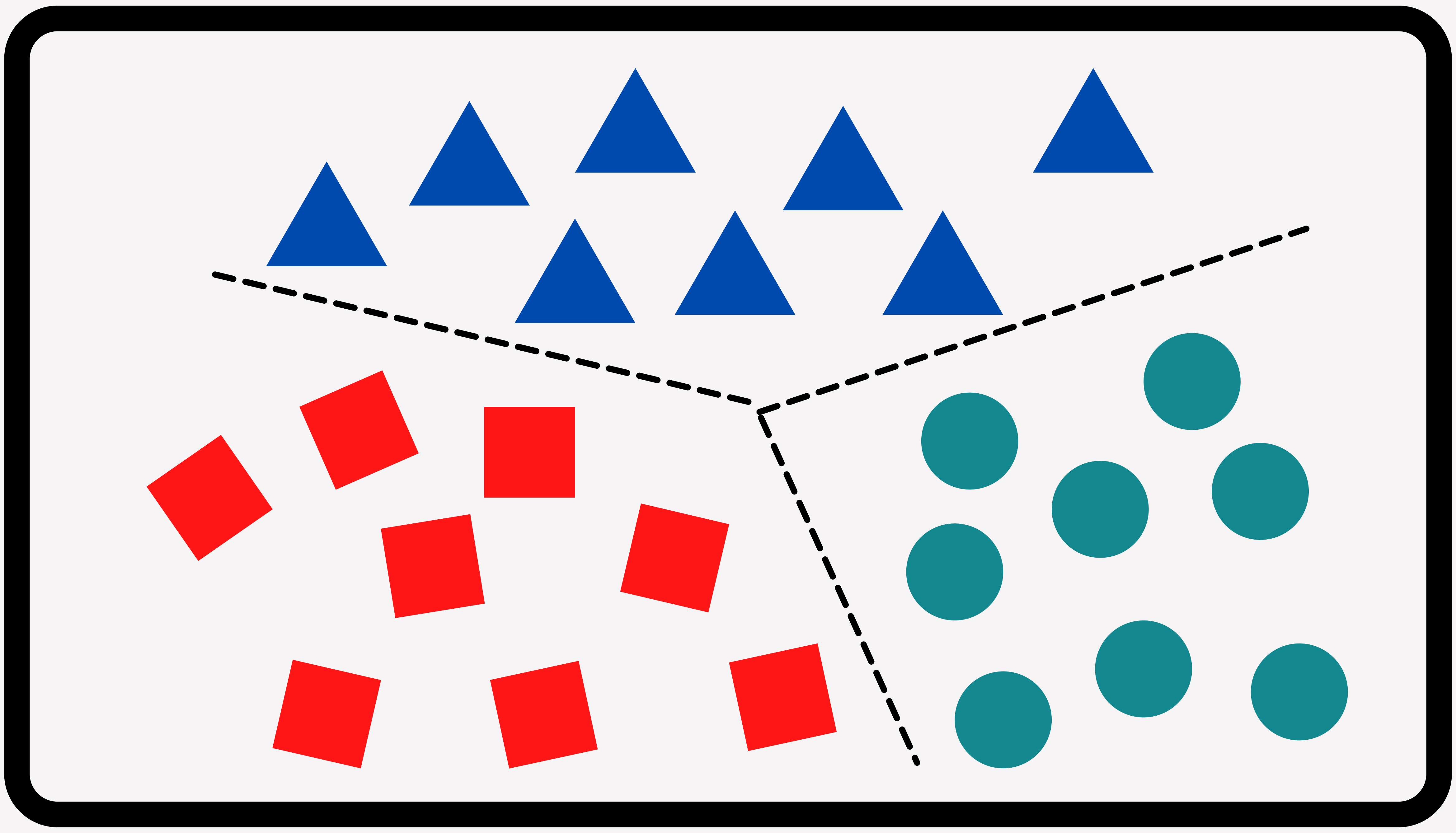
Classification

A point is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.

Value of K

Value of K is chosen in such a way that it gives the smallest errors on the validation set. Small K [Low Bias, High Variance]. Large K [High Bias, Low Variance].

NAIVE BAYES CLASSIFIER



Naive Bayes is a simple probabilistic classifiers based on applying Bayes Theorem with strong (naive) independence assumptions between the dataset features.

The likelihood which is the probability of data given class.

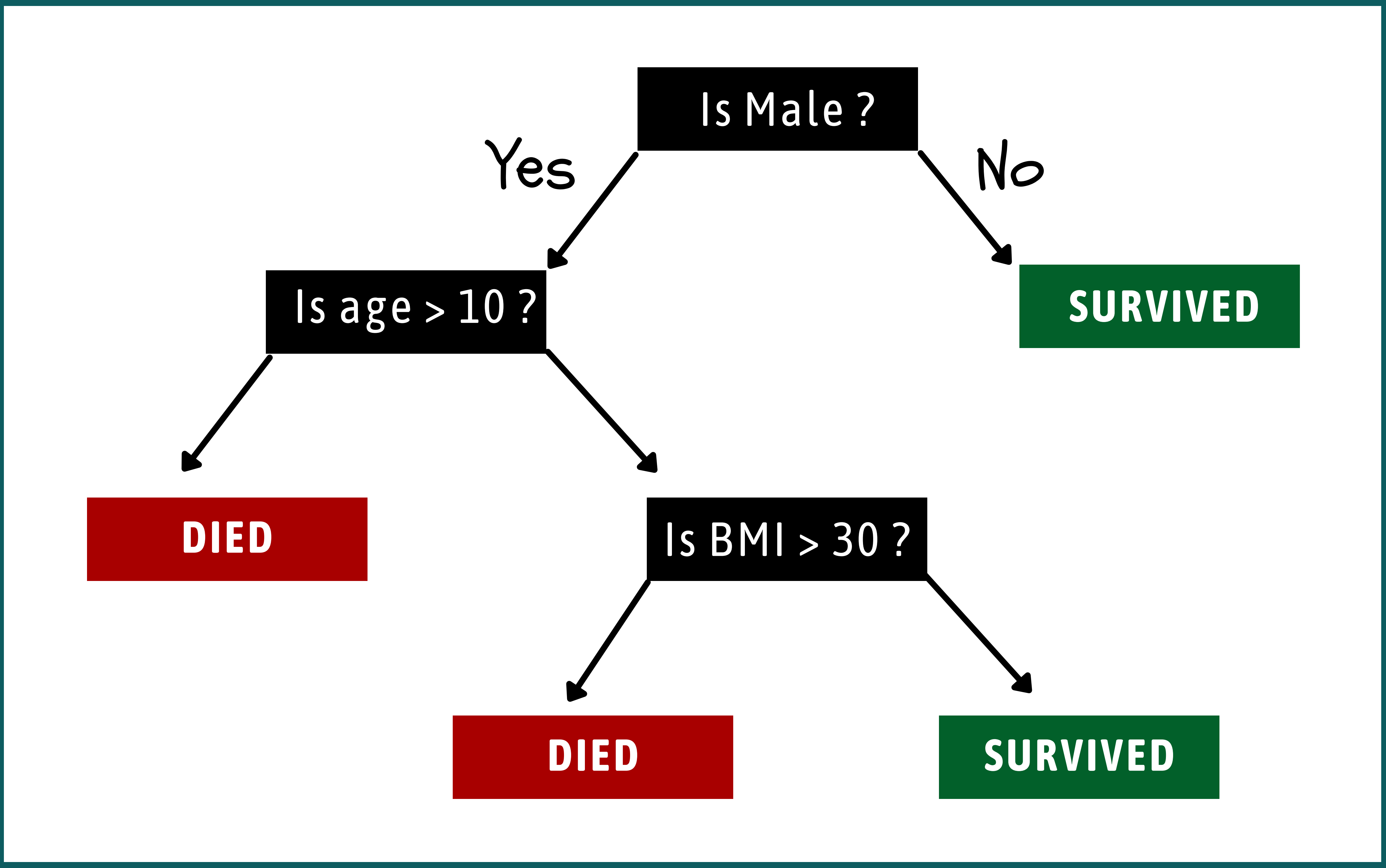
The prior probability of class.

$$P(C|A) = \frac{P(A|C) \times P(C)}{P(A)}$$

The posterior probability of class given attribute/data.

The prior probability of data/attribute.

DECISION TREE



A Decision tree used for classification is a flowchart like tree structure drawn upside down.

Each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node (terminal node) holds a class label.

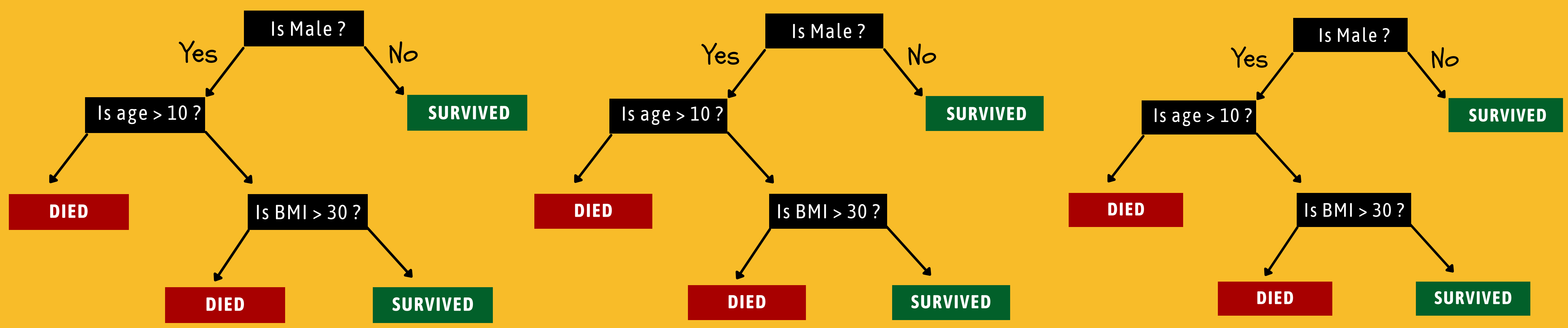
The example of decision tree above denotes whether the passenger died or survived.

Output Lables

Based upon attributes/columns from the data set namely sex, age and BMI.

RANDOM FOREST

Multiple Decision Trees are created using random subsets of features and bootstrapped dataset.



Each tree then votes, predicting the target class.

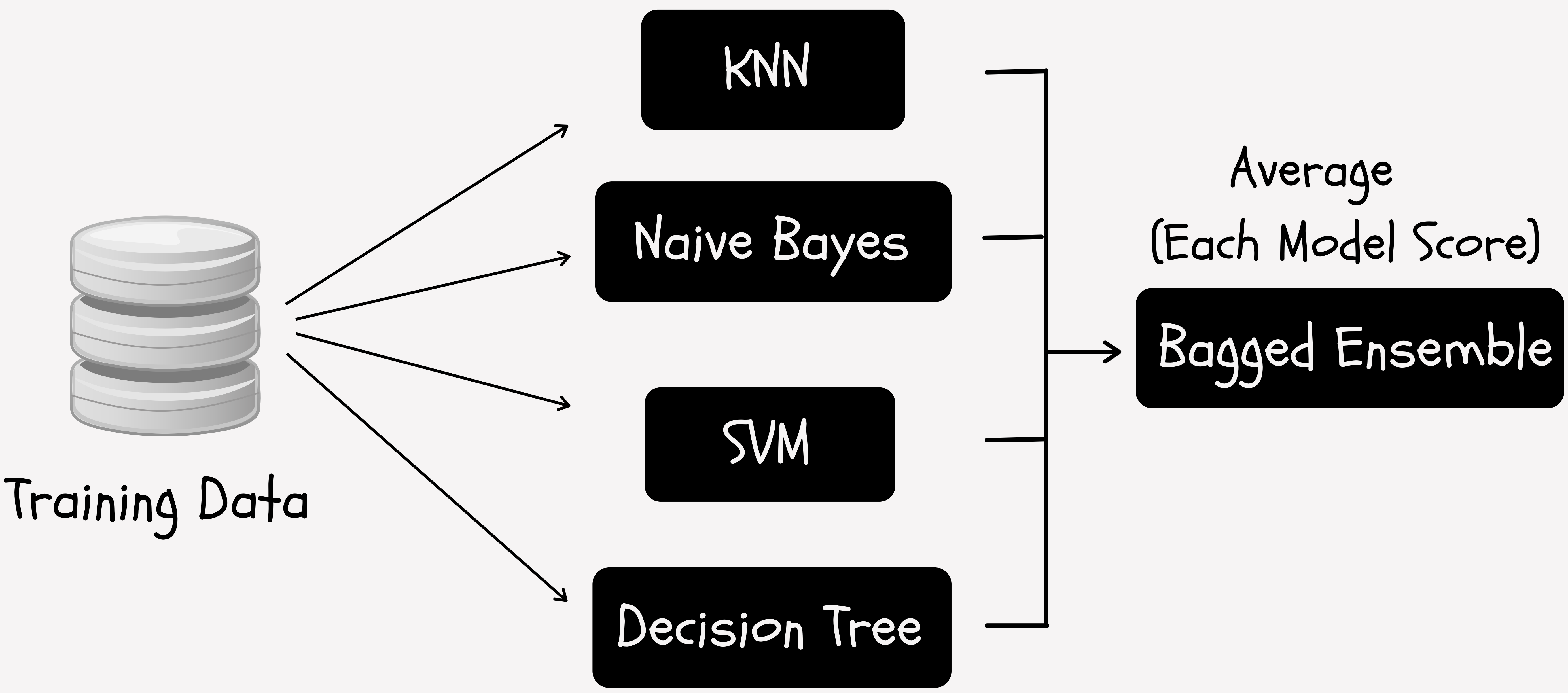
Majority Vote to a class, or average voting is used to get the

Final Prediction

XPERTUP.COM

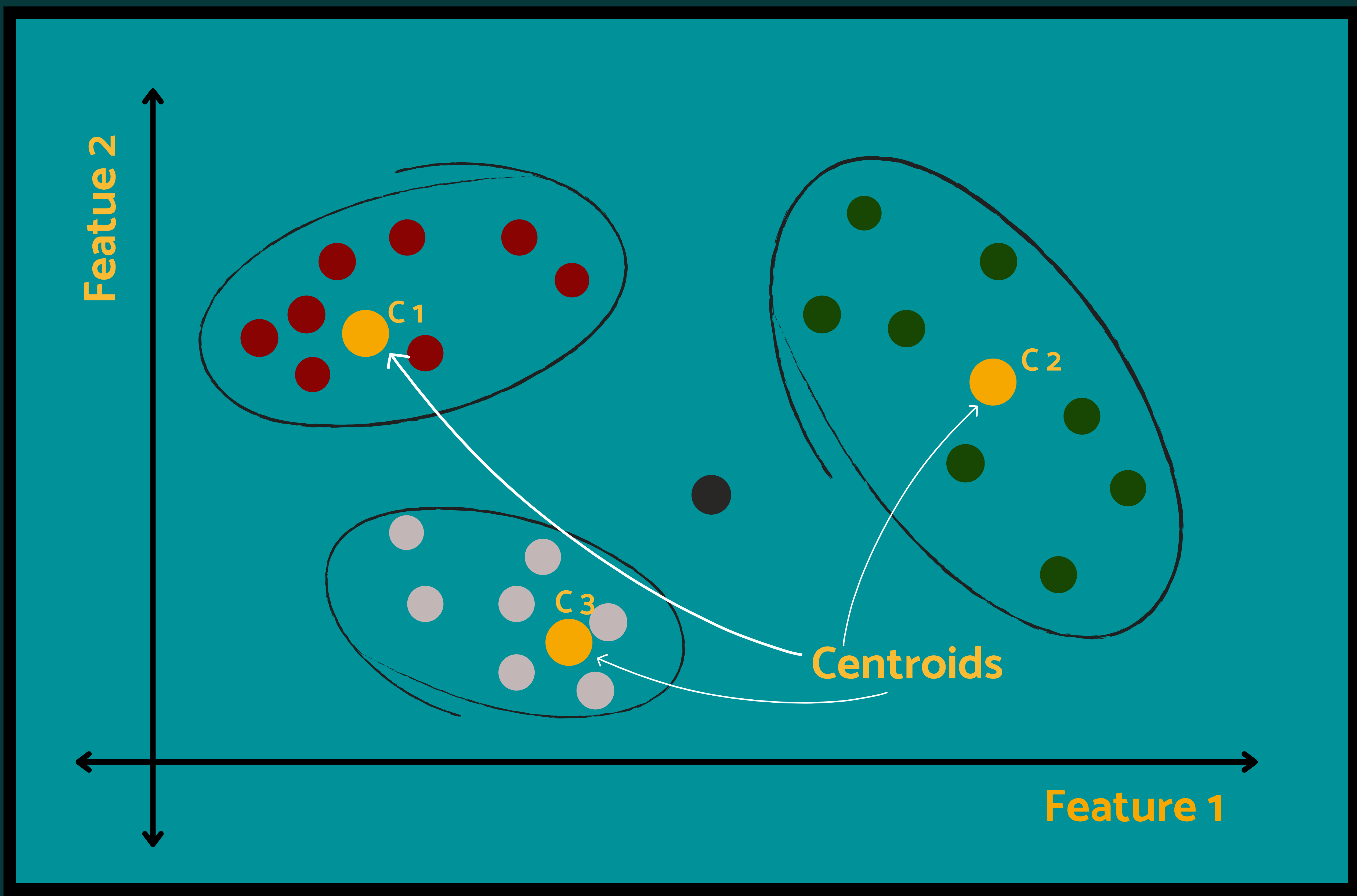
ENSEMBLE METHOD

When multiple models are trained separately and majority voting or average of the predictions from individual models are used to determine the final prediction, such a method is called Ensemble.



Bagging takes homogeneous weak learners, learns them independently from each other in parallel and combines them following some kind of deterministic averaging process.

K MEANS CLUSTERING



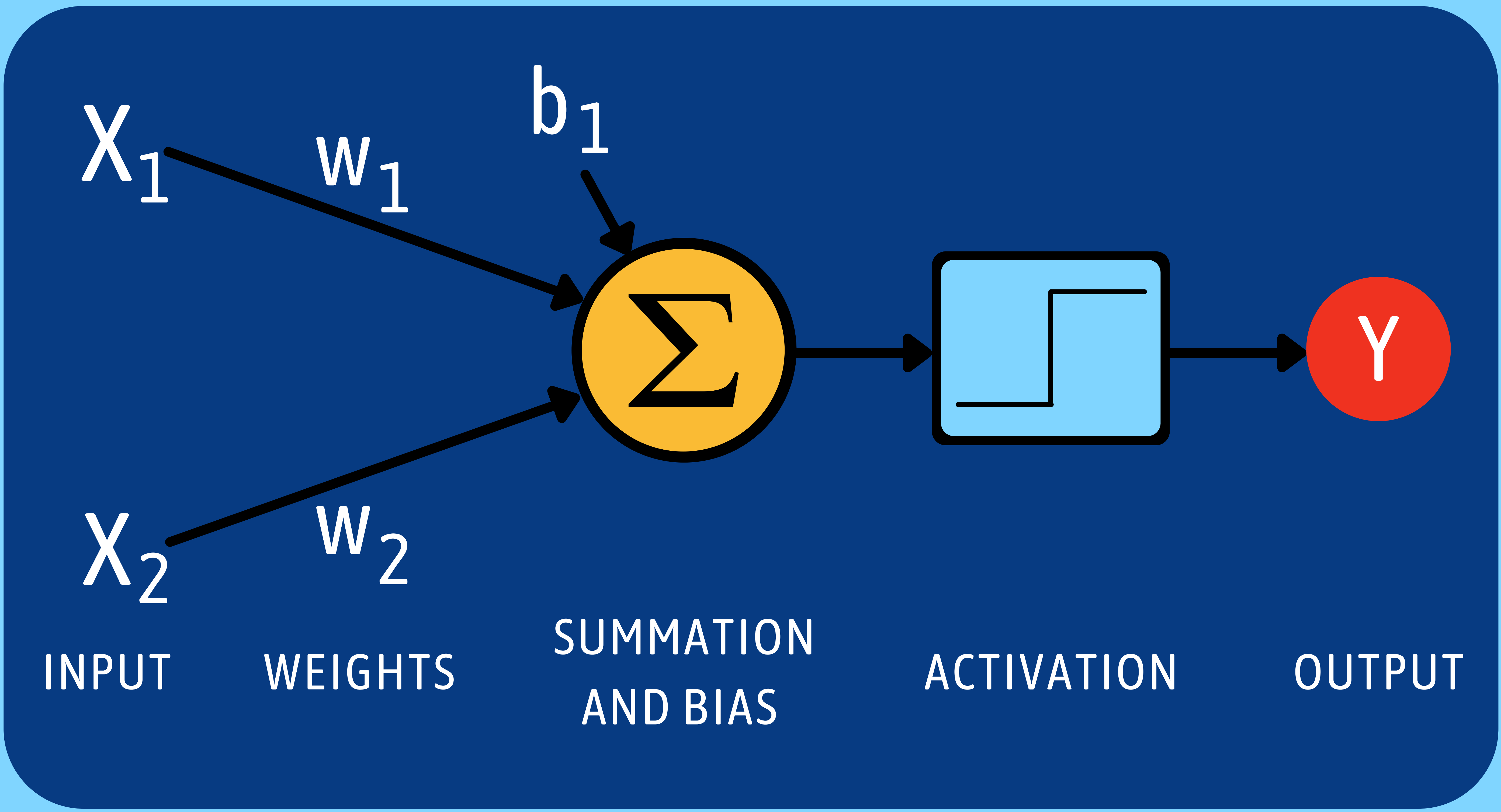
Clustering is the process of dividing the entire data into groups (also known as clusters) based on the patterns in the data.

The main objective of the K-Means Clustering is to minimize the sum of distances between the points and their respective cluster centroid (C1, C2, C3).

Stopping criteria used for K-means:

1. Centroids of newly formed clusters do not change.
2. Points remain in the same cluster.
3. Maximum no of iterations are reached.

PERCEPTRON



Perceptrons are the building blocks of a single layer in a neural network, made up of 4 parts:

- Input Values (X)
- Weights(w) and Bias (b)
- Net sum
- Activation function

1. Inputs (X) are multiplied with their weights w.
2. Add all the multiplied values and call them Weighted Sum.
3. Apply that weighted sum to the Activation Function for output (y)

ACTIVATION FUNCTIONS

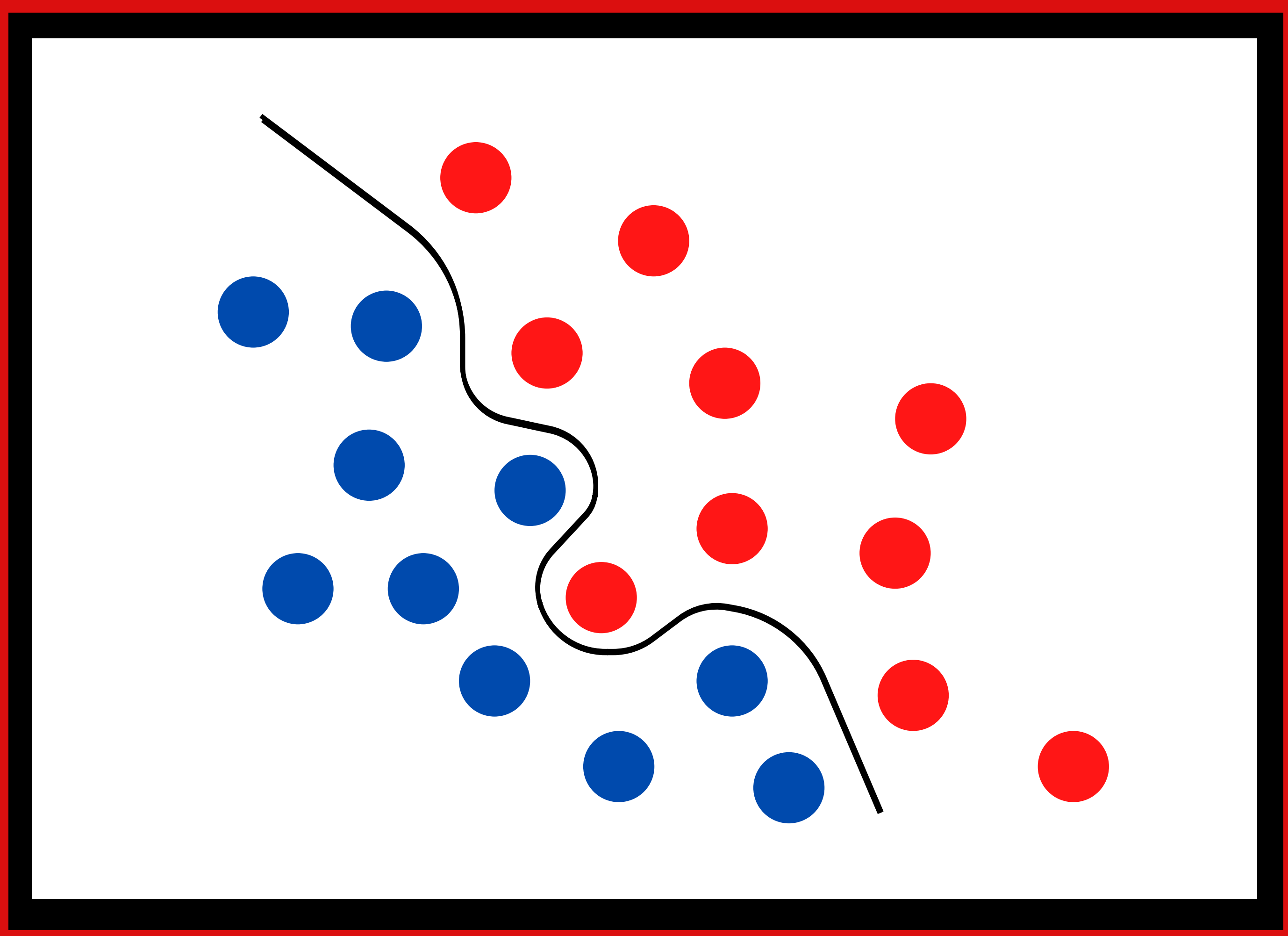
Activation functions is attached to each neuron in the network, and determines whether it should be activated ("fired") or not, based on whether each neuron's input is relevant for the model's prediction.

Activation functions also helps to normalize the output of each neuron to a range between 1 and 0 or between -1 and 1.

Types of Activation Functions

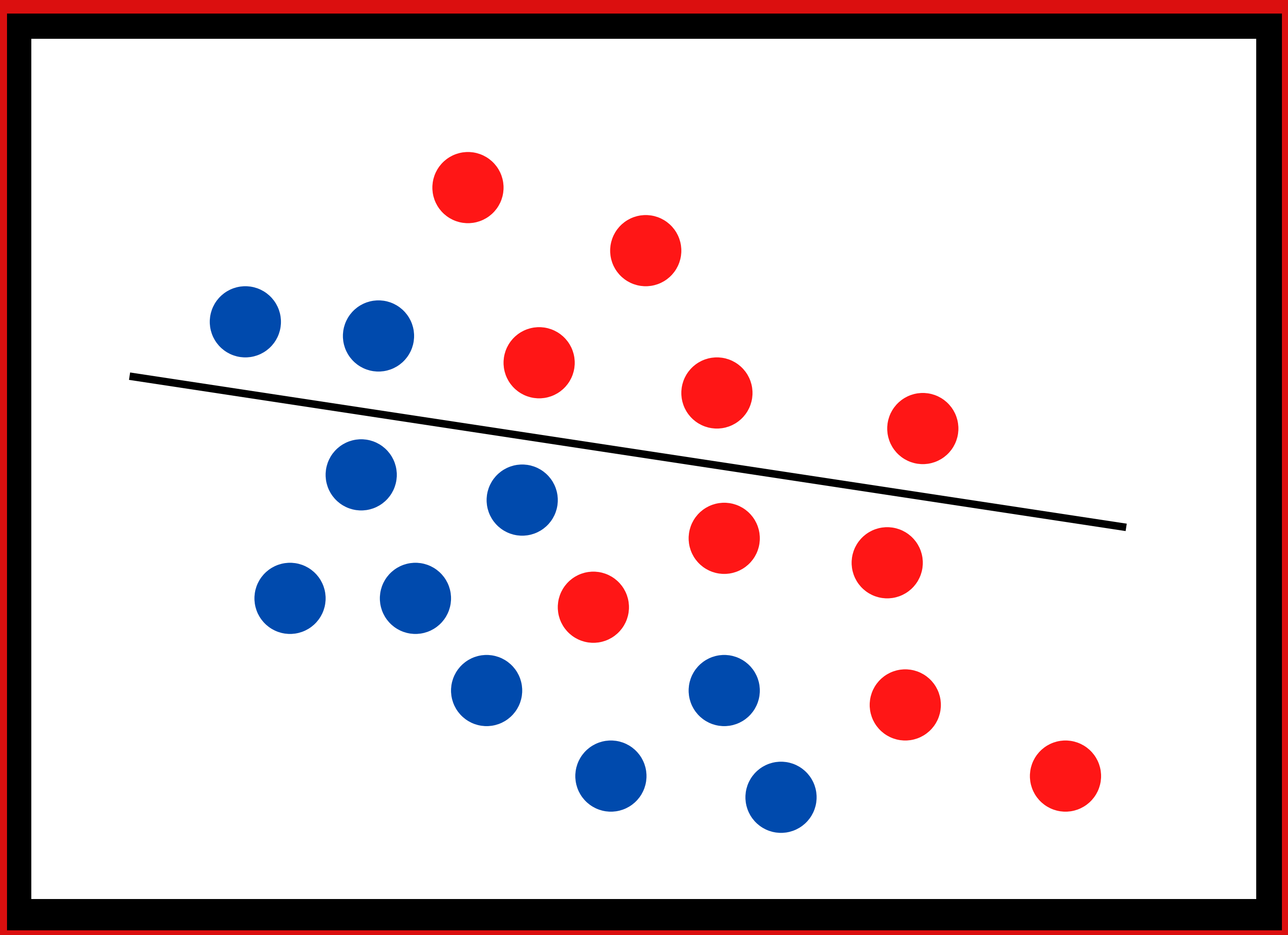
Linear	f(x) = ax
Binary	f(x) = 0 if x<0 else return 1
Sigmoid	f(x) = 1/(1+e^-x)
Tanh	f(x) = 2 sigmoid(2x)-1
ReLU	f(x) = max(0,x)
Leaky ReLU	f(x) = 0.01x if x<0 else x
Swish	f(x) = x*sigmoid(x)

MODEL FITTING



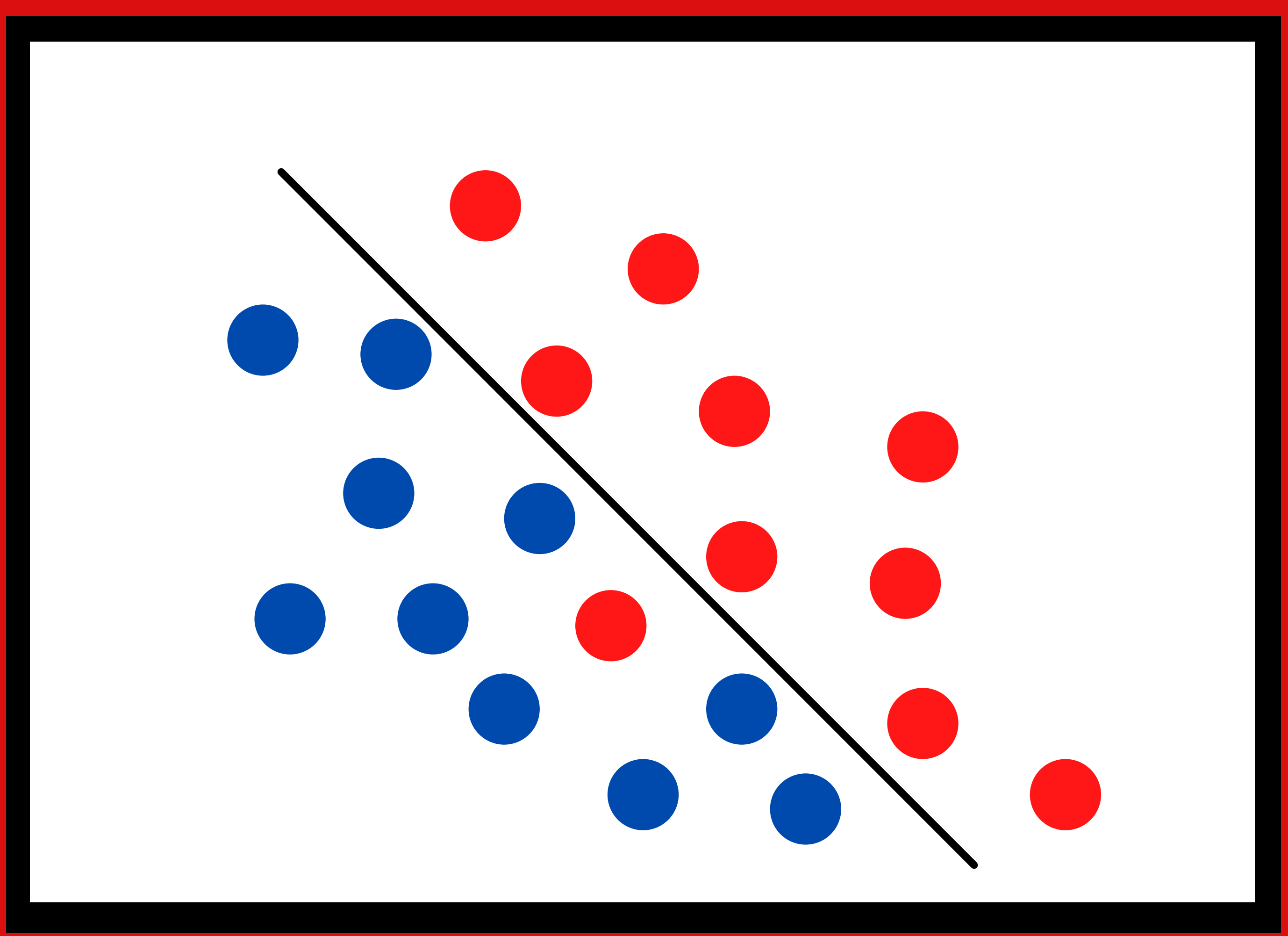
OverFit

High Variance!!



UnderFit

High Bias!!



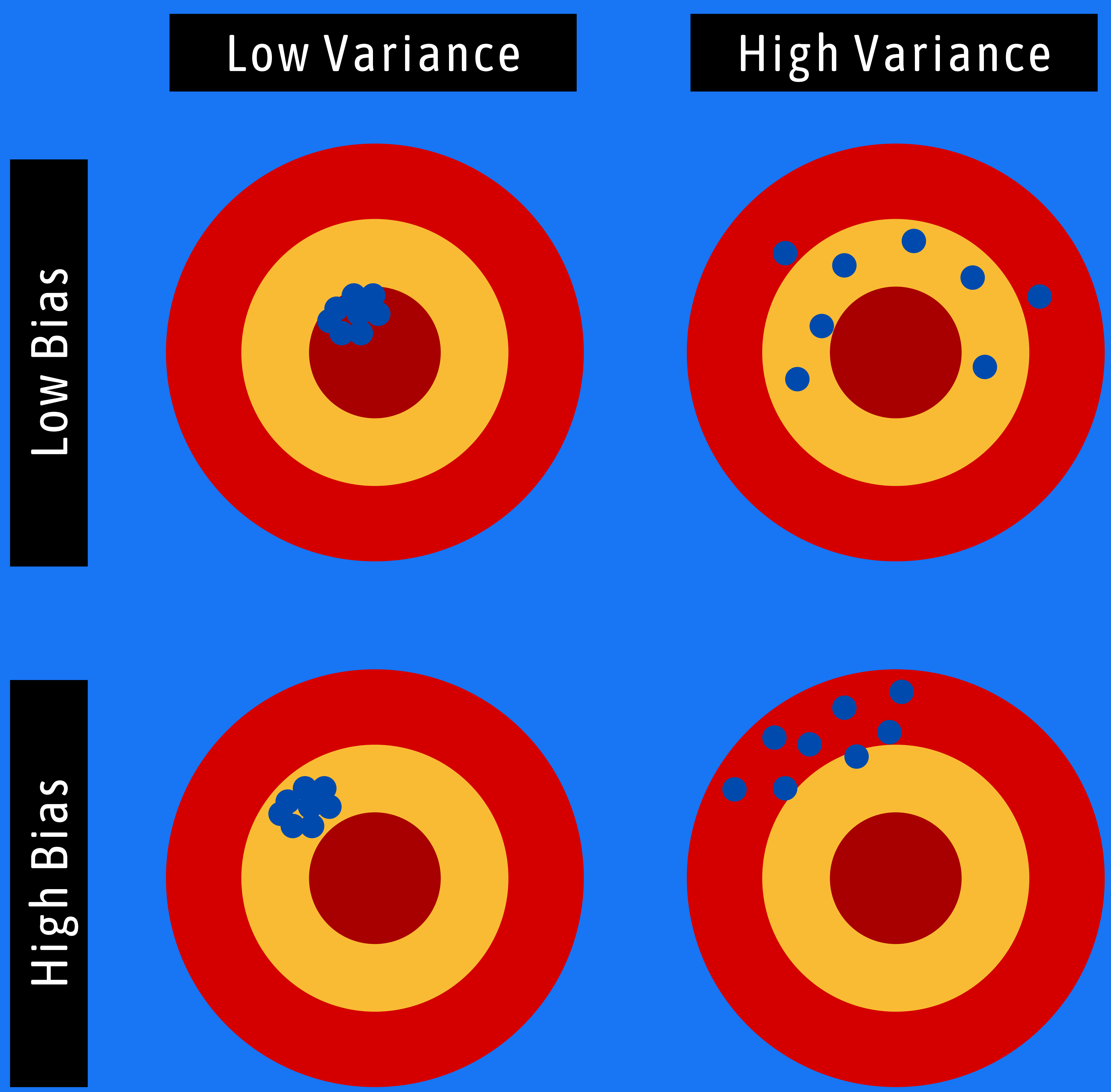
GoodFit

Optimal
Bias-Variance

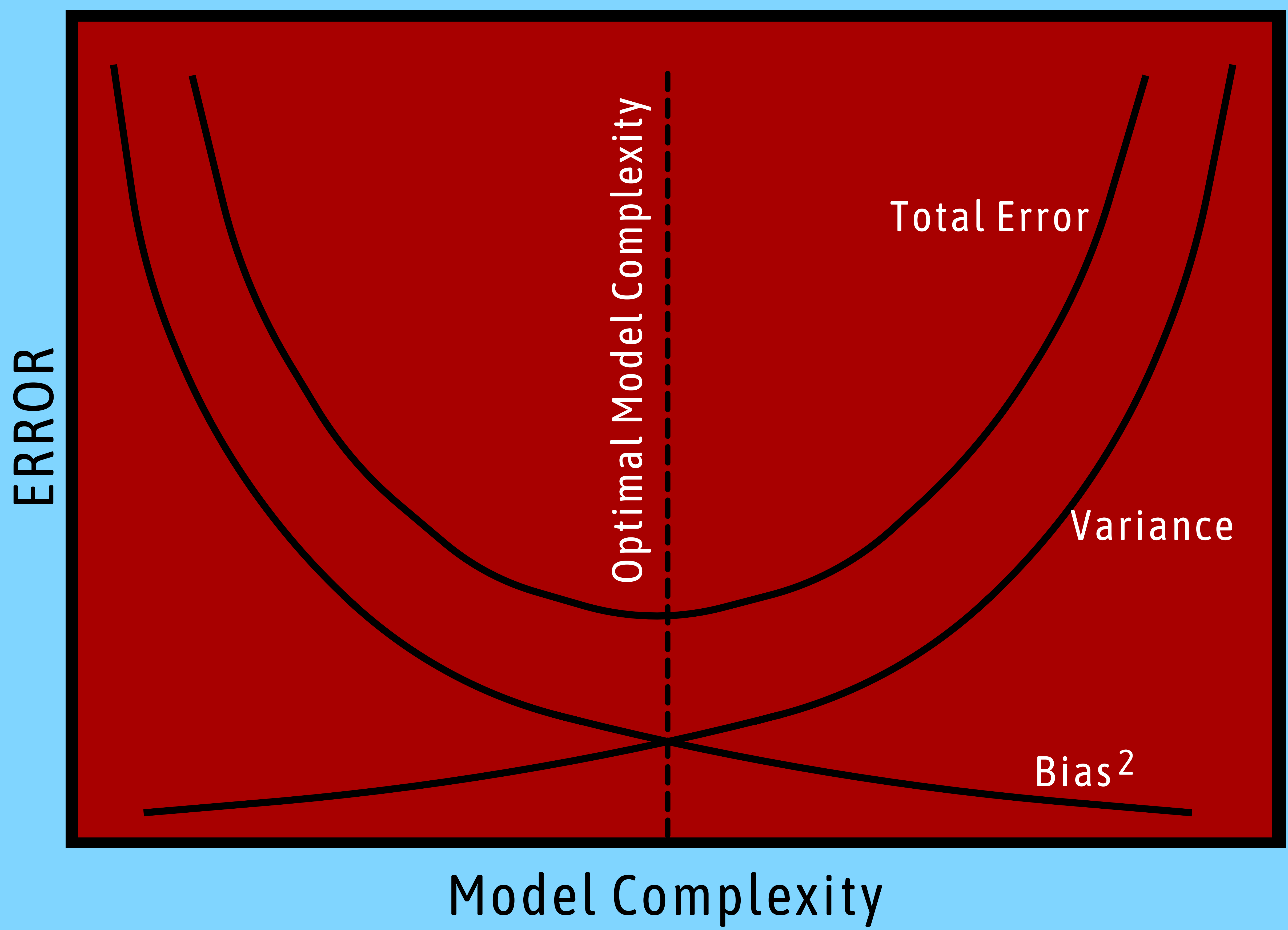
BIAS AND VARIANCE

Bias is the difference between the average prediction of our model and the correct value which we are trying to predict. Model with high bias means it is not fitting well on the training data.

Variance is the variability of model prediction for a given data point which tell us the spread of our data. Model with high variance means it is not able to make accurate predictions.



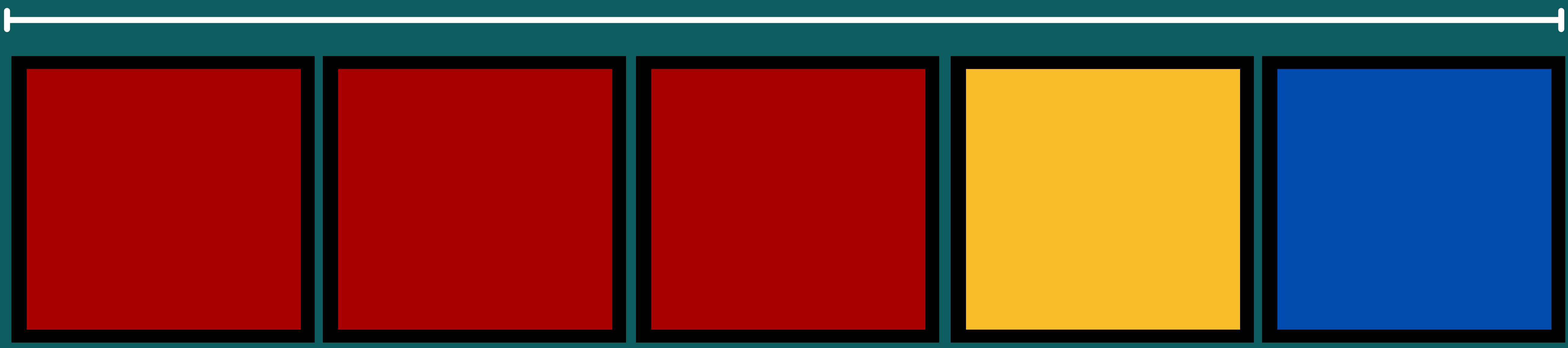
BIAS-VARIANCE TRADEOFF



Error = Bias^2 + Variance + Irreducible Error

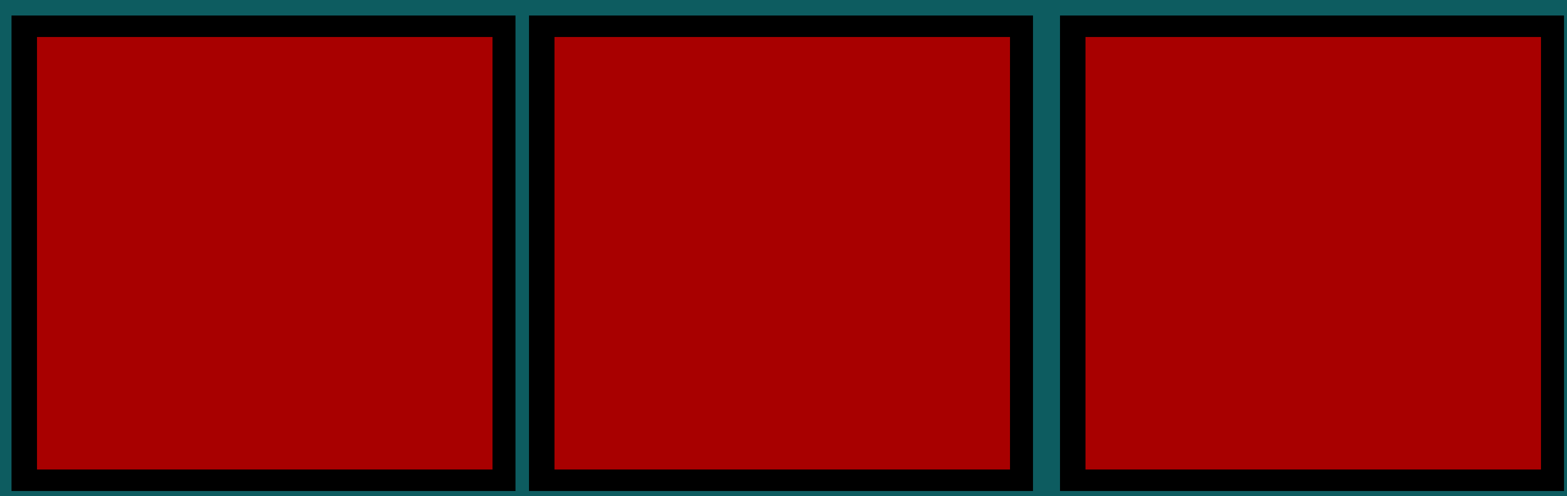
DATA SPLITTING

Complete Dataset



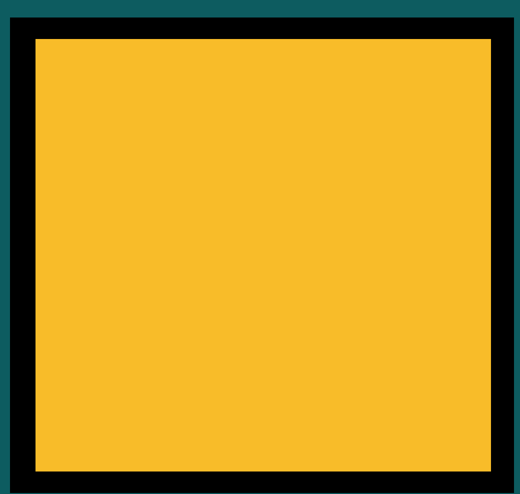
Divide dataset into into three parts to avoid overfitting and model selection bias.

Training Dataset



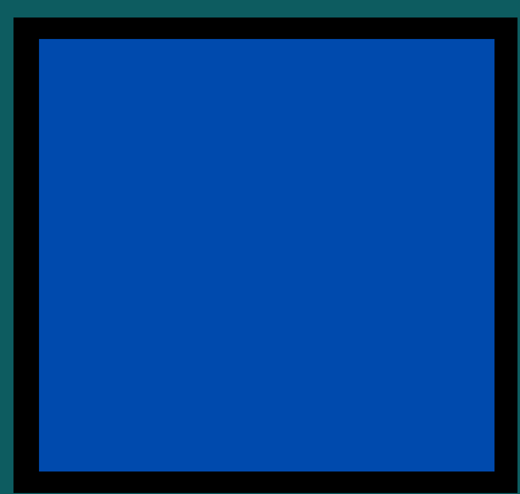
Sample of dataset used to train the model.

Validation Dataset



Evaluates accuracy while training the model and fine-tuining hyperparameters.

Testing Dataset

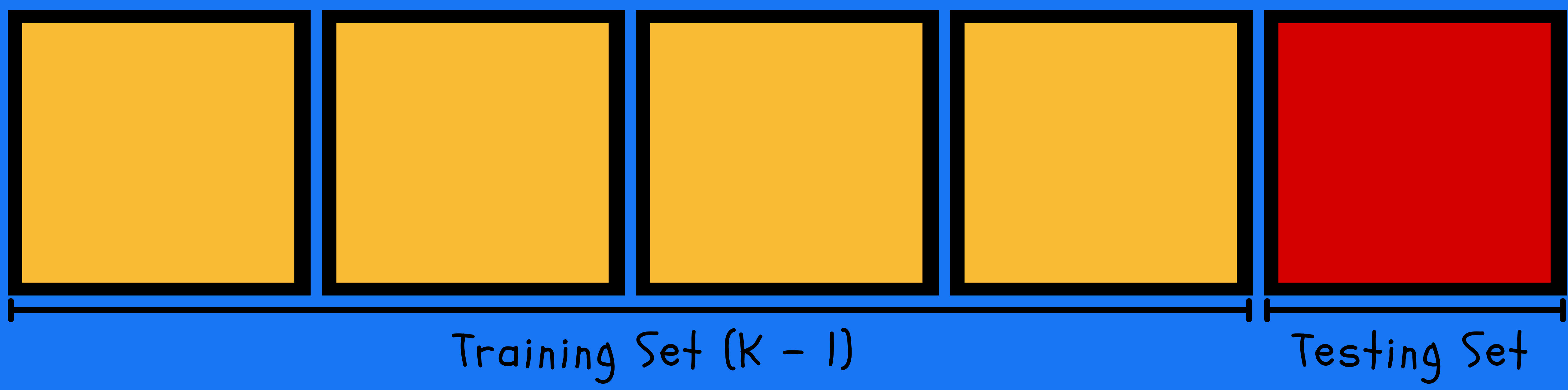


Model has never seen this set. Used for unbiased final evaluation of the model.

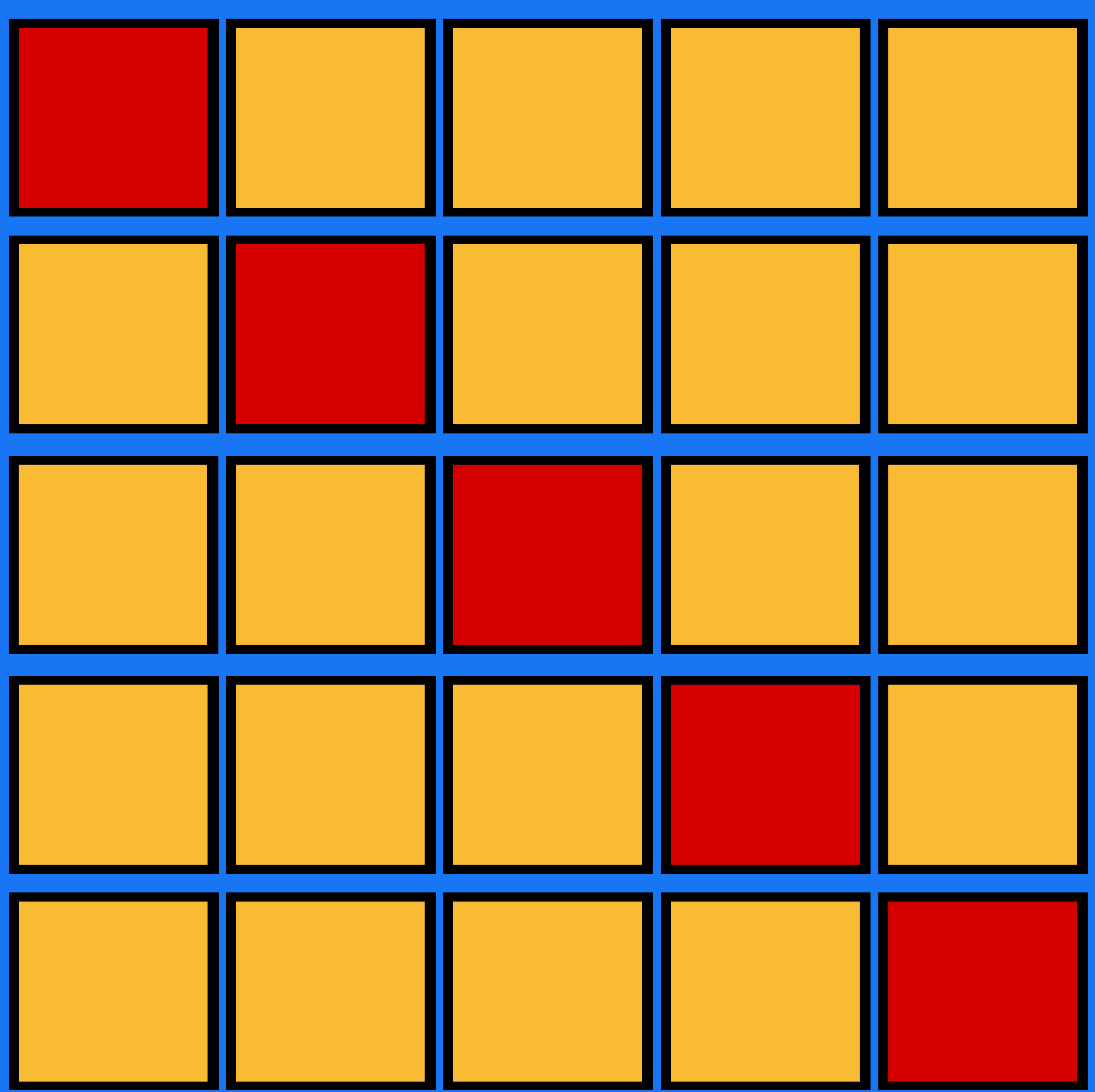
K-FOLD VALIDATION

Whole Dataset

Split Dataset - K Parts



Iterate



For each iteration, k-1 parts become training set (pink) used to train the model and remaining one part is the test set (purple) which evaluates the model.

Every part gets a chance to become test set.

Average of Evaluation score for each iteration is taken as the final evaluation score of the model.

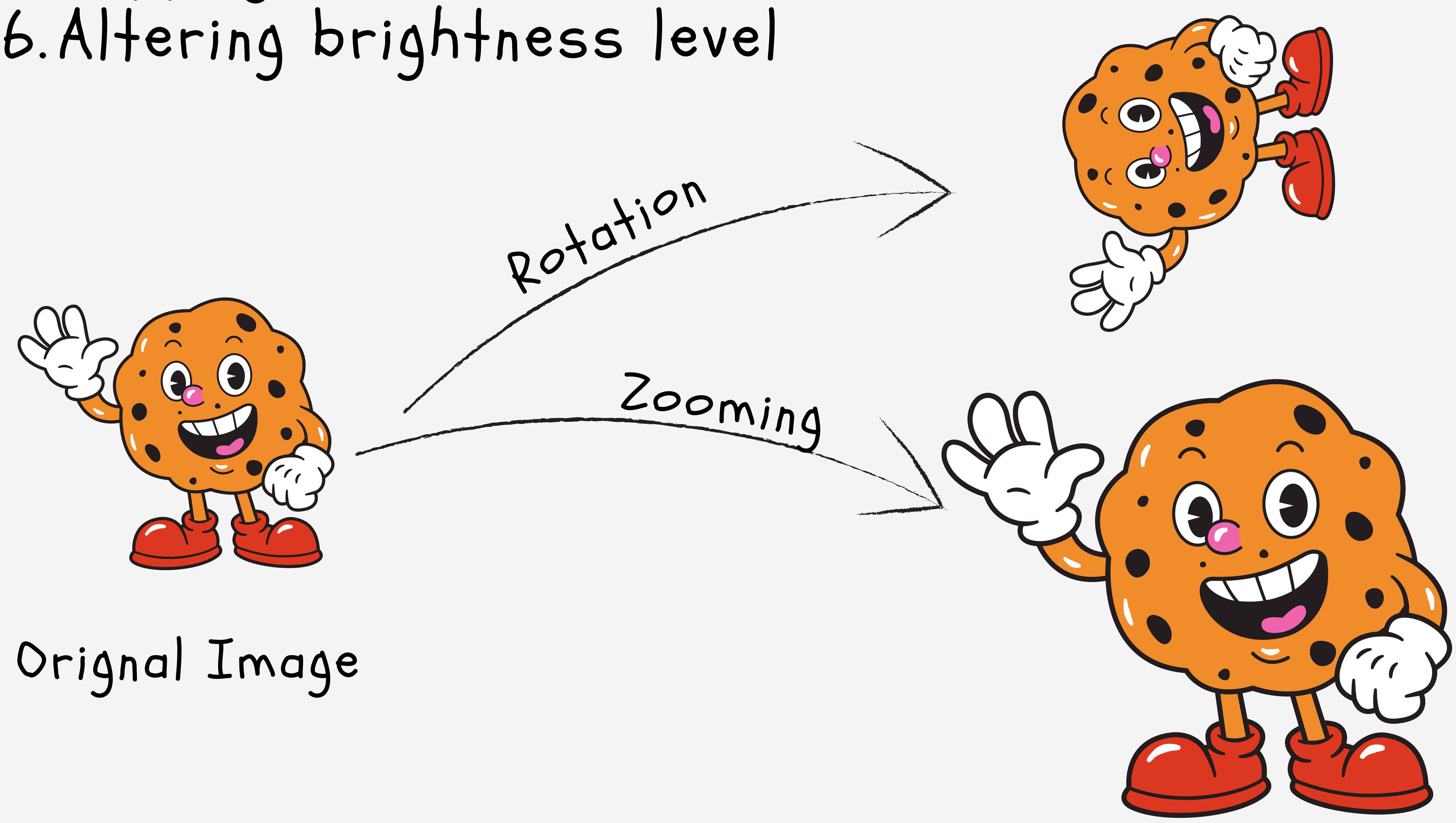
Data Augmentation

Large dataset usually helps in reducing overfitting.

Data augmentation is the process of increasing the amount of data. In this no new data is collected, rather we transform the already available data.

Eg. Image Dataset size can be increased by:

1. Rotation
2. Shearing
3. Zooming
4. Cropping
5. Flipping
6. Altering brightness level



XPERTUP.COM

L1 / L2 Regularization

Regularization techniques are used to reduce overfitting in the learning process.

Large weights in a neural network are a sign of a more complex network that has overfit the training data.

L1

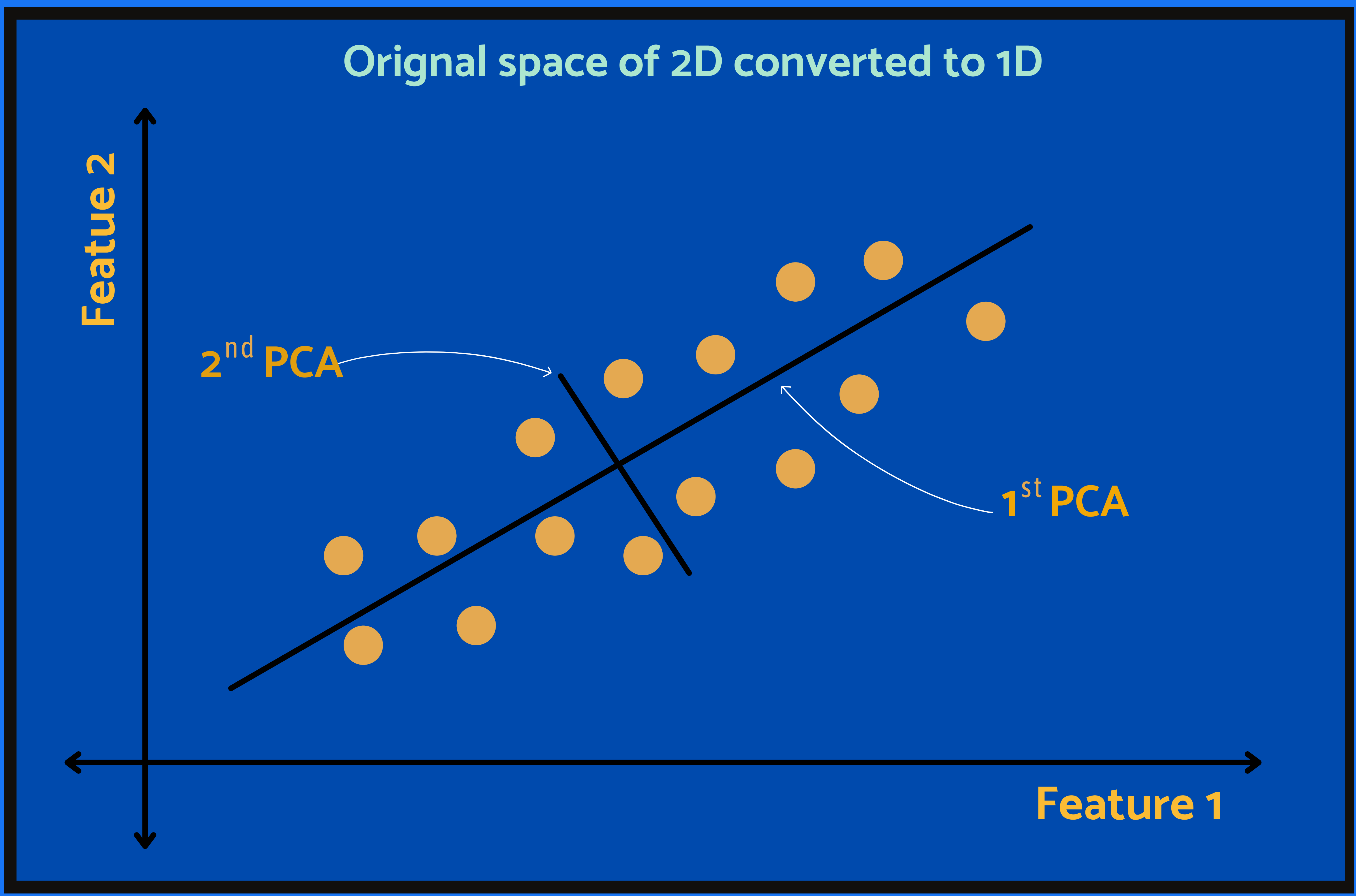
L1 regularization adds the penalty term in cost function by adding the absolute value of weight parameters.

L2

L2 regularization adds the penalty term squared value of weights in the cost function.

Difference between them is that L1 regularization tries to estimate the median of the data while the L2 regularization tries to estimate the mean of the data to avoid overfitting.

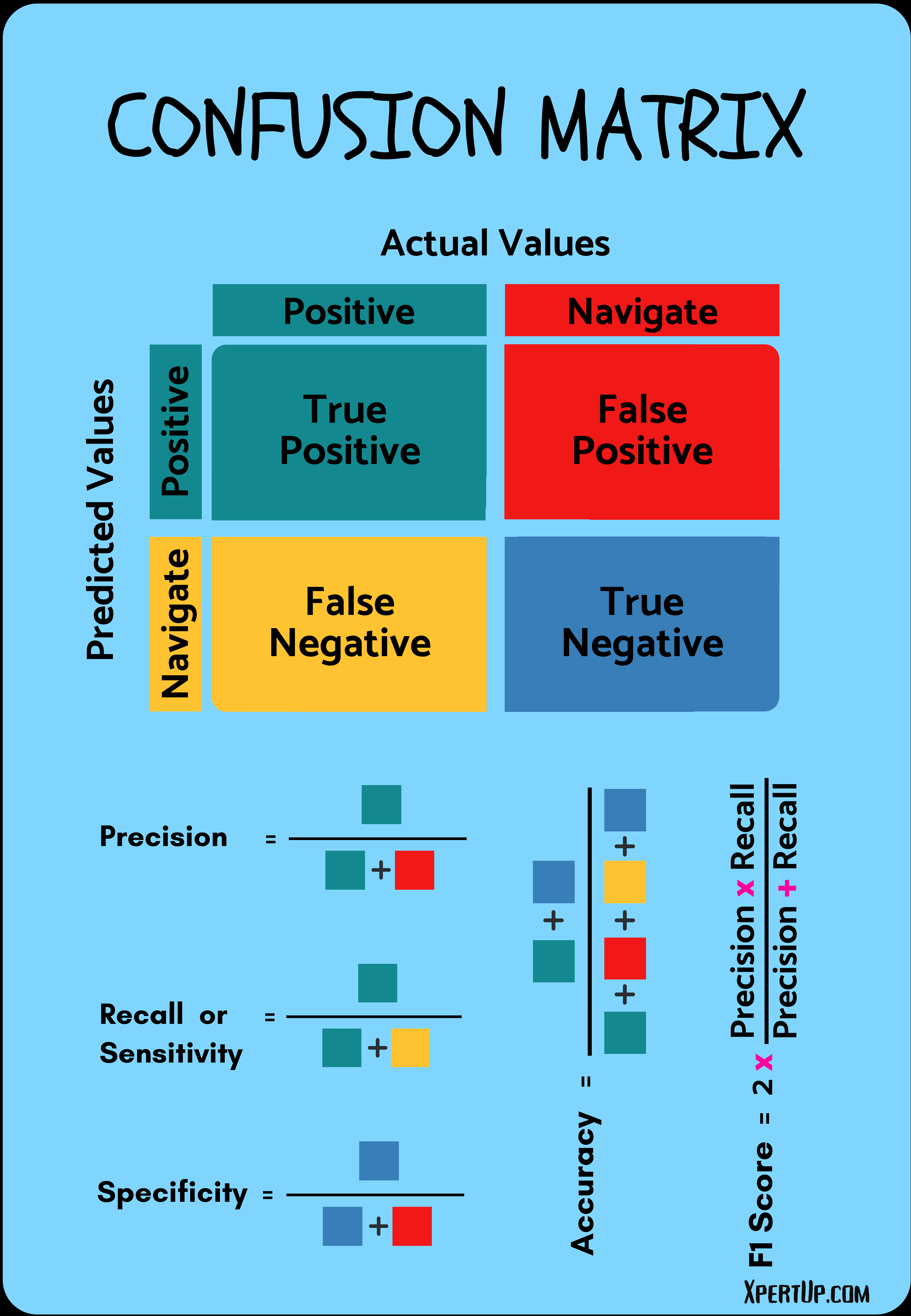
Principal Component Analysis



High dimensionality means that the dataset has a large number of features.

PCA is an unsupervised technique primarily used for features dimensionality reduction while losing only small information.

Problem associated with high dimensionality in the machine learning field is model overfitting, which reduces the ability to generalize beyond the examples in the training set.


CONFUSION MATRIX
Actual Values
Positive
Navigate
Predicted Values
Positive
Navigate
True Positive
False Positive
False Negative
True Negative
Precision =
+
Recall or Sensitivity =
+
Specificity =
+
Accuracy =
+
+
+
+
F1 Score = 2 x Precision x Recall / Precision + Recall
XPERTUP.COM