5 ALGORITHMS YOU MUST BE AWARE OF BEFORE GOING TO DEEP LEARNING

FOR A BETTER FUTURE ...

Ghosh 4 AI

Ghosh4AI

Some common challenges

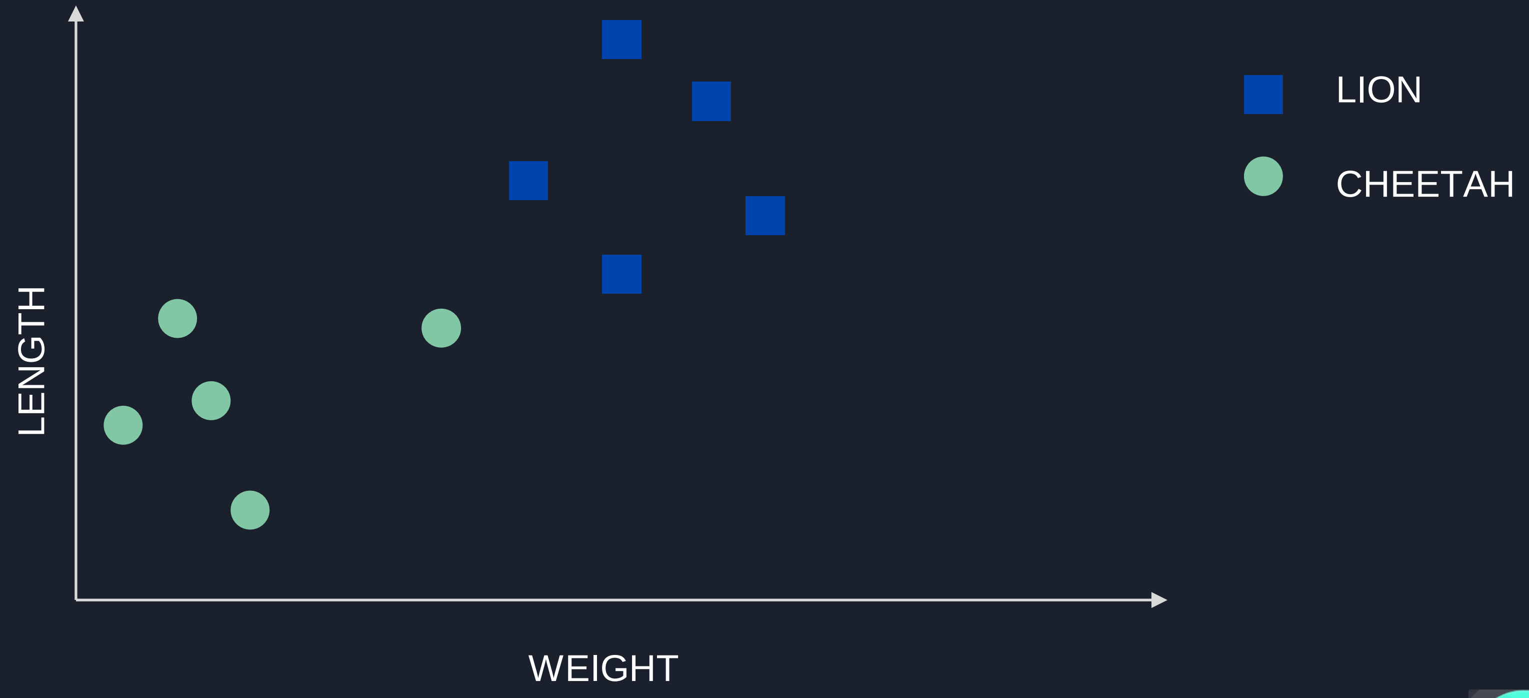
- Classification (KNN, Decision Trees, Neural Nets, SVM)
- Clustering (K-Means, Hierarchical, Fuzzy C Means)
- Density Estimation (Gaussian Mixture Models)
- Regression (Linear or Logistic Regression)
- Dimensionality Reduction (PCA, LDA, SVD)

and many more

K-Nearest Neighbor (KNN)

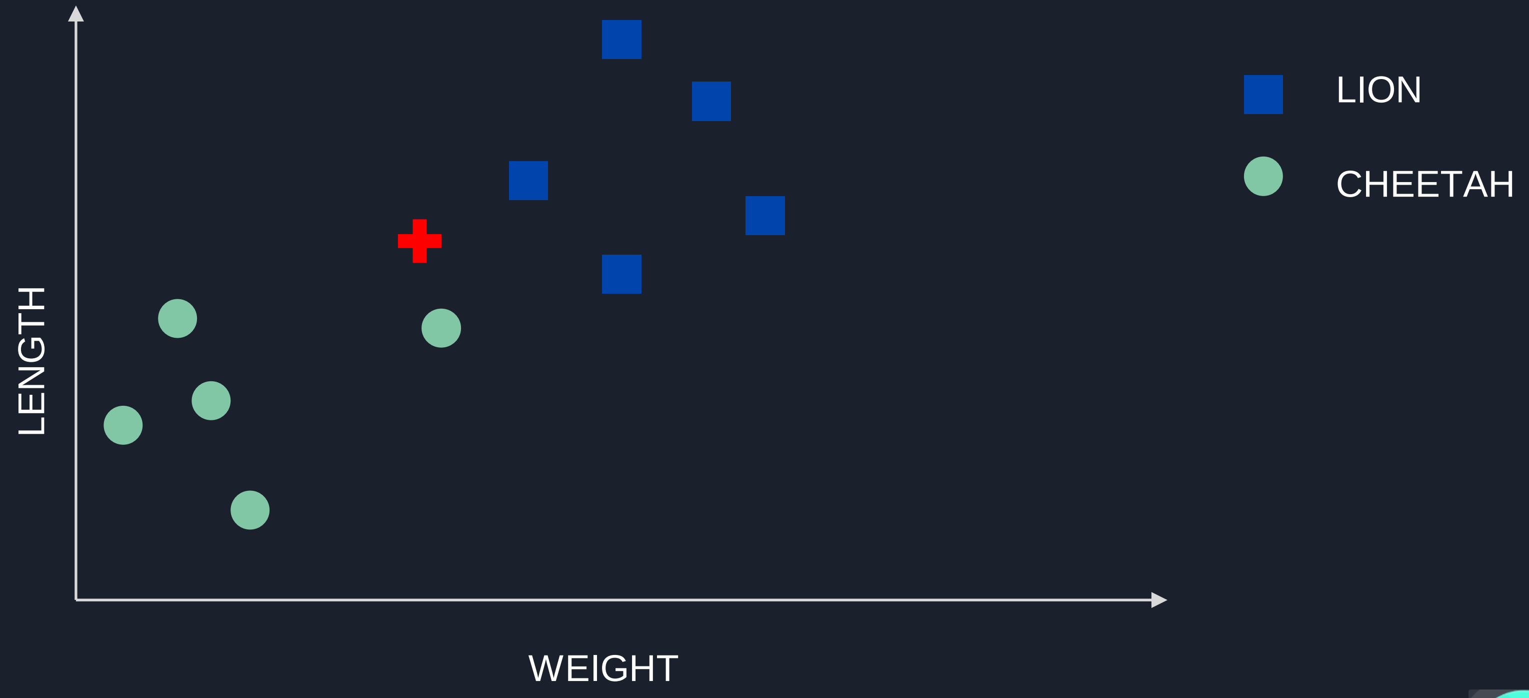
Ghosh4AI

The simplest idea is often the best idea



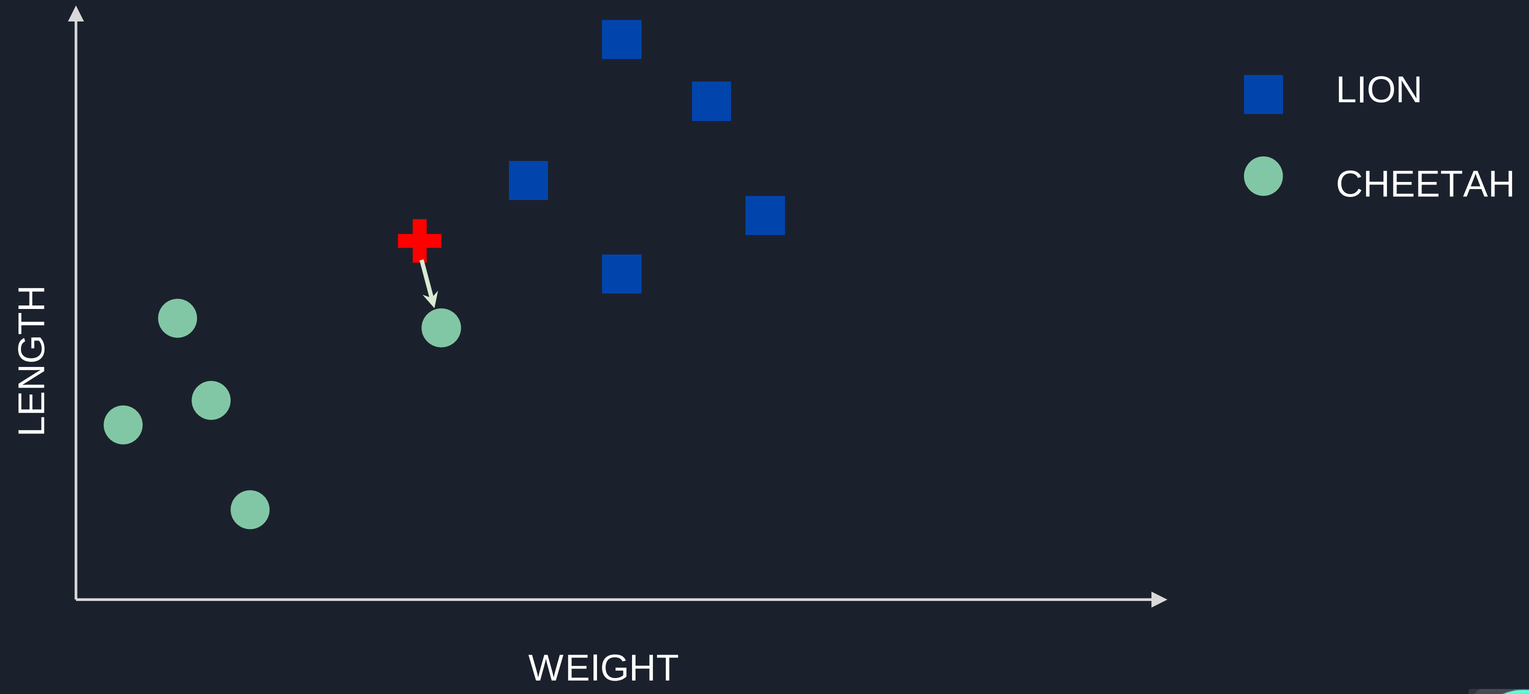
Ghosh4AI

The simplest idea is often the best idea



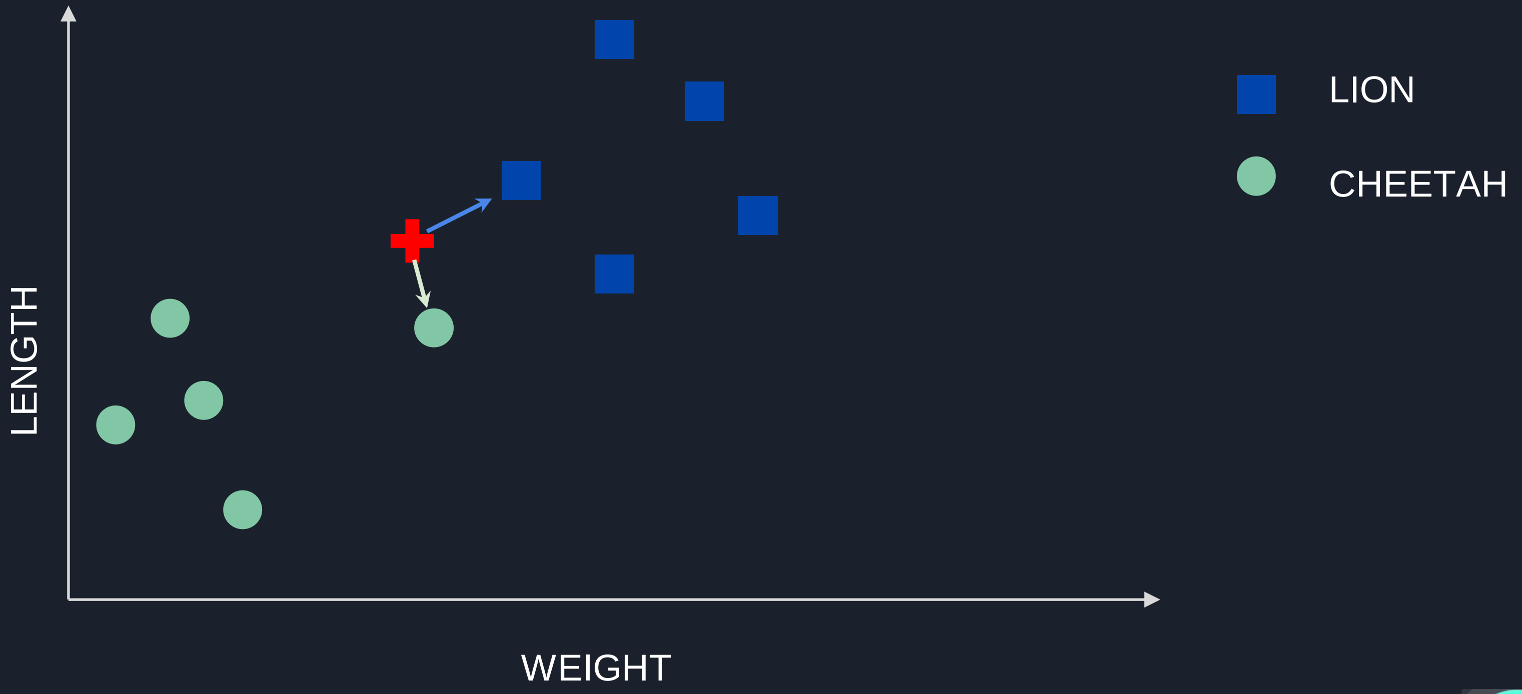
Ghosh4AI

The simplest idea



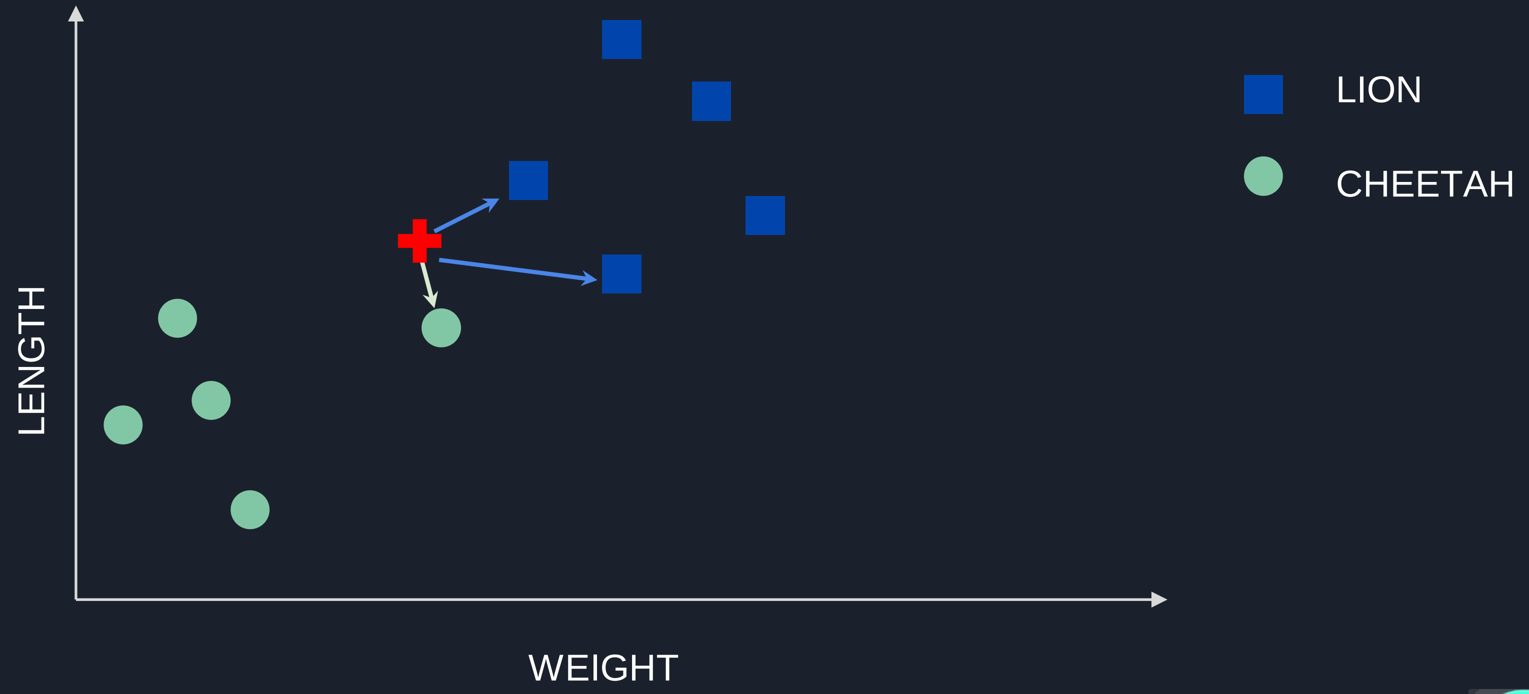
Ghosh4AI

Ambiguity



Ghosh4AI

More neighbours... Better predictions



Ghosh4AI

Further References

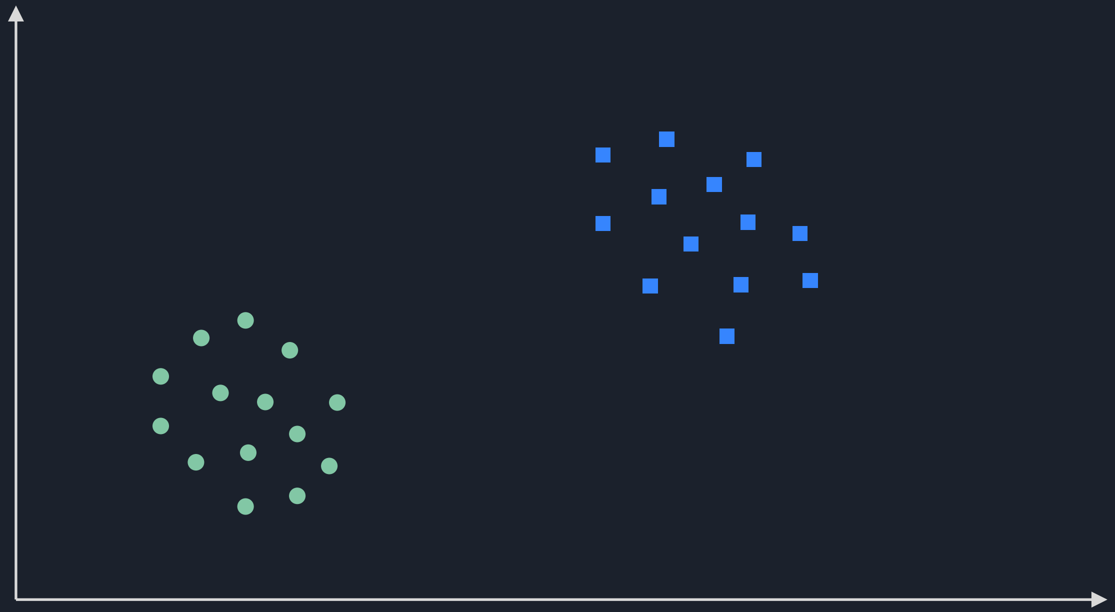
- Wikipedia - k-nearest neighbors algorithm - Wikipedia
- Blogs :
  - Introduction to KNN, K-Nearest Neighbors : Simplified - Analytics Vidhya
  - A Quick Introduction to K-Nearest Neighbors Algorithm - Medium
  - K-Nearest Neighbours - GeeksforGeeks
- Demo :
  - http://vision.stanford.edu/teaching/cs231n-demos/knn/
- API
  - sklearn.neighbors.KNeighborsClassifier — scikit-learn 0.19.2
  - k-nearest neighbor classification - MATLAB
  - Weka 3 - Data Mining with Open Source Machine Learning Software
- Books :
  - Machine Learning – T. Mitchell
  - Pattern Recognition and Machine Learning – C.M. Bishop
- Related Papers
  - **K-nearest neighbour**
  - A fuzzy **k-nearest neighbor** algorithm
  - The distance-weighted **k-nearest-neighbor** rule
  - When is “**nearest neighbor**” meaningful?

Ghosh4AI

K-Means Clustering

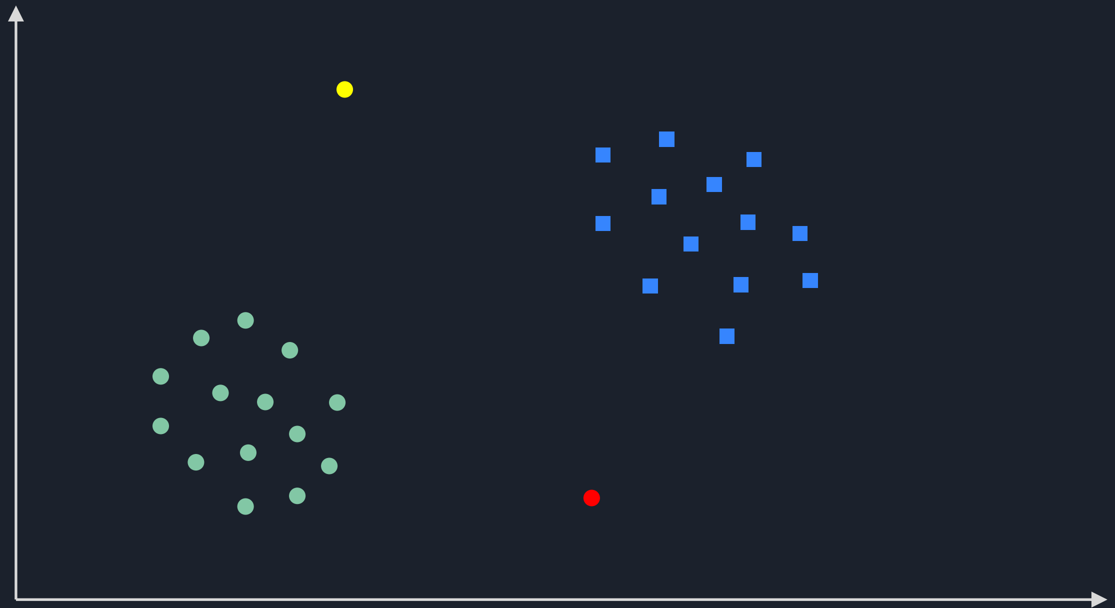
Ghosh4AI

Clustering : Dividing samples into groups



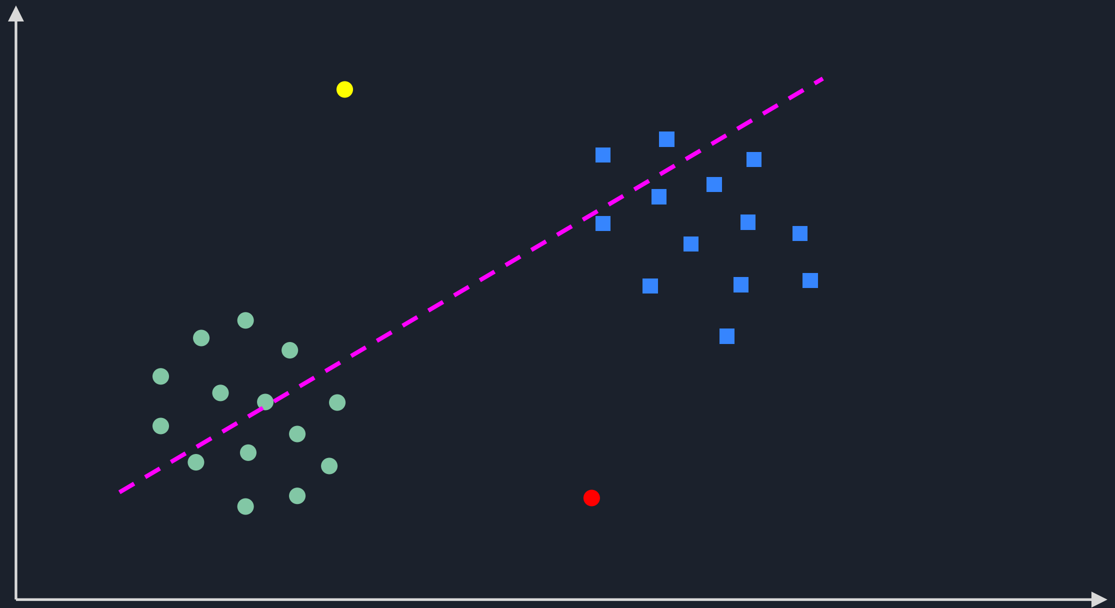
Ghosh4AI

2 classes 2 means



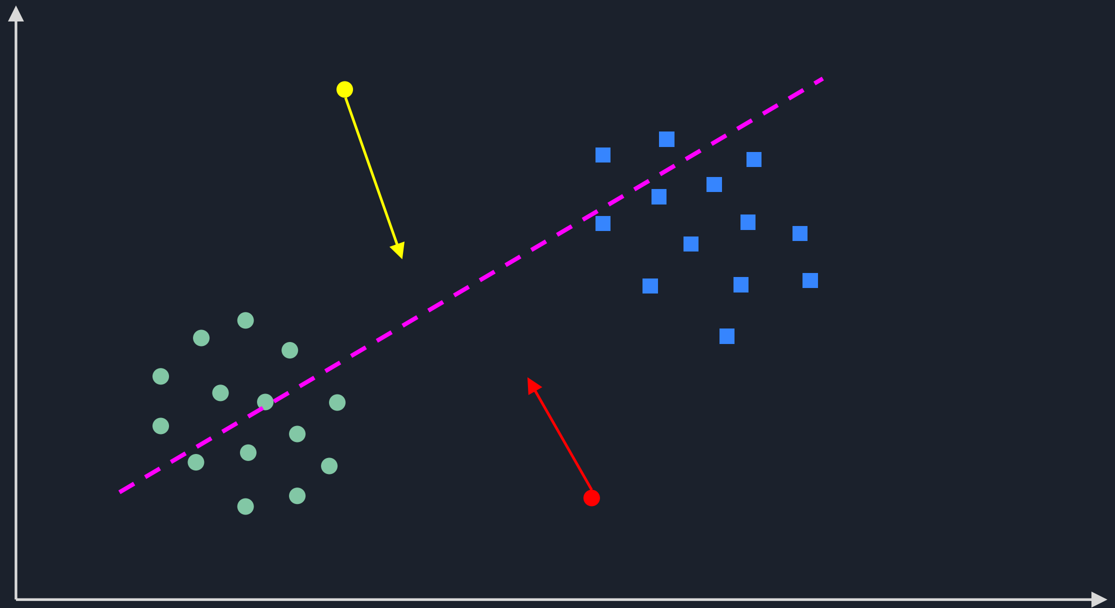
Ghosh4AI

Who belongs to whom



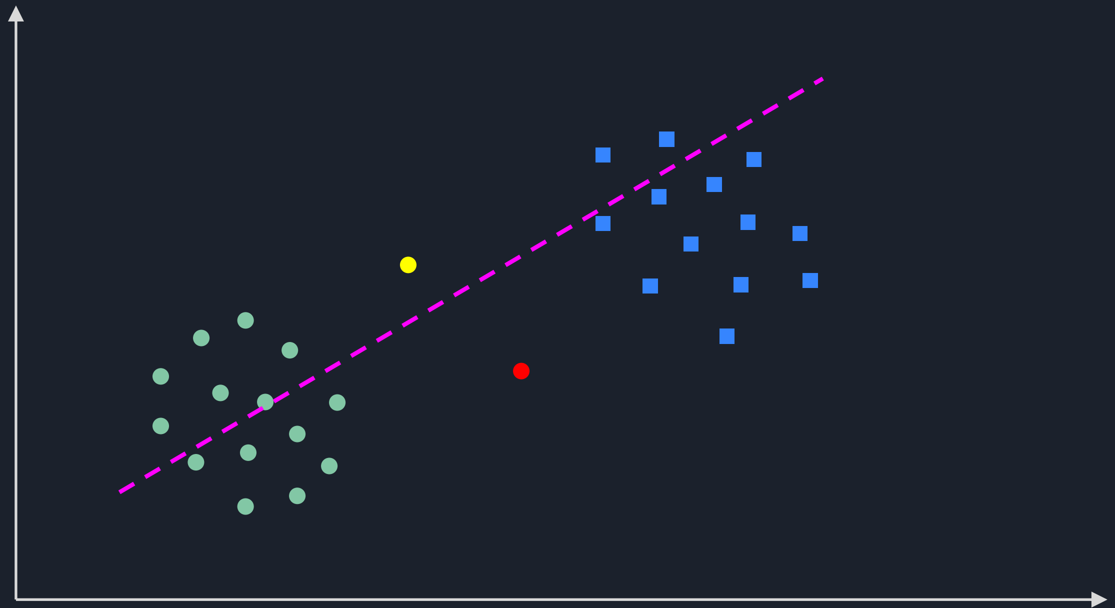
Ghosh4AI

Means need to move to the centre of their members

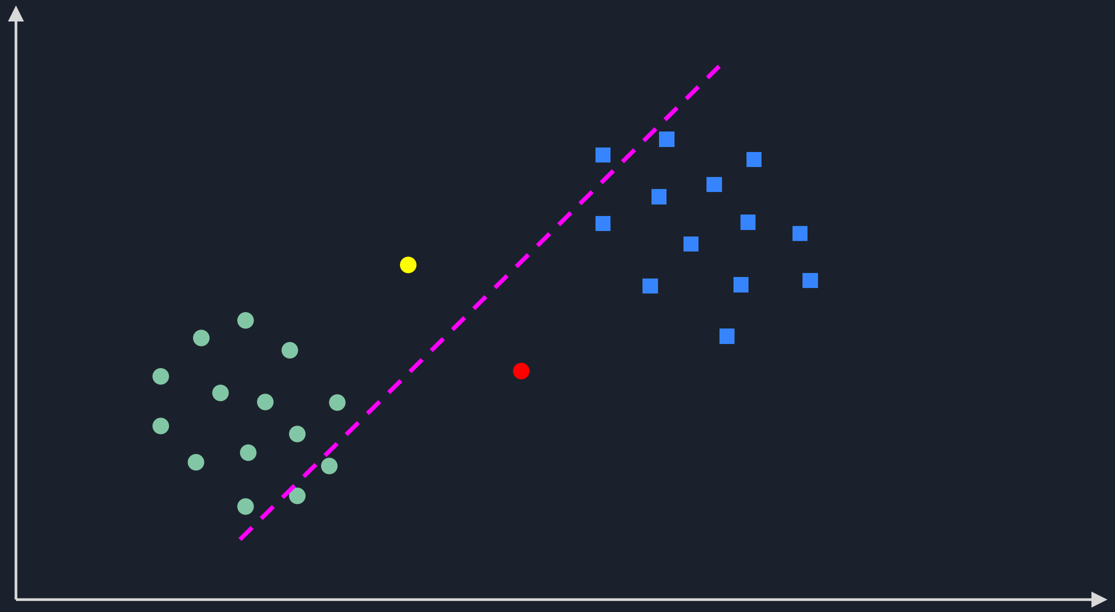


Ghosh4AI

Means have moved and samples need to decide again

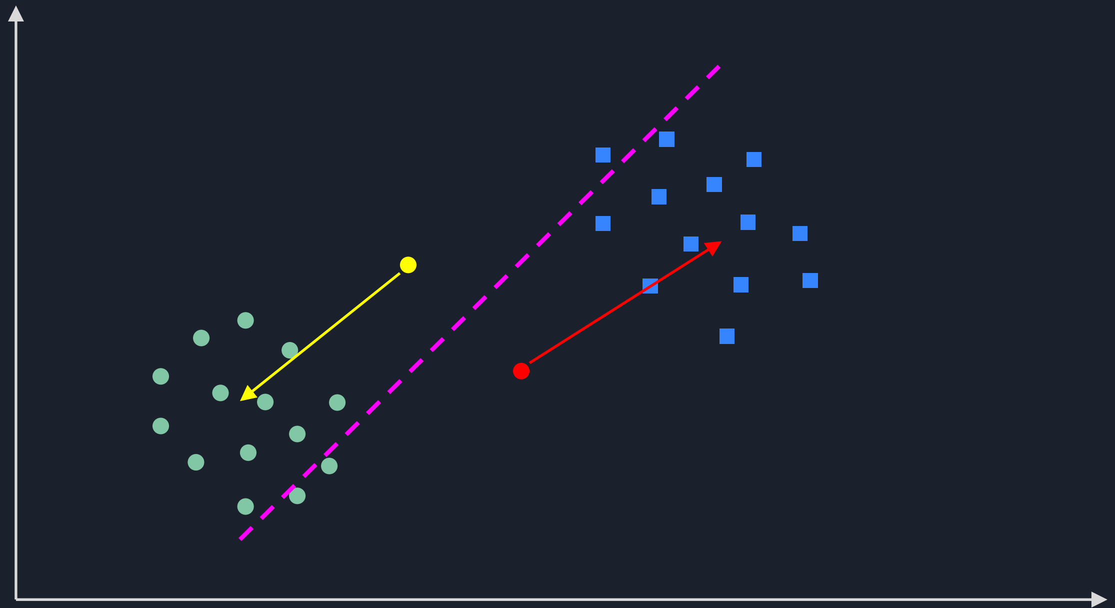


Switching sides



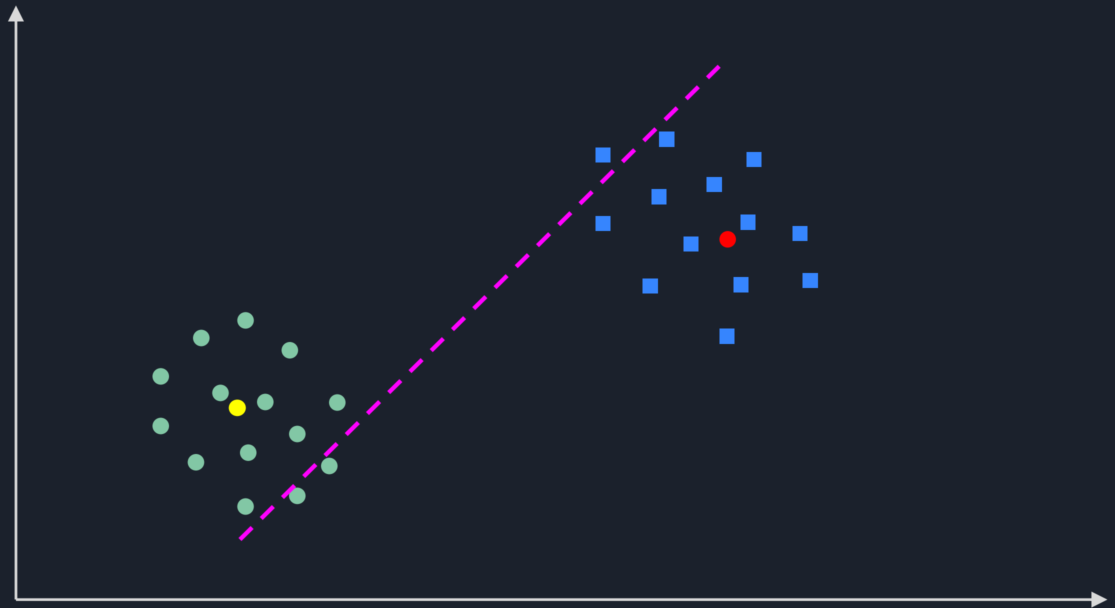
Ghosh4AI

Mean needs to move again



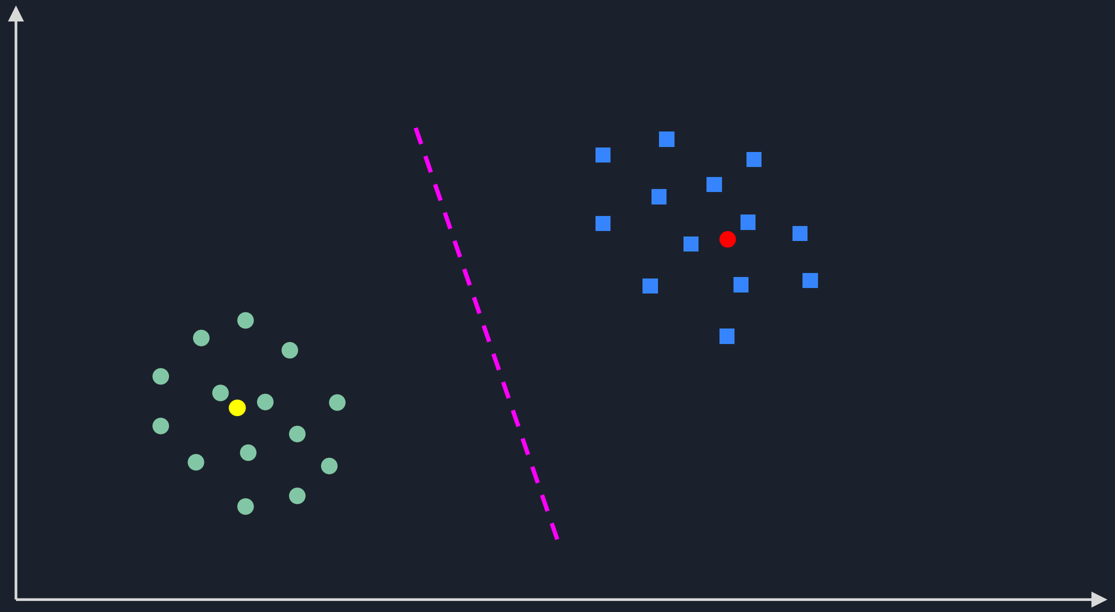
Ghosh4AI

Samples decide again



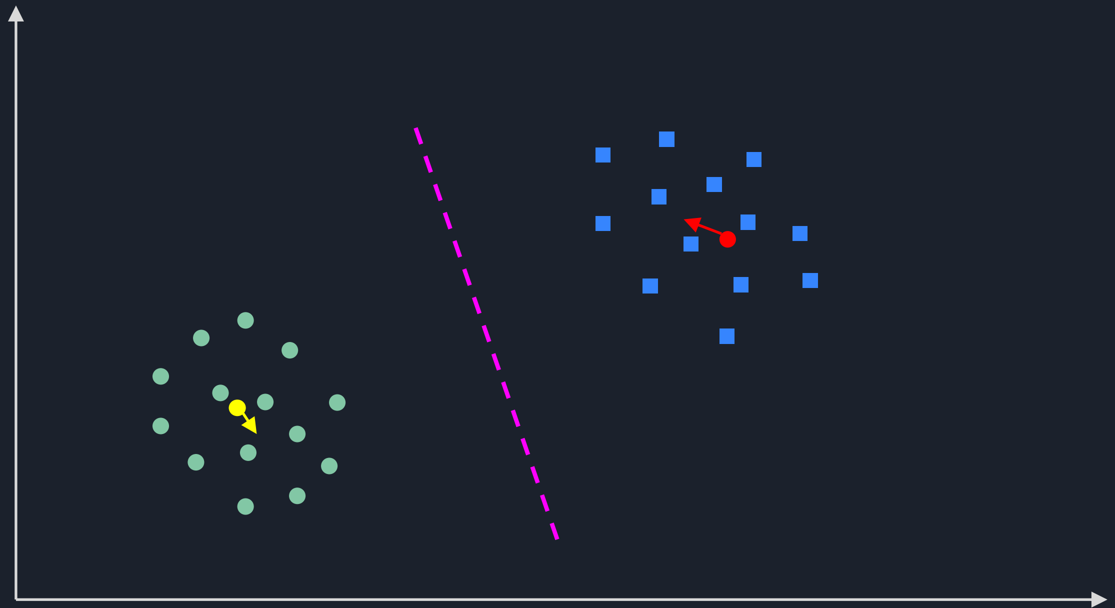
Ghosh4AI

Switching sides



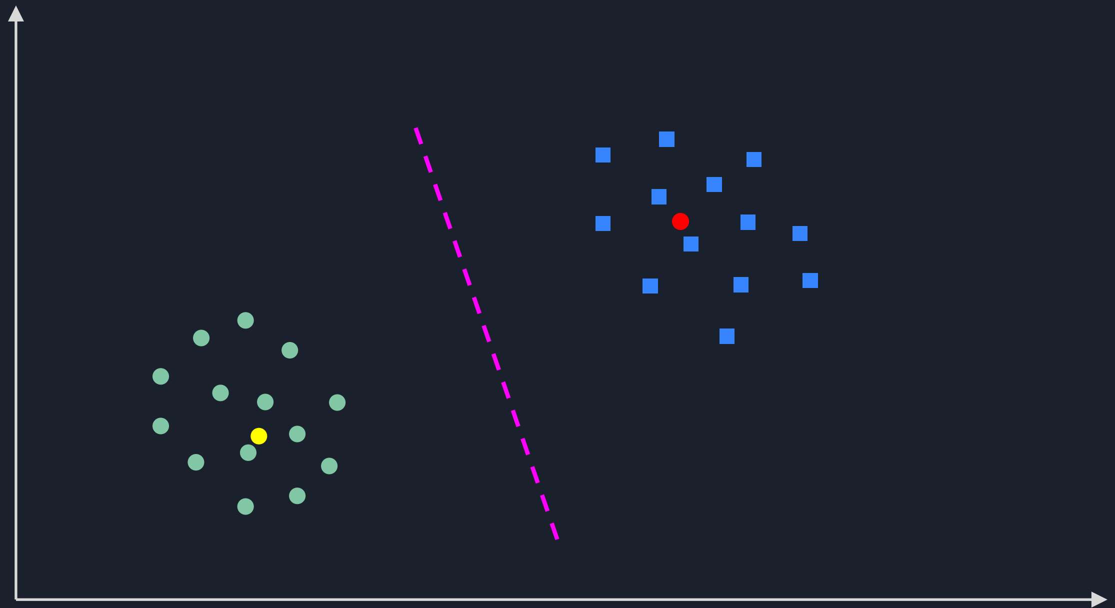
Ghosh4AI

Means move again

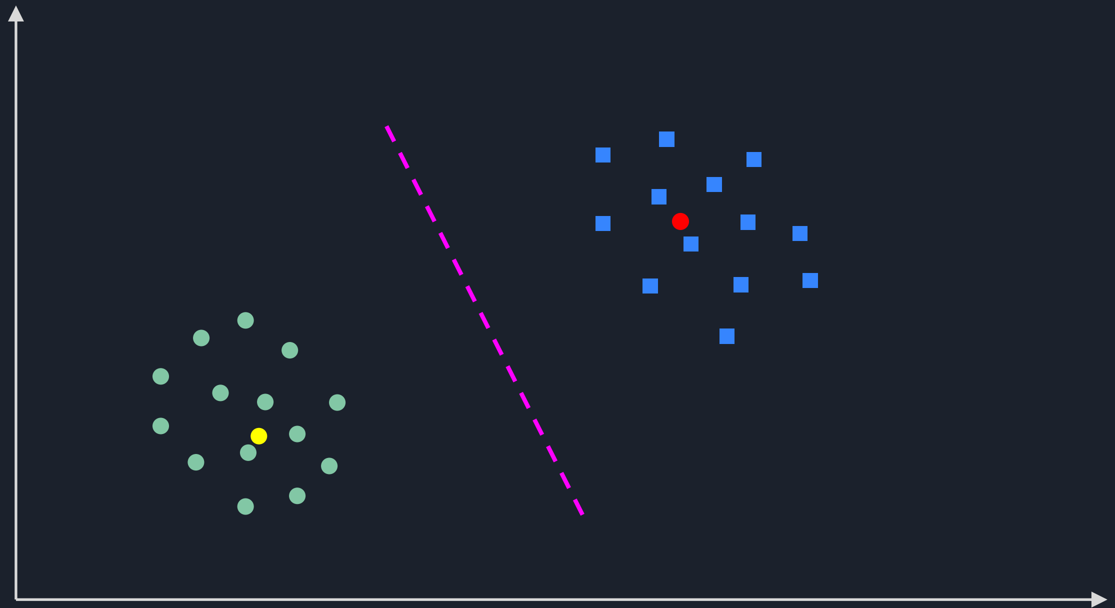


Ghosh4AI

Samples need to decide again .. But ...



No need to switch anymore



Ghosh4AI

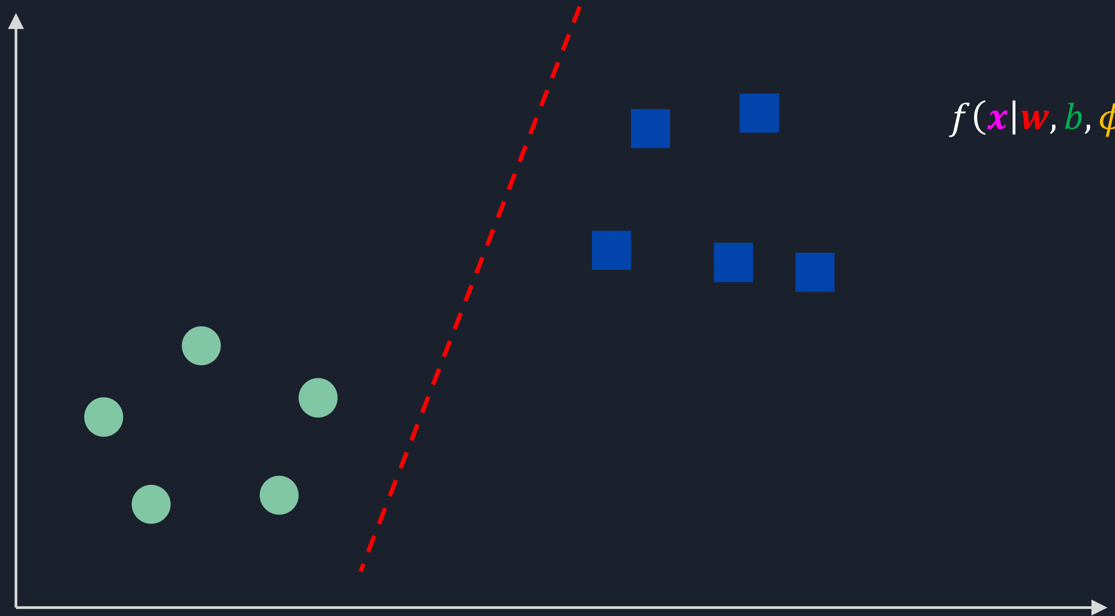
Further References

- Wikipedia -k-means clustering - Wikipedia
- Blogs :
  - Introduction to K-means Clustering - DataScience.com
  - Clustering using K-means algorithm – Towards Data Science
  - K means Clustering - Introduction - GeeksforGeeks
- Demo :
  - http://web.stanford.edu/class/ee103/visualizations/kmeans/kmeans.html
- API
  - sklearn.cluster.KMeans — scikit-learn 0.19.2
  - k-means clustering – MATLAB
  - Weka 3 - Data Mining with Open Source Machine Learning Software
- Books :
  - Pattern Recognition and Machine Learning – C.M. Bishop
- Related Papers
  - An efficient **k-means clustering** algorithm
  - Refining Initial Points for **K-Means Clustering.**
  - An efficient **k-means clustering** algorithm: Analysis and implementation
  - A **k-means clustering** algorithm

Ghosh4AI

Support Vector Machines (SVM)

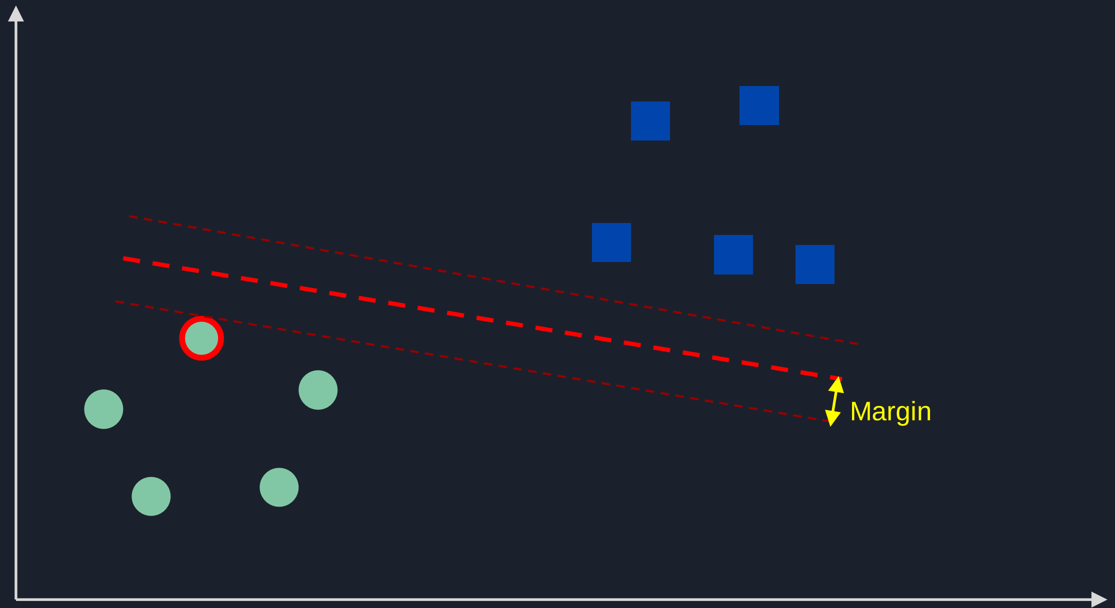
It all comes down to drawing a straight line



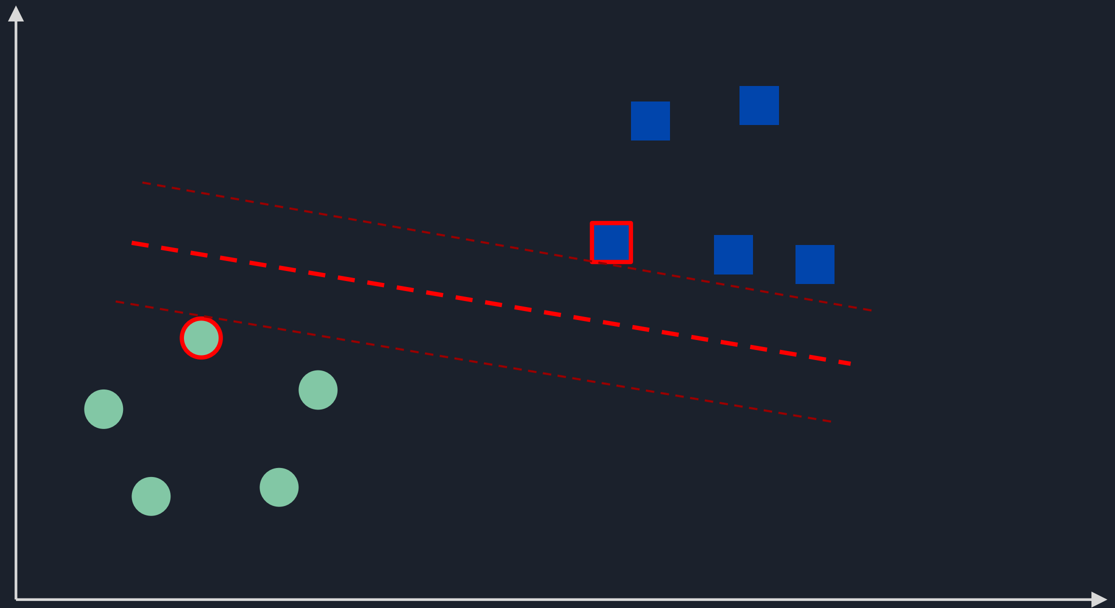
$$f(\boldsymbol{x}|\boldsymbol{w}, b, \phi) = \boldsymbol{w}^T \phi(\boldsymbol{x}) + b$$

Ghosh4AI

But where to position the line

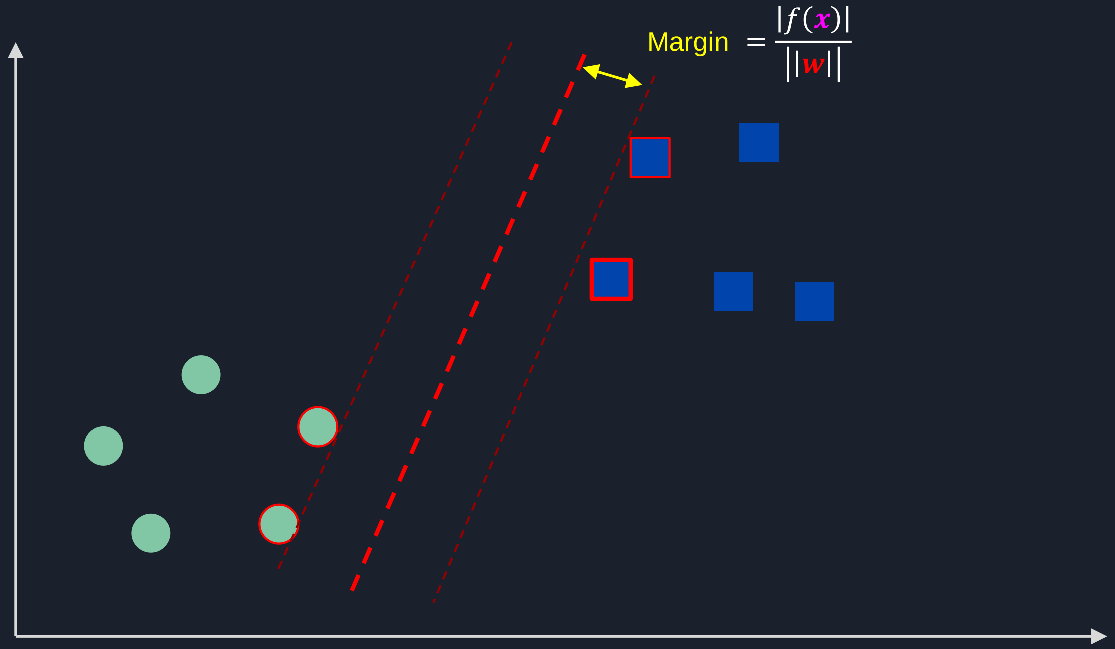


But lots of options



Ghosh4AI

But lots of options



Ghosh4AI

A little bit of maths and voila

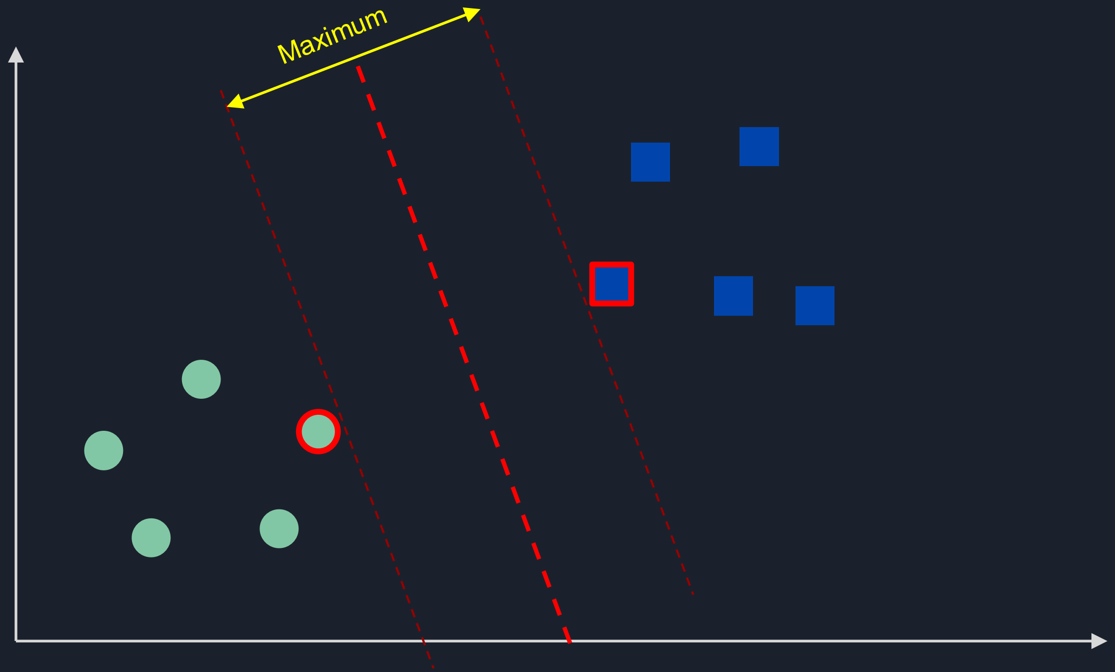
$$\frac{|f(\boldsymbol{x})|}{||\boldsymbol{w}||} = \frac{|\boldsymbol{w}^T\boldsymbol{x} + b|}{||\boldsymbol{w}||}$$

$$= \frac{|\kappa\boldsymbol{w}^T\boldsymbol{x} + \kappa b|}{||\kappa\boldsymbol{w}||}$$

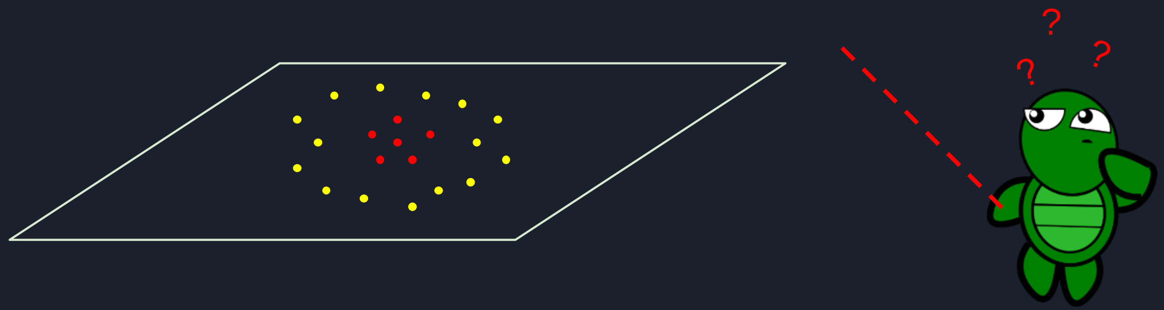
Set κ such a way so that $f(x)$ is 1 or -1 for the points closest to the line.
Then Maximum margin can be obtained by

$$\arg\max_{\boldsymbol{w}} \frac{1}{||\boldsymbol{w}||}$$

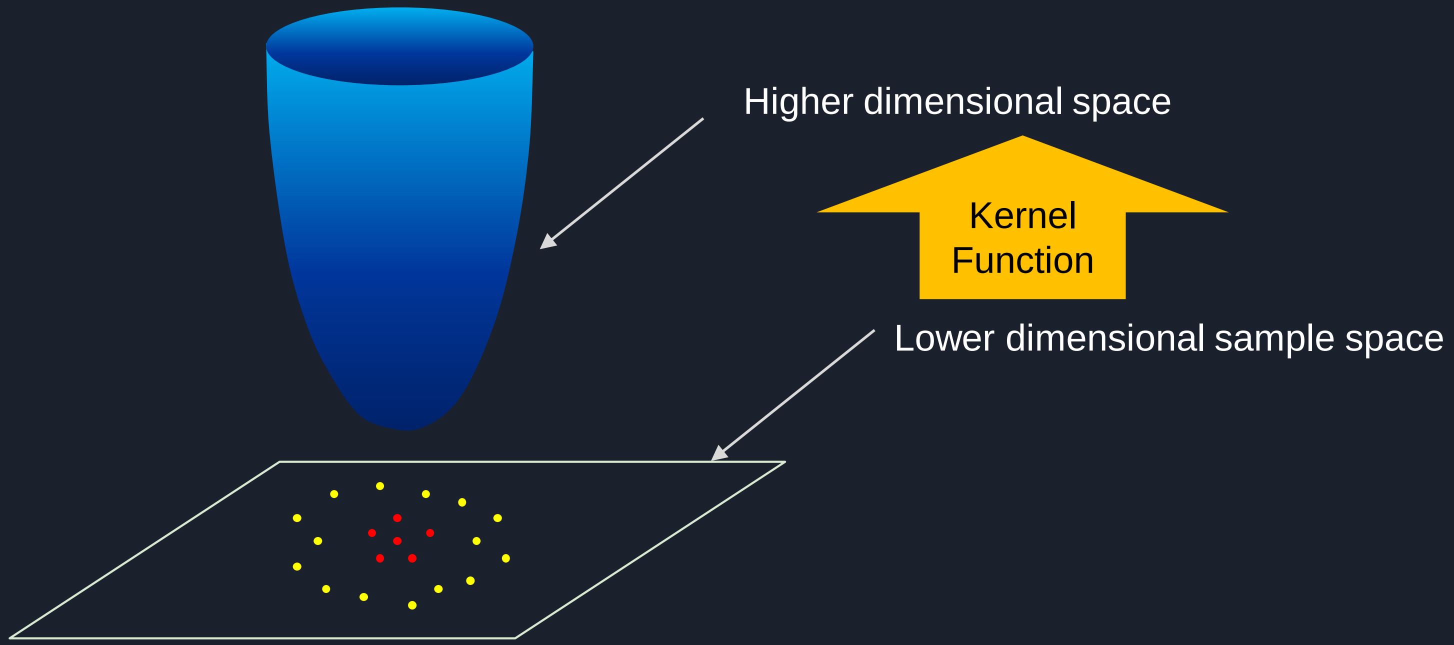
Fat is not always bad...



Is a straight line enough ?

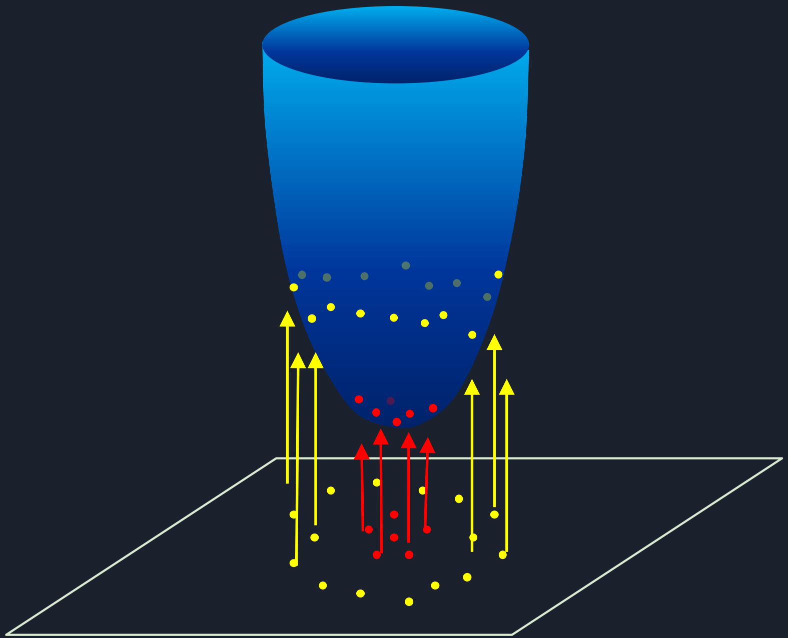


Not in this dimension.. We need more dimensions...



Ghosh4AI

Map values to higher dimensional space



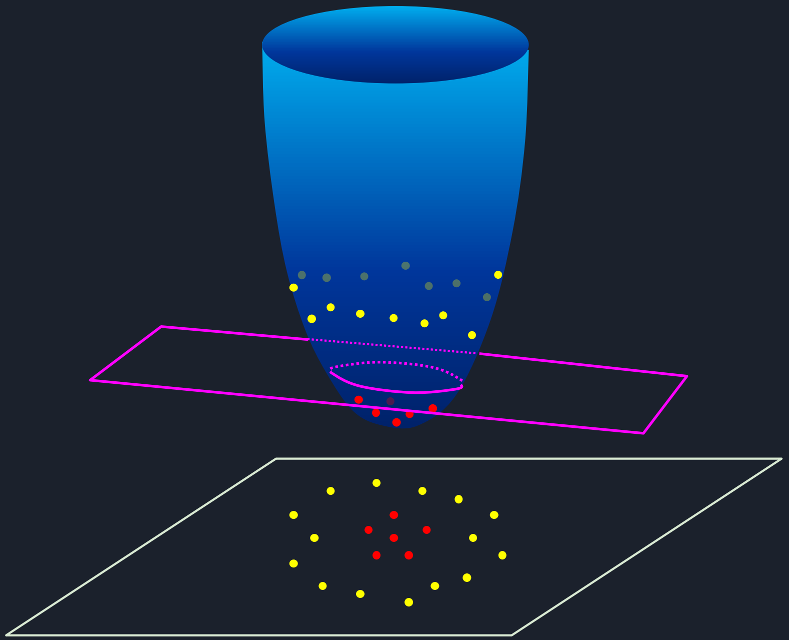
$x \rightarrow \kappa(x)$

κ = Kernel Function

Linear, Polynomial, RBF, Periodic, Gaussian ...

Ghosh4AI

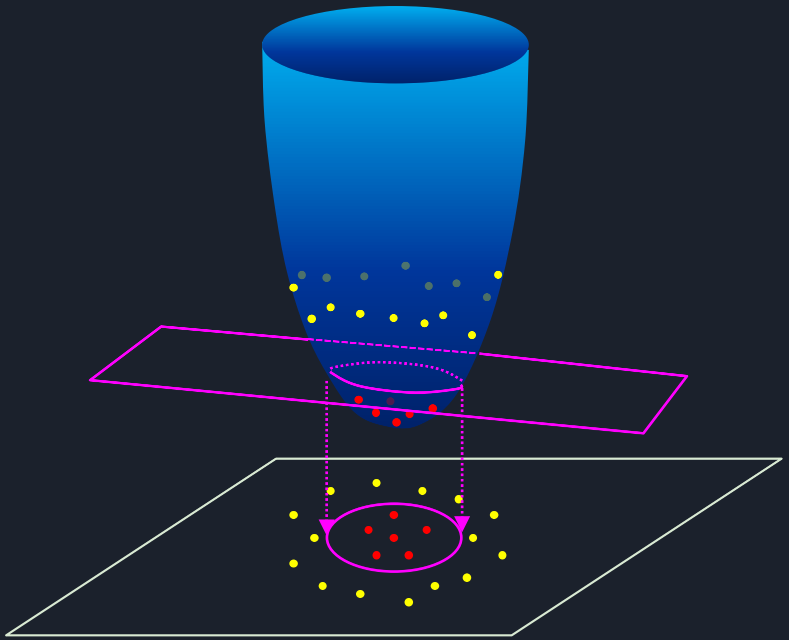
Create a hyperplane in the higher dimension



$$f(\boldsymbol{x}|\boldsymbol{w}, b, \kappa) = \boldsymbol{w}^T \kappa(\boldsymbol{x}) + b$$

Ghosh4AI

Map values back to the lower dimensional space



$$g(\boldsymbol{x}) = \kappa^{-1}(f(\boldsymbol{x}))$$

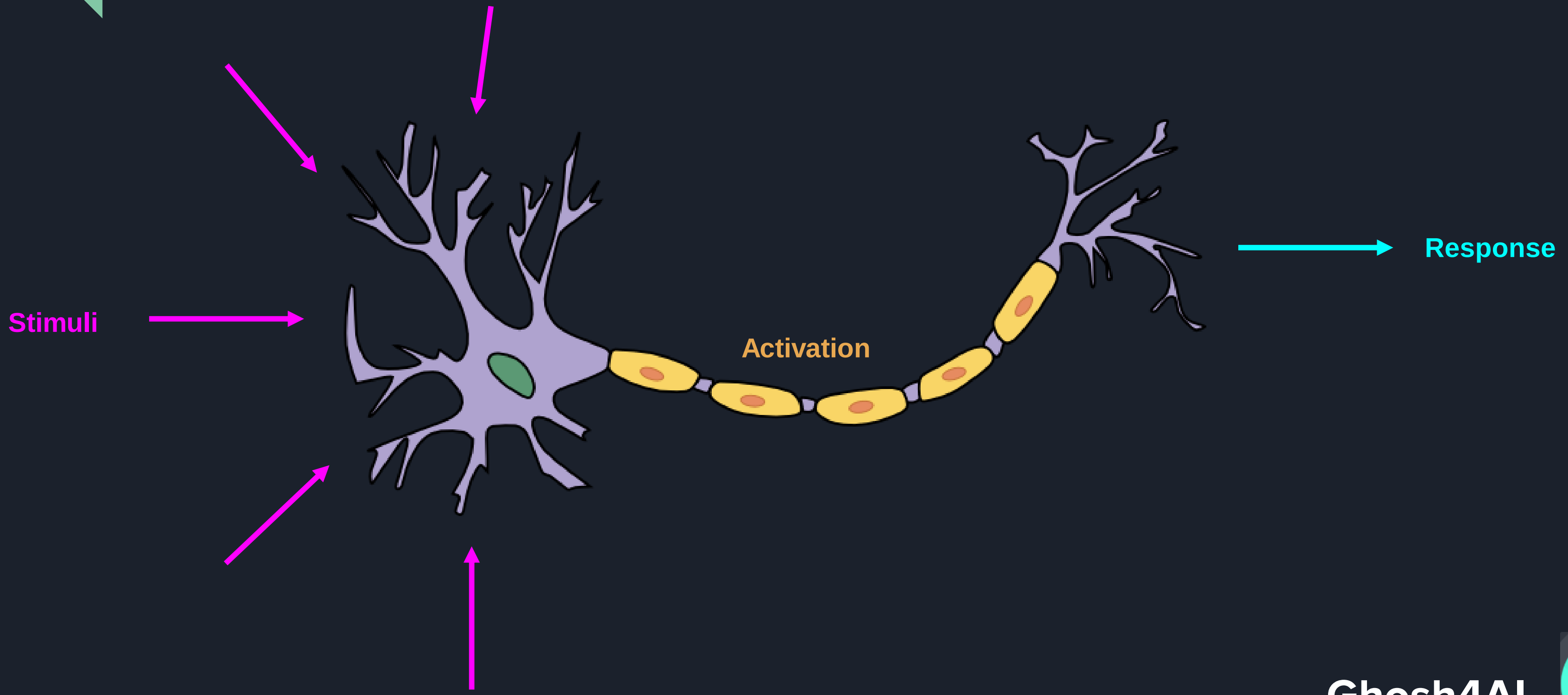
Further References

- Wikipedia -Support vector machine - Wikipedia
- Blogs :
  - Understanding Support Vector Machine algorithm from examples
  - Chapter 2 : SVM (Support Vector Machine) — Theory
  - Support Vector Machines for Machine Learning
- API
  - Support Vector Machines — scikit-learn 0.19.2
  - Support Vector Machines for Binary Classification – MATLAB
  - Weka 3 - Data Mining with Open Source Machine Learning Software
  - LIBSVM -- A Library for Support Vector Machines
- Books :
  - Pattern Recognition and Machine Learning – C.M. Bishop
  - Learning with kernels - Bernhard Schölkopf
- Related Papers
  - Support vector machines
  - Measuring the VC-dimension of a learning machine

Multilayer Perceptrons (MLP)

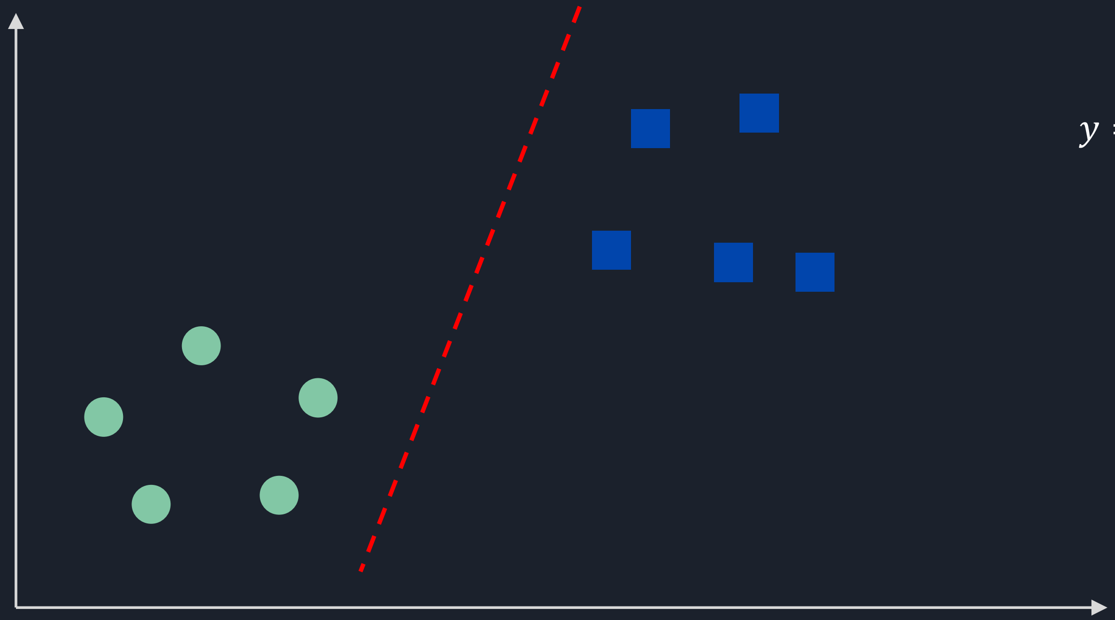
Ghosh4AI

The real neuron



Ghosh4AI

Decision as a straight line



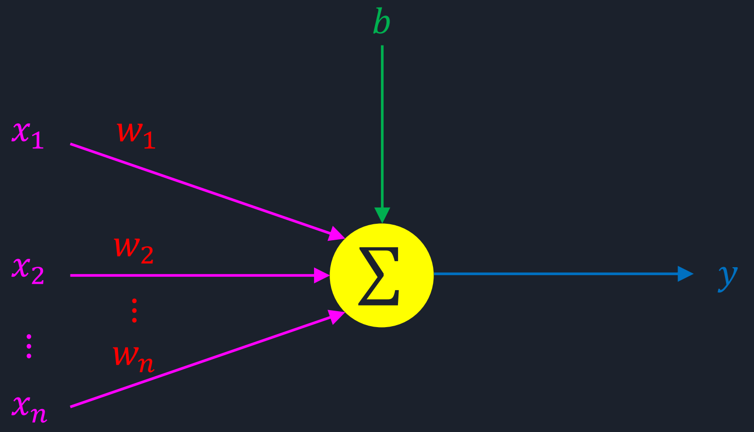
$$y = \boldsymbol{w}^T \boldsymbol{x} + b$$

The straight line

$$y = \boldsymbol{w}^T \boldsymbol{x} + b$$

$$y = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} [x_1 \quad x_2 \quad \dots \quad x_n] + b$$

$$y = \sum_{i}^{n} w_i x_i + b$$



Ghosh4AI

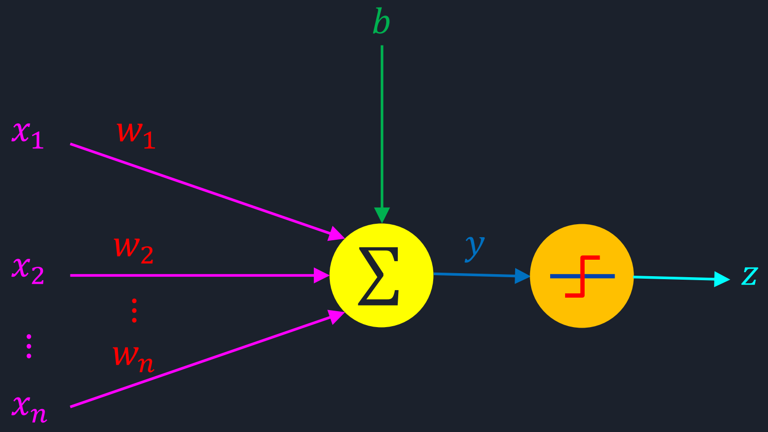
The artificial Neuron making decisions

$$y = \boldsymbol{w}^T \boldsymbol{x} + b$$

$$y = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} [x_1 \ \ x_2 \ \ \ldots \ \ x_n] + b$$

$$y = \sum_{i}^{n} w_i x_i + b$$

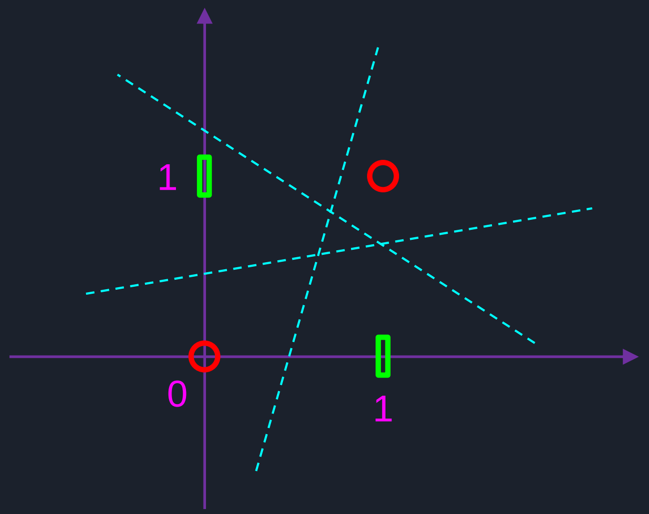
$$z = sign\left(\sum_{i}^{n} w_i x_i + b\right)$$



Ghosh4AI

Is a straight line enough ?

$$z = sign(w^T \boldsymbol{x} + b)$$

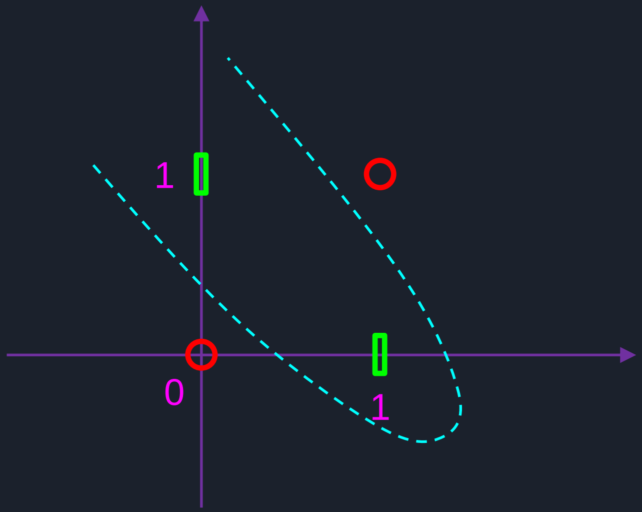


Ghosh4AI

Bend the line

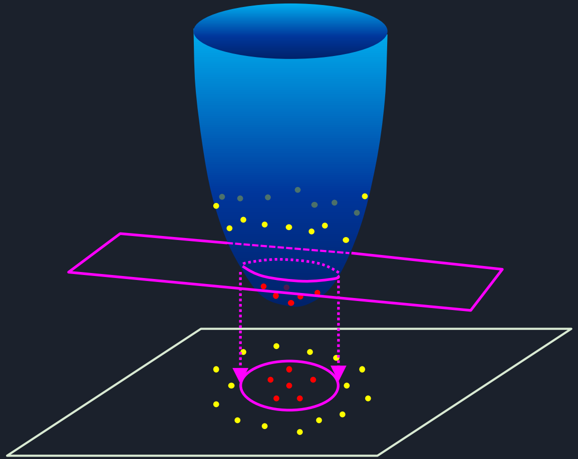
$$z = f_{NL}(\boldsymbol{w}^T \boldsymbol{x} + b)$$

Non Linear Function

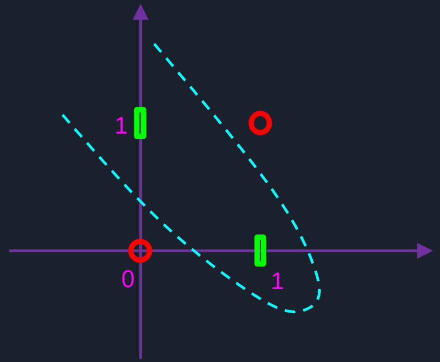


Ghosh4AI

SVM vs Neural Networks



$z = w^T \phi(x) + b$

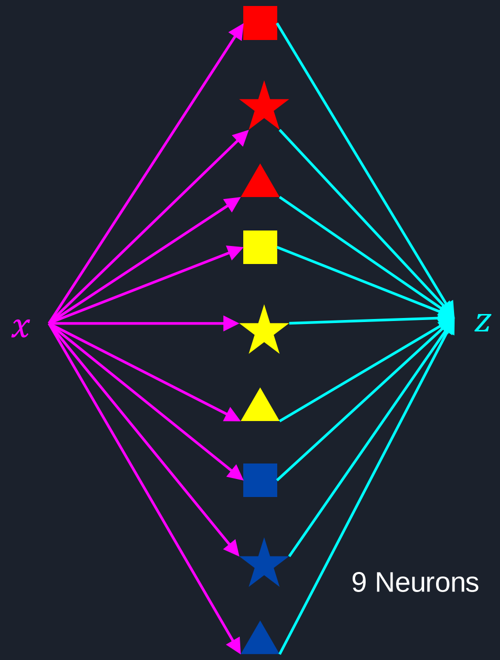


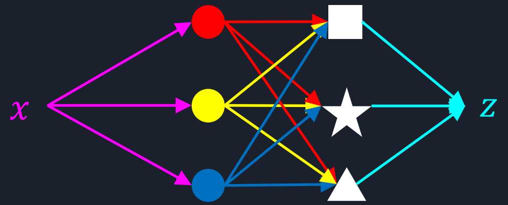
$z = f_{NL}(w^T x + b)$

Ghosh4AI

Stack them up...

Colours = { red, yellow, blue }, Shapes = { Square, Triangle and Star }

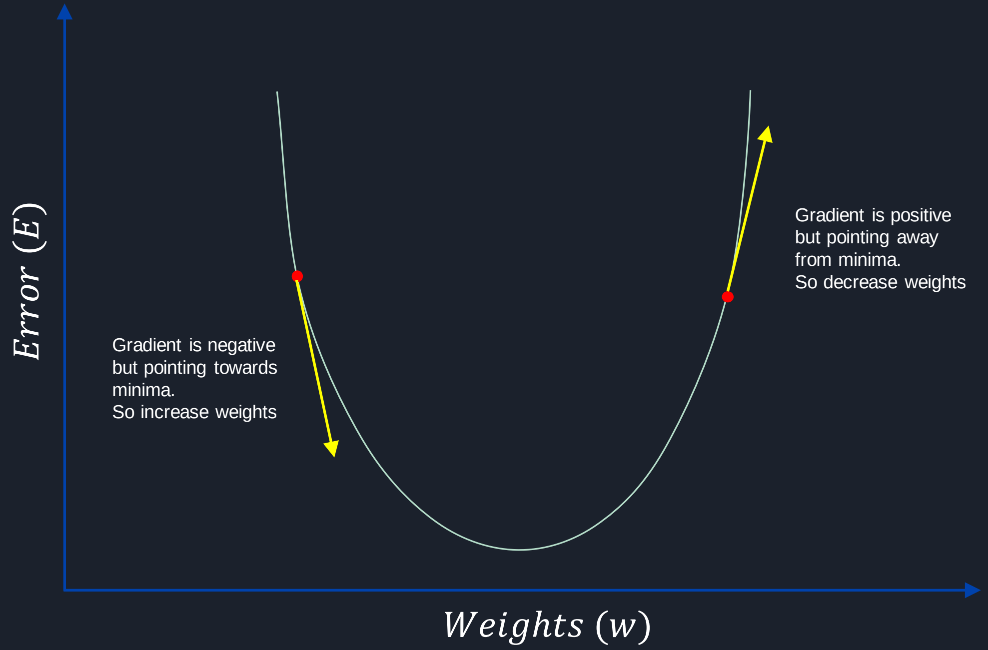




6 Neurons

Ghosh4AI

The gradient



Ghosh4AI

How to calculate weights ?

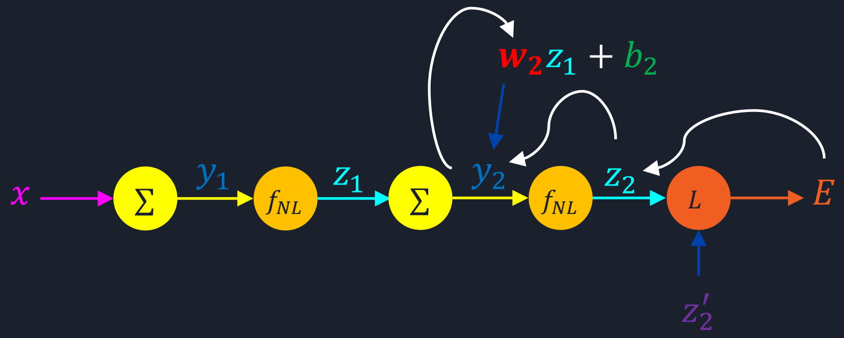
Change in error
with respect to
change in weights

$$\boldsymbol{w} = \boldsymbol{w} - \eta \frac{dE}{d\boldsymbol{w}}$$

$$E = L(z_{observed}, z_{actual})$$

Ghosh4AI

Derivative to chain of partial derivatives

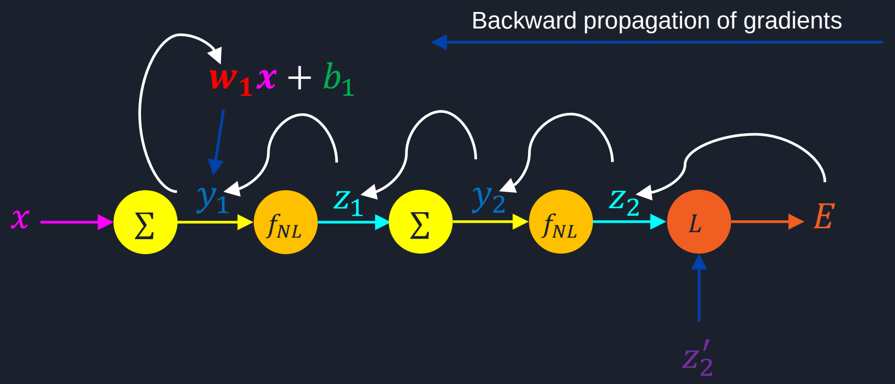


$$E = L(z_2, z_2')$$

$$\frac{dE}{dw_2} = \frac{\delta E}{\delta z_2} \cdot \frac{\delta z_2}{\delta y_2} \cdot \frac{\delta y_2}{\delta w_2}$$

Ghosh4AI

Backpropagation through multiple layers



$$E = L(z_2, z_2')$$

$$\frac{dE}{dw_1} = \frac{\delta E}{\delta z_2} \cdot \frac{\delta z_2}{\delta y_2} \cdot \frac{\delta y_2}{\delta z_1} \cdot \frac{\delta z_1}{\delta y_1} \cdot \frac{\delta y_1}{\delta w_1}$$

Ghosh4AI

Further References

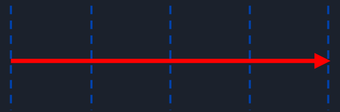
- Wikipedia -Artificial neural network - Wikipedia
- Blogs :
  - Neural networks and deep learning
  - How to build your own Neural Network from scratch in Python
  - What is a neural network? - Introduction to deep learning
- Demo :
  - A Neural Network Playground
  - Neural Network demo — Preset: Binary Classifier for XOR
- API
  - Neural network models (supervised) — scikit-learn 0.19.2
  - Neural Network Toolbox - MATLAB – MathWorks
  - Weka 3 - Data Mining with Open Source Machine Learning Software
  - TensorFlow
  - PyTorch
- Books :
  - Pattern Recognition and Machine Learning – C.M. Bishop
  - Neural networks- Simon S. Haykin
- Related Papers
  - Learning internal representations by error propagation
  - The perceptron: a probabilistic model for information storage and organization in the brain.
  - A learning algorithm for continually running fully recurrent neural networks
  - 30 years of adaptive neural networks

Ghosh4AI

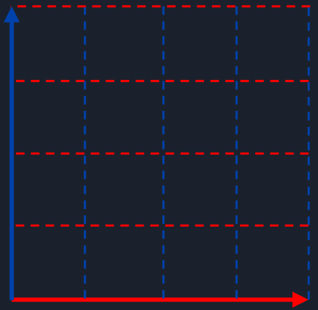
Principal Component Analysis (PCA)

Ghosh4AI

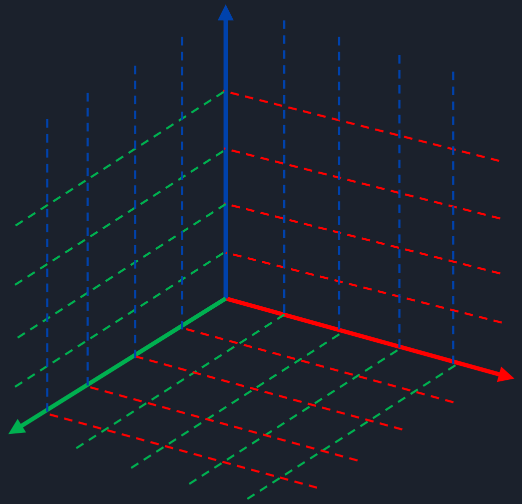
The curse of dimensionality



4 regions



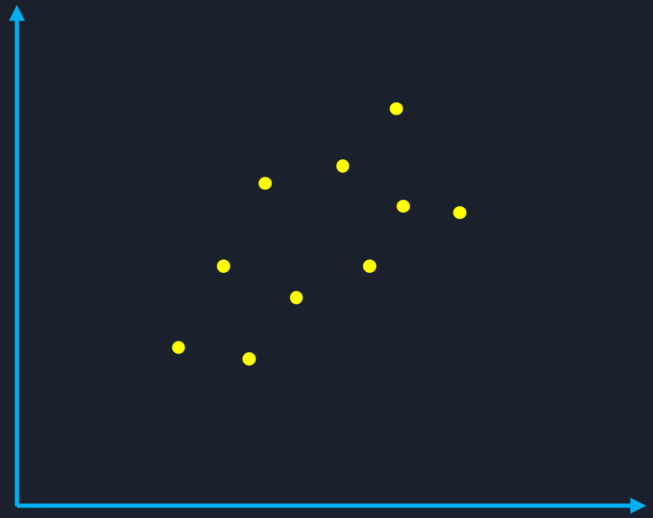
16 regions



64 regions

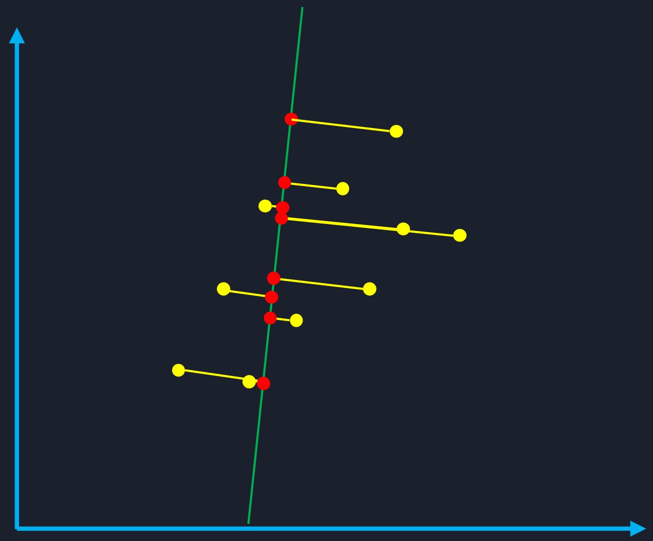
Ghosh4AI

2 dimensions to represent sample space



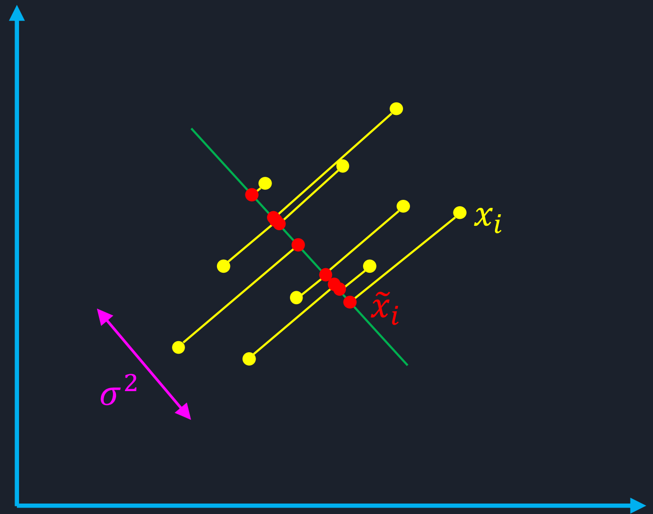
Ghosh4AI

Mapping from 2 dimension to a single dimension.



Ghosh4AI

We want the samples to be distinct in the new dimension… so, more variance

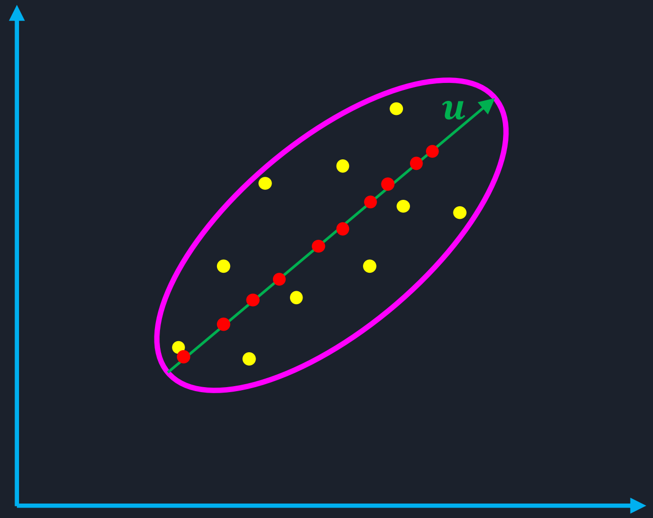


Goal :

- $Maximize\ \sigma^2$ or
- $Minimize\ \sum_i (x_i - \tilde{x}_i)^2$.

Ghosh4AI

Along the major axis of an ellipse



u is the eigen vector corresponding to the largest eigen value

Further References

- Wikipedia -Principal component analysis - Wikipedia
- Blogs :
  - A One-Stop Shop for Principal Component Analysis – Towards Data Science
  - Understanding Principal Component Analysis – Rishav Kumar - Medium
  - Practical Guide to Principal Component Analysis (PCA) in R & Python
- Demo :
  - http://setosa.io/ev/principal-component-analysis/
- API
  - sklearn.decomposition.PCA — scikit-learn 0.19.2
  - Principal component analysis (PCA) on data - MATLAB princomp
  - Weka 3 - Data Mining with Open Source Machine Learning Software
- Books :
  - Pattern Recognition and Machine Learning – C.M. Bishop
  - **Principal components analysis – GH Dunteman**
- Related Papers
  - **Principal component analysis**