Python For Machine Learning

UE19EC353

Department of ECE

PES University

Finally, Machine Learning !!

Some Machine Learning basics

- Deep Learning Ian Goodfellow et al.

What is learning?

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

- Mitchell

The Task *T*

- Learning is our means of attaining the ability to perform the task.
- Described in terms of how ML system should process an ***example.***
- ***Example :*** Collection of ***features*** quantitatively measured from an object/event

$$x \in R^n$$

$$where\ x_i\ is\ feature\ of\ the\ vector$$

Some Important / Common Tasks

- **Classification** : specify which of k categories some input belongs to.

$$f: R^n \rightarrow \{1, \ldots, k\}$$

- **Regression** : to predict a numerical value given some input.

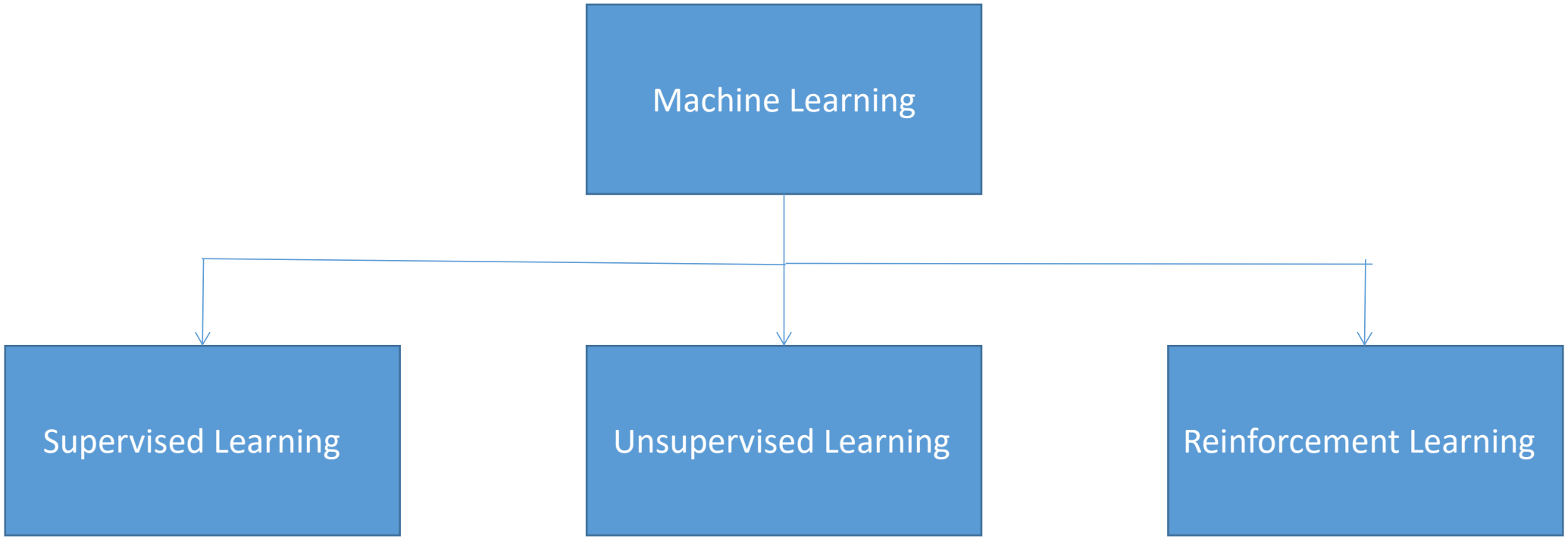
$$f: R^n \rightarrow R$$

- **Anomaly Detection :** sift through a set of events or objects and flags some of them as being unusualor atypical.
- **Machine Translation** : convert input sequence of symbols in some language into a sequence of symbols in another language.

Performance Measure *P*

- To evaluate the abilities of a machine learning algorithm
- Eg: For classification, *P* is Accuracy ,i.e, proportion of examples for which the model produces the correct output
- Measuring ***Error Rate or Expected 0-1 loss.***
- Evaluate performance measures on ***test data***

Experience *E*



Experience *E*

- ***Unsupervised Learning:***
  - experience a dataset containing manyfeatures, then learn useful properties of the structure of this dataset.
  - involves observing several examplesof a random vectorxand attempting to implicitly or explicitly learn the proba-bility distribution p(x), or some interesting properties of that distribution
  - Eg : Clustering, Density Estimation
- ***Supervised Learning :***
  - a dataset containing features,but each example is also associated with a label or target. Eg: Iris Dataset.
  - target *y* being provided by an instructor or teacher who shows the machine learning system what to do
  - Classification, Regression

Experience *E*

- ***Reinforcement Learning:***
  - algorithms interact with an environment
  - there is a feedback loop between the learning system and its experiences
  - Q-learning, temporal-difference learning
- ***Semi Supervised Learning :***
  - supervised learning where the training data contains very few labeled examples and a large number of unlabeled examples.
  - make effective use of all of the available data, not just the labelled data like in supervised learning.

Common Error / Loss Functions

- Regression:
  - Mean Squared Error
  - Mean Absolute Error
- Classification:
  - Cross Entropy Loss

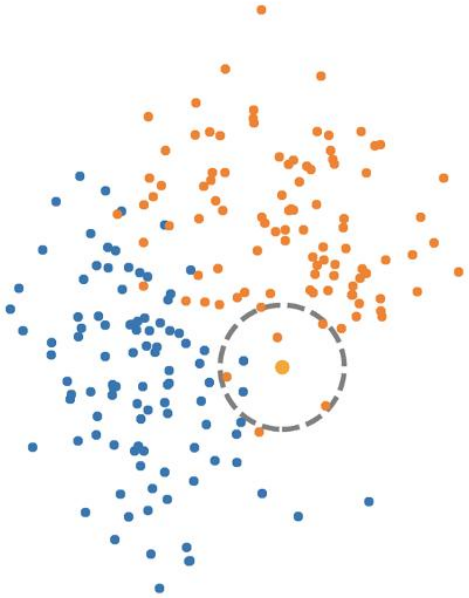
Jupyter Notebook

Side by Side Explanation

KNN Classifcation

- Supervised Classsification Algorithm
- the process of finding the closest point to the input point from the given data set
- Steps:
  - first step is to transform data points into their mathematical vectors
  - algorithm finds distance (Euclidean) betweeen vectors
  - computes the distance between each data point and the test data and then finds the probability of the points being similar to the test data
  - Classification done by highest probabilities

KNN Classifcation



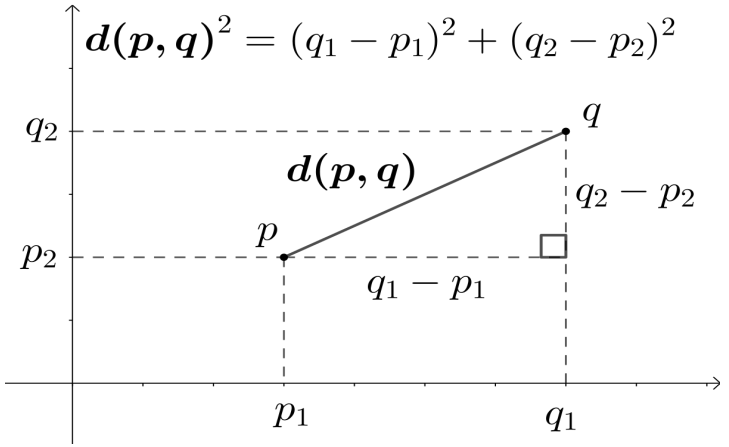
a) A labeled dataset has three distict groups

b) We want to classify a new data point

C) We find the k-nearest neighbors

D) The new point is the most prominent class

Class 1
Class 2



$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in \mathcal{A}} I(y^{(i)} = j)$$

*credits to wiki and medium

Preprocessing : Feature Scaling

- unscaled data lead to difficulties in visualizations
- degrade the predictive performance of many machine learning algorithms.
- Scaling :
  - Standard Scaler
  - Min max scaler
  - Max Abs Scaler
  - Robust Scaler

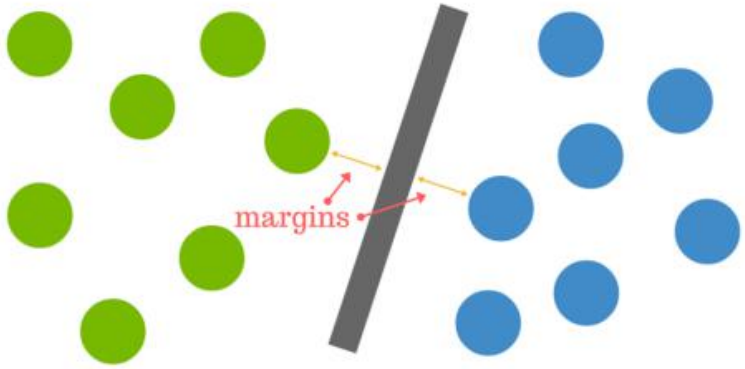
Cross Validation (K Fold)

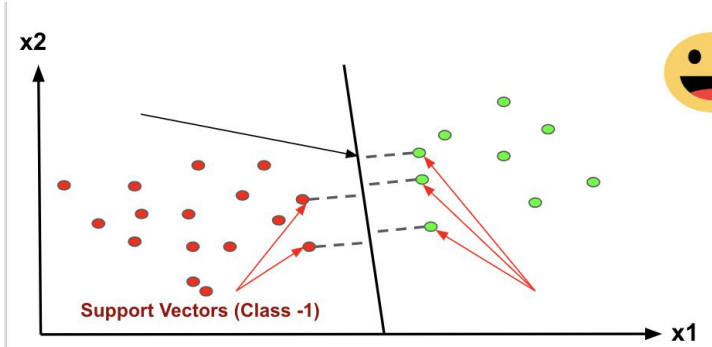
- Never enough data for training : removing for validation will cause underfitting
- Reducing training data : lose important info/insights
- Data into K subsets.
- each time, one of the k subsets is used as the test set/ validation set and the other k-1 subsets are put together to form a training set.
- error estimation is averaged over all k trials to get total effectiveness of our model.
- significantly reduces bias as we are using most of the data for fitting, and also significantly reduces variance as most of the data is also being used in validation set.

Support Vector Machines

- Both Classification and regression
- finding a hyperplane that best divides a dataset into two classes
- Support Vectors : data points nearest to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane
- Intuitively, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified
- The distance between the hyperplane and the nearest data point from either set is known as the margin.
- The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly.

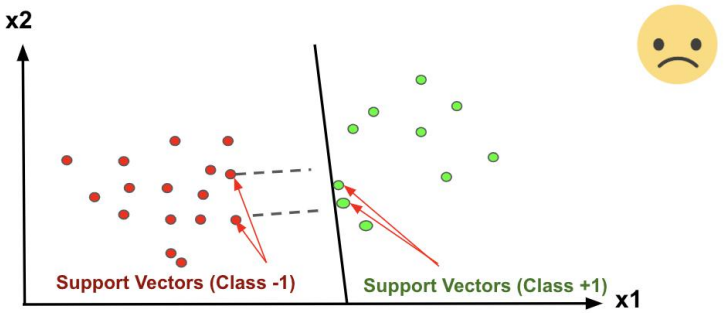
Support Vector Machines





Good Margin

- all sector vectors have the same distance with the maximum margin hyperplane

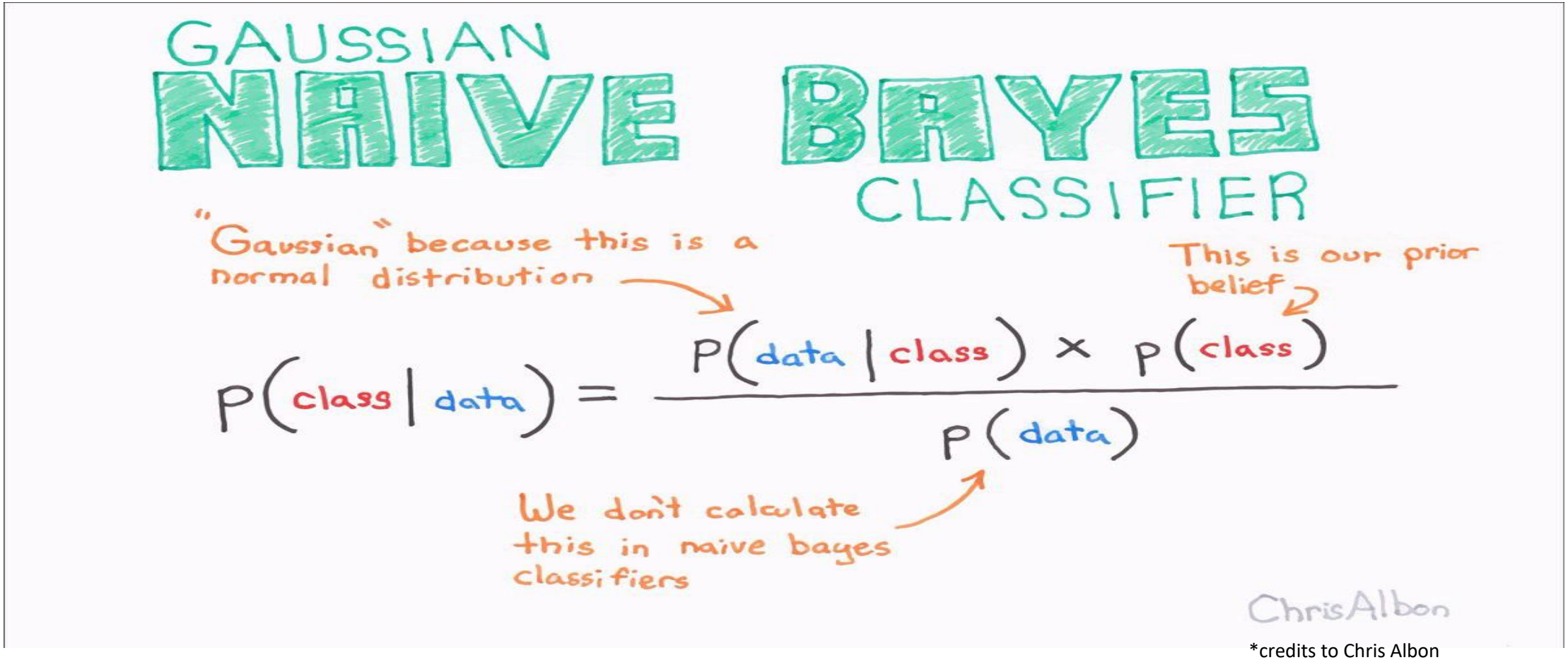


Bad Margin

- very close to either class -1 support vectors or class +1 support vectors

*credits to kdnuggets and velocity business solutions

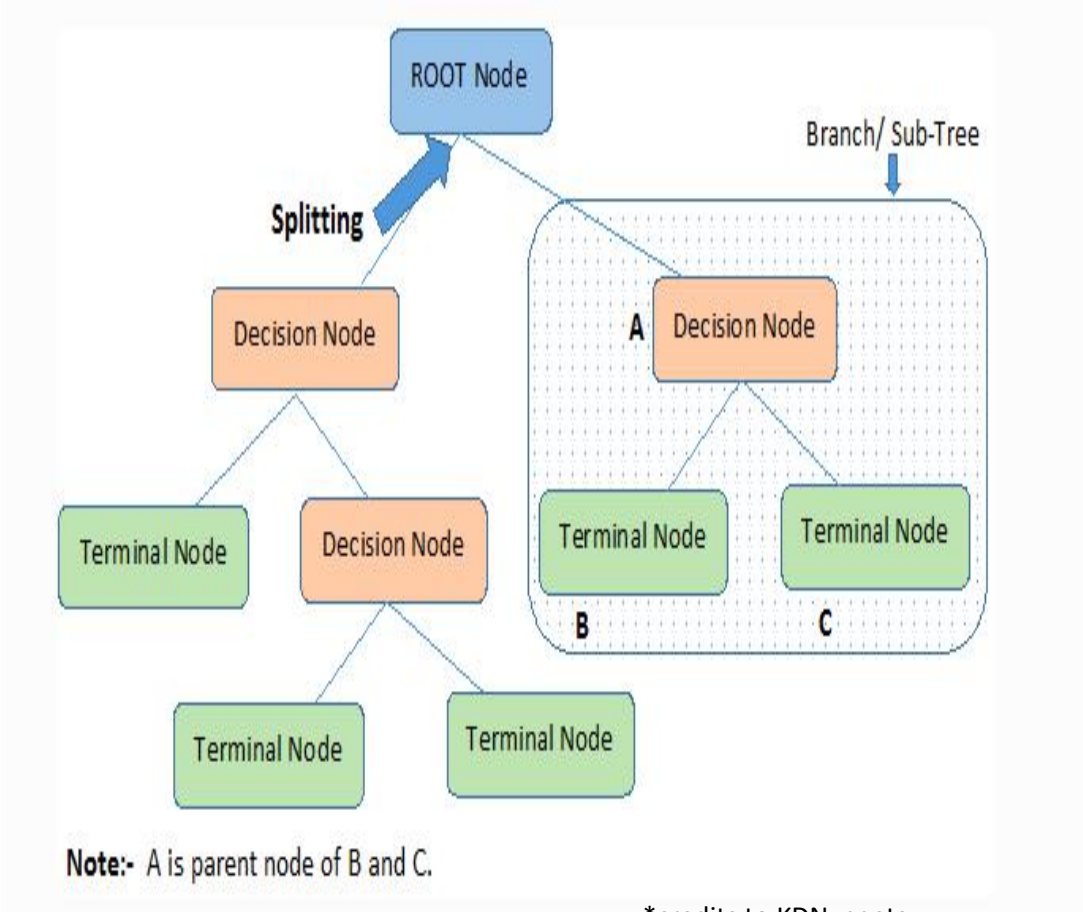
Naive Bayes Classifier



*credits to Chris Albon

Decision Tree Classifier

- Tree structure based classification/regression (CART)
- Ensemble learning
- multiple algorithms to decide to split a node into two or more sub-nodes
- Gini Index (0 to 1): calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly.



*credits to KDNuggets

Open CV: Working with Images

Computer Vision : seeing is beleiving??

Images Processing

- Process, analyse and extract useful information from images
- ***Computer Vision:*** process, load, transform and manipulate images
- Images are represented as pixel arrays
- m x n x dim

Color image = m x n x 3

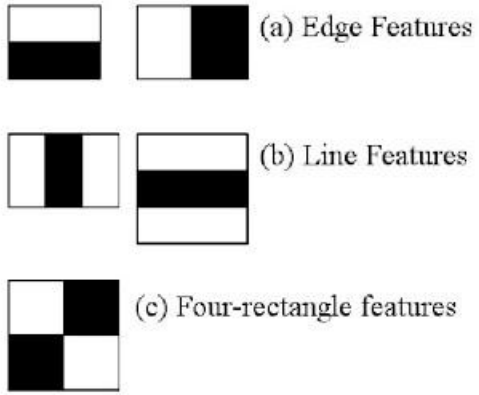
Grayscale image = m x n x 1

Canny Edge Detection

- Edge detection operation
- Multi step algorithm
  - filter based on the derivative of a Gaussian in order to compute the intensity of the gradients
  - Gaussian reduces the effect of noise present in the image
  - edge pixels are kept or removed using hysteresis thresholding on the gradient magnitude.

Face Detection using Haar Algorithm

- Viola-Jones Face detection algo (2001)



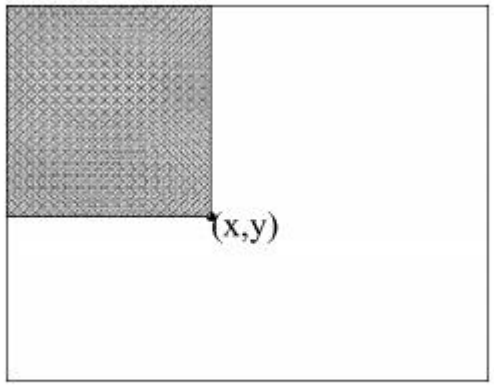
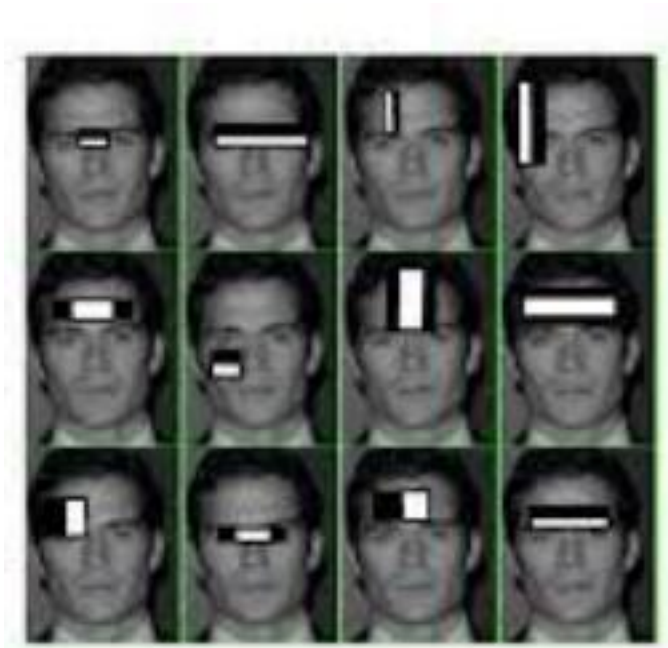
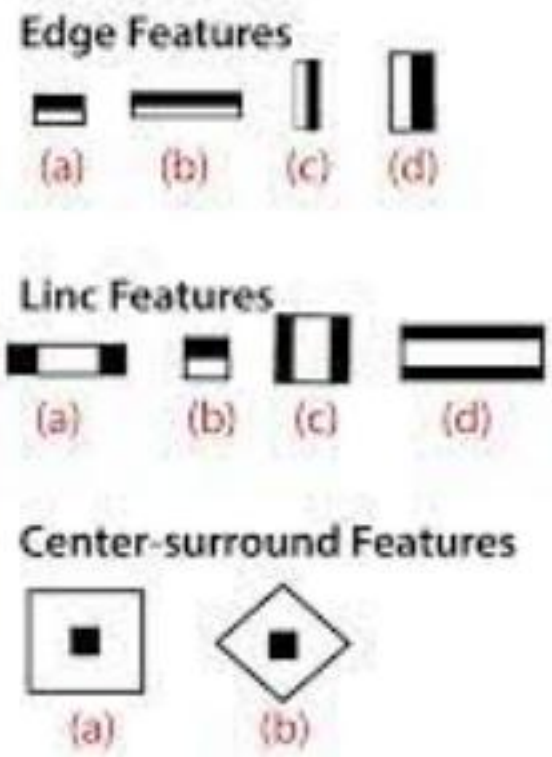


Figure 2. The value of the integral image at point (x, y) is the sum of all the pixels above and to the left.





Credits to various online sources