CV: Facial recognition and emoji generation

Synopsis:

We can generalize the emotion detection steps as follows:
1) Dataset pre-processing

2) Face detection

3) Feature extraction

4) Classification based on the features

1. Pre-processing

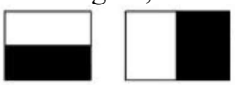
Pre-processing involves operation with images at lowest level of abstraction.
The pre-processing steps that are implemented are

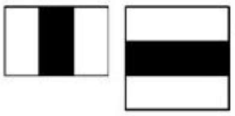
- *Reduce the noise*
  Noise is an unwanted component of the image. The existence of noise in a face image can degrade the accuracy of a face recognition. The best method to overcome noise in the image is to use smoothing (filter). We can use the Gabor filter and Non-Negative Matrix Factorization (NMF). The noises encountered consist of impulse noise (salt-and-pepper), additive noise (Gaussian) and multiplicative noise (speckle).
- *Covert the image to binary/grayscale*
  Binary/grayscale images are black and white images. We use binary/grayscale images because coloured images contain a lot of information which may not be required for our algorithm and thus increasing the time of execution.
  Binary/grayscale images allow easy separation of an object from the background
- *Pixel brightness transformation*
  These modify pixel brightness.
  There are 2 types of brightness transformation:
  1. Brightness correction
     They consider the original brightness
     They depend on the position of pixel in the image
  2. Grey scale transformation
     They do not depend on the position of the pixel
     They are mostly used if the result is viewed by a human
     The final image has equally distributed brightness levels over the whole brightness scale
- *Geometric transformation*
  These transformations are done when the image is distorted when it is captured

2.Face Registration :

- **Haar Features :**

All human faces share some similar features, like eye region is darker than upper check region, nose region is brighter than eye region.


(a) Edge Features


(b) Line Features

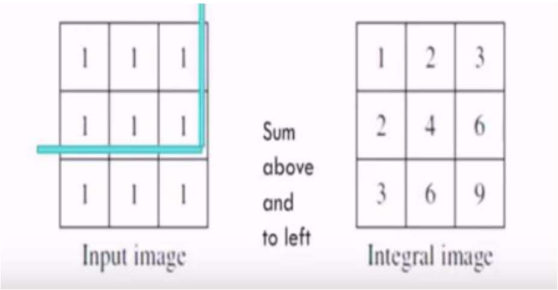

(c) Four-rectangle features

Here are some Haar feature, using those we can say there is a face or not.
Haar feature signifies that black region is represented by +1 and white region is represented by -1 .
It uses a 24X24 window for an image. Each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle. Now all possible sizes and locations of each kernel is used to calculate plenty of features. For each feature calculation, we need to find sum of pixels under white and black rectangles. For 24X24 window, there will be 160000+ Haar features, which is a huge number. To solve this, they introduced the integral images.

- **Integral Images:**

The basic idea of integral image is that to calculate the area. So, we do not need to sum up all the pixel values rather than we have to use the corner values and then a simple calculation is to be done.
The integral image at location x, y contains the sum of the pixels above and to the left of x, y, inclusive :

$$ii(x, y) = \sum_{x' \le x, y' \le y} i(x', y'),$$



- **Adaboost :**

Adaboost is used to eliminate the redundant feature of Haar. A very small number of these features can be combined to form an effective classifier.

Using adaboost we can determine which are relevant out of 160000+ feature. After finding all the features, a weighted value is added to it which is which is used to evaluate a given window is a face or not.
$F(x) = a1f1(x) + a2f2(x) + a3f3(x) + a4f4(x) + a5f5(x) + \ldots.$

F(x) is strong classifier and f(x) is weak classifier.
Weak classifier always provide binary value i.e. 0 and 1. If the feature is present the value will be 1, otherwise value will be 0.

- **Cascading :**

Suppose, we have an input image of 640X480 resolution. Then we need to move 24X24 window throughout the image and for each window 2500 features are to be evaluated. Taking all 2500 features in a linear way it checks weather there is any threshold or not and then decide it is a face or not.
But instead of using all 2500 features for 24X24 times we will use cascade. Out of 2500 features, 1st 10 features are classified in one classifier, next 20-30 features are in next classifier, then next 100 in another classifier.
The advantage is we can eliminate non face from 1st step instead of going through all 2500 features for 24X24 window.

3. Classification

1. **Support Vector Machines (SVM)**
   The object u want to classify is represent as a point in n-dimensional space. The co-ordinates of these points are usually called features. SVMs perform the classification test by drawing a hyperplane i.e. a line in 2d or a plane in 3d in such a way that all points of similar category are on one side of the hyperplane and all point of second category are on the other side of the hyperplane. As there can be multiple planes SVMs try to find

the hyperplane which separates these category in the sense that it maximizes the distance to the points in either category. This distance is call the margin and the points that fall on the margin are called the supporting vectors.

To maximize the distance between the planes we want to minimize $\|\mathbf{w}\|$

We also have to prevent data points from falling into the margin, we add the following constraint: for each i either

$$\mathbf{w}^\mathsf{T}\mathbf{x}_i - b \geq 1\,,\ \text{if}\ y_i = 1, \quad \text{or} \quad \mathbf{w}^\mathsf{T}\mathbf{x}_i - b \leq -1\,,\ \text{if}\ y_i = -1.$$

In case of emotion detection, usually a multi-class SVM is used instead of a binary to detect emotions such as anger, contempt, disgust, fear, happy, sadness and surprise [1]. K-fold cross-validation is used to remove any variances in the database and to compare different machine learning algorithms [9]. In k-fold cross validation, the dataset is divided k times into k slices, and prediction results are averaged over all iterations.

2. Hidden Markov Models (HMM)

Hidden Markov Models (HMM) is based on statistics, and are useful for finding hidden structure of data. They are also very popular for emotion detection through speech . Input is a sequence of observed features and there are hidden states corresponding to consecutive events. HMM is expressed as follows:

$\lambda = (A, B, \pi)$

where,

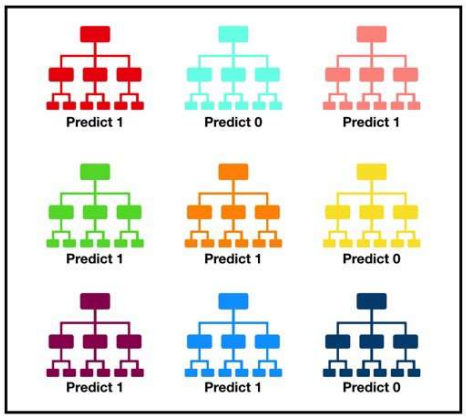
A = (a_{ij}) transition probability matrix between the hidden states

B = (b_{ij}) observation symbols probability from a state

Π = initial probability of states

3. Random Forest Classifiers

Random forests are based on decision trees, but instead of just one classifier, use more forests or classifiers to decide the class of the target variable. Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes become our model's prediction. **Ensemble methods** use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone



Tally: Six 1s and Three 0s
Prediction: 1

4. **K-nearest neighbors (kNN)**

   K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.

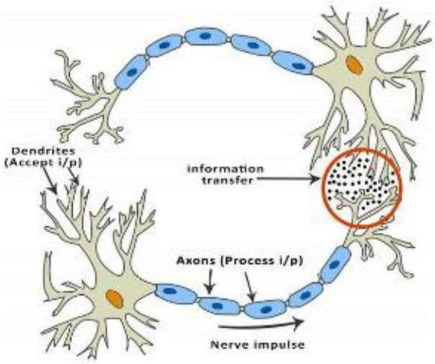
   K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

   K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

   Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

5. **Artificial Neural Networks**:

   The idea of ANNs is based on the belief that working of human brain by making the right connections, can be imitated using silicon and wires as
   living neurons and dendrites. The human brain is composed of 86 billion nerve cells called neurons. They are connected to other thousand cells by Axons. Stimuli from external environment or inputs from sensory organs are accepted by dendrites. These inputs create electric impulses, which quickly travel through the neural network. A neuron can then send the message to other neuron to handle the issue or does not send it forward.



ANNs are composed of multiple nodes, which imitate biological neurons of human brain. The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value**.** Each link is associated with weight**.** ANNs are capable of learning, which takes place by altering weight values.

Conclusion:

I chose the Haar Cascade method for face extraction. Benefits of this algorithm are: they're very fast at computing Haar-like features due to the use of integral images (also called summed area tables. They are also very efficient for feature selection through the use of the AdaBoost algorithm. They can detect faces in images regardless of the location or scale of the face. The are also capable of running in real-time.

I selected the artificial neural networks algorithm. The approach is based on the assumption that a neutral face image corresponding to each image is available to the system. Each neural network is trained independently with the use of on-line back propagation. A research paper showed that it produced 97% training accuracy, 85% testing accuracy, and 95% overall accuracy for the four basic (Angry, Happy, Neutral, and Sad emotions. . The diagonal components reveal that all the emotions can be recognized with more than 75.00 percent of accuracy. It is the most advanced method to recognize emotions and is also helpful for a larger database. We can use multilayer neural network which can give a proper emotional state.