

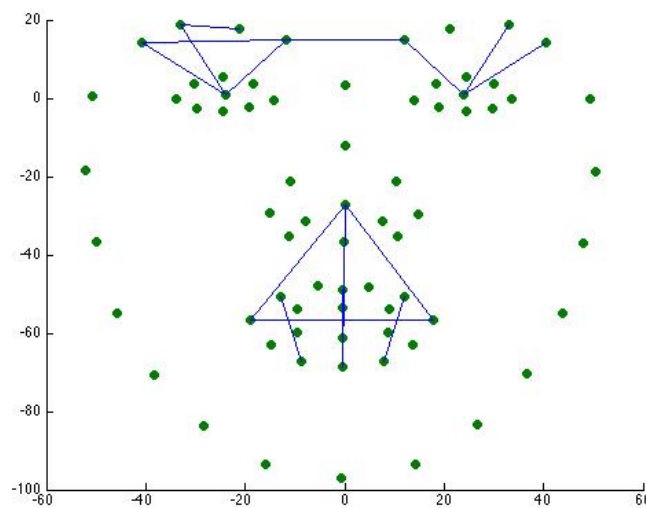
CS/ECE/ME532 Assignment 4

1. The goal of this problem is to design a classifier that will predict if a person, represented by measurements of their face, is happy or angry. A key to any classification task is to use good features that discriminate between the two categories.

Consider the two faces below. What features help you determine whether the person is happy or angry? The mouth, eyes, and brow certainly seem to be important clues.



The image below depicts a set of landmarks that can be automatically measured in a face image. These include points corresponding to the eyes, the brows, the nose, and the mouth. We will use $n = 9$ distances between pairs of these points to classify whether the image represents someone who is happy or angry.



Features extracted from $m = 128$ face images (like the two shown above) provided in the m -by- n matrix \mathbf{X} in the file `face_emotion_data.mat`. This file also includes the $m \times 1$ vector of labels \mathbf{y} . Here happy faces are labeled $+1$ and angry faces are labeled -1 . Your task is to find the weights for a linear classifier that will use the features to predict whether the emotion displayed on a face image is happy or angry.

Define a feature vector $\mathbf{x}_i = [x_{1i} \ x_{2i} \ \cdots \ x_{9i}]^T$ and classifier weights $\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_9]^T$ so that the label, $y_i \approx \mathbf{x}_i^T \mathbf{w}$.

- a) Use the training data \mathbf{X} and \mathbf{y} and a least squares problem to train your classifier weights.
- b) Explain how to use the weights you found to classify a new face image as happy or angry?
- c) Which features seem to be most important? Justify your answer. Note that the nine columns of the training data feature matrix \mathbf{X} have been normalized to have the same two-norm.
- d) Design a classifier based on three of the nine features. Which three should you choose? Describe the procedure for designing your classifier.
- e) What percent of the training labels are incorrectly classified using all nine features? What percent of the training labels are incorrectly classified using your reduced set of three features?
- f) Now use cross validation to assess your classifier performance. Divide the available data in to eight subsets of sixteen samples (e.g., examples $1-16, 17-32, \dots, 113-128$). Use seven sets to design your classifier weights, then use the remaining hold-out set to evaluate the classifier performance. Compute the number of misclassifications made on this hold-out set and divide that number by 16 (the size of the set) to estimate the error rate for that hold-out set. Repeat this process eight times using the eight different possible divisions between training and hold-out sets and average the error rates to obtain a final performance estimate.