Prasanna Kumar Srinivasachar
Sham Prasad PS
Vibhor Mishra

Project Baseline reports

Project:

- Gender prediction based on handwriting
    - Dataset : Set of images having handwritten texts and a csv file that has writer id and gender(labels) which need to be predicted. (https://www.kaggle.com/c/icdar2013-gender-prediction-from-handwriting/data)
    - **Techniques explored**: For the baseline, we have used KNN for the classification task. We also tried other algorithms like SVM with a linear kernel (LinearSVC)) and SVC with rbf kernel. We used 200 training samples and 50 test samples. The accuracy reported are as follows :

	KNN (neighbors = 10)	LinearSVC	SVC(kernel=rbf,gamma=1.5,C=2)
Training data	0.74	0.73125	0.7600
Test data	0.615	0.6600	0.6700

- Emotion detection: We try to predict the emotion expressed by a person in the given image.
    - Dataset: csv files containing emotions column along with pixel values of facial images (labels) . (https://inclass.kaggle.com/c/facial-keypoints-detector/data)
    - We segmented facial parts using facial landmark detector, an implementation of the "One Millisecond Face Alignment with an Ensemble of Regression Trees", included in the dlib library. As features, we have used width and length of these facial parts, i.e. both eyes, both eyebrows and lips, as of now which is doing pretty well and giving us a decent accuracy.
    - After performing gridSearch with a cross validation of 5. We got 41 neighbours as the best param and attained an accuracy of 60.531. The following graph shows

mean of the accuracies for this 5-fold cross validation versus the number of nearest neighbours.