Name: David Hatcher

Guided Learning #2

CAP4103 / CAP6101

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Caltech*** | **Classifier** | **Parameters** | **Additional Changes** | **Accuracy** | **Observations** |
| **Model 1** | SVC | Kernal=’linear’  C=.75 | None | 0.11 | Going to try changing the C value |
| **Model 2** | SVC | Kernal=’linear’  C=1.5 | Added code to view individual accuracy to see if maybe there was something specific for certain people that could be addressed | 0.12 | Time complexity seems to be going up, might not be efficient enough to continue with SVC |
| **Model 3** | SVC | Kernal=’linear’  C=2 | None | 0.13 | Time complexity very high, switching kernal |
| **Model 4** | KNN | K = 9 | Changed landmarks to 68 after looking at the people that were having lots of misses | 0.47 | Doing much better but I probably need to dial the k value in a little more |
| **Model 5** | KNN | K = 7 | Landmarks = 68 | 0.50 | Might need a lower value since one label has only 5 images |
| **Model 6** | KNN | K = 5 | Landmarks = 68 | 0.57 | K might need to be lower because of the 5 image outlier |
| **Model 7** | KNN | K = 3 | Landmarks = 68 | 0.65 | Trying k = 1 but might need try a different distance metric as k = 1 seems weak |
| **Model 8** | KNN | K = 1 | Landmarks = 68 | 0.72 | I don’t really feel k = 1 is really good as it leaves a lot of room for error, trying different distance metrics |
| **Model 9** | Gaussian Naïve Bayes | Default | Landmarks = 68 | 0.78 | Going to adjust smoothing to see if I can get better accuracy out of that |
| **Model 10** | Gaussian Naïve Bayes | var\_smoothing = .5 | Landmarks = 68 | 0.93 | Might be able to get higher with more tweaking of the smoothing |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***SoF*** | **Classifier** | **Parameters** | **Additional Changes** | **Accuracy** | **Observations** |
| **Model 1** | Gaussian Naïve Bayes | var\_smoothing = .5 | Landmarks = 68 | 0.86 |  |
| **Model 2** | Gaussian Naïve Bayes | Var\_smoothing = 1 | Landmarks = 68 | 0.87 | Accuracy seems to be capping out as changing the smoothing in either direction isn’t getting any better. Switching classifier |
| **Model 3** | Gaussian Naïve Bayes | var\_smoothing = 1 | Increased brightness of the images  Landmarks = 68 | 0.88 | Going to try increasing brightness more and possibly adjusting resize of the images |
| **Model 4** | Gaussian Naïve Bayes | var\_smoothing = 1 | Increased resize of images from 100x100 to 200x200  Landmarks = 68 | 0.95 | Looking at the images many seem dark, going to try adjusting brightness more |
| **Model 5** |  |  |  |  |  |
| **Model 6** |  |  |  |  |  |
| **Model 7** |  |  |  |  |  |
| **Model 8** |  |  |  |  |  |
| **Model 9** |  |  |  |  |  |
| **Model 10** |  |  |  |  |  |

After trying the following, with different parameters in each, I was unable to achieve any more increases:

* KNN
* SVM linear
* SVM poly
* DecisionTree
* RandomForest
* MLP
* Categorical Naïve Bayes
* Complement Naïve Bayes

**Did the classification model generalize to the SoF dataset? If yes, explain why you think it did generalize. If no, explain why you think it did not.**

The classification model that the best fit I could find for the Caltech dataset was a good start, 87% accuracy, but not the best accuracy I found for the SoF dataset. One of the biggest reasons I believe this to be the case was that many of the images for SoF dataset were larger than those of the Caltech dataset and the faces tended to be further away. As the original image was set to 100x100 pixels and only upsampled 1 time. I believe this made it harder for the detector to detect faces in these images which lead to a much smaller set of training samples therefore a much less accurate model. Once I increased the image sizes to 200x200 I was able to achieve a much higher accuracy. I also believe like the brightness was an issue as well, as many of these images were quite dark.

**Did the classification model generalize to the Caltech dataset? If yes, explain why you think it did generalize. If no, explain why you think it did not.**

The classification model that was best for SoF was about the same for Caltech as, Caltech was for SoF, I was able to achieve 84% accuracy with the SoF model when applied to Caltech. I believe the major issue that caused this was the increase in brightness that was being applied in SoF. Most of the Caltech images were well lit and I believe the added brightness was obscuring some of the images and causing the incorrect classifications.