Problem 1

Summary of train and test accuracies:

Spambase and Spambase abridged:
Linear kernel: Train accuracy: 97.19% Test accuracy: 87.5%
RBF kernel : Train accuracy: 95.0% Test accuracy: 90.0%
Polynomial kernel: Train accuracy: 99.38% Test accuracy: 83.75%

fake-data1:
Linear kernel: Train accuracy: 98.17% Test accuracy: 100.0%
RBF kernel: Train accuracy: 100.0% Test accuracy: 100.0%
Polynomial kernel : Train accuracy: 100.0% Test accuracy: 100.0%

fake-data2:
Linear kernel: Train accuracy: 54.07% Test accuracy: 44.19%
RBF kernel: Train accuracy: 100.0% Test accuracy: 100.0%
Polynomial kernel : Train accuracy: 100.0% Test accuracy: 100.0%

As seen with how the SVM classifier performed on the spam dataset, the training accuracy for the linear kernel was at around 97 percent, with testing accuracy at nearly 88 percent. With the RBF kernel, both train and test accuracies were 90 percent or higher, and with the polynomial kernel, train accuracy was nearly at 100 percent, whereas test accuracy was at around 84 percent.

Moreover, as seen with the spambase and spambase abridged datasets, the kernel with the highest test accuracy was the RBF kernel, whereas the kernel with the highest train accuracy was the polynomial kernel. However, the kernel with the lowest train accuracy was the RBF kernel, and the lowest test accuracy was the linear kernel. Looking at the other datasets, it is clear that overall, the linear kernel lowered the train and test accuracies the most, while the RBF had a slightly better performance overall (across datasets).

Problem 2

As mentioned in the article, I think that they should limit their use of facial recognition technologies, particularly for the time being, given that they are clearly incredibly inaccurate when applied to individuals of different demographics and races. Until they have been refined and thoroughly tested, it is best that they are not used as the sole source of information when

making an arrest based off of video surveillance. The technologies should only really be used in extreme scenarios (such as violences), given that they have the potential to really disrupt lives. I also think that, in order to further prevent wrongful accusations, police should also refer to other forms of evidence, such as eyewitness testimonies, phone proof, as well as location data. Overall, more than one form of evidence should be referred to when investigating a crime.

Therefore, I believe that the companies producing these facial recognition technologies should focus on training their models on individuals of color, rather than mostly white men, because clearly, their technologies function unfairly on individuals of different demographics. If they are able to improve their algorithm so that it produces more accurate matches, this could prevent future wrongful accusations.

Thus, I do believe that increasing the racial diversity of the images fed into the algorithm would help. It is possible that there were fewer racially diverse images fed into the algorithm while it was being trained, which could have resulted in the accuracies being significantly lower for African American and Asian individuals. Having more data of both racial groups would likely help increase accuracies.

Additionally, I also believe that low-quality images should be banned from being used, given that they would only likely result in extremely inaccurate matches. The lower the quality of the image, the more difficult it is for the algorithm to produce a correct match. Thus, this would only increase the chances of there being a wrongful accusation made. For this reason, I believe that these algorithms should only be used in law enforcement in instances of extreme violence— certainly not for smaller crimes such as theft. Since facial recognition can never be 100 percent accurate, it is best for law enforcement agencies to hold off on using the technology for the time being.