Pr. 9. (sol/hsjt3)

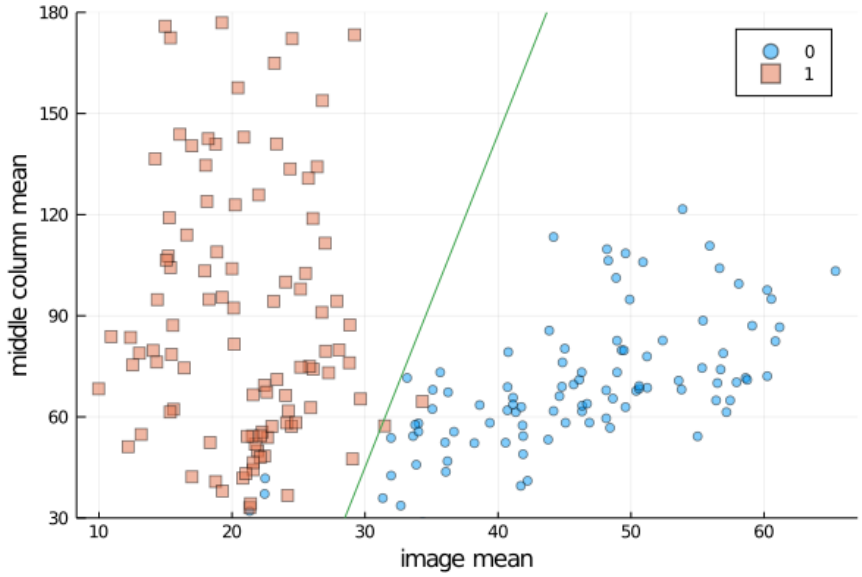
Part 1 Code for a `data_to_features` function (it is fine to use a loop too):

```
feature1_fun = (d) -> [mean(d); mean(d[Int(end/2),:]); 1]
"""
features = data_to_features(d3)

Extracts three hand-crafted features from the given data.

In:
    d3 : a stack of images, i.e., a data array of size [nx,ny,n],
            where n is the number of images in the stack

Out:
- An array of size [n,3], where the first column is the mean of each image,
        the 2nd column is the middle-column-mean of each image,
        and the third column is the ```1``` feature
"""
function data_to_features(d3)
    tmp = mapslices(feature1_fun, d3, dims=1:2)
    return reshape(tmp, size(tmp,1), size(d3,3))'
end
```

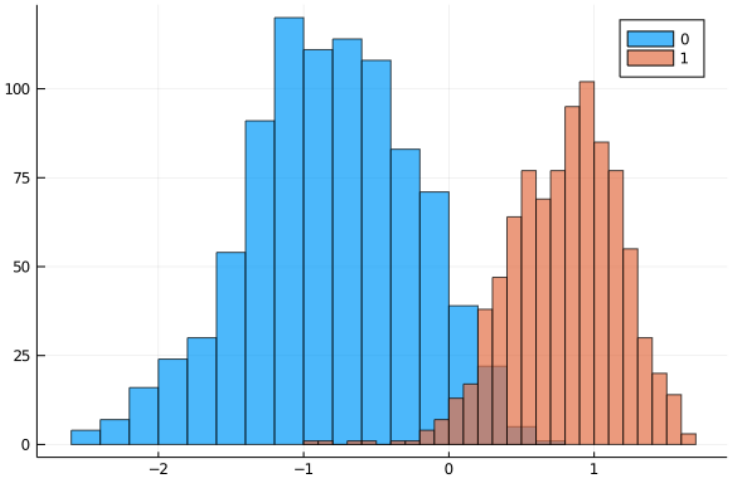
Part 2 LS estimates: $\hat{\boldsymbol{x}} = \begin{bmatrix} -0.057 \\ 0.006 \\ 1.456 \end{bmatrix}$

Coefficient x_1 is negative because the 0 digits tend to have a larger mean value than the 1 digits. Consider using $\boldsymbol{x}$ to classify an image as $\text{sign}(\boldsymbol{a}'\boldsymbol{x})$. The larger mean of the 0 images leads to a larger magnitude negative value for the first sum in the inner product compared to the 1 digit images, making it more likely that the sign of the inner product with a 0 digit image is negative, which improves the tendency to correctly classify a 0 digit.

Scatter plot with separating line (decision boundary):

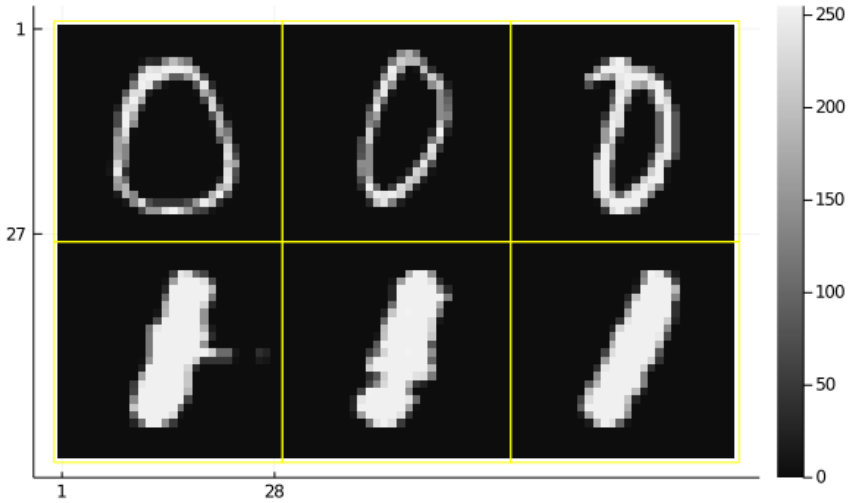


Part 3 Histogram of inner products $\{\boldsymbol{a}_i'\boldsymbol{x}\}$:



Percent of correctly classified digits:
Percent 0 correct = 92.6%
Percent 1 correct = 98.1%

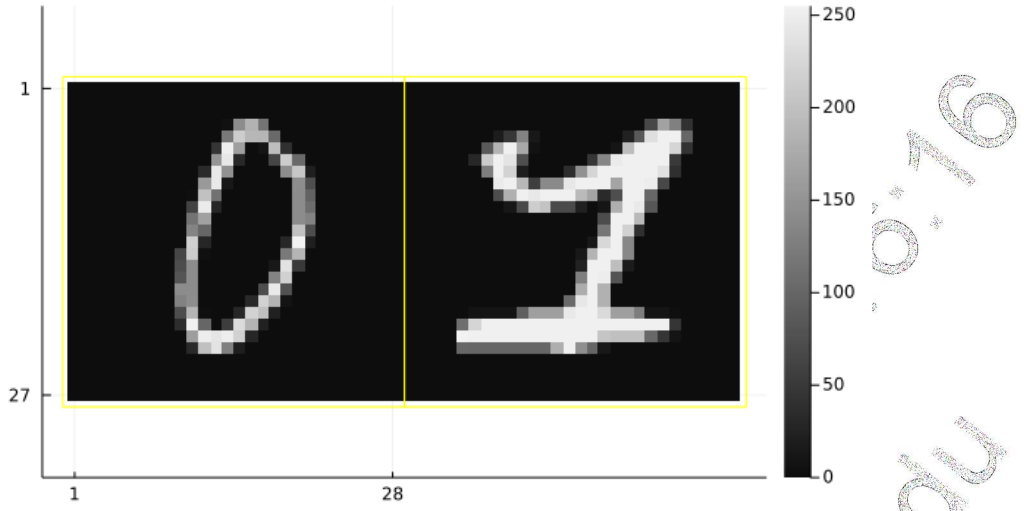
Part 4 Show some of the misclassified images:



Comment on why these were misclassified:

These zero digit images are very "faint" (lower norm than typical) wheras the 1 digit images have higher mean values than usual.

Part 5 Show the "worst classified" "0" and "1" images when using Tikhonov Regularization with a regularization parameter $\beta = 10$:



How do you think the classifier will perform as $\beta \to \infty$?

As the regularization parameter tends to infinity, the minimizer of the cost function approaches $\hat{\boldsymbol{x}} \to \mathbf{0}_3$. The classifier will find an $\hat{\boldsymbol{x}}$ whose norm is approximately 0, instead of one that minimizes $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2$, which will cause the classifier to perform poorly.

In this case our $\boldsymbol{A}$ here is $M \times 3$ where M is quite large and $\boldsymbol{A}$ has linearly independent columns. Tikhonov regularization is especially helpful when $\boldsymbol{A}$ is wide or has (nearly) linearly dependent columns.