Problem 1

(i) The distance between the parallel lines is the distance between the intersecting points on the line $y = -\frac{x}{w}$

$$wx - b = -\frac{x}{w}$$

$$wx + \frac{x}{w} - b = 0$$

$$\left(w + \frac{1}{w}\right)x - b = 0$$

$$x = \frac{bw}{w^2 + 1}$$

$$-wy = \frac{y + b}{w}$$

$$y = \frac{-b}{w^2 + 1}$$

$$for\ y = 1, \qquad b = -1 - w^2$$

$$for\ y = -1, \qquad b = w^2 + 1$$

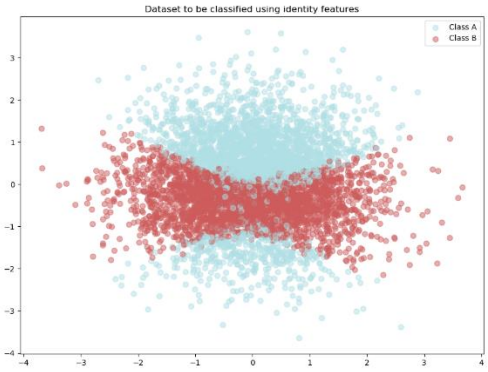
$$d = sqrt\left(\left(\frac{(w^2 + 1)w - (-1 - w^2)w}{w^2 + 1}\right)^2 + \left(1 - (-1)\right)^2\right)$$

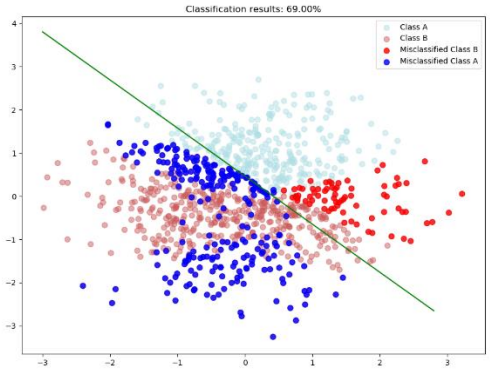
$$d = 2w$$

(ii) See code

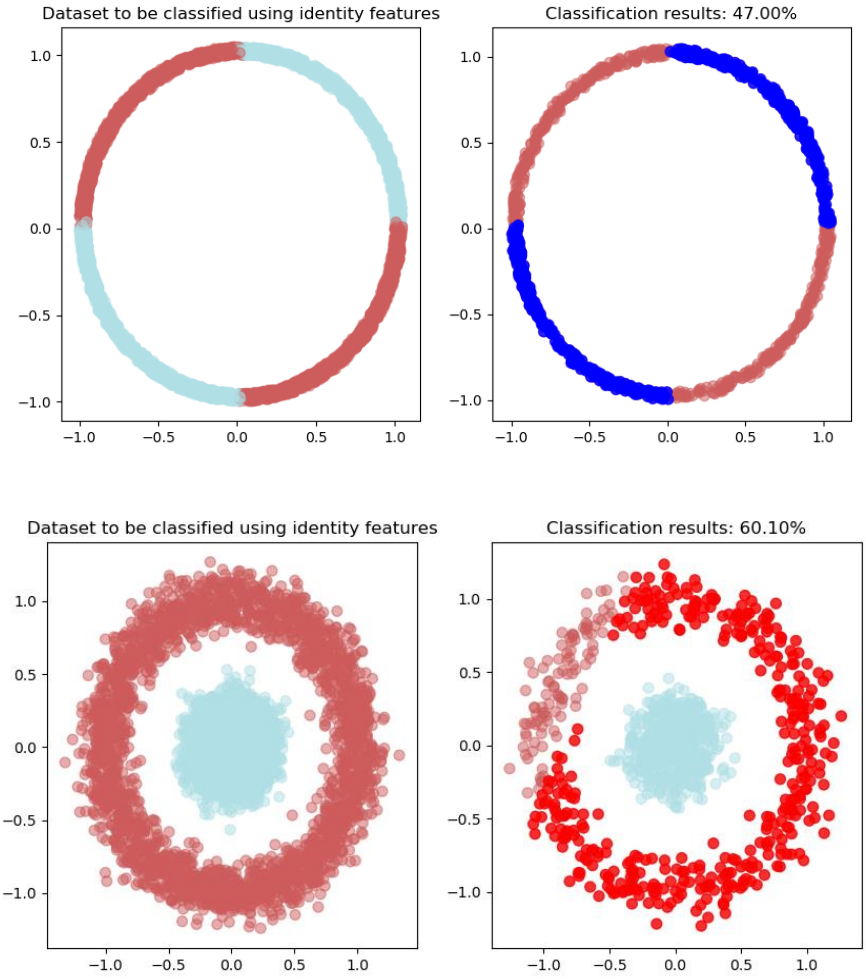
(iii) The SVM does very well since the data is separable.

(iv) The SVM doesn't do very well since the data is not separable with the current mapping resulting in consistent misclassification

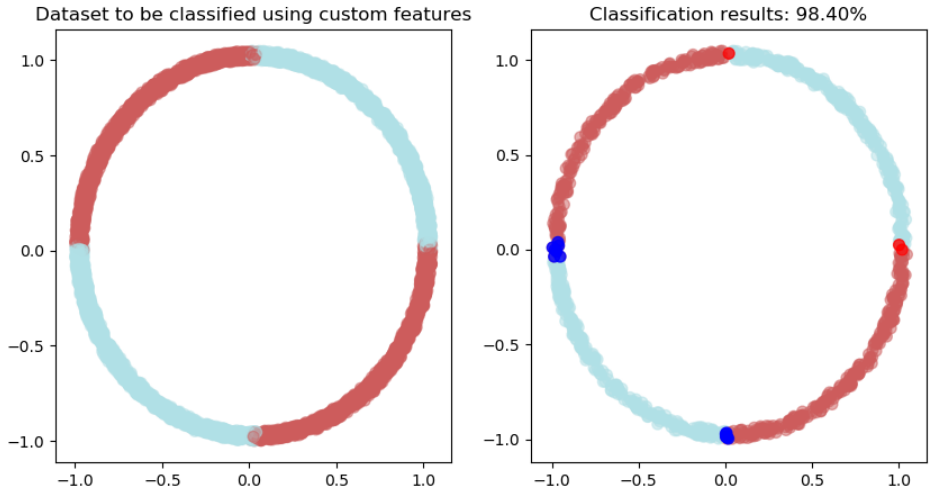


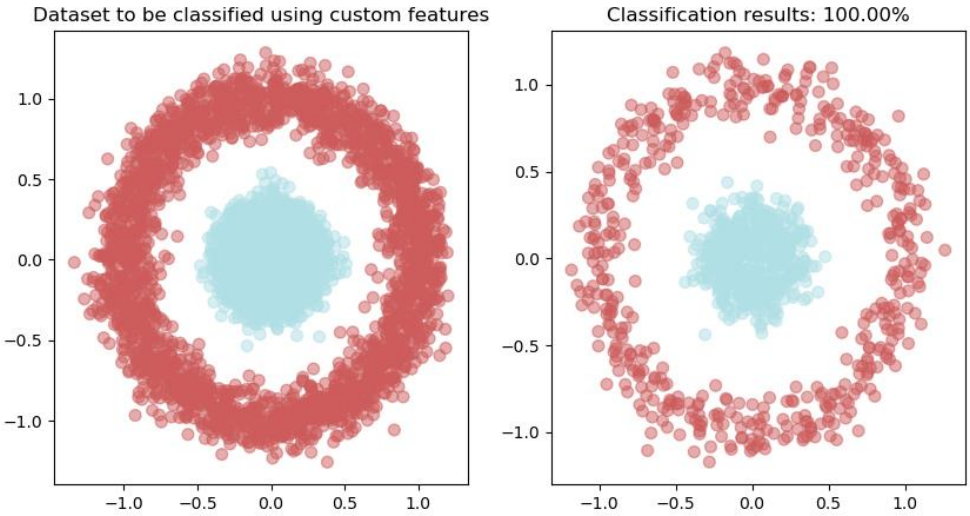


(v) Like the nonlinear case, the circle and inner circle cannot be sufficiently classified with our linear set of basis functions. For the circle, it appears the data can be separated by a tangent relationship, that is a ratio of $\frac{x_2}{x_1}$. The data in the 1st and 3rd quadrants of the circle would have a positive ratio and the 2nd and 4th would have a negative ratio, providing a degree of separation. For the inner circle, the equation of a circle could be used as a third dimension, where the innermost data would have a smaller radius and the outermost data would have a larger radius.


Dataset to be classified using identity features
Classification results: 47.00%
Dataset to be classified using identity features
Classification results: 60.10%

(vi) See plots

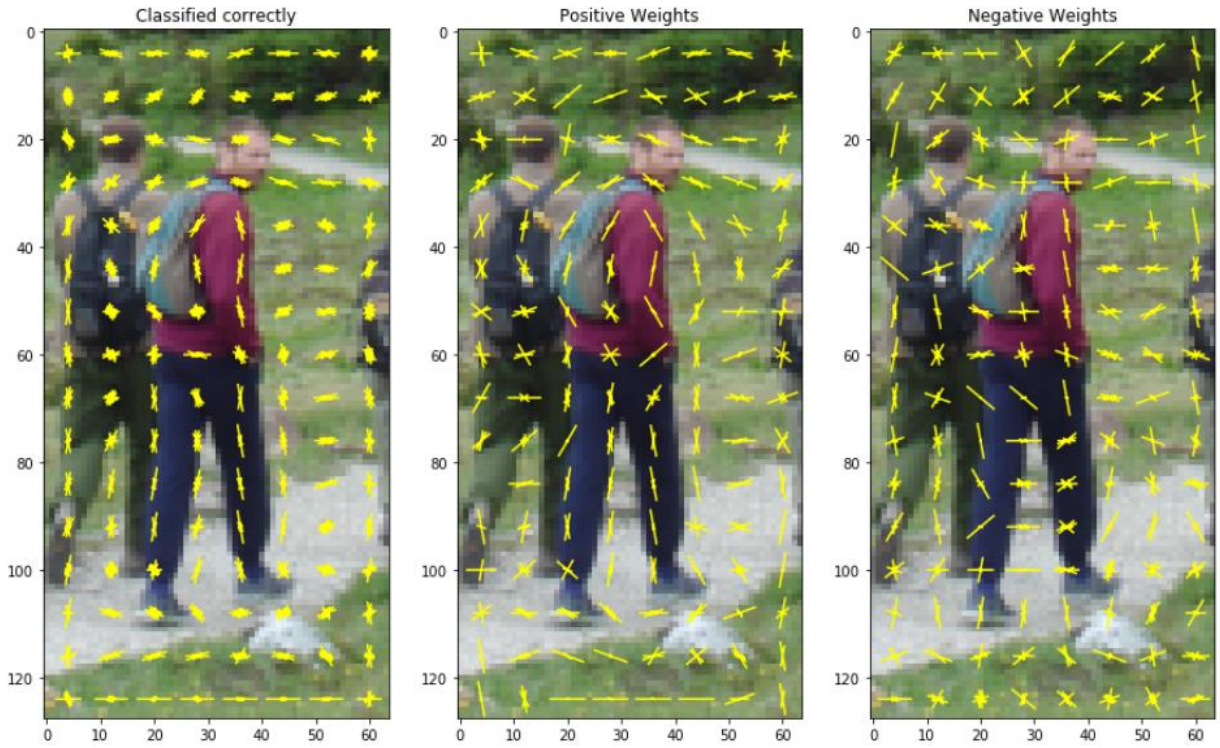




(vii) SVM is a linear classifier because it finds a linear mapping of the features in the form xW – b. If x is a basis function, the features can be nonlinear, and still SVM can produce a linear mapping for the separation of the data.

(viii) See code

(ix) The classification accuracy was 97%.



Problem 2

(i) See code
(ii) See code
(iii) See code
(iv)
(v) You may not train on all available features in the images, since you are getting the images at a later stage of the model.
(vi) See code
(vii) See code
(viii)
(ix) It feeds into the global average pooling. It takes the average of the regions
(x) See code
(xi)
(xii) See code
(xiii)

Problem 3

(i)
(ii)
(iii)

(iv)

(v) There are two Dense Layers. For activations, ReLu is used on the first layer and there is also a softmax to get a distribution of probabilities

(vi) See code

(vii)