EECS 445 HW4

1.

Code is provided.

1d. For each pixel in our uncompressed image, RGB can take any value. Thus, we have 0-255 values for each of R, G, and B. Each of these can be represented with 8 bits, so we have a total of

$$numerator = 512\ x\ 512\ x\ 24\ bits$$

to store the uncompressed image.

For the compressed image, RGB can only take one of 16 values. Of course, if we want to store this, we need to create a mapping from each point to its nearest centroid, which will require a constant C bits to store that mapping, in addition to the 4-bits required to store the data for the centroids. Thus, for the compressed image, we have a total of

$$denominator = (512\ x\ 512\ x\ 4) + \ C\ bits$$

Thus, our compression factor is

$$\frac{numerator}{denominator} = \frac{512\ x\ 512\ x\ 24}{(512\ x\ 512\ x\ 4) + C} = 6x\ factor\ approximately$$

2. (a) From previous hw,

$$\nabla_w E(\vec{w}, b) = \sum_{i=1}^{N} \frac{1}{N}\vec{w} - C\, I\left[t^{(i)}(w^T x^{(i)} + b) < 1\right] t^{(i)} x^{(i)}$$

$$\frac{\partial}{\partial b} E(\vec{w}, b) = \sum_{i=1}^{N} -C I\left[t^{(i)}(w^T x^{(i)} + b) < 1\right] t^{(i)}$$

Update rules. of SGD

$w^* \leftarrow 0 \qquad b^* \leftarrow 0$

for $j=1$ to NumIterations do

  for $i=1$ to N do

    $w_{grad} \leftarrow \nabla_w E^{(i)}(w^*, b^*)$

    $b_{grad} \leftarrow \frac{\partial}{\partial b} E^{(i)}(w^*, b^*)$

    $w^* \leftarrow w^* - \alpha(j)\, w_{grad}$

    $b^* \leftarrow b^* - 0.01\, \alpha(j)\, b_{grad}$

  end

end

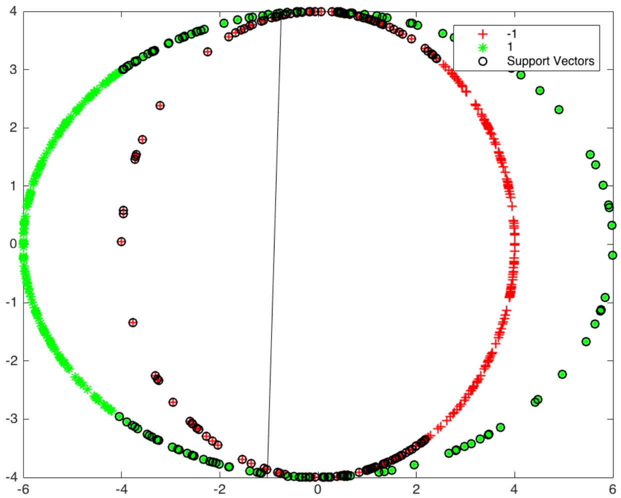
return w^*

```
-------------------q2b---------------------------
0.100000 is the optimal value of C
0.500000 is the optimal value of n0
9.000000 is the percent error on testing data
-------------------------------------------------
```

```
-------------------q2c---------------------------
1.000000 is the optimal value of C
1.000000 is the optimal value of n0
8.500000 is the percent error on testing data
-------------------------------------------------
```

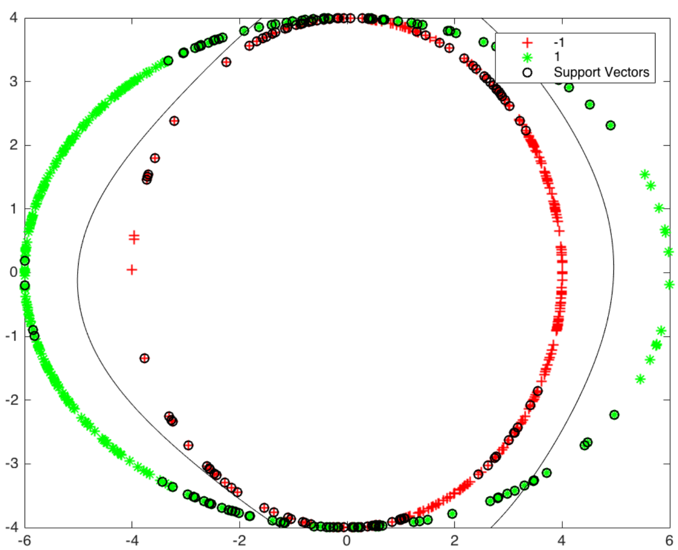
Q3a.

88.0 percent classification accuracy



Q3b.

100.0 percent classification accuracy



Q3c.

Using a Gaussian kernel function allows us to more richly classify our data from not using a kernel, thus making it easier to classify data using an SVM model.

Q3d.

Optimal value of sigma = 0.2
Accuracy = 100%

```
------------------q3c--------------------------
0.200000 is the optimal value of sigma
-------------------------------------------------
```

4. (a) $\vec{h} = \tanh(b + W\vec{x})$

(i)

$$\hat{t} = \sigma(b_4 + \theta^T\vec{h})$$

$$E(x,\hat{t}) = -(t\log(\hat{t}) + (1-t)\log(1-\hat{t}))$$

Want to find $\nabla_\theta E(x,\hat{t})$, by chain rule, we know $= \frac{\partial \hat{t}}{\partial \theta} \cdot \frac{\partial E}{\partial \hat{t}}$

First, compute $\frac{\partial E}{\partial \hat{t}}$

$$\Rightarrow \frac{\partial}{\partial t} E(x,t) = \frac{\partial}{\partial t} -(t\log(\hat{t}) + (1-t)\log(1-\hat{t}))$$

$$= \frac{-\partial}{\partial t} t\log(\hat{t}) - \frac{\partial}{\partial t}(1-t)(\log(1-\hat{t})) = \frac{1-t}{1-\hat{t}} - \frac{t}{\hat{t}} = \frac{\hat{t}-t}{\hat{t}(1-t)}$$

Next, compute $\frac{\partial \hat{t}}{\partial \theta} = \frac{\partial}{\partial \theta}\sigma(b_4 + \theta^T\vec{h}) = \sigma(b_4 + \theta^T\vec{h})(1-\sigma(b_4+\theta^T\vec{h}))\vec{h}$

Multiplying, we get $= \hat{t}(1-\hat{t})\vec{h}$

$$\nabla_\theta E(x,t) = \cancel{\hat{t}(1-\hat{t})}\vec{h} \cdot \frac{\hat{t}-t}{\cancel{\hat{t}(1-\hat{t})}} = \boxed{\vec{h}(\hat{t}-t)}$$

(ii) Same procedure as (a), except we want to compute $\frac{\partial \hat{t}}{\partial b_4} \cdot \frac{\partial E}{\partial \hat{t}}$

$$\frac{\partial \hat{t}}{\partial b_4} = \frac{\partial}{\partial b_4}\sigma(b_4+\theta^T\vec{h}) = \sigma(b_4+\theta^T\vec{h})(1-\sigma(b_4+\theta^T\vec{h}))$$
$$= (\hat{t})(1-\hat{t})$$

$$\Rightarrow \frac{\partial}{\partial b_4} E(x,t) = \cancel{(\hat{t})(1-\hat{t})} \cdot \frac{\hat{t}-t}{\cancel{(\hat{t})(1-\hat{t})}} = \boxed{\hat{t}-t}$$

☒

Scanned by CamScanner

4 (b) Want to find

(i)

$\nabla_W E(\vec{x}, t)$, by chain rule, $= \frac{\partial E}{\partial h} \cdot \frac{\partial h}{\partial W}$

First, compute

$$\frac{\partial E}{\partial h} = \frac{\partial \hat{t}}{\partial h}\left[\frac{\partial}{\partial \hat{t}} E(\vec{x}, t)\right] = \frac{\partial}{\partial h}\left[\sigma(b_q + \theta^T \vec{h})\right]\left[\frac{\hat{t} - t}{\hat{t}(1-\hat{t})}\right]$$

$$= \sigma(b_q + \theta^T \vec{h})(1 - \sigma(b_q + \theta^T \vec{h}))\, \theta^T \left[\frac{\hat{t} - t}{\hat{t}(1-\hat{t})}\right]$$

$$= \cancel{(\hat{t})(1-\hat{t})}\, \theta^T \left[\frac{\hat{t} - t}{\cancel{\hat{t}(1-\hat{t})}}\right] = \theta^T(\hat{t} - t)$$

Next, compute

$$\frac{\partial h}{\partial W} = \frac{\partial}{\partial W} \tanh(\vec{b} + Wx) = [1 - \tanh^2(\vec{b} + Wx)]\, \vec{x}$$

$$= [1 - h^2]\, \vec{x}$$

Thus, $\nabla_W E(\vec{x}, t) = \theta^T(\hat{t} - t)(1 - \vec{h}^2)\vec{x}$

(ii) Similar to (i), except we want

$$\nabla_b E(\vec{x}, t) = \cancel{\frac{\partial E}{\partial h} \frac{\partial b}{\partial W}} \quad \frac{\partial E}{\partial h} \cdot \frac{\partial h}{\partial b}$$

Thus,

$$\cancel{\frac{\partial E}{\partial b} \frac{\partial b}{\partial W}} = \frac{\partial h}{\partial b} = \frac{\partial}{\partial b} \tanh(\vec{b} + Wx) = 1 - \tanh^2(\vec{b} + W\vec{x})$$

$$= 1 - h^2.$$

Thus, $\nabla_b E(\vec{x}, t) = \theta^T(\hat{t} - t)(1 - \vec{h}^2)$ $\boxtimes$