1.1

Features Used :

1) Perimeter

2) Area

3) Compactness.

~~Perimeter:~~

Motivation : Area and perimeter are a natural and the most intuitive features that one can think of.

Since the variance in the images given was quite less, area and perimeter seem to do most of our job.

However, to push the accuracy further, I used compactness $\propto$ Area / Perimeter^2

Compactness is the measure of circularity of a shape, i.e. It measures how close a shape is to a circle.

With these three feature I was able to get a test-accuracy of 96%.

Compactness is useful in distinguishing the ~~star and triangle~~ star and triangle shapes from the other two.

Q01

a) Let $K = 2$.

$$E(w) = -\frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j=1}^{2} y_j^{(i)} \times \log\left(P(Y=j \mid w_j, \phi(x^{(i)}))\right)$$

since, there are only ~~two~~ two classes,

if $y_0^{(i)} = 1$ And $y_1^{(i)} = 1$

$y_1^{(i)} = 0$ $\quad y_0^{(i)} = 0$

$$\therefore \quad y_1^{(i)} = 1 - y_0^{(i)}$$

More over, since there are only two classes,

$$P(Y=0 \mid w_0, \phi(x^{(i)})) + P(Y=1 \mid w_1, \phi(x^{(i)})) = 1$$

$$\therefore \quad E(w) = -\frac{1}{N} \sum_{i=1}^{N} y_0^{(i)} \log\left(P(Y=0 \mid w_0, \phi(x^{(i)}))\right)$$

$$+ (1-y_0^{(i)}) \cdot \log\left(1 - P(Y=0 \mid w_0, \phi(x^{(i)}))\right)$$

$$\therefore \quad E(w) = -\frac{1}{N} \sum_{i=1}^{N} y_0^{(i)} \cdot \log(p^{(i)}) + (1-y^{(i)}) \cdot \log(1-p^{(i)})$$

Which is the same as the binary cross entropy loss.

$\therefore$ It is a special case for the logistic Regression Loss.

2.1) b)

We will calculate the gradient for a single example and add them later for getting the batch gradient.

$$E(w)^{(i)} = -\sum_{k=1}^{K} y_k^{(i)} \times \log(P(Y=k \mid w_k, \phi(x^{(i)})))$$

$$= -\sum_{k=1}^{K} y_k^{(i)} \times \log\left(\frac{e^{-w_k^T \phi(x^{(i)})}}{\sum_j e^{-w_j^T \phi(x^{(i)})}}\right)$$

$$= -\left[\sum_{k=1}^{K} y_k^{(i)} (-w_k^T \phi(x^{(i)})) - y_k^{(i)} . \log\left(\sum_j e^{-w_j^T \phi(x^{(i)})}\right)\right]$$

$$= \log\left(\sum_j e^{-w_j^T \phi(x^{(i)})}\right) \sum_{k=1}^{K} y_k^{(i)} + \sum_{k=1}^{K} y_k^{(i)} (w_k^T) . \phi(x^{(i)})$$

$$= \log\left(\sum_j e^{-w_j^T \phi(x^{(i)})}\right) + \sum_{k=1}^{K} y_k^{(i)} w_k^T . \phi(x^{(i)})$$

$$\frac{\partial E(w)^{(i)}}{d w_{mn}} = \frac{\partial \log\left(\sum_j e^{-w_j^T \phi(x^{(i)})}\right)}{\partial w_{mn}} + \frac{\partial}{\partial w_{mn}} \sum_{k=1}^{K} y_k^{(i)} w_k^T . \phi(x^{(i)})$$

$$\frac{\partial (E(w)^{(i)})}{\partial w_{mn}} = \frac{-e^{-w_n^T \phi(x^{(i)})} \times \phi(x^{(i)})_m}{\sum_J e^{-w_j^T \phi(x^{(i)})}} + y_n^{(i)} . \phi(x^{(i)})_m .$$

where w_i represents ith col of W

$\phi(x^{(i)})$ represents ith row of $\phi(x)$

$y^{(i)}$ represents ith row of i.

$\therefore$ Let $P_{ij} = \dfrac{e^{-w_j^T \phi(x^{(i)})}}{\sum_K e^{-w_k^T \phi(x^{(i)})}}$ represent the softmax matrix $[P_{ij}]$

$\therefore$ We have; rewriting subscripts for clarity.

$$\frac{\partial E(w^{(i)})}{\partial w_{mn}} = \left(-P_{in} \cdot \phi(x)_{im} + Y_{in} \cdot \phi(x)_{im}\right)$$

$$= (Y_{in} - P_{in})\,\phi(x)_{im}$$

Now, Adding up gradients for all examples.

$$\left(\frac{\partial E(w)}{\partial w_{mN}}\right) = \frac{1}{N} \cdot \sum_i (Y_{in} - P_{in})\,\phi(x)_{im}$$

$$= \frac{1}{N} \sum_i \phi^T(x)_{mi} \cdot (Y_{in} - P_{in})$$

Since, $C_{mn} = \sum_i a_{mi} \cdot b_{in}$, where $C = A \cdot B$,

we have:

$$\left(\frac{\partial E(w)}{\partial w}\right) = \frac{1}{N}\, \phi^T(x) \cdot (Y - P)$$

~~where~~ [NOTE: I haven't followed the convention that $\frac{\partial E}{\partial w} = (\nabla_w E)^T$,

I have simply used $\frac{\partial E}{\partial w} = \nabla_w E$.]

2.2 b) TEST ACCURACY OBTAINED: 0.86.

Ans.

TEST ACCURACY FOR THE DESCRIBED ALGORITM : 0.84

No, Accuracy is Not a good choice for evaluation metric.

Because accuracy does not take care of "class - imbalance".

i.e. If 99% of the samples would be class A and only 1% would be class B, based on the underlying distribution on the samples;-

we may get 99% accuracy by saying that all samples given to us belong to class 'A'.

2.2 c) Yes, F_1 score is a good evaluation metric.

F_1 Score achieved: 0.30

F_1 score is better than Accuracy because, the algorithm was Not as good, and it wasn't able to do tremendously good than assigning all the data points to class '0', Even then we achieved an accuracy of 0.86.

However, F_1 score takes into consideration the imbalance in the classes ~~assigning~~ and gives a better picture.

2.2 e) Perceptron Test Accuracy: 0.786

LOGISTIC regression Test Accuracy : 0.84

Clearly Logistic regression achieved more test accuracy, because it is more complex than perceptron algorithm,

Moreover, it acts truely in another dimension and can do more than linearily separate class domains,

it can learn complex separating planes, because of the transfer function.