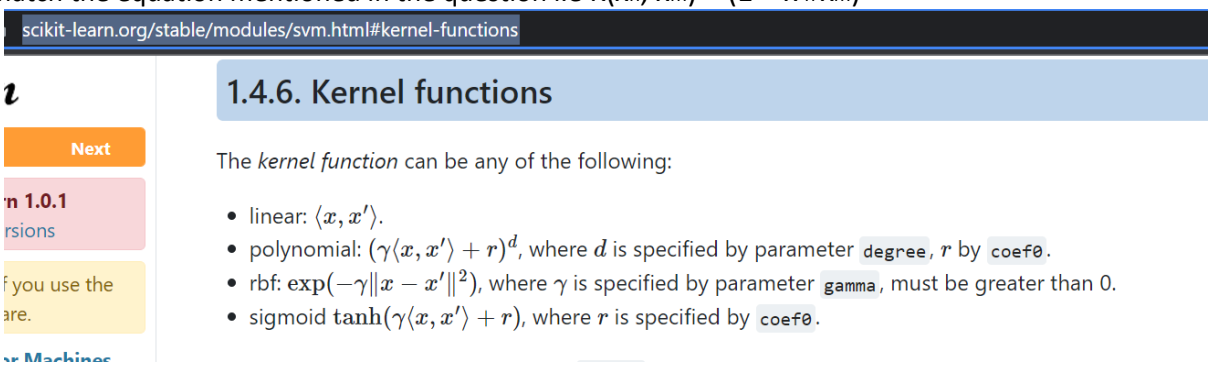
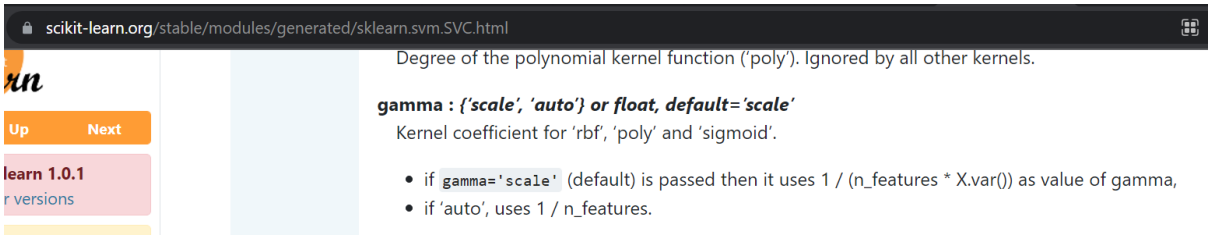
NOTE 1 **SCALING** - as per the link (http://scikit-learn.org/stable/modules/svm.html) in assignment , scaling is very important for SVM. Hence I have done modelling without scaling as well as with Min Max scaling & Standard Scaling.

NOTE 2 : Gamma values used in my code :

As per the documentation at – https://scikit-learn.org/stable/modules/svm.html#kernel-functions I am setting gamma = 1 for both poly & rbf kernels & coeff =1 even when it is not explicitly specified to match the equation mentioned in the question i.e $K(x_n; x_m) = (1 + x^T_n x_m)^q$



Also, https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html mentions gamma is a relevant parameter for poly kernels -



Q 4.(a)- Entire Test Set-

Without Scaling - the number of support vectors : [14 14]
test accuracy : 0.9787735849056604
With Min Max Scaling –
the number of support vectors : [44 45]
0.9787735849056604
With Std Scaling –
the number of support vectors : [18 18]
0.9811320754716981

Q. 4. (b) The results – as we can see from the Excel screenshot below, the number of support vectors increases with bigger data size, accuracy nearly remains same.

WITHOUT SCALING –

for train data set size : (50,)-Total no of support vectors : 2 i.e. for each class [1 1] ; accuracy for training data size 50 is :0.9811320754716981 //

for train data set size : (100,)- Total no of support vectors : 4 i.e. for each class [2 2]; accuracy for training data size 100 is :0.9811320754716981//

for train data set size : (200,)- Total no of support vectors : 8 i.e. for each class [4 4]; accuracy for training data size 200 is :0.9811320754716981//

for train data set size : (800,)--Total no of support vectors : 14 i.e. for each class [7 7]; accuracy for training data size 800 is :0.9811320754716981// --Rest of the data reported below -

Test Data Size	Results	Without Scaling	Min Max Scaling	Standard Scaling
50	Accuracy	0.981132075	0.976415094	0.971698113
100	Accuracy	0.981132075	0.981132075	0.981132075
200	Accuracy	0.981132075	0.981132075	0.981132075
800	Accuracy	0.981132075	0.981132075	0.981132075
50	No of Support Vectors	[1 1] = 2 total	[6 6] = 12 total	[2 2] =4 total
100	No of Support Vectors	[2 2] =4	[10 10] = 20 total	[3 3] = 6 total
200	No of Support Vectors	[4 4] = 8 total	[16 17] = 33 total	[5 6] = 11
800	No of Support Vectors	[7 7]= 14 total	[30 31] = 61 total	[10 10] =20

Q 4. (c)

As per NOTE 2 at the start of the document, I have set gamma =1, coef =1 :

Since the Scaling is not mentioned , I will answer this question from the modelling with data with no scaling .

i. **Train Error at C=0.0001**

Training error for C =0.0001 and degree =2 is :0.008968609865470878

Training error for C =0.0001 and degree =5 is :0.004484304932735439

Hence Error is higher at Q=2. Given **1st statement** is **FALSE**.

1st statement is FALSE

Justification – A more complex model (e.g. higher degree) will fit the training data better

(same conclusion i.e. FALSE using Min max & Standard Scaling)

ii. **When C = 0.001,**

Polynomial_Scaling_None_C_0.001_degree_2- the number of support vectors : [38 38]

Polynomial_Scaling_None_C_0.001_degree_5- the number of support vectors : [13 12]

Hence

2nd statement is TRUE

(same conclusion i.e. TRUE using Min max & Standard Scaling)

iii. **When C= 0.01, Training Error** : *Using no scaling*

Training error for C =0.01 and degree =2 is :0.004484304932735439

Training error for C =0.01 and degree =5 is :0.0038436899423446302

Using STANDARD scaling -

Training error for C =0.01 and degree =2 is :0.006406149903907754

Training error for C =0.01 and degree =5 is :0.005124919923126248;

Using Min Max Scaling –

Training error for C =0.01 and degree =2 is :0.007046764894298563

Training error for C =0.01 and degree =5 is :0.0038436899423446302

Hence , in all 3 cases, training error is lower at Q=5 i.e. **3rd statement is FALSE**

Justification – A more complex model (e.g. higher degree) will fit the training data better

3rd statement is FALSE

iv. WITHOUT Scaling –

Test error for C =1 and degree =2 is :0.018867924528301883

Test error for C =1 and degree =5 is :0.021226415094339646;

Using STANDARD SCALING –

Test error for C =1 and degree =2 is :0.021226415094339646

Test error for C =1 and degree =5 is :0.02594339622641506

Using Min Max Scaling,

Test error for C =1 and degree =2 is :0.02358490566037741

Test error for C =1 and degree =5 is :0.02358490566037741 (i.e. they both are same for Min Max Scaling case)

Hence the test error is higher for Q=5, **4th statement is FALSE** (likely due to overfitting)

4th statement is FALSE

NOTE – the 4.d. answer changes to True **if we do not set** gamma =1, coef =1.

Q 4.(d).

All the following results are from using gamma = 1

WITHOUT SCALING –

Training error for C =0.001 is :0.356181934657271; Test error: 0.37735849056603776

Training error for C =1 is :0.004484304932735439; Test error: 0.021226415094339646

Training error for C =100 is :0.0032030749519538215; Test error: 0.018867924528301883

Training error for C =10000 is :0.002562459961563124; Test error: 0.02358490566037741

Training error for C =1000000 is :0.0006406149903908087; Test error: 0.02358490566037741

Lowest Training Error – for C = 10^6 – WITHOUT ANY SCALING

Lowest Test Error – for C = 100 (Same for other Scaling)

USING STANDARD SCALING

Training error for C =0.001 is :0.356181934657271; Test error: 0.37735849056603776

Training error for C =1 is :0.004484304932735439; Test error: 0.021226415094339646

Training error for C =100 is :0.0032030749519538215; Test error: 0.018867924528301883

Training error for C =10000 is :0.002562459961563124; Test error: 0.02358490566037741

Training error for C =1000000 is :0.0006406149903908087; Test error: 0.02358490566037741

Lowest Training Error – for C = 10^6 – WITH STANDARD SCALING

Lowest Test Error – for C = 100

USING MIN MAX SCALING –

Training error for C =0.001 is :0.356181934657271; Test error: 0.37735849056603776

Training error for C =1 is :0.004484304932735439; Test error: 0.021226415094339646;

Training error for C =100 is :0.0032030749519538215; Test error: 0.018867924528301883

Training error for C =10000 is :0.002562459961563124; Test error: 0.02358490566037741

Training error for C =1000000 is :0.0006406149903908087; Test error: 0.02358490566037741

Lowest Training Error – for C = 10^6 – WITH MINMAX SCALING

Lowest Test Error – for C = 100