

Chapter 6

Recognition of Online Handwritten Assamese Characters: Using Qualitative Feature Vectors

This chapter presents a discussion on Online Handwritten Assamese Characters recognition using qualitative features. Qualitative features are computed from the qualitative representation of online handwritten Assamese characters. The qualitative representation of online handwritten Assamese characters is discussed in Chapter 5. Two qualitative features vectors namely, QFS: 1 and QFS: 2 are formed. Character recognition experiments are performed for online handwritten Assamese characters using these two feature vectors and the corresponding recognition results are reported in this chapter.

6.1 Feature Set

The qualitative feature vectors are composed of Qualitative Orientations and Qualitative Curvature Component Signatures. Qualitative Orientation is derived from the orientation angle of strokes and Qualitative Curvature Component Signature is derived from qualitative curvature components. Each online handwritten Assamese character is represented as a sequence of Qualitative symbols. These qualitative symbols are derived from the orientation angle of strokes. The qualitative curvature components are the shape primitives with which the outline of a character is approximated.

6.1.1 Feature Computation

Preprocessed strokes are inputs to feature computation modules. Schemes for feature computation are discussed below.

6.1.1.1 Qualitative Encoding

The representation of an online handwritten Assamese character in terms of the sequence or chain of qualitative symbols is termed as qualitative encoding. A series of qualitative symbols is obtained for each stroke; all such qualitatively encoded strokes of a character make the qualitative encoding of a character. The algorithm for finding the series of orientation symbol for a stroke is presented in Algorithm 6.1.

Algorithm 6.1

Input: Stroke S

Output: Qualitative Encoding of Stroke S

N: Number of points in S

Angle: Array which stores the angle computed at the
middle point of three consecutive points in S

QArray: Array of length N-2 which stores the Qualitative
symbols

i: Integer variable

1. $i := 1$
2. Perform the following steps from 3 to 15 till $i \leq N-1$
3. If $(\text{Angle}(i) == 0)$ Then $\text{QArray}(i) = 'P'$
4. Else If $(\text{Angle}(i) > 0 \ \&\& \ \text{Angle}(i) \leq 45)$ Then
 $\text{QArray}(i) = 'Q'$
5. Else If $(\text{Angle}(i) > 45 \ \&\& \ \text{Angle}(i) < 90)$ Then
 $\text{QArray}(i) = 'R'$
6. Else If $(\text{Angle}(i) == 90)$ $\text{QArray}(i) = 'S'$
7. Else If $(\text{Angle}(i) > 90 \ \&\& \ \text{Angle}(i) \leq 135)$ $\text{QArray}(i) = 'T'$
8. Else If $(\text{Angle}(i) > 135 \ \&\& \ \text{Angle}(i) < 180)$ $\text{QArray}(i) = 'U'$
9. Else If $(\text{Angle}(i) == 180)$ $\text{QArray}(i) = 'G'$

```

10. Else If (Angle(i)>180 && Angle(i)<=225)QArray(i)='V'
11. Else If (Angle(i)>225 && Angle(i)<270) QArray(i)='W'
12. Else If (Angle(i)==270)QArray(i)='X'
13. Else If (Angle(i)>270 && Angle(i)<=315)QArray(i)='Y'
14. Else If (Angle(i)>315 && Angle(i)<360)QArray(i)='Z'
15. End

```

6.1.1.2 Qualitative Curvature Component Signature

The representation of an online handwritten Assamese character in terms of the sequence of Qualitative Curvature Component (QCC) is termed as qualitative curvature component signature for the character. A series of qualitative curvature components is obtained for each stroke. The algorithm for finding the QCC Signature of an Assamese character is presented in Algorithm 6.2.

Algorithm 6.2

Input: QE - Qualitative Encoding of a Character

Output: QCC Signature

N: Number of symbols in QE

1. Initialize *QCC Signature String* to empty string.
2. Create sub-strings of symbols which are neighbours in the Orientation Model for each stroke of QE;
3. Call each such sub-string a *strokelet*.
4. Store all the *strokelets* in the sequence they appear.
5. For each *strokelet* S_i
 - Look-up *strokelet* to QCC transformation.
 - If transformation exists
 - Add QCC symbol to *QCC Signature String*
 - Else
 - Combine S_i with next *strokelet* S_{i+1}
 - Call that S_i.
 - Go to Look-up
6. Repeat until all *strokelets* are done.

6.1.2 Qualitative Feature Vectors

The qualitative feature vector QFS:1 is formed with only Qualitative Orientations. The qualitative feature vector QFS:2 consists of both Qualitative Orientations and Qualitative Curvature Component Signatures. The feature vectors are computed for the numerals, basic alphabetic characters and Juktakhors taken from the dataset of Online Handwritten Assamese Characters.

6.1.2.1 Qualitative feature vector QFS:1

In feature vector QFS:1, the attributes are the orientation symbols which are derived from the orientation angle of strokes. The orientation symbols are derived from the resampled online handwritten Assamese characters using the algorithm 6.1. Here, all of the characters have been resampled to a fixed number of 100 points. Three consecutive points are considered for the computation of the direction or orientation angle. The direction angle is the one which is formed between two straight lines joining three consecutive points and the direction angle is measured at the middle point of these three consecutive points. Therefore, we have derived 98 direction angles for 100 resampled points in each character in the dataset. For each direction angle belonging to a certain range, we assign one unique symbol to it using the Algorithm 6.1; feature vector is 98 attributes where each attribute represents an orientation symbol.

6.1.2.2 Qualitative feature vector QFS:2

Each online handwritten character can be represented by a series of qualitative curvature components as described in chapter 5. To form the feature vector QFS:2 an online handwritten Assamese character is represented as a sequence of qualitative curvature components arranged as encountered through strokes. The feature vector QFS:2 is an extension of the feature vector QFS:1. Feature vector QFS:2 is formed with qualitative orientation symbols and qualitative curvature components. In feature vector QFS:2 the first 98 attributes are the qualitative symbols representing qualitative orientation of strokes followed by the sequence

of qualitative curvature components representing the online handwritten Assamese character.

6.2 *Classification*

Here we present the classification results of online handwritten Assamese characters, which include numerals, basic alphabetic characters and conjunct consonants (Juktakkhors). Character recognition experiments were performed on the 10 numeral classes (total of $45 \times 10 = 450$ numerals), 52 classes of basic alphabetic characters (total of $45 \times 52 = 2340$ basic alphabetic characters) and 121 classes of conjunct consonants (total of $45 \times 121 = 5445$ conjunct consonants) available in the dataset. In order to conduct the classification, we used Support Vector Machine (SVM).

6.2.1 *SVM based Classifier*

Experiments on Classification using linear kernel, polynomial kernel and Gaussian radial basis functions, which are the common choices for kernel functions in SVM, are performed for the classification of online handwritten Assamese numerals, basic alphabetic characters and Juktakhors. A 10-fold cross validation procedure was used.

6.2.2 *Classification Results*

We present the classification results of online handwritten Assamese numerals, basic alphabetic characters and Juktakkhors. Classification results are based on the feature vectors QFS:1 and QFS:2. Table 6.1 shows the classification results of online handwritten Assamese numerals for the feature sets QFS:1 and QFS:2. Table 6.2 shows the classification results of online handwritten Assamese basic alphabetic characters for the feature sets QFS:1 and QFS:2. Similarly, Table 6.3 shows the classification results of online handwritten Assamese conjunct consonants for the feature sets QFS:1 and QFS:2. Additionally, we present the

overall recognition results in Table 6.4 considering all the 8235 online handwritten Assamese characters (consisting of numerals, basic alphabetic characters and Juktakkhors) based on feature set QFS:2.

6.2.2.1 Classification Results: Numerals

The classification results for online handwritten Assamese numerals are presented in the Table 6.1.

Table 6.1. Average recognition rates of numerals (using qualitative feature vectors)

<i>Online handwritten Assamese numerals</i>					
Type of SVM kernel	Total Number of Instances	Correctly Classified Instances	Average recognition rate (Across Classes)	Standard Deviation (SD) of recognition rate (Across Classes)	Average \pm SD
Feature Set QFS:1					
Linear	450	383	85.11	8.70	85.11 \pm 8.70
Polynomial (C=1,E=4)	450	403	89.56	5.67	89.56 \pm 5.67
RBF (C=1,Gamma=0.01)	450	389	86.44	6.76	86.44 \pm 6.76
Feature Set QFS:2					
Linear	450	430	95.56	2.36	95.56 \pm 2.36
Polynomial (C=1, E=4)	450	420	93.33	4.56	93.33 \pm 4.56
RBF (C=1,Gamma=0.01)	450	418	92.89	5.02	92.89 \pm 5.02

Linear Kernel

The average recognition rates achieved for the online handwritten Assamese numerals using linear kernel were 85.11% (based on QFS:1) and 95.56% (based on QFS:2) with a 10 fold cross validation process. A total of 450 characters were used as samples in the numeral recognition experiment (refer to Table 6.1). The kernel parameter settings C = 1 and E = 1 are default associated with the linear kernel.

Polynomial Kernel

The polynomial kernel parameter setting ($C = 1$ and $E = 4$) for the feature sets QFS:1 and ($C = 1$ and $E = 4$) for feature set QFS:2 were obtained by grid search operation. The average recognition rates achieved for the online handwritten Assamese numerals using polynomial kernel were 89.56% (based on QFS:1) and 93.33% (based on QFS:2) with a 10 fold cross validation process. A total of 450 characters were used as samples in the numeral recognition experiment (refer to Table 6.1).

RBF Kernel

- *Parameter Settings: Grid Search*

The kernel parameter settings ($C=1$, $\gamma=0.01$) for the feature set QFS:1 and ($C=1$, $\gamma=0.01$) for the feature set QFS:2 were obtained by grid search operation. The plots of grid search are shown in the Figure 6.1 and Figure 6.2.

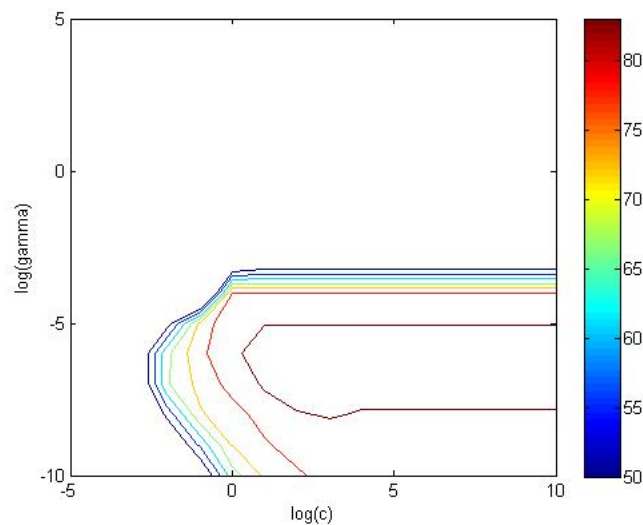


Figure 6.1 Grid search plot of C and gamma for RBF kernel (using QFS:1 for numerals)

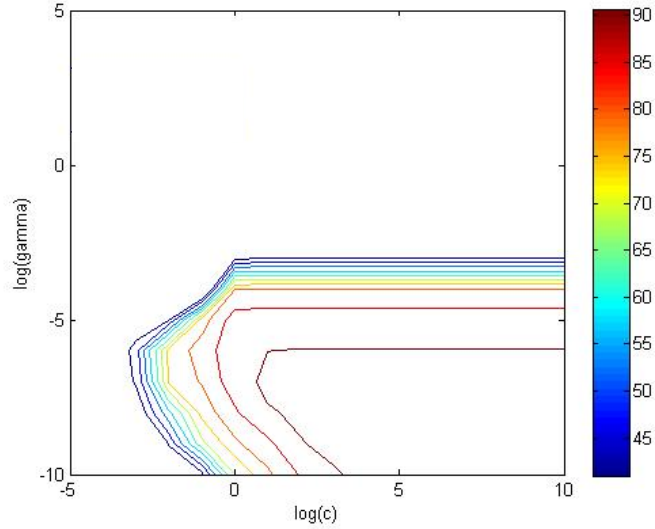


Figure 6.2 Grid search plot of C and gamma for RBF kernel (using QFS:2 for numerals)

- *Experimental Results*

The average recognition rates achieved for the online handwritten Assamese numerals using RBF kernel were 86.44% (based on QFS:1) and 92.89% (based on QFS:2) with a 10 fold cross validation process. A total of 450 characters were used as samples in the numeral recognition experiment (refer to Table 6.1).

6.2.2.2 Classification Results: Basic Alphabetic Characters

The classification results for online handwritten Assamese *basic alphabetic characters* are presented in the Table 6.2.

Table 6.2. Average recognition rates of basic alphabetic characters (using qualitative feature vectors)

<i>Online handwritten Assamese basic alphabetic characters</i>					
Type of SVM kernel	Total Number of Instances	Correctly Classified Instances	Average recognition rate (Across Classes)	Standard Deviation (SD) of recognition rate (Across Classes)	Average \pm SD
Feature Set QFS:1					
Linear	2340	1250	53.42	21.71	53.42 \pm 21.71
Polynomial (C=1,E=4)	2340	1511	64.57	17.32	64.57 \pm 17.32
RBF (C=2,Gamma=0.03)	2340	1450	61.97	18.50	61.97 \pm 18.50
Feature Set QFS:2					
Linear	2340	2223	95.00	3.64	95.00 \pm 3.64
Polynomial (C=1, E=4)	2340	1987	84.92	8.76	84.92 \pm 8.76
RBF (C=1,Gamma=0.01)	2340	2067	88.33	9.19	88.33 \pm 9.19

Linear Kernel

The average recognition rates achieved for the online handwritten Assamese basic alphabetic characters using linear kernel were 53.42% (based on QFS:1) and 95.00% (based on QFS:2) with a 10 fold cross validation process. A total of 2340 characters were used as samples in the basic alphabetic characters recognition experiment (refer to Table 6.2). The kernel parameter settings $C = 1$ and $E = 1$ are default associated with the linear kernel.

Polynomial Kernel

The polynomial kernel parameter setting ($C = 1$ and $E = 4$) for the feature sets QFS:1 and ($C = 1$ and $E = 4$) for feature set QFS:2 were obtained by grid search operation. The average recognition rates achieved for the online handwritten Assamese basic alphabetic characters using polynomial kernel were 64.57% (based on QFS:1) and 84.92% (based on QFS:2) with a 10 fold cross validation

process. A total of 2340 characters were used as samples in the basic alphabetic characters recognition experiment (refer to Table 6.2).

RBF Kernel

- *Parameter Settings: Grid Search*

The kernel parameter settings ($C=2$, $\gamma=0.03$) for the feature set QFS:1 and ($C=1$, $\gamma=0.01$) for the feature set QFS:2 were obtained by grid search operation. The plots of grid search are shown in the Figure 6.3 and Figure 6.4.

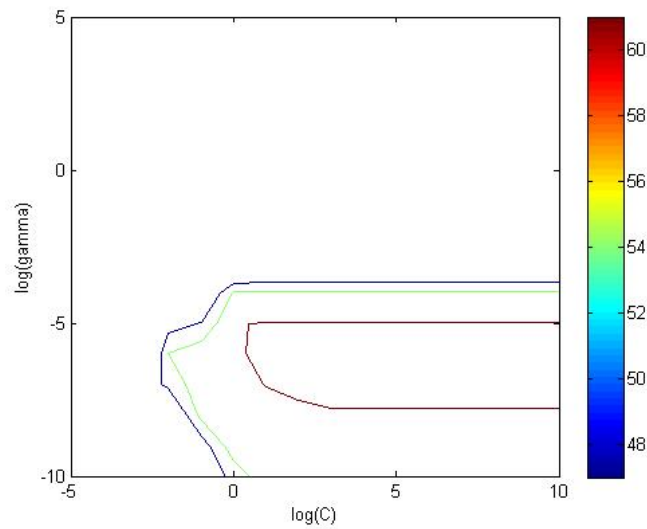


Figure 6.3 Grid search plot of C and γ for RBF kernel (using QFS:1 for basic alphabetic characters)

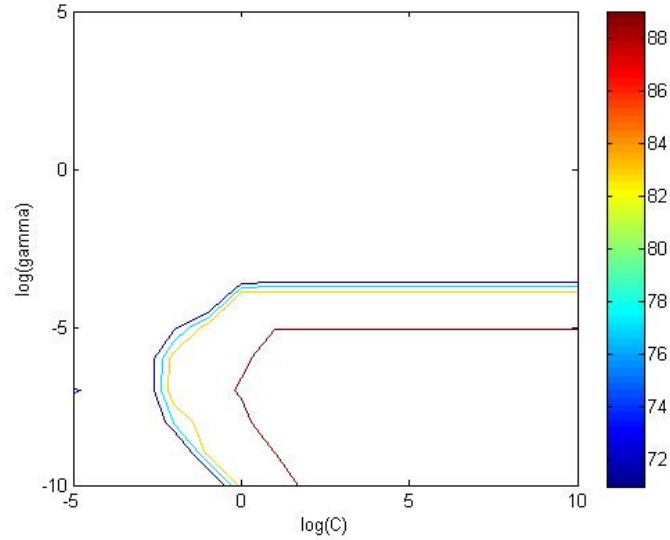


Figure 6.4 Grid search plot of C and gamma for RBF kernel (using QFS:2 for basic alphabetic characters)

- *Experimental Results*

The average recognition rates achieved for the online handwritten Assamese basic alphabetic characters using RBF kernel were 61.97% (based on QFS:1) and 88.33% (based on QFS:2) with a 10 fold cross validation process. A total of 2340 characters were used as samples in the Assamese basic alphabetic characters recognition experiment (refer to Table 6.2).

6.2.2.4 Classification Results: Conjunct Consonants (*Juktakkhors*)

The classification results for online handwritten Assamese *Juktakkhors* are presented in the Table 6.3.

Table 6.3. Average recognition rates of Juktakkhors (using qualitative feature vectors)

<i>Online handwritten Assamese Conjoint Characters (Juktakkhors)</i>					
Type of SVM kernel	Total Number of Instances	Correctly Classified Instances	Average recognition rate (Across Classes)	Standard Deviation (SD) of recognition rate (Across Classes)	Average \pm SD
Feature Set QFS:1					
Linear	5445	1882	34.57	19.38	34.57 \pm 19.38
Polynomial (C=1,E=4)	5445	2615	48.03	19.13	48.03 \pm 19.13
RBF (C=2,Gamma=0.03)	5445	2185	40.13	19.17	40.13 \pm 19.17
Feature Set QFS:2					
Linear	5445	5180	95.13	6.48	95.13 \pm 6.48
Polynomial (C=1,E=4)	5445	4538	83.34	10.25	83.34 \pm 10.25
RBF (C=1,Gamma=0.01)	5445	4838	88.85	9.54	88.85 \pm 9.54

Linear Kernel

The average recognition rates achieved for the online handwritten Assamese Juktakkhors using linear kernel were 34.57% (based on QFS:1) and 95.13% (based on QFS:2) with a 10 fold cross validation process. A total of 5445 characters were used as samples in the basic alphabetic characters recognition experiment (refer to Table 6.3). The kernel parameter settings $C = 1$ and $E = 1$ are default associated with the linear kernel.

Polynomial Kernel

The polynomial kernel parameter setting ($C = 1$ and $E = 4$) for the feature sets QFS:1 and ($C = 1$ and $E = 4$) for feature set QFS:2 were obtained by grid search operation. The average recognition rates achieved for the online handwritten Assamese Juktakkhors using polynomial kernel were 48.03% (based on QFS:1) and 83.34% (based on QFS:2) with a 10 fold cross validation process. A total of

5445 characters were used as samples in the Assamese Juktakkhors recognition experiment (refer to Table 6.3).

RBF Kernel

- *Parameter Settings: Grid Search*

The kernel parameter settings ($C=2$, $\gamma=0.03$) for the feature set QFS:1 and ($C=1$, $\gamma=0.01$) for the feature set QFS:2 were obtained by grid search operation. The plots of grid search are shown in the Figure 6.5 and Figure 6.6.

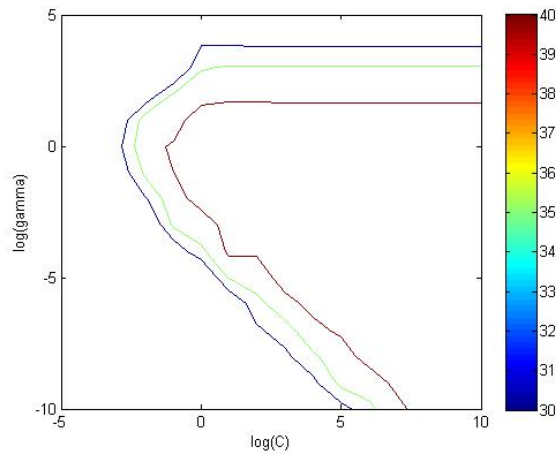


Figure 6.5 Grid search plot of C and gamma for RBF kernel (using QFS:1 for Juktakkhors)

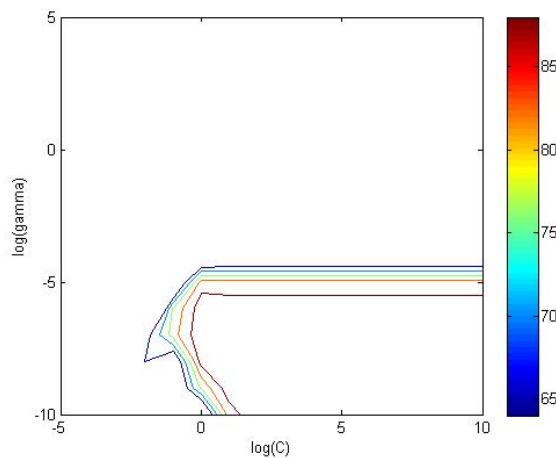


Figure 6.6 Grid search plot of C and gamma for RBF kernel (using QFS:2 for Juktakkhors)

- *Experimental Results*

The average recognition rates achieved for the online handwritten Assamese Juktakkhors using RBF kernel were 40.13% (based on QFS:1) and 88.85% (based on QFS:2) with a 10 fold cross validation process. A total of 5445 characters were used as samples in the Assamese Juktakkhors recognition experiment (refer to Table 6.3).

6.2.3.3 Overall Recognition Rate

In the previous sections we presented the individual recognition rates of online handwritten Assamese numerals, basic alphabetic characters and Juktakkhors for three SVM kernels using qualitative feature vectors QFS:1 and QFS:2. Here we present the overall recognition rates of all the 183 characters (a total of 8235 samples which include numerals, basic alphabetic characters and Juktakhors) using the feature vector QFS:2 for linear, polynomial and RBF kernels of SVM classifier. Table 6.4 presents the overall recognition rates of online handwritten Assamese characters for the combined set of numerals, basic alphabetic characters and Juktakkhors.

Table 6.4. Overall recognition rates of online handwritten Assamese combined set of numerals, basic alphabetic characters and Juktakkhors (using qualitative feature vectors)

<i>Online handwritten Assamese characters (numerals, basic alphabetic characters & Juktakkhors)</i>					
Type of SVM kernel	Total Number of Instances	Correctly Classified Instances	Average recognition rate (Across Classes)	Standard Deviation (SD) of recognition rate (Across Classes)	Average \pm SD
Feature Set QFS:2					
Linear	8235	7735	93.93	6.40	93.93 \pm 6.40
Polynomial (C=1, E=3)	8235	7163	86.98	8.62	86.98 \pm 8.62
RBF (C=1, Gamma=0.01)	8235	7106	86.29	11.10	86.29 \pm 11.10

Linear Kernel

The overall recognition rate achieved for the online handwritten Assamese combined set of character using linear kernel is 93.93% (based on QFS:2), with a 10 fold cross validation process. All the 8235 characters of the dataset were used as samples in the character recognition experiment (refer to Table 6.4). The kernel parameter settings $C = 1$ and $E = 1$ are default associated with the linear kernel.

Polynomial Kernel

The polynomial kernel parameter setting $C = 1$ and $E = 3$ for the feature set QFS:2 was obtained by grid search operation. The overall recognition rate achieved for the online handwritten Assamese combined set of characters using polynomial kernel is 86.98% (based on QFS:2) with a 10 fold cross validation process. All the 8235 characters of the dataset were used as samples in the character recognition experiment (refer to Table 6.4).

RBF Kernel

- *Parameter Settings: Grid Search*

The kernel parameter settings $C=1$ and $\gamma=0.01$ for the feature set QFS:2 was obtained by grid search operation. The plot of grid search operation is shown in the Figure 6.7.

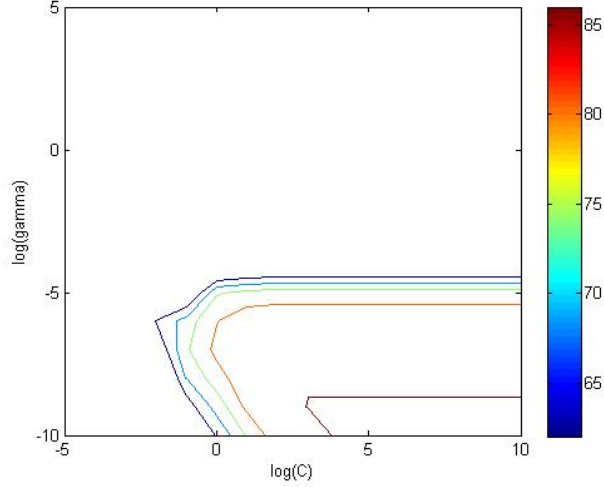


Figure 6.7 Grid search plot of C and gamma for RBF kernel (using QFS:2 for the combined set of characters)

- *Experimental Results*

The overall recognition rate achieved for the online handwritten Assamese combined set of characters using RBF kernel was 86.29% (based on QFS:2) with a 10 fold cross validation process. All the 8235 characters of the dataset were used as samples in the character recognition experiment (refer to Table 6.4).

6.3 Conclusion

Using a *Qualitative Orientation Model* qualitative direction of strokes for a character is proposed. Further, a qualitative analysis of 8,235 characters is done based on *curvature components*; and a qualitative curvature component signature for online handwritten Assamese character is evolved. Two different combinations of features based on above qualitative representation are used for classification. The best recognition result obtained on the entire dataset of 8235 characters for the feature set FS:3 is 73.21% corresponding to polynomial kernel. Similarly, the best recognition result obtained on the entire dataset of 8235 characters for the feature set QFS:2 is 93.93% . This result is obtained corresponding to linear kernel. Based on this result, we can conclude that for the

recognition of online handwritten Assamese characters the qualitative features are more effective than quantitative features.