Support Vector Machine (SVM) Results

I combined the training and cross validation data subsets into one and used that for training. I tested with the subset designated for testing. I used as kernels :

linear	$x_1^T x_2$
gaussian	$exp\left(-\frac{\|x_1-x_2\|^2}{2\sigma^2}\right)$
polynomial	$(\gamma x_1 * x_2 - r)^d$

I did very limited testing with the linear kernel as the results where really dissapointing. I spent most of time experimenting with the gaussian kernel and with the polynomial. Here are tables with the results,

Table 1: Results for the "gaussian" kernel function

Result	C	sigma
72.00	10	1.0
76.13	30	3.0
72.40	50	5.0
77.20	30	5.0
75.33	30	7.0
78.00	30	4.0
76.40	30	2.0
76.26	30	3.5
78.40	30	4.5
79.06	30	4.75
75.46	30	4.85
78.66	16	4.75
76.13	20	4.75
77.46	24	4.75
77.33	27	4.75
77.60	29	4.75
77.33	32	4.75
79.87	**2**	**4.75**
79.07	2	4.75
78.93	0.5	4.75
78.53	4	4.75
78.13	1	4.75
77.73	4	4.75
79.87	**8**	**4.75**
80.00	**9**	**4.75**
78.40	9	4.85
76.53	9	4.80
77.87	9	4.70

Table 2: Results for the "polynomial" kernel function

Result	C	γ	r	d
58.00	9.00	1.00	16.00	3.00
56.00	9.00	1.00	1.00	3.00
52.80	9.00	1.00	1.00	4.00
62.00	**9.00**	**1.00**	**1.00**	**2.00**
50.00	32.00	1.00	1.00	2.00
55.07	4.00	1.00	1.00	2.00
60.00	4.00	1.00	2.00	2.00
66.13	**4.00**	**1.00**	**4.00**	**2.00**
61.33	4.00	1.00	8.00	2.00
49.07	4.00	1.00	6.00	2.00
52.67	4.00	0.03	8.00	2.00
54.67	4.00	0.06	4.00	2.00
61.87	**4.00**	**0.12**	**4.00**	**2.00**
54.13	4.00	1.50	4.00	2.00

From the results above it is obvious that gaussian kernel hands down provides for the best accuracy. From the experimentation it is also apparent that σ is the factor that affects the learning the most in contrast with C. I think that it also points to the problem of achieving a better result. From the following figure, provided by Graham, we can see that two of the three clusters are well defined, whereas the one closest to the origin is somewhat problematic. Actually I do not know if it gets treated as two clusters, one well defined at the top and another more dispersed at the bottom, or just one. Also notice how the orientation for the rightmost two is northwest to southwest versus the other one(s) being east to west. These should provide some clues about improving the accuracy but *I have no clue* how to go about achieving that. Of course we should always be cautious about overfitting the existing training data.

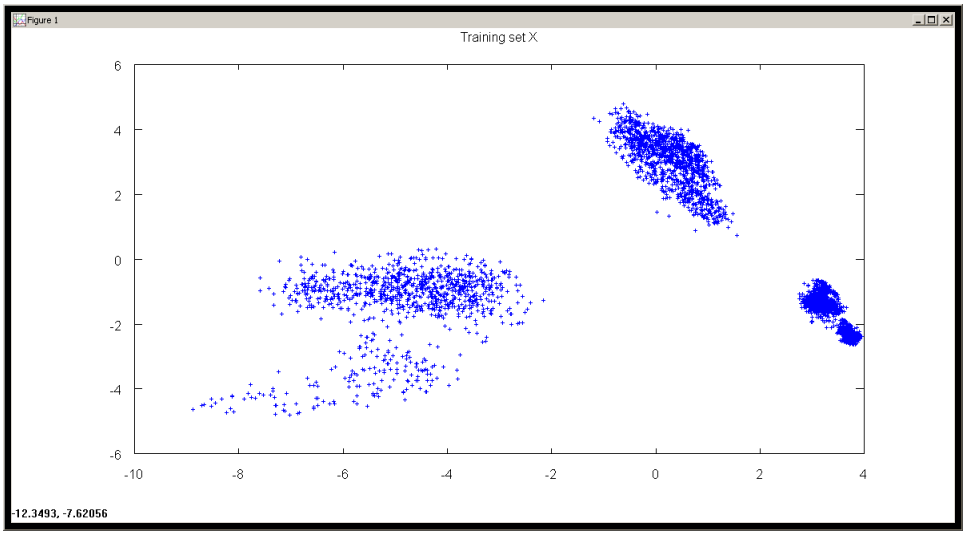

Figure 1: Plot of training data set

Caveat regarding "polynomial" kernel

I have to admit that I did not do nowhere near to exhaustive experimentation with all four (4) parameters involved in the "polynomial" kernel C, γ, r and d. So I have to allow for the possibility that better results could be achieved by another untested combination of these parameters!

Usefull links

A Practical Guide to Support Vector Classification

SVM Software