1)
The SVM learner has been implemented exactly as described in the specification. Stochastic sub-gradient descent iterates over all the examples updating weight vector at each step. Gradient is conditionally calculated as per the equations, as is the rate as well as the weight update conditions.

2)
The data sets are transformed to a set consisting of all products of features. i.e. the transformed set consists of only degree 2 points, where each point is a multiplication with each of the other features in the set.

The distance calculated from the origin is the Euclidean distance, that is, root of summation of squares of each of the features.

Max distance astro original 619.213510713
Max distance astro scaled 1.91259469966
Max distance astro original transformed 362017.38657
Max distance astro scaled transformed 2.89316599301

3)
For the 3rd experiment, the ρ0 ∈ {0.001, 0.01, 0.1, 1} and C ∈ { 1, 10, 20, 30}. I removed few of the smaller values, as the code was running too long then, moreover experiments showed that larger values of C were favored. For the margin, I am reporting the correctly classified point closest to the margin, as well as the incorrectly classified point farthest from the margin(as I believe it gives insight into the data).

Data Set	ρ_o	C	Avg Accuracy
Data 0	0.001	30	1
Astro original	0.001	30	0.827
Astro scaled	0.001	30	0.795
Astro original transformed	1	1	0.852
Astro scaled transformed	1	30	0.805

I believe that, when the question says to show the best 5 parameters, it means for one parameter for each of the training datas. But in case you wish to see 5 best parameters per data set, the last page of the report consists of the entire output of the script.

DataSet	Accuracy on Test	Margin (smallest positive)	Most badly classified negative

Data 0	1	0.022	-
Astro original	0.779	0.0129	-22.057
Astro scaled	0.773	9.2355	-0.364
Astro original transformed	0.795	0.233	-2597.807
Astro scaled transformed	0.826	6.032	-0.365

Output :
-----Data0 data set----------
Finding best hyper parameters for SVM from given list...
sampling dataSpace Accuracy for best Hyper param (0.001, 30) : 1.0
top 5 hyper params
((0.001, 30), 1.0)
((0.001, 20), 0.999)
((0.1, 20), 0.983)
((1, 30), 0.982)
((0.01, 30), 0.9810000000000001)
accuracy on test set 1.0
lowest positive margin 0.0229361436631
most misclassified point 0
-----Data0 data set----------
-----astro original data set ----------
Finding best hyper parameters for SVM from given list...
sampling dataSpace Accuracy for best Hyper param (0.001, 30) : 0.827440843946
top 5 hyper params
((0.001, 30), 0.8274408439456984)
((0.01, 20), 0.8224498802168705)
((1, 30), 0.8020363132013619)
((0.1, 1), 0.797640062203169)
((1, 1), 0.7875583154709368)
{1: 2.044446839941048, 2: 2.318043690470369, 3: -0.42746599692692144, 4: -1.3433489537933583}
accuracy on test set 0.779
accuracy on train set 0.837164130787
lowest positive margin 0.0129016703186
most misclassified point -22.0575428397
-----astro original data set ----------

-----astro scaled data set ----------
Finding best hyper parameters for SVM from given list...
sampling dataSpace Accuracy for best Hyper param (0.001, 30) : 0.795322153575
top 5 hyper params

((0.001, 30), 0.7953221535745808)
((0.001, 20), 0.7723395536502333)
((0.01, 20), 0.7723237927121421)
((1, 30), 0.7720358929096791)
((0.1, 20), 0.7687450090362712)
{1: -0.8407862261902388, 2: 0.9630785149530378, 3: -0.5193528150824164, 4: 1.7515971994873016}
accuracy on test set 0.773
accuracy on train set 0.821625121399
lowest positive margin 9.23557439558e-05
most misclassified point -0.364188101788
-----astro scaled data set ----------

-----astro original transformed data set ----------
Finding best hyper parameters for SVM from given list...
sampling dataSpace Accuracy for best Hyper param (1, 1) : 0.85231580717
top 5 hyper params
((1, 1), 0.852315807170176)
((0.001, 1), 0.8257386626318665)
((0.01, 20), 0.8193575841634095)
((0.1, 10), 0.8193344681208761)
((1, 30), 0.8105745387298786)
{1: -0.6315697600986191, 2: 0.9442353751739168, 3: -0.14519975113039613, 4: 0.9193023485979736, 5: 0.2565091749371683, 6: -0.36328419688785507, 7: 0.010323325907163818, 8: -0.008538170731143081, 9: -0.7109159975319991, 10: -0.33478283830666267}
accuracy on test set 0.7955
accuracy on train set 0.827128520557
lowest positive margin 0.233440924149
most misclassified point -2597.80757817
-----astro original transformed data set ----------

-----astro scaled transformed data set ----------
Finding best hyper parameters for SVM from given list...
sampling dataSpace Accuracy for best Hyper param (1, 30) : 0.805061362586
top 5 hyper params
((1, 30), 0.8050613625856344)
((0.1, 30), 0.8047524481990502)
((1, 20), 0.7804553860379102)
((0.01, 20), 0.7736771319295591)
((0.001, 20), 0.7710702727693017)
{1: 1.096352770710169, 2: -0.7299607308260675, 3: 0.3252042692935899, 4: -1.3306177783925832, 5: -1.125566548990456, 6: 0.5427539018668696, 7: -0.5316422140003362, 8: -0.33164044385169333, 9: -0.4343181792630138, 10: 0.9093297311752503}
accuracy on test set 0.82625

accuracy on train set 0.838782777598
lowest positive margin 6.03249738836e-05
most misclassified point -0.365675163617
-----astro scaled transformed data set ----------