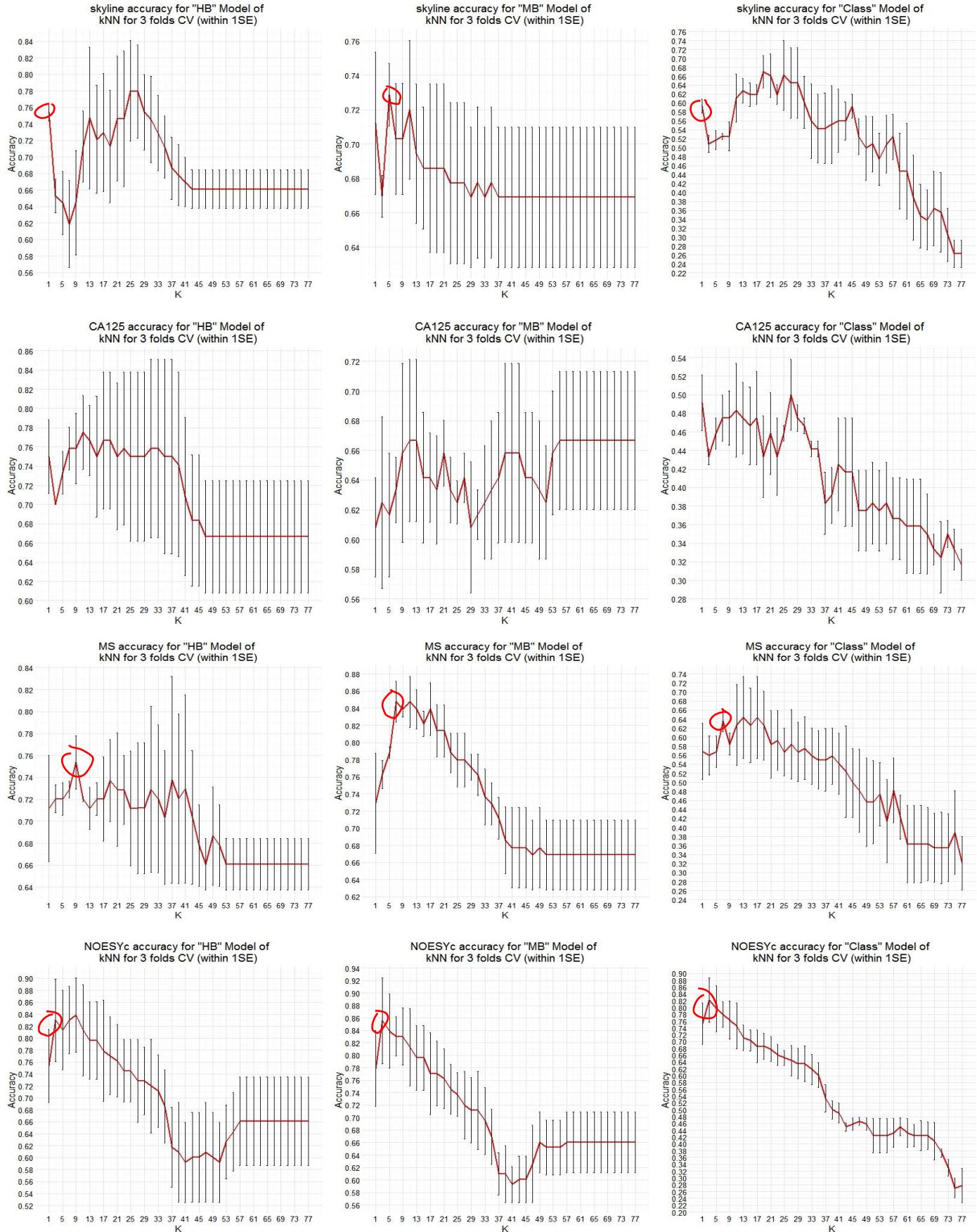


Updates on data analysis for early-stage epithelial ovarian cancer and the comparison to the previous results thereof

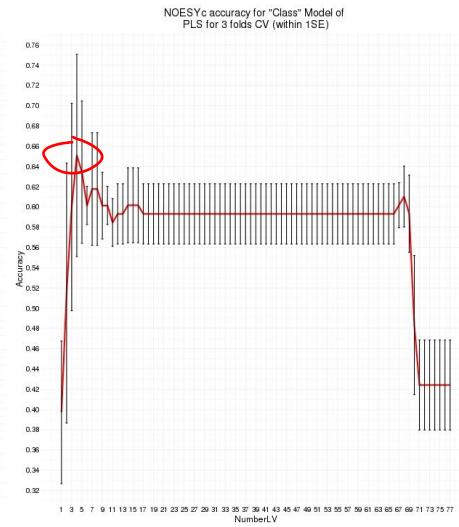
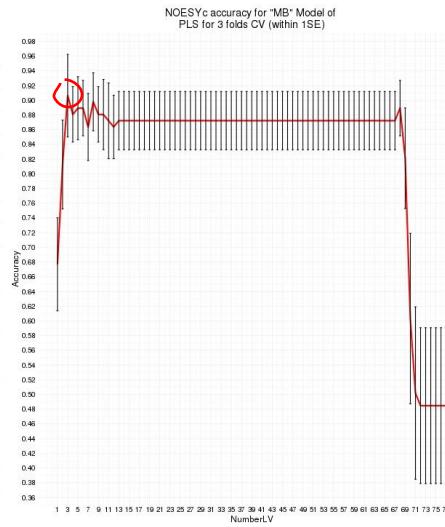
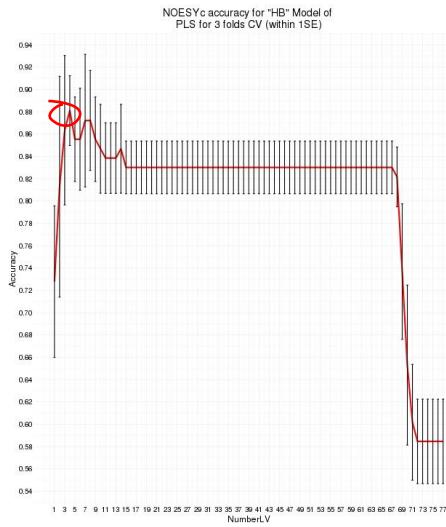
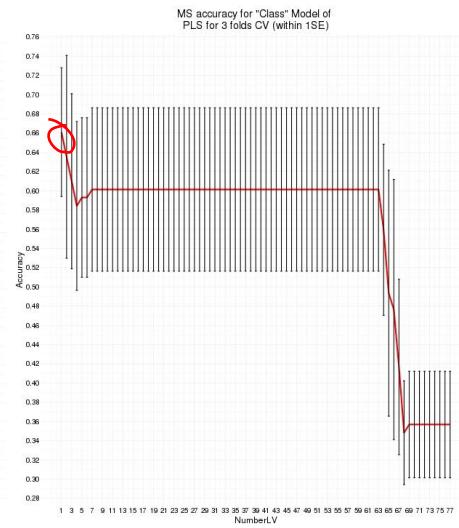
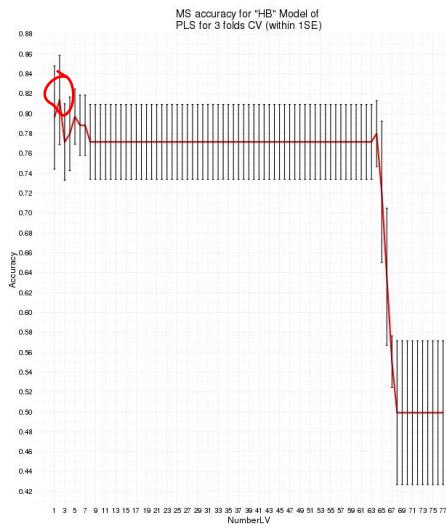
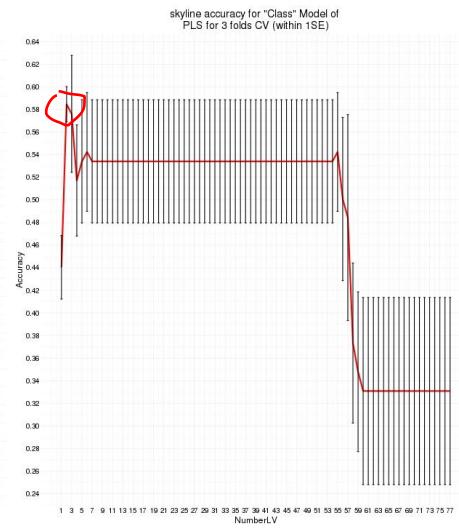
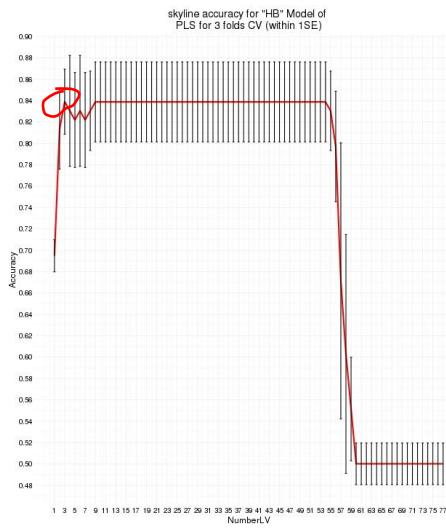
Fan Zhang

1. Model selection

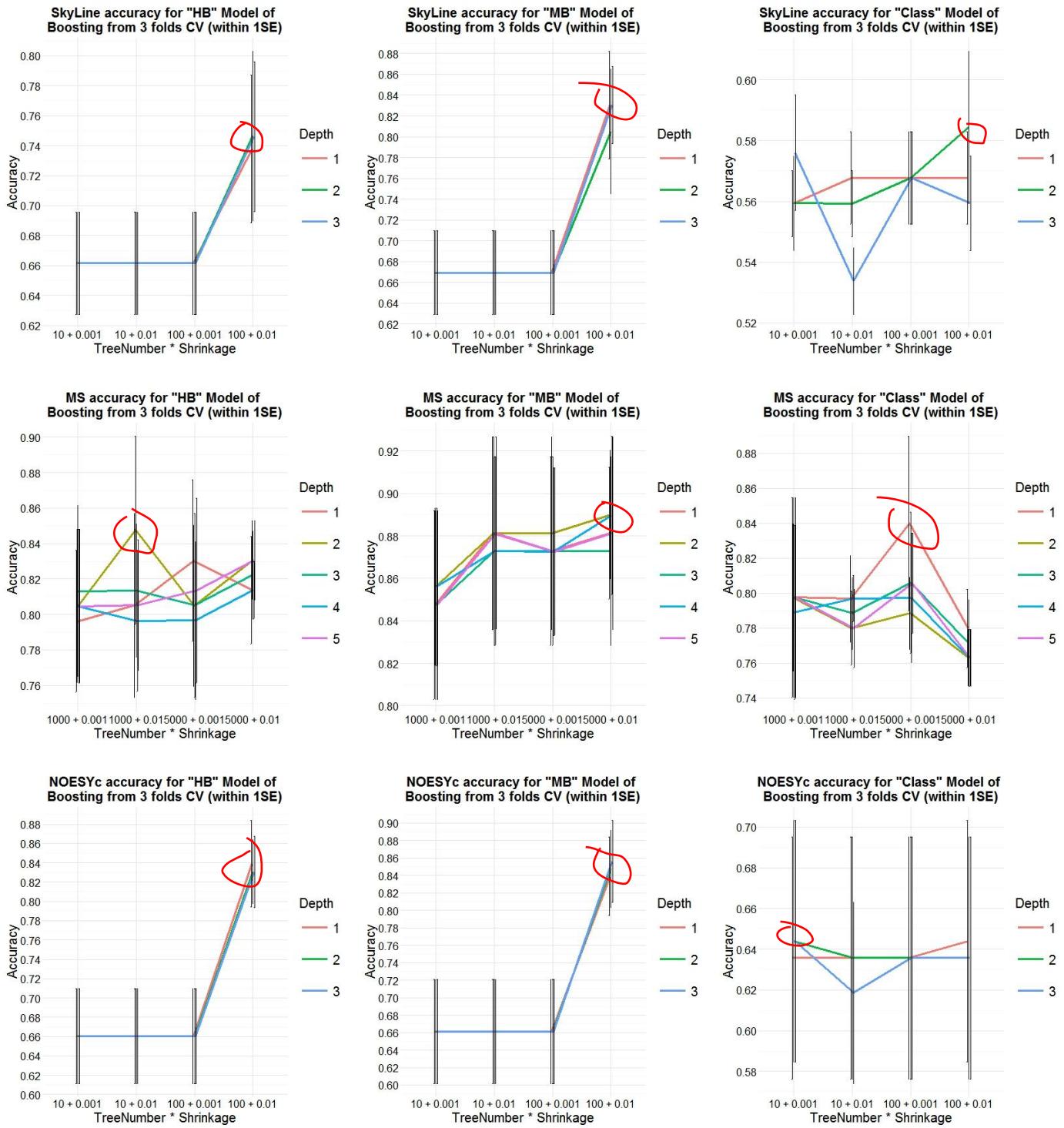
1.1 kNN



1.2 PLS



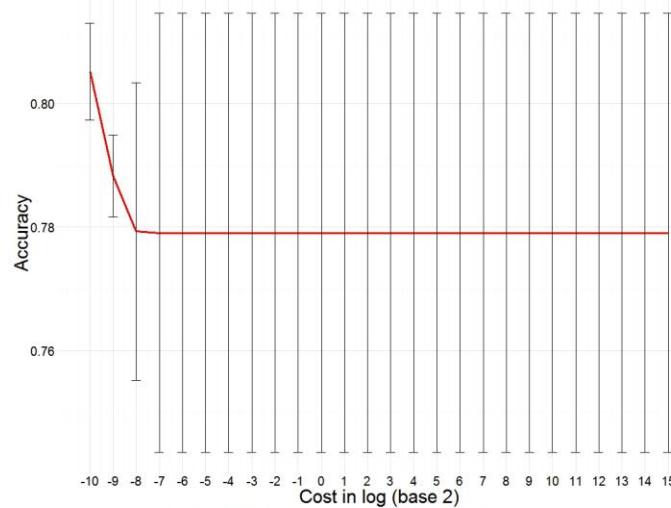
1.3 Boosting



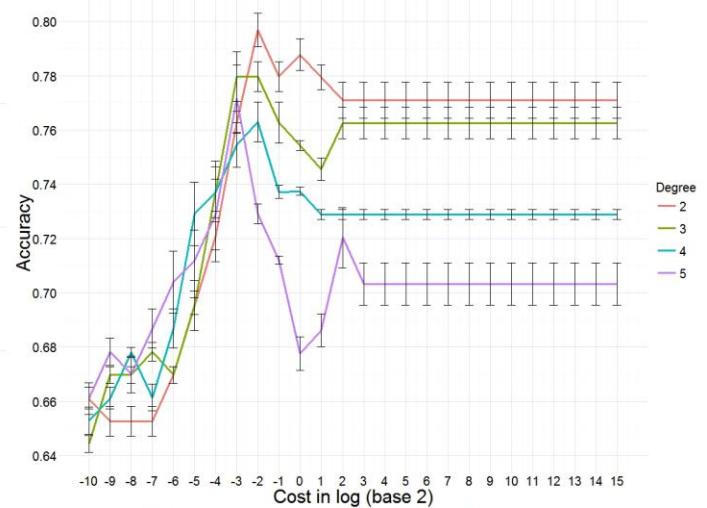
1.4 SVM

1.4.1 MS

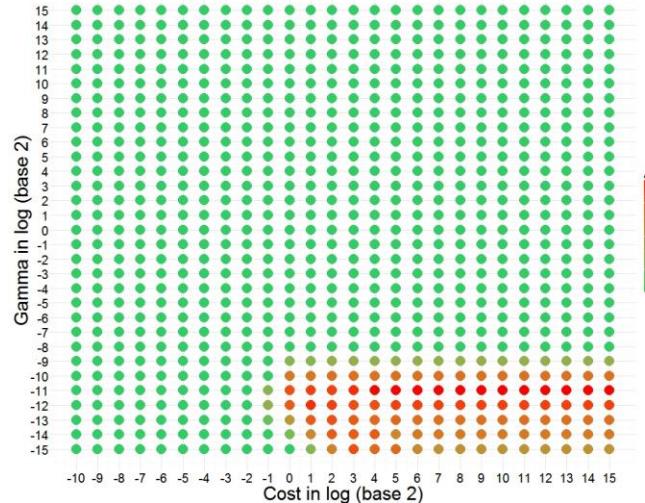
MS accuracy for "MB" Model of SVM Linear from 3 folds CV (within 1SE)



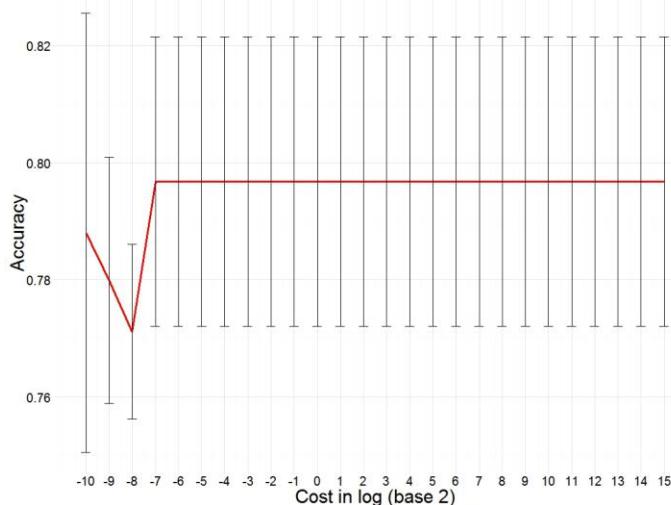
MS accuracy for "MB" Model of SVM Polynomial from 3 folds CV (within 0.25SE)



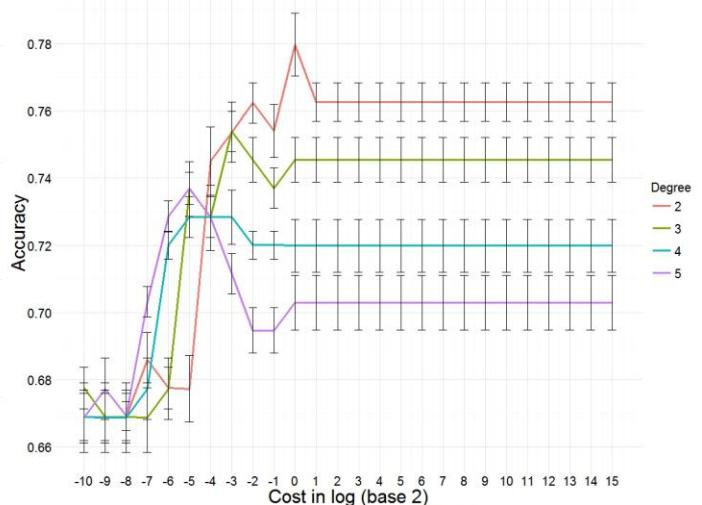
Grid of MS accuracy for "MB" Model of SVM Radial from 3 folds CV



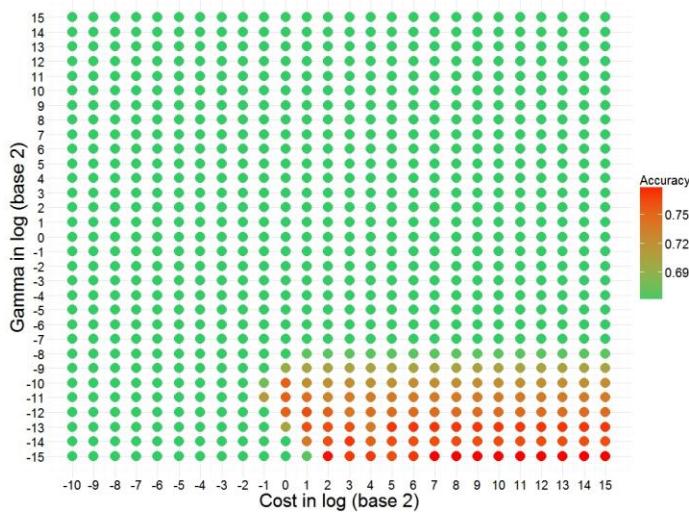
MS accuracy for "HB" Model of SVM Linear from 3 folds CV (within 1SE)



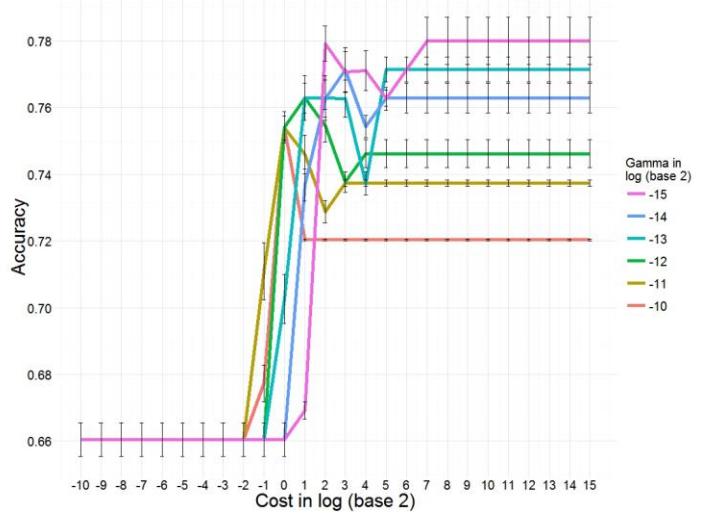
MS accuracy for "HB" Model of SVM Polynomial from 3 folds CV (within 0.25SE)



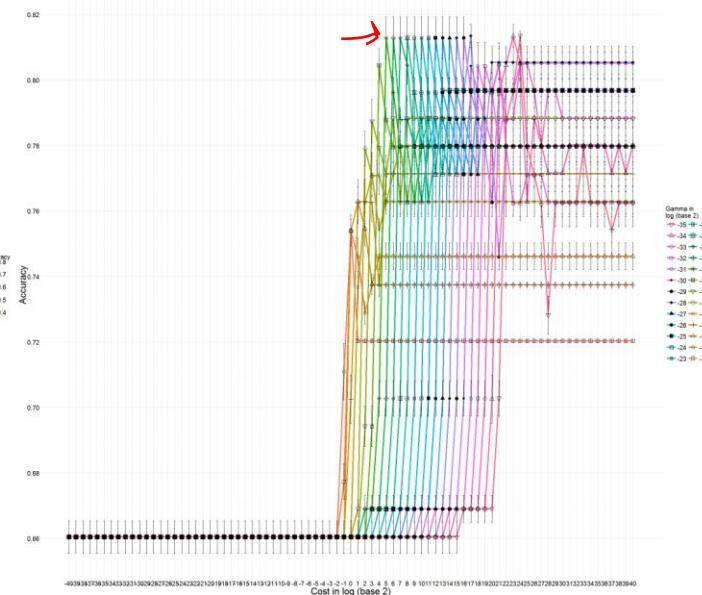
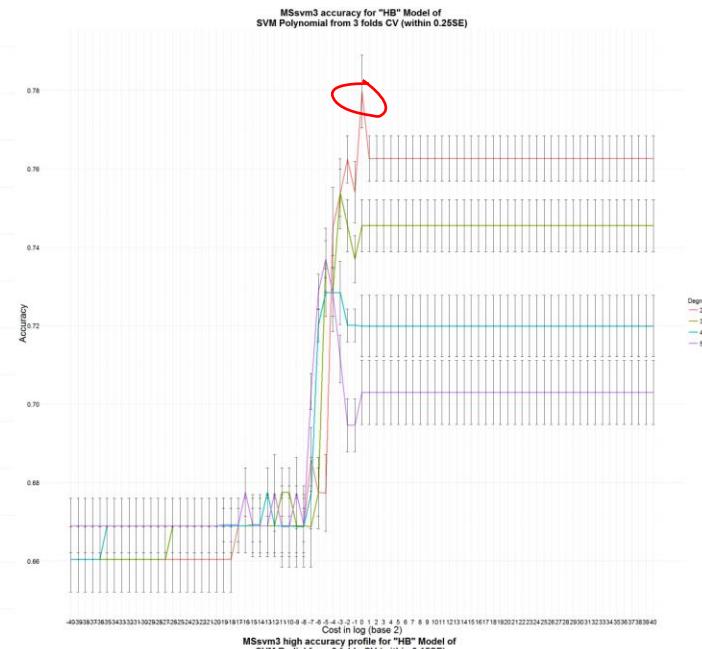
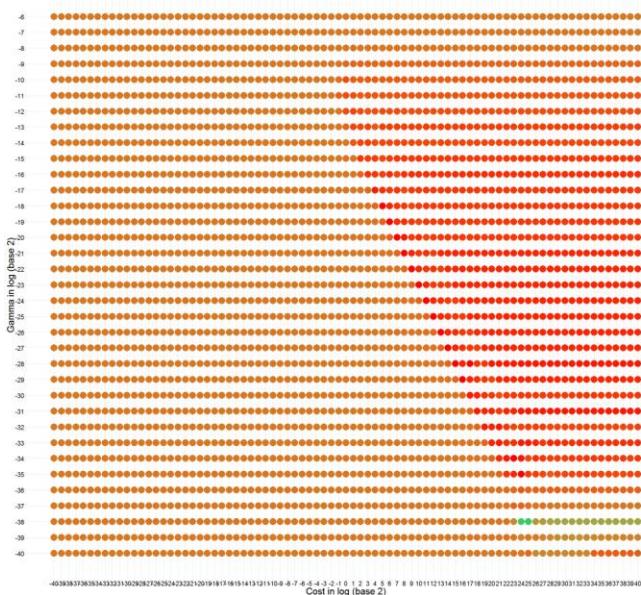
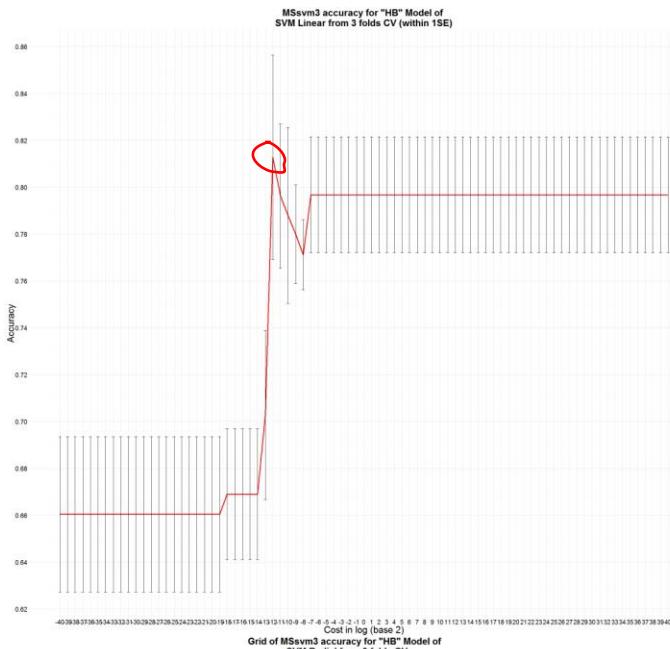
Grid of MS accuracy for "HB" Model of SVM Radial from 3 folds CV



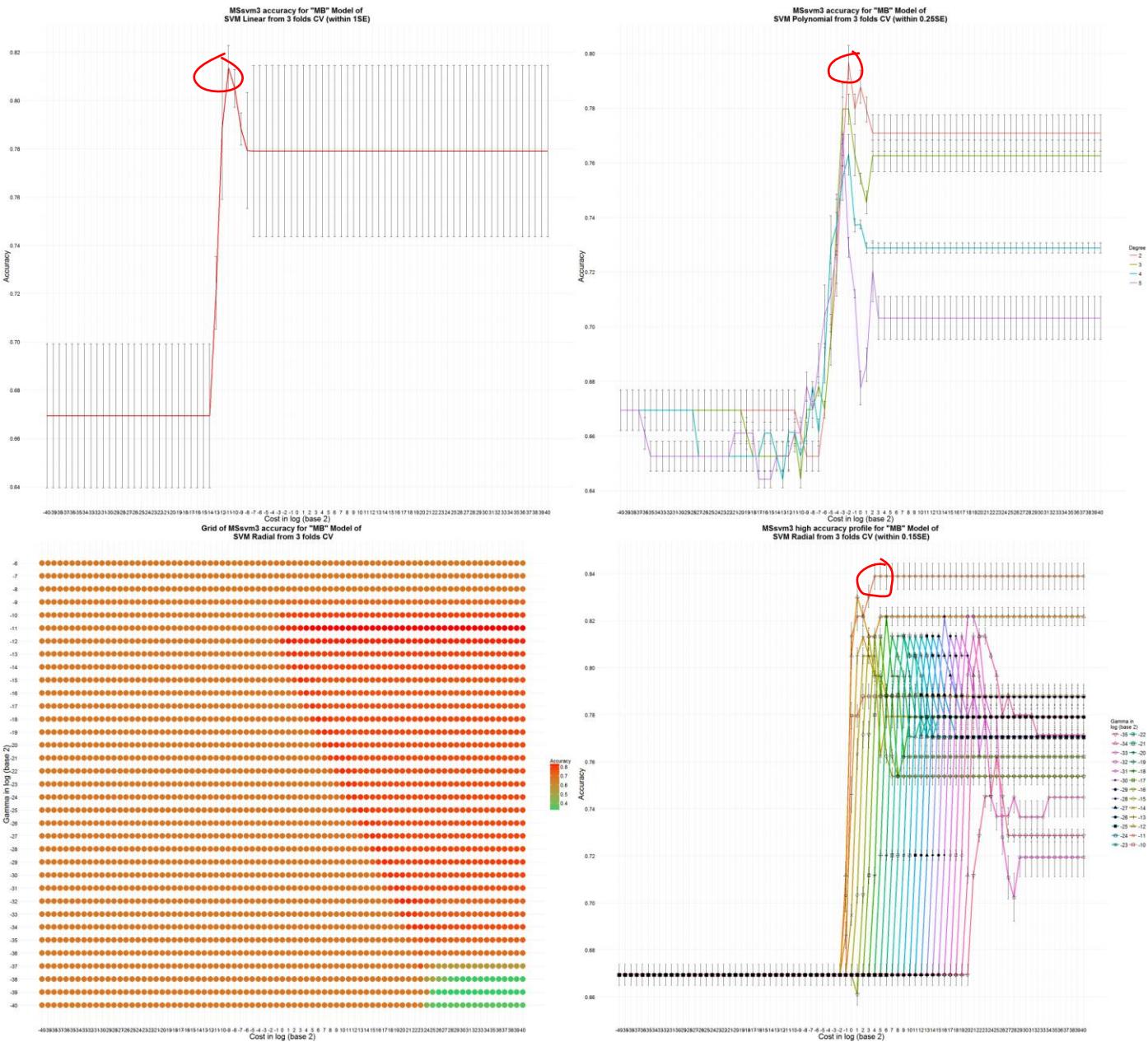
MS high accuracy profile for "HB" Model of SVM Radial from 3 folds CV (within 0.15SE)



Original parameter setting plot for HB of MS

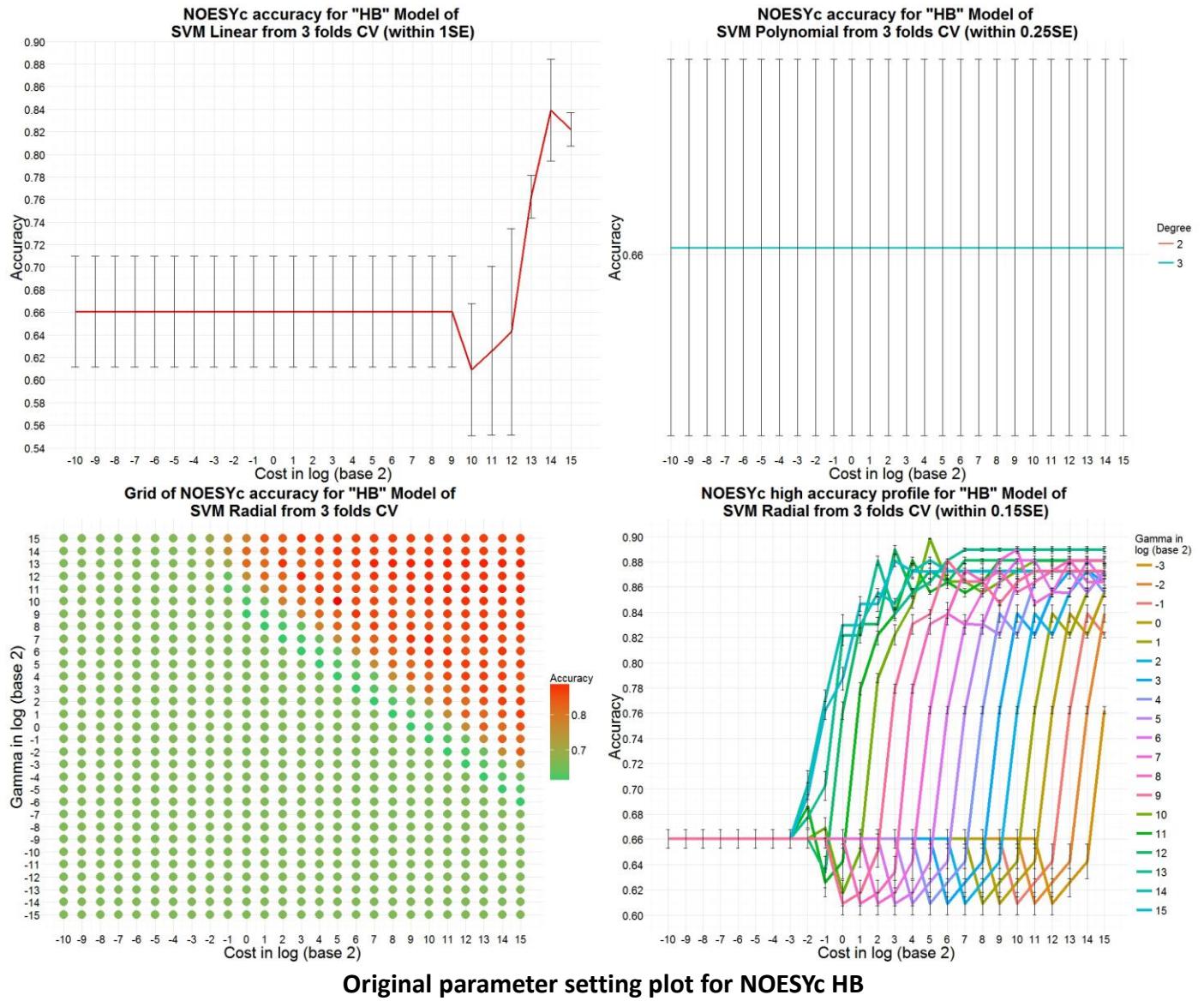


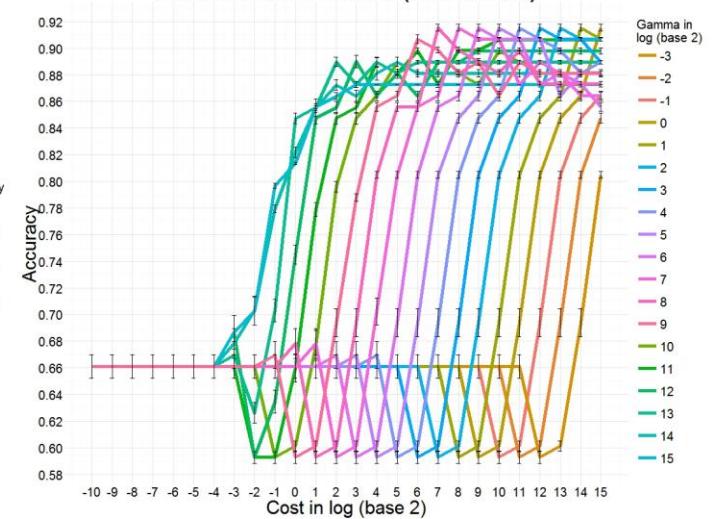
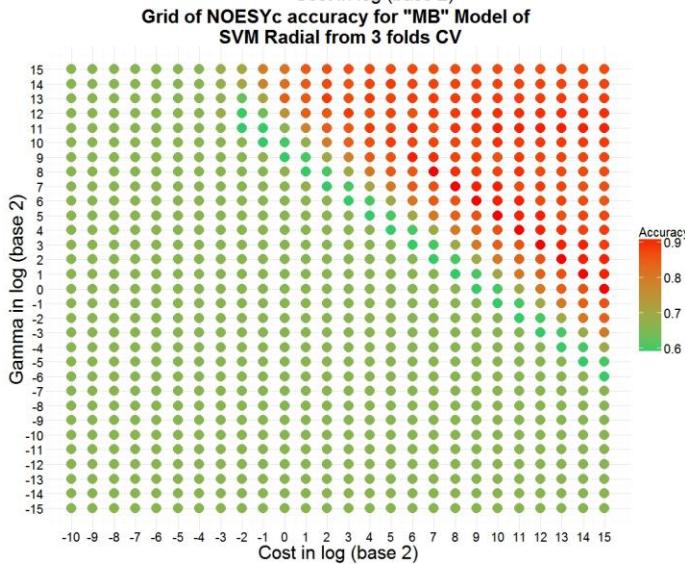
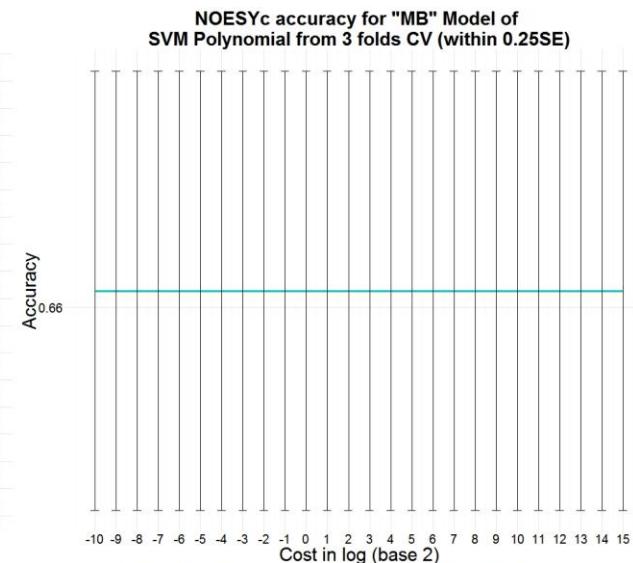
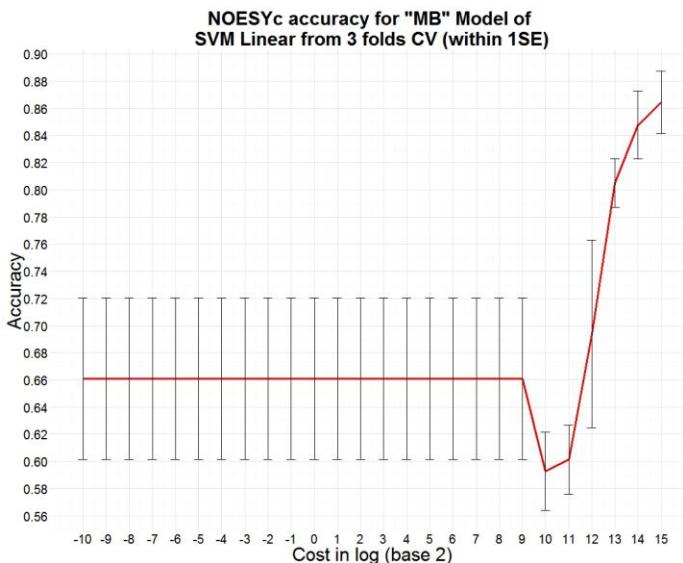
Updated parameter setting plot for HB of MS



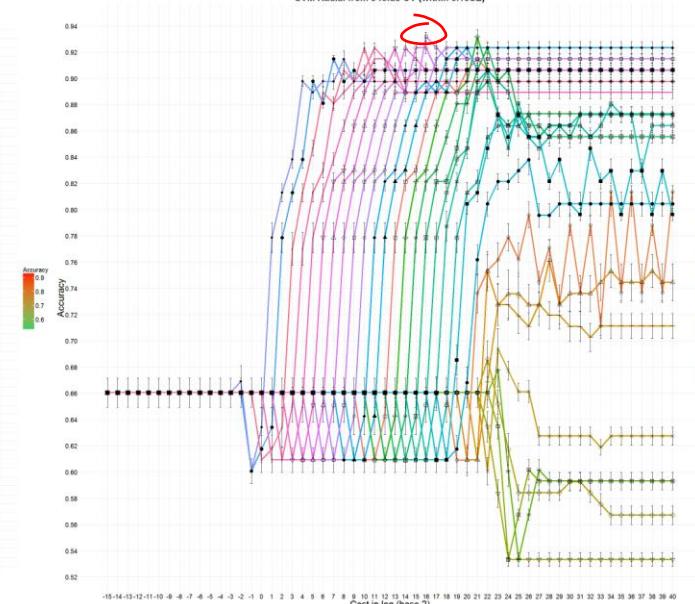
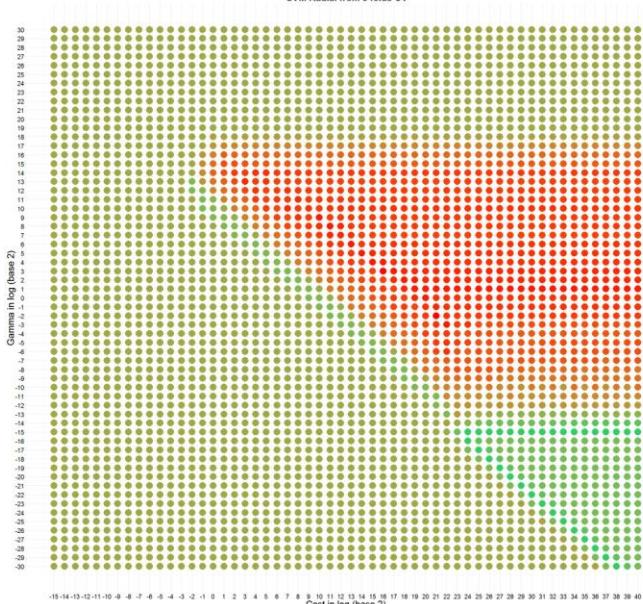
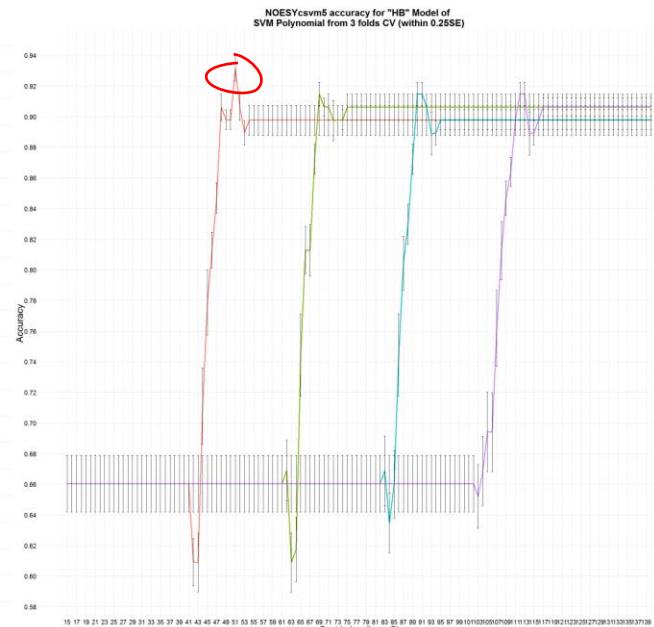
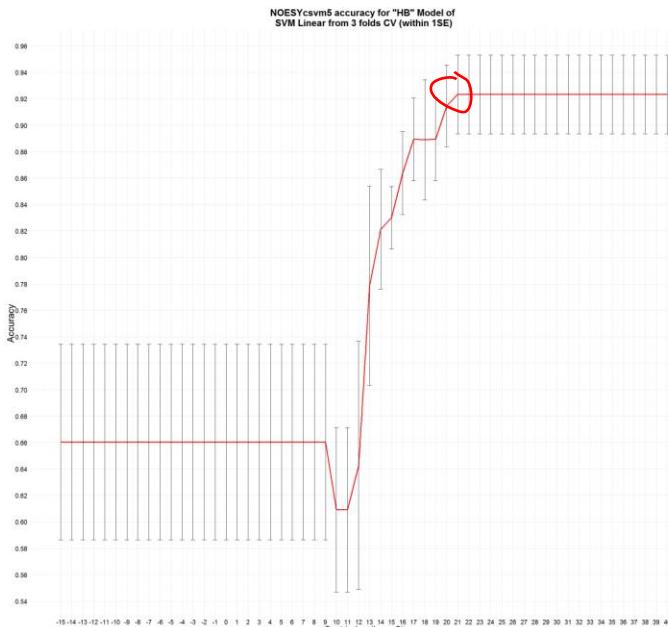
Updated parameter setting plot for MB of MS

1.4.2 NOESYc

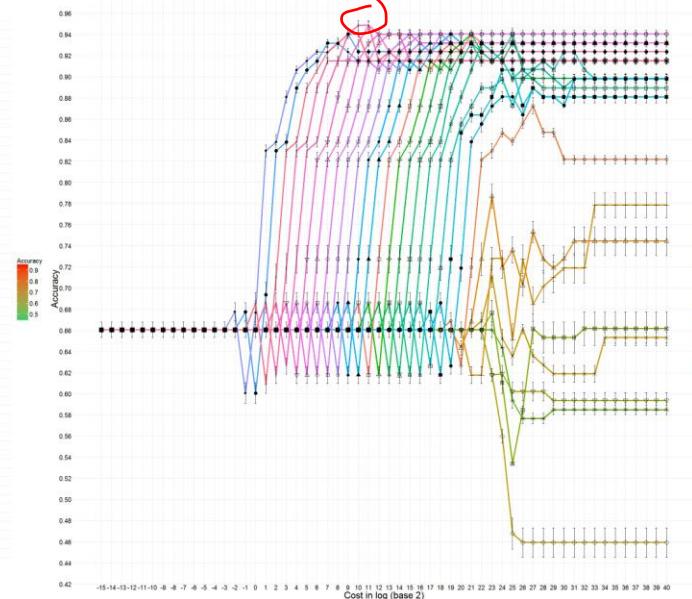
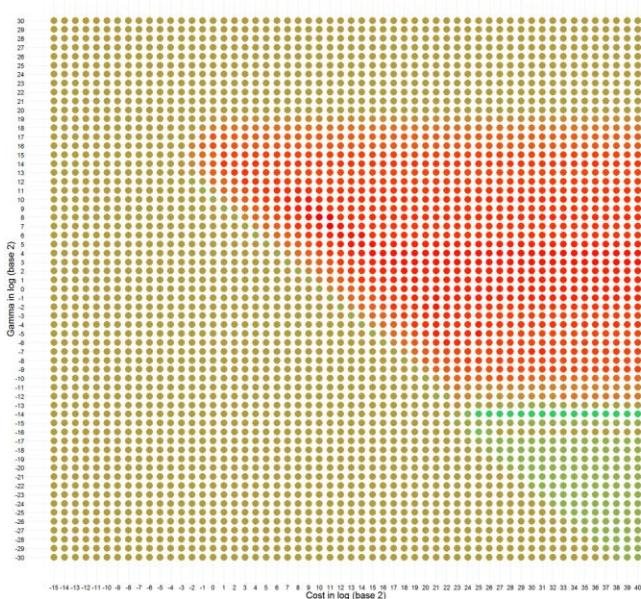
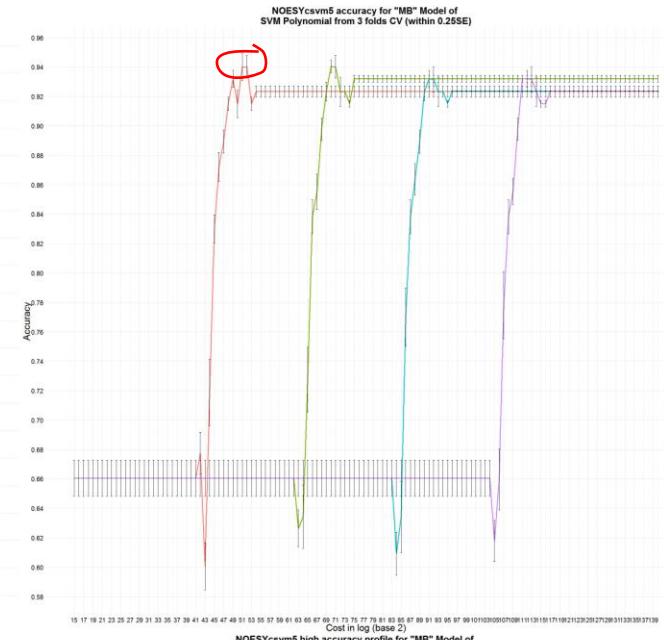
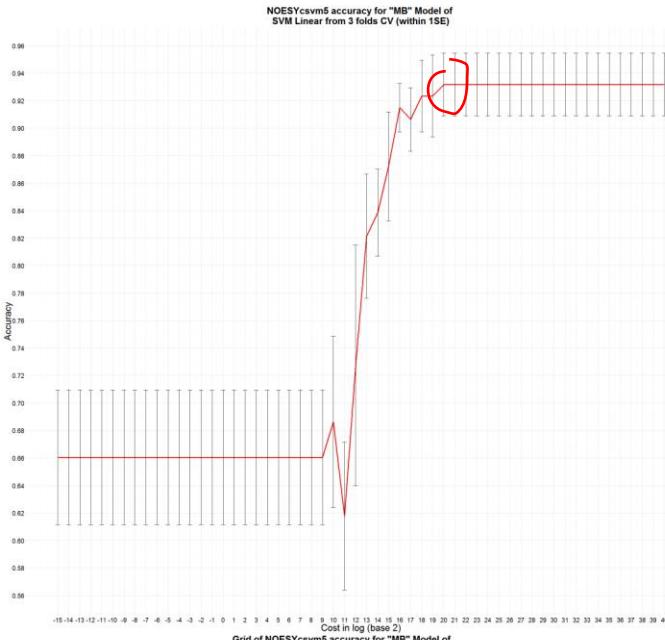




Original parameter setting plot for NOESYc MB



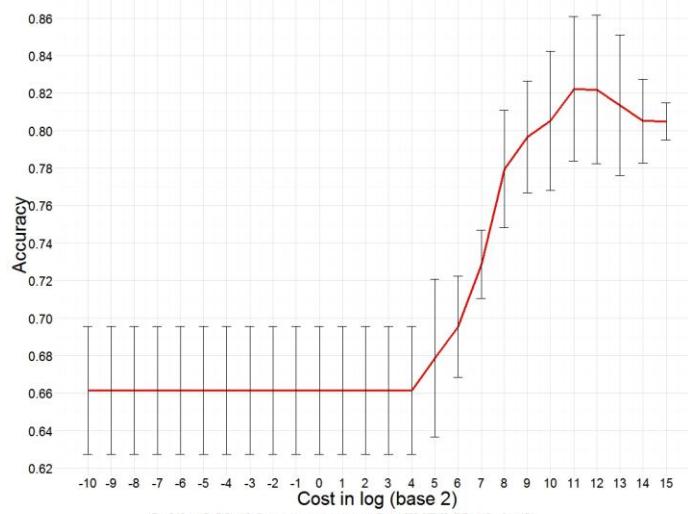
Updated parameter setting plot for NOESYc HB



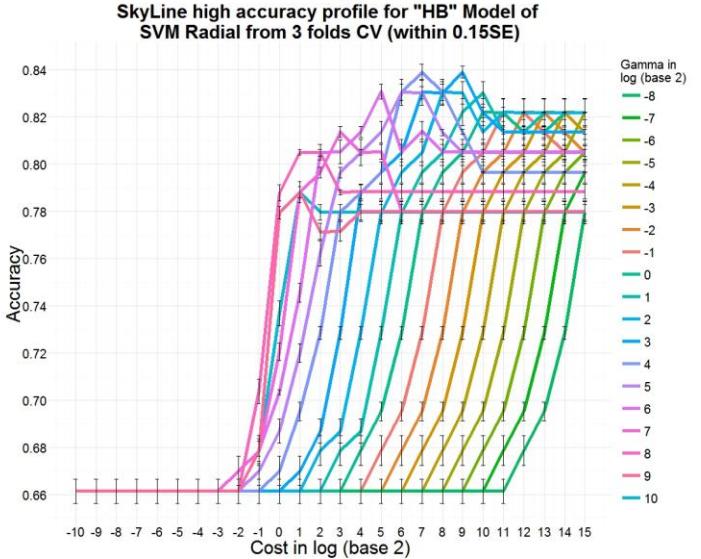
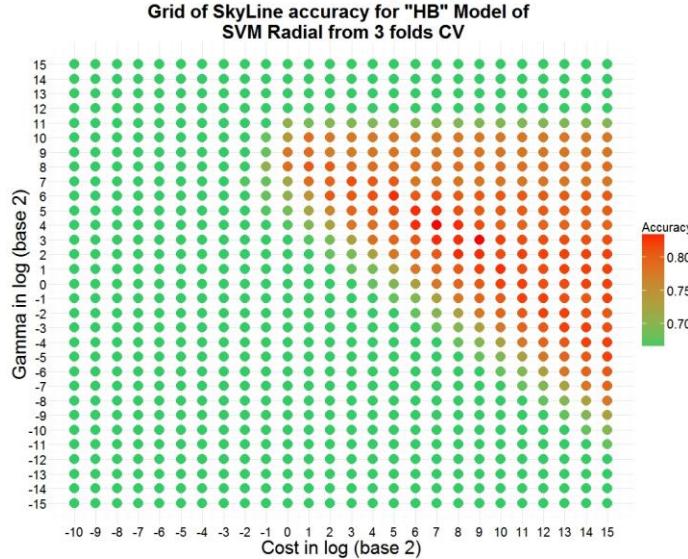
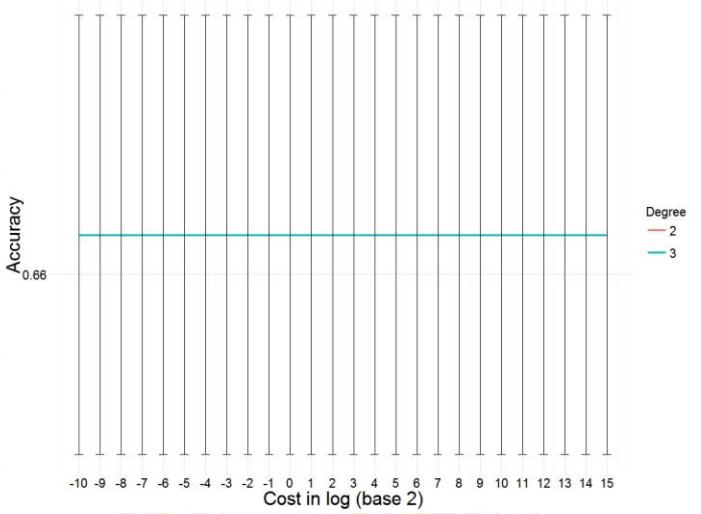
Updated parameter setting plot for NOESYc MB

1.4.3 Skyline

SkyLine accuracy for "HB" Model of SVM Linear from 3 folds CV (within 1SE)

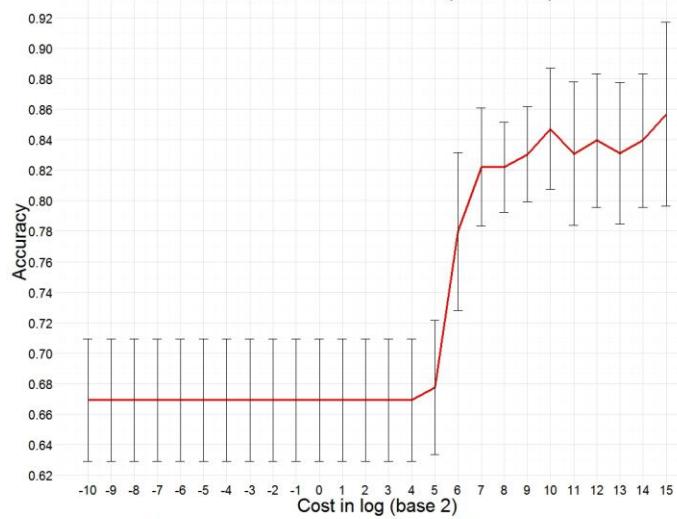


SkyLine accuracy for "HB" Model of SVM Polynomial from 3 folds CV (within 0.25SE)

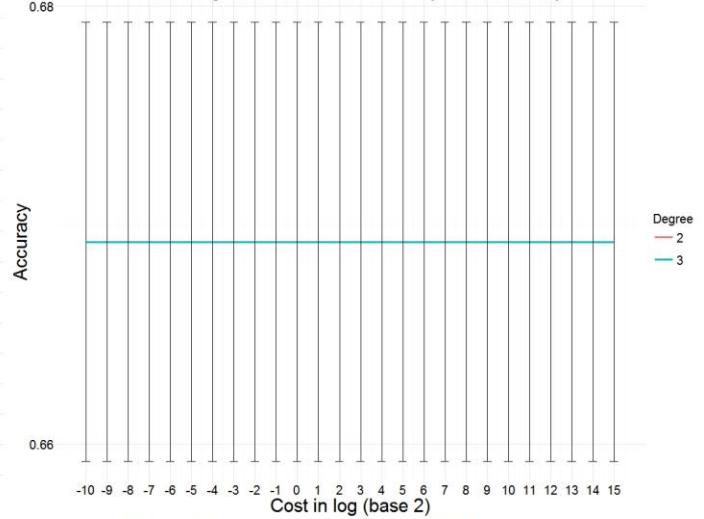


Original parameter setting plot for Skyline HB

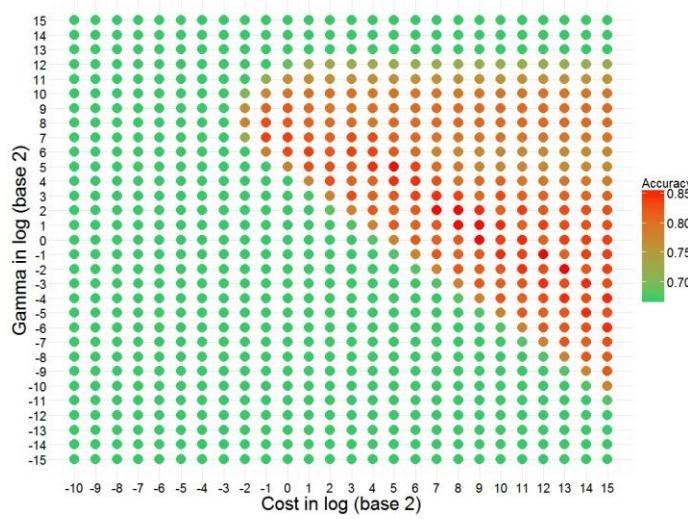
SkyLine accuracy for "MB" Model of SVM Linear from 3 folds CV (within 1SE)



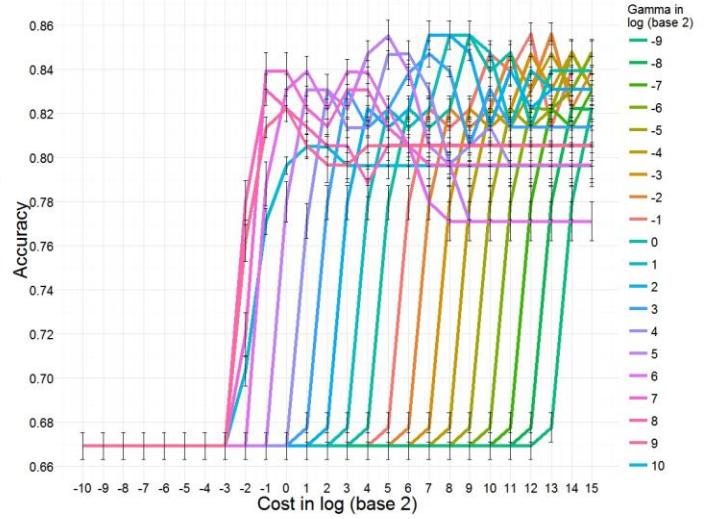
SkyLine accuracy for "MB" Model of SVM Polynomial from 3 folds CV (within 0.25SE)



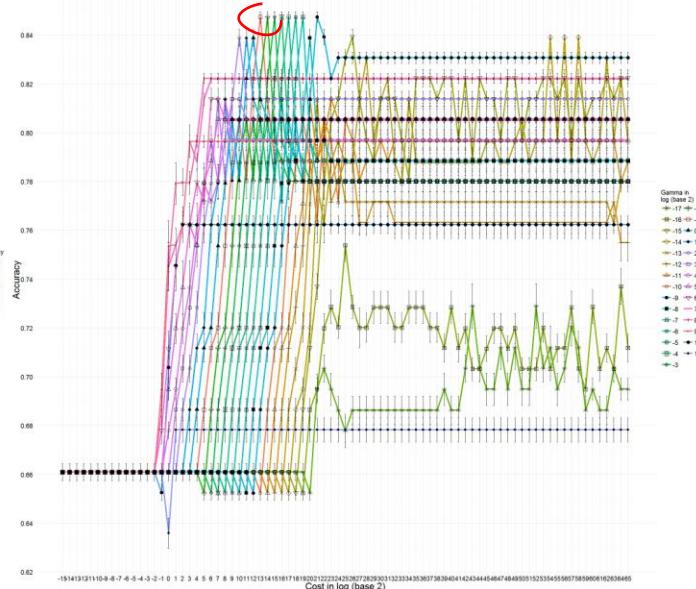
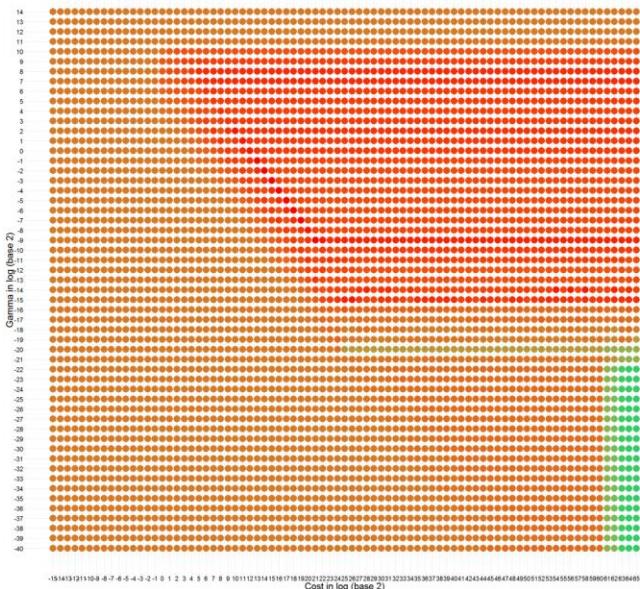
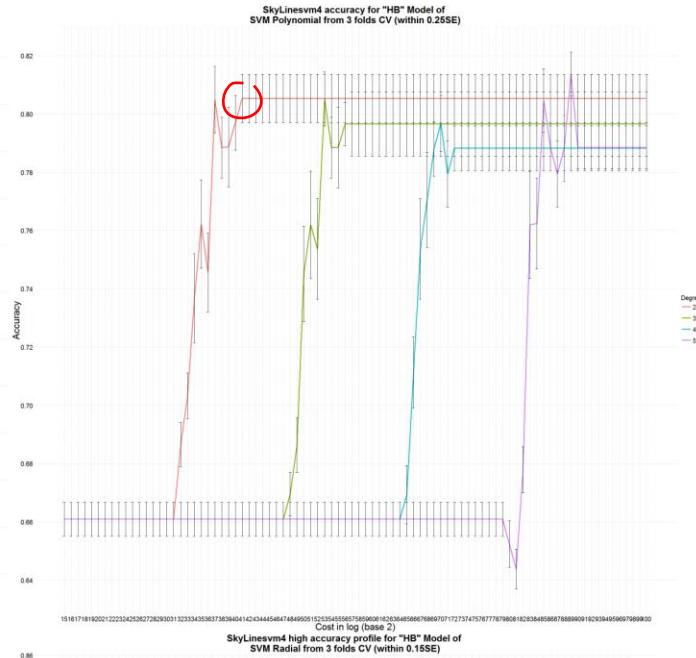
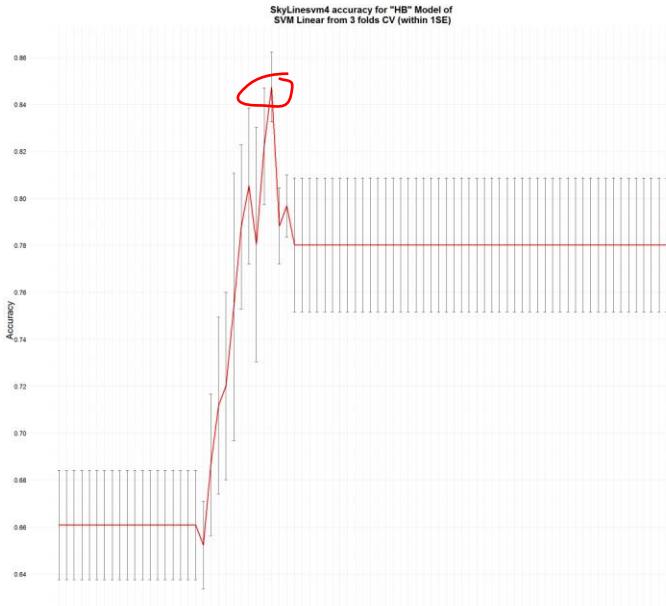
Grid of SkyLine accuracy for "MB" Model of SVM Radial from 3 folds CV

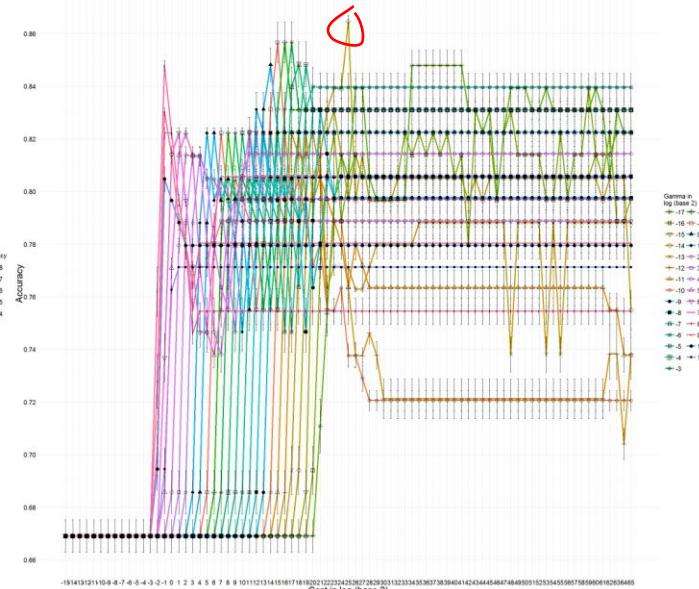
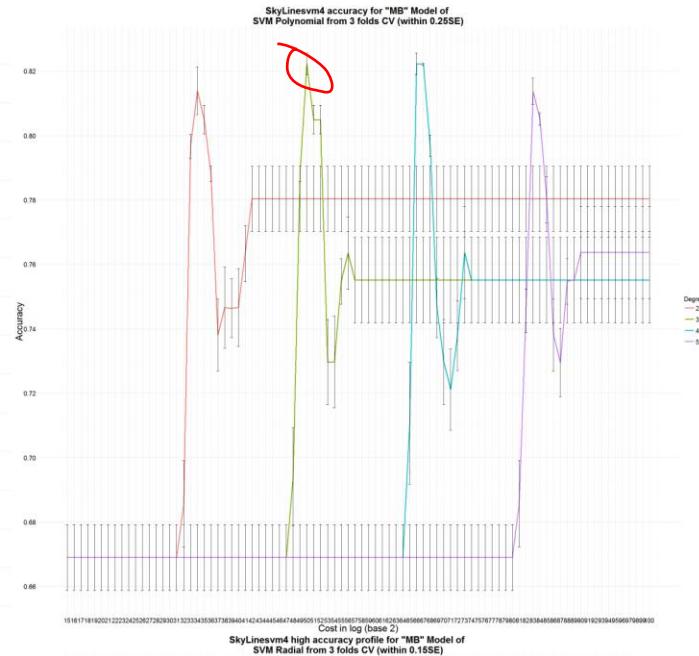
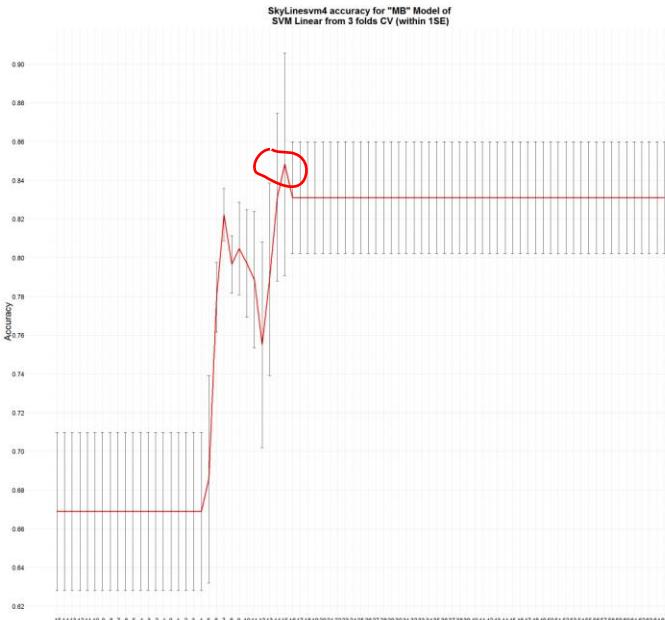


SkyLine high accuracy profile for "MB" Model of SVM Radial from 3 folds CV (within 0.15SE)



Original parameter setting plot for Skyline MB





Updated parameter setting plot for Skyline MB

1.5 Conclusion

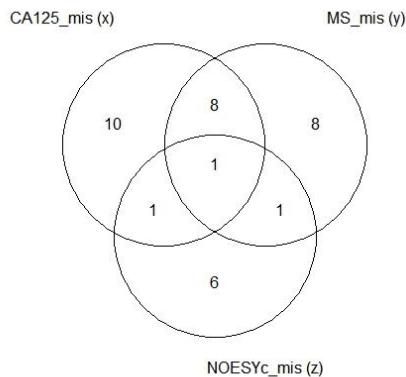
Selected models and the associated parameters (in red)

Method	Dataset			Model
	CA125	MS (parameter/accuracy)	NOESYC (parameter/accuracy)	
kNN	--	K=9 / 0.75	K=3 / 0.83	HB
	--	K=7 / 0.85	K=3 / 0.855	MB
	--	K=7 / 0.64	K=3 / 0.82	3-Class
PLS	--	Cp=2 / 0.81	Cp=4 / 0.88	HB
	--	Cp=1 / 0.855	Cp=3 / 0.905	MB
	--	Cp=1 / 0.66	Cp=4 / 0.65	3-Class
Bagging	--	--	--	HB
	--	--	--	MB
	--	--	Tree = 3000 / 0.881	3-Class
RF	--	--	--	HB
	--	--	--	MB
	--	--	--	3-Class
Boosting	--	1000 + 0.01 + 2 / 0.85	100 + 0.01 + 1 / 0.84	HB
	--	5000 + 0.01 + 2 / 0.89	100 + 0.01 + 3 / 0.85	MB
	--	5000 + 0.001 + 1 / 0.84	10 + 0.001 + 2 / 0.645	3-Class
SVM-L	--	- / 0.81	Cost=2^21 / 0.922	HB
	--	- / 0.81	Cost=2^20 / 0.93	MB
	--	--	--	3-Class
SVM-P	--	- / 0.78	2^51 + 2 / 0.93	HB
	--	- / 0.795	2^70 + 3 / 0.94	MB
	--	--	--	3-Class
SVM-R	--	- / 0.815	2^16 + 2^3 / 0.93	HB
	--	- / 0.838	2^10 + 2^8 / 0.95	MB
	--	--	--	3-Class
Logistic	Logistic	--	--	HB
	Logistic	--	--	MB
	Logistic	--	--	3-Class

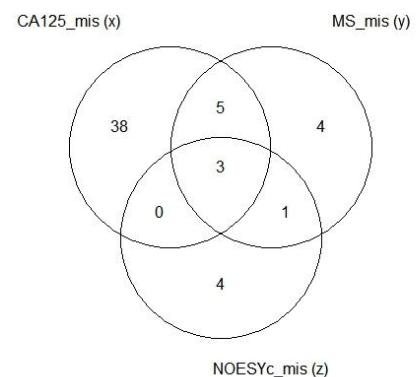
2. Venn Diagram

2.1 Venn results

Venn Diagram for HB misclassification



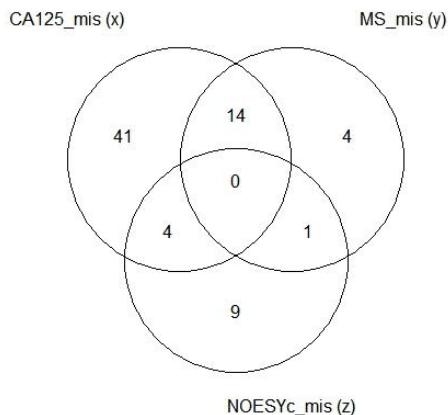
Venn Diagram for MB misclassification



N unique: xyz = 35; x = 20; y = 18; z = 9

N unique: xyz = 55; x = 46; y = 13; z = 8

Venn Diagram for 3-Class misclassification



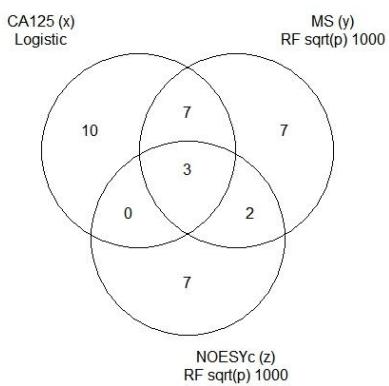
N unique: xyz = 73; x = 59; y = 19; z = 14

Annotation for 3-way Misclassified Samples, via Venn Diagram.

Model	ID	Diagnosis	Age	Stage
HB	G7	Cancer	58	1
	G113	Normal	36	--
MB	G124	Normal	48	--
	G210	Normal	53	--
3Class	--	--	--	--

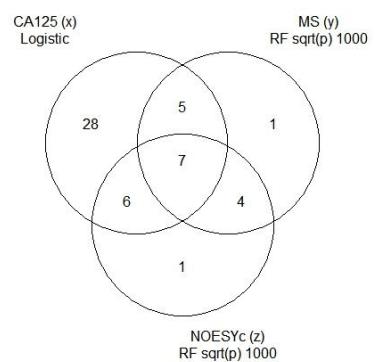
2.2 Comparison to outputs using Joselle's settings

Venn Diagram HB



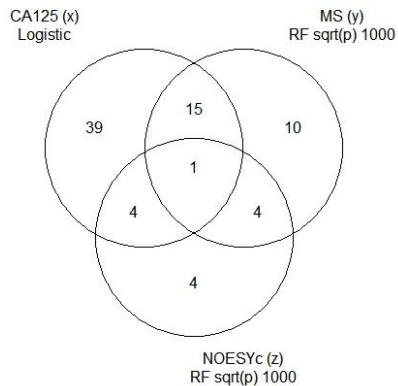
N unique: xyz = 36; x = 20; y = 19; z = 12

Venn Diagram MB



N unique: xyz = 52; x = 46; y = 17; z = 18

Venn Diagram 3-Class

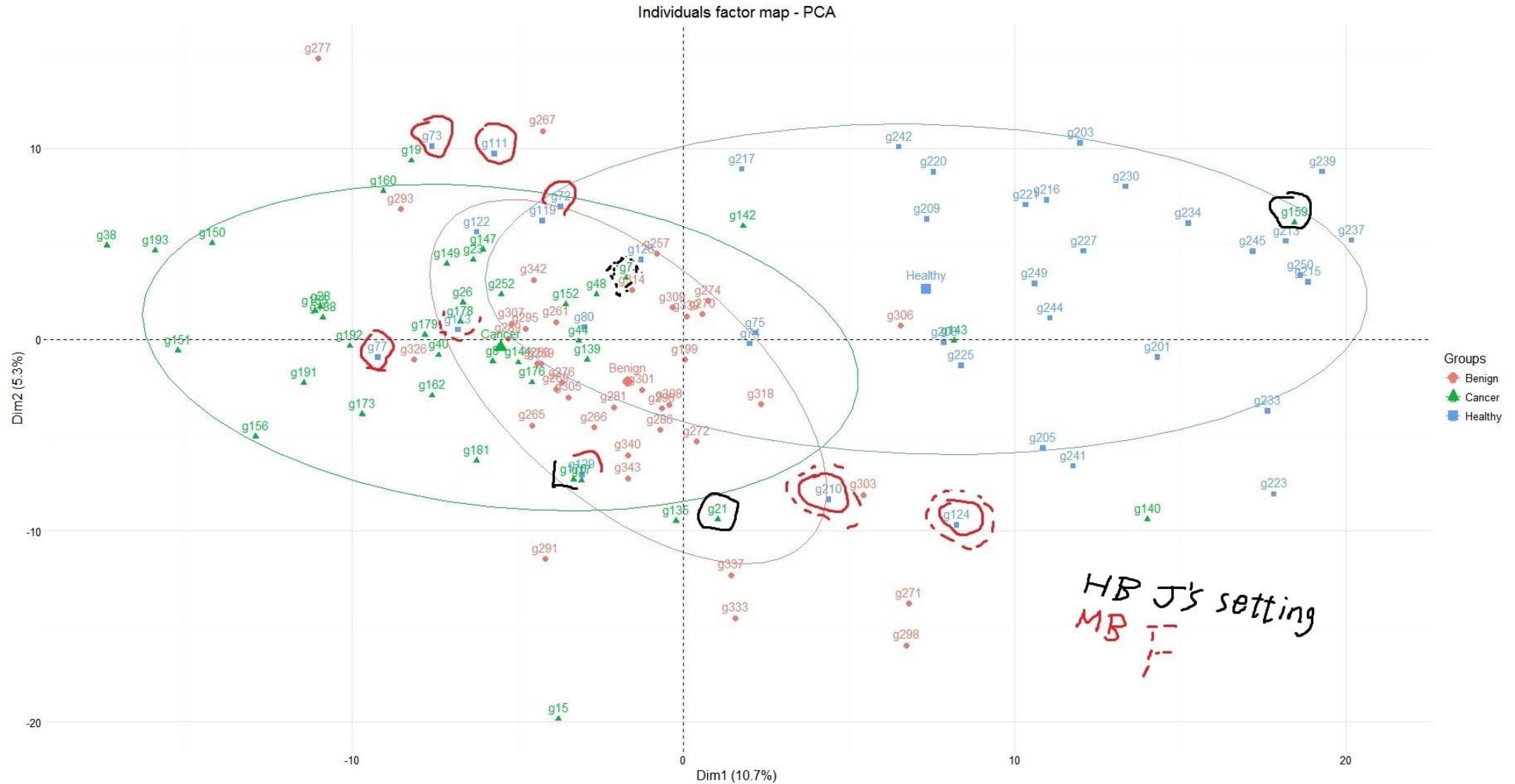


N unique: xyz = 77; x = 59; y = 30; z = 13

Annotation for 3-way Misclassified Samples, via Venn Diagram, using RF (sqrt(p) & 1000 trees) for MS and NOESYc and Logistic for CA125

Model	ID	Diagnosis	Age	Stage
HB	g159	Cancer	65	1A
	g21	Cancer	51	1C
	g17	Cancer	50	1B
MB	g77	Normal	49	
	g111	Normal	49	
	g73	Normal	43	
	g129	Normal	46	
3Class	g124	Normal	48	
	g210	Normal	53	
	g72	Normal	45	
3Class	g159	Cancer	65	1A

2.2.3 Discussion



Red circles represent 3-way misclassified subjects from MB model while the black ones represent HB model. Solid circles indicate Joselle's result and dashed ones are mine. In HB model, the right two groups will be considered as one group indicating non-cancer. Under this circumstance, those subjects who are diagnosed as cancer (green subjects) but locate in "HB" group (black circled) will be most likely predicted as "non-cancer" according to the statistical models. This kind of misclassification may also include the scenario that patients may be diagnosed as cancer due to conservative concern from doctors. On the other hand, under MB model, the left two groups will be considered as one group. Similarly, those subjects who are actually healthy (blue subjects) but relatively locate on the left side will be predicted as "cancer" (red circled). Under this situation, the above "conservative" concern will not likely result in a misclassification because both the "concern" and the model will lead to the same outcome as "cancer".