

Q2 ) a) i)

Dataset1-9 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	5.275e-28	480.897
	w	[-7.01477582 3.20265861 -2.01056618 4.61891474 -8.48679639 5.34513234 -1.36854253 -20.00142649 13.2641012 3.11232438]			
		L1(w)= 68.425	L2(w)= 27.802	Spars = 0	
LASSO	2.0	1067.508	1492.756	14.107	233.383
	w	[ 0.12578696 2.26001059 0. -3.34237423 -0.5.01163416 0. -5.93509725 -0. 1.43300028]			
		L1(w)= 18.107	L2(w)= 8.870	Spars = 4	
RIDGE	4.0	-840.630	628.977	18.117	264.804
	w	[-0.24139412 2.5665958 -0.28155627 -1.71113314 -1.61199141 2.81003838 2.21325862 -3.03423719 -2.75553818 1.62177076]			
		L1(w)= 18.847	L2(w)= 6.669	Spars = 0	

## Dataset2-9 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	86.336	112.651
	w	[ 0.433, 2.397, 0.568, -3.870, 0.855, 2.250, 2.041, -6.177, -1.804, 1.254]			
		L1(w)= 21.654	L2(w)= 8.613	Spars = 0	
LASSO	0.0	723.813	1269.935	87.635	110.196
	w	[0.43208693 2.33887343 0.43071582 -2.94652499 0.2.36208926 1.92436118 -6.33525333 -1.61782688 1.14532181]			
		L1(w)= 19.533	L2(w)= 8.238	Spars = 1	
RIDGE	6.0	-121.861	27.571	89.147	111.420
	w	[0.45904623 2.25533004 0.55844399 -2.57037539 -0.32269209 2.23048119 2.05571123 -4.14875114 -3.77234633 1.17294188]			
		L1(w)= 19.546	L2(w)= 7.371	Spars = 0	

Dataset3-9 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	98.213	109.124
	w	[ 1.715, 1.904, 0.412, -3.172, 0.253, 4.872, -0.252, -8.712, 0.805, 0.891]			
		L1(w)= 22.994	L2(w)= 10.864	Spars = 0	
LASSO	-2.0	674.273	1094.949	98.45	109.072
	w	[1.70180095 1.88836961 0.37895217 -2.91361253 0.460921066 0. -7.90337403 -0. 0.87016587]			
		L1(w)= 20.265	L2(w)= 9.977	Spars = 3	
RIDGE	2.0	-102.137	2.223	98.222	108.986
	w	[1.71554948 1.90414688 0.41169323 -3.16239173 0.24355474 4.83535943 -0.21617519 -8.46183322 0.55511326 0.8910176 ]			
		L1(w)= 22.396	L2(w)= 10.628	Spars = 0	

Q2 ) a ) ii)

- 1) Test MSE is best obtained with L1 (LASSO) regularizer. For Dataset 1&2, Test MSE was better for Ridge regularizer compared to no regularizer.
- 2) Yes, each regularizer lowers the corresponding norm of w. As training data increases, the difference in corresponding norms starts decreasing (with  $N_{tr} \rightarrow 1000+$ , it's almost insignificant). This is because, with few training samples, our weight vector takes higher values for Linear regression with no regularizer classifier. As the training samples increases, our Linear regression model with no regularizer performs well and takes lower weight values.
- 3) We obtain sparsity with only LASSO regularizer. As training data increases, our best fit Lambda value ( $\lambda$ ) decreases. In general, Sparsity is seen more with lower sized training dataset.



**Q2 ) b) i)**

Dataset4-2 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	95.38019904643618	163.48761227397387
	w	[ 6.77265711 -2.4928513 7.23801612]			
		L1(w)= 16.503524531665292	L2(w)= 10.221157922129628	Spars=0	
LASSO	2.0	256.25146952709184	249.16970913724862	98.92934420027042	134.48864379047166
	w	[5.53704387 0. 4.00141645]			
		L1(w)= 9.538460320163804	L2(w)= 6.831558271694046	Spars=1	
RIDGE	4.0	-196.337	50.48813742731631	102.92593770240387	127.90734292133521
	w	[4.61680832 1.55139276 2.16748412]			
		L1(w)= 8.335685196859789	L2(w)= 5.331015470702115	Spars=0	

Dataset5-2 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	87.12261767437649	114.70433167932684
	w	[4.05510307 2.74884213 -0.29784002]			
		L1(w)= 7.101785217832998	L2(w)= 4.908024311927913	Spars=0	
LASSO	2.0	139.29912344057894	75.98257910583324	88.72353370540213	105.57083378205994
	w	[3.7532937 2.47956376 0. ]			
		L1(w)= 6.232857452937665	L2(w)= 4.498383042243042	Spars=1	
RIDGE	6.0	-115.603	14.631848794699655	88.86631773419617	106.00172830717882
	w	[ 3.79650626 2.41954902 -0.18839922]			
		L1(w)= 6.4044545005619415	L2(w)= 4.505904073798311	Spars=0	

Dataset6-2 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	101.35833777888638	101.44570933707959
	w	[1.21077089 2.30240071 0.23152547]			
		L1(w)= 3.74469706915353	L2(w)= 2.6116315269679835	Spars=0	
LASSO	-10.0	137.055342060517	63.836299120708816	101.35833791468285	101.44595275106839
	w	[1.2107991 2.30236633 0.23141756]			
		L1(w)= 3.744582986648264	L2(w)= 2.6116047303958663	Spars=0	
RIDGE	6.0	-112.144	4.0246773291225395	101.47660067151448	101.30261474052405
	w	[1.22012634 2.2180724 0.24090883]			
		L1(w)= 3.6791075627301373	L2(w)= 2.542949176508624	Spars=0	

Dataset7-2 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	25.417551693358597	116.51141337592615
	w	[ 1.6193184 4.35846137 -2.05316003]			
		L1(w)= 8.030939797861016	L2(w)= 5.082700433707281	Spars=0	
LASSO	0.0	80.56417433998845	88.39221271457387	26.522508540648026	108.51285152596137
	w	[1.49398093 3.84541442 -1.4616377 ]			
		L1(w)= 6.801033058092814	L2(w)= 4.37670833943422	Spars=0	
RIDGE	2.0	-70.890	13.911391429017938	26.617548498116584	109.29761312532439
	w	[ 1.59580358 3.78666035 -1.49465134]			
		L1(w)= 6.877115268463562	L2(w)= 4.372569988665177	Spars=0	



Dataset8-2 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	95.15432277075584	109.24257017878496
	w	[3.58068323 1.91863829 0.60434473]			
		L1(w)= 6.10366625566634	L2(w)= 4.107030295969647	Spars=0	
LASSO	0.0	155.94480304519337	88.25024220759911	95.20146796600213	109.46763164804094
	w	[3.51599517 1.88223308 0.593483 ]			
		L1(w)= 5.991711258559851	L2(w)= 4.032027469140141	Spars=0	
RIDGE	6.0	-115.911	10.698199618203887	95.90349733909382	110.51198296942115
	w	[3.20005109 1.41174703 0.97725369]			
		L1(w)= 5.589051809257596	L2(w)= 3.6315811189650824	Spars=0	

Dataset9-2 features

	Model Selection			Performance	
	Best Param $\log_2 \lambda$	Mean of MSE	STD of MSE	MSE on Train	MSE ON Test
Least Square	-	-	-	83.3240152571577	111.41165530589785
	w	[ 4.11404128 3.04009919 -0.51630424]			
		L1(w)= 7.670444709675 4755	L2(w)= 5.141411166840 956	Spars=0	
LASSO	0.0	123.0456877361 9322	57.15983617651 494	83.90836860287 891	109.50398425369 923
	w	[4.10151715 2.50447238 0. ]			
		L1(w)= 6.605989527505 749	L2(w)= 4.8057075252116 49	Spars=1	
RIDGE	4.0	-90.854	3.3415442744863 117	83.44263614295 154	110.3600950184 0029
	w	[ 4.10848705 2.79878652 -0.28354144]			
		L1(w)= 7.190815012245 55	L2(w)= 4.97928384123978 2	Spars=0	

**Q2 ) b) ii) --- plots are in the pdf file containing code part. Please refer over there.**

**Q2) b) iii)**

- 1) The plots have 2 figures each, MSE AND regularizer constraints drawn over a range of  $w_1$  and  $w_2$  for a particular value of  $w_0$ (for the best  $\lambda$ ). We can see that MSE with no regularizer intercepts the constraints contour at greater values of  $w_1$  and  $w_2$ . Also, LASSO (l1 regularizer)

gives us sparsity which can be observed by the intersection of MSE (lowest) with constraints contour at either  $w_1=0$  or  $w_2=0$ .

- 2) With lower training data size, the regularizer has a larger effect on MSE. Regularized MSE takes a lower value compared with unregularized MSE.
- 3) With the 'special case' datasets, the LASSO regularization does not give any sparsity with  $N_{tr}=10$  and  $N_{tr}=30$  but it does give sparsity for  $N_{tr}=100$  which is not the case in set of datasets 4,5,6 where we got sparsity for  $N_{tr}=10$ (dataset 4) and  $N_{tr}=30$ (dataset 5) but did not get sparsity for  $N_{tr}=100$ (dataset 6). In general, Sparsity is observed in lower sized training datasets as its easier for model to overfit training data with lesser samples and obtain poor generalization. This was opposite with Datasets 7,8,9 (Special Case).