Assignment 2

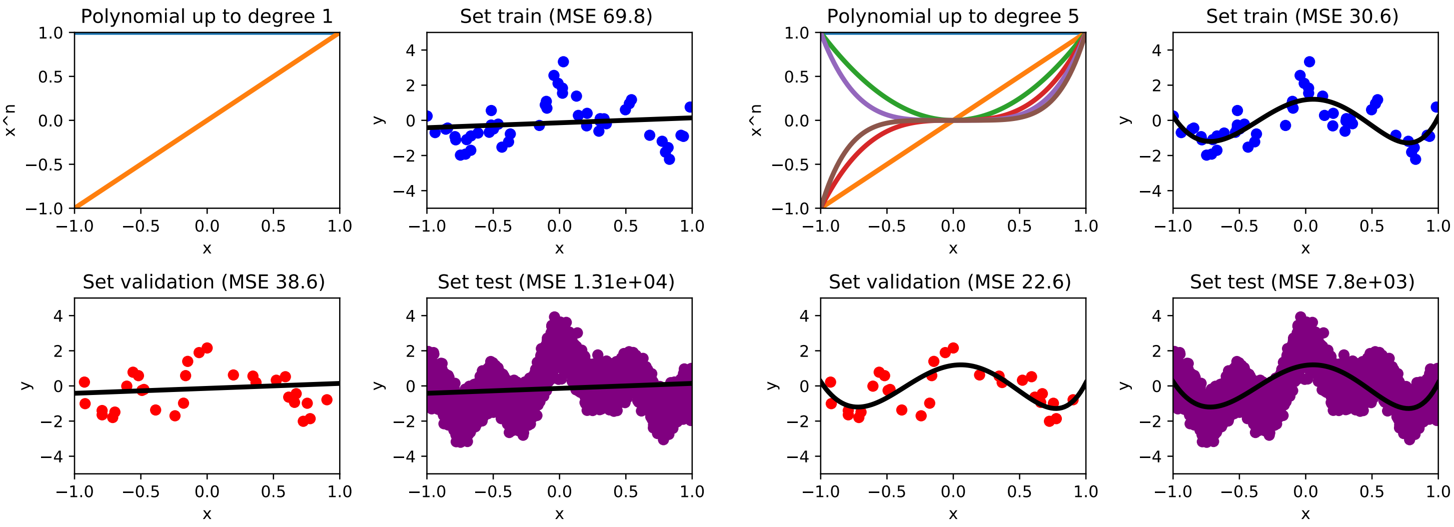
Computational Intelligence, SS2020

Team Members		
Last name	First name	Matriculation Number
Blöcher	Christian	01573246
Bürgener	Max	01531577

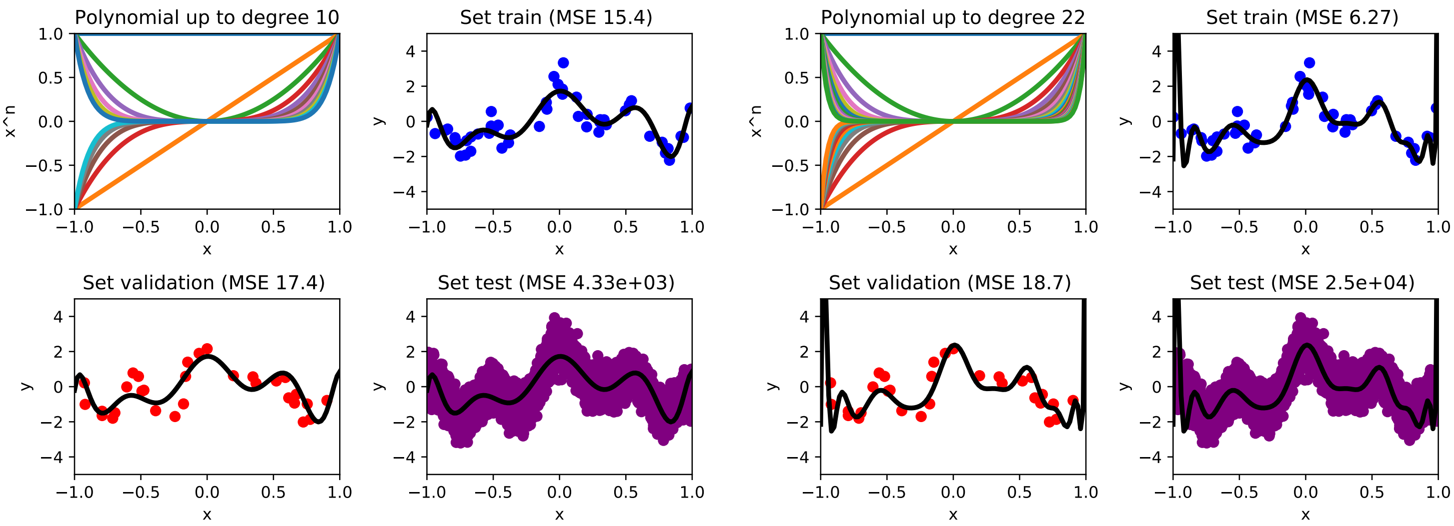
1 Linear regression

1.1 Derivation of Regularized Linear Regression

1.2 Linear Regression with polynomial features



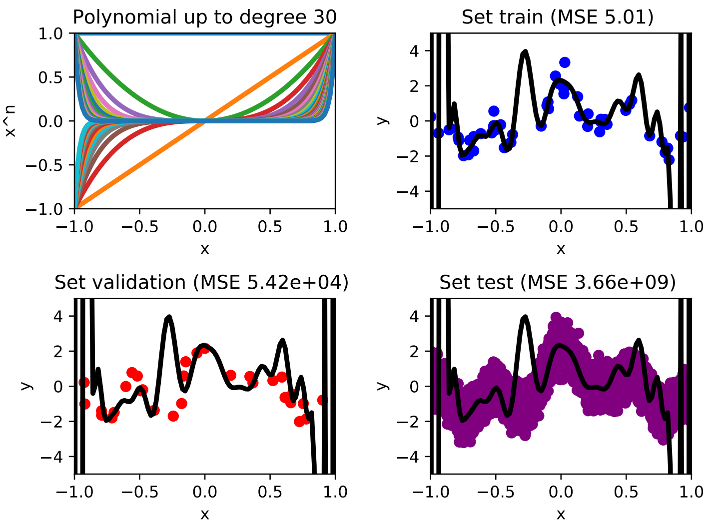
(a) $n = 1$ (b) $n = 5$



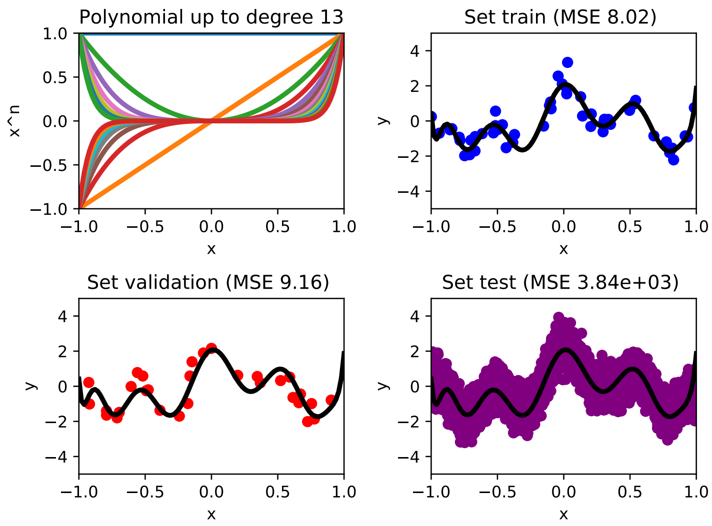
(c) $n = 10$ (d) $n = 22$

Figure 1: Results of Linear Regression for varying polynomial degree n.

As can be seen in figure 3 the training cost decreases with increasing polynomial degree n. That means the lowest cost on the training set can be achieved by using polynomials up to the highest degree $n = 30$ (s. figure 2), but then the testing cost is maximised. This is due to overfitting: By finding parameters that suit the training data best, the solution becomes too specific for the validation and testing sets, leading to greater errors. Looking at figure 2 all training data points are in the close vicinity of the polynomial but the output function barely resembles the data. The best results can be obtained with degree $n = 13$, which minimises the validation cost (s. figure 3) and leads to a very low testing cost. Having a validation set in addition to the training set helps in finding the right degree for the linear regression process and greatly improves the quality of the solution.



(a) Training set, $n = 30$



(b) Validation set, $n = 13$

Figure 2: Results with lowest cost on training and validation set.

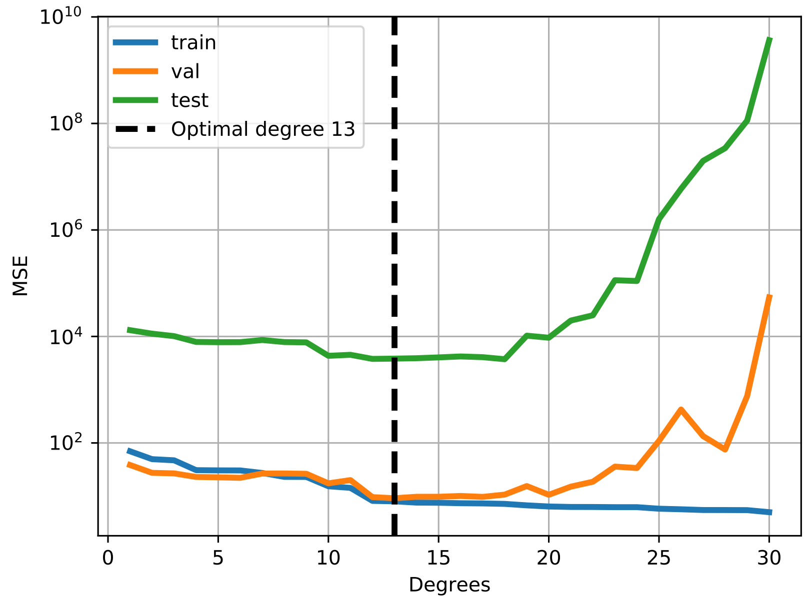
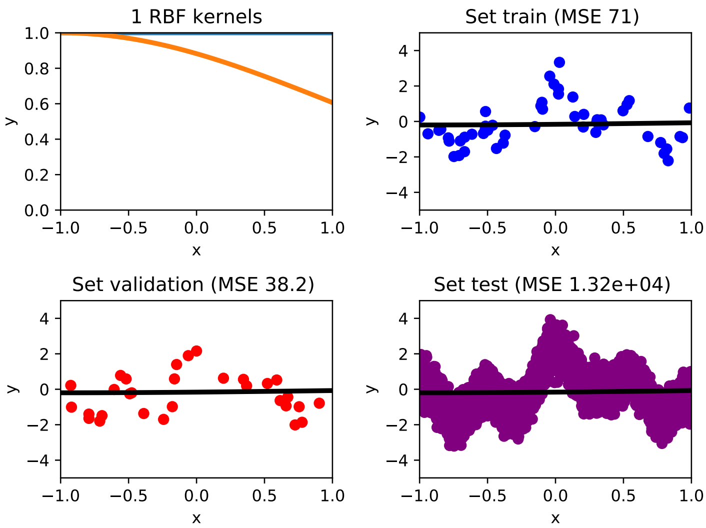


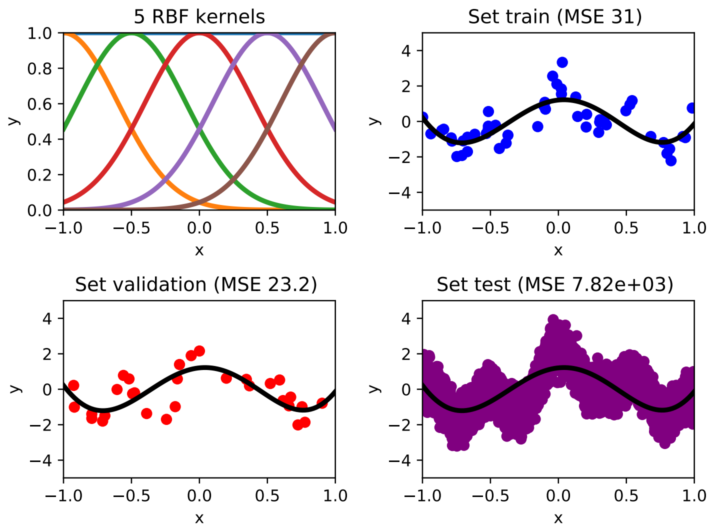
Figure 3: Training, validation and testing costs as a function of the polynomial degree n.

1.3 (Bonus) Linear Regression with radial basis functions

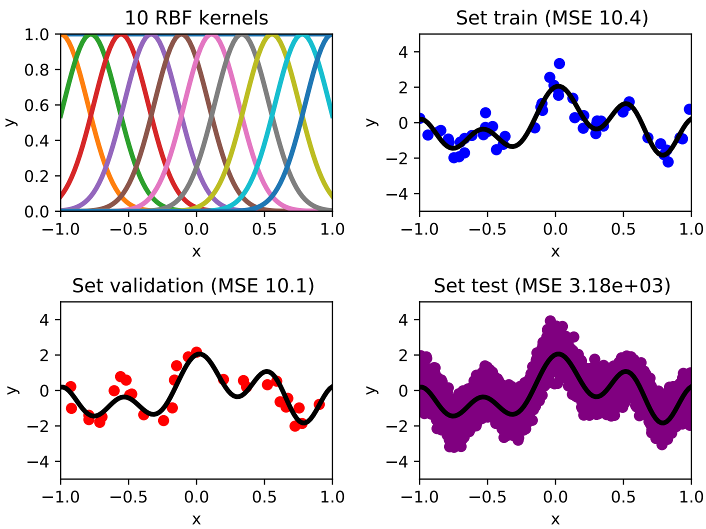
The results of the Linear Regression approach with radial basis functions are similar to the ones with polynomial functions: Increasing the degree/number of kernels l leads to lower training cost (s. figure 6) - minimised for $l = 40$ (s. figure 5) - but choosing l too high results in overfitting (s. figure 4). The cost function of the validation set is minimised for degree $l = 9$ (s. figure 5), leading to a low but not quite minimised testing cost. The actual testing cost minimum is found with degree $l = 10$. Improving on the results of Linear Regression with polynomial basis functions it is about 20% lower than the testing cost minimum with (higher) degree $n = 13$ (s. figures 4 and 2).



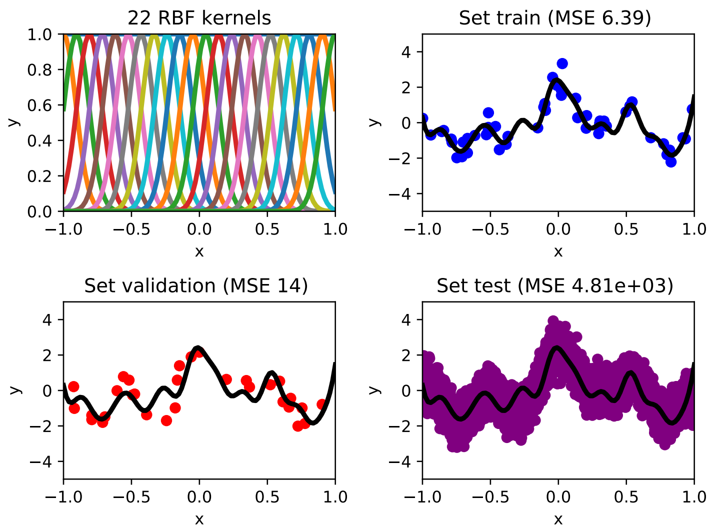
(a) $l = 1$



(b) $l = 5$

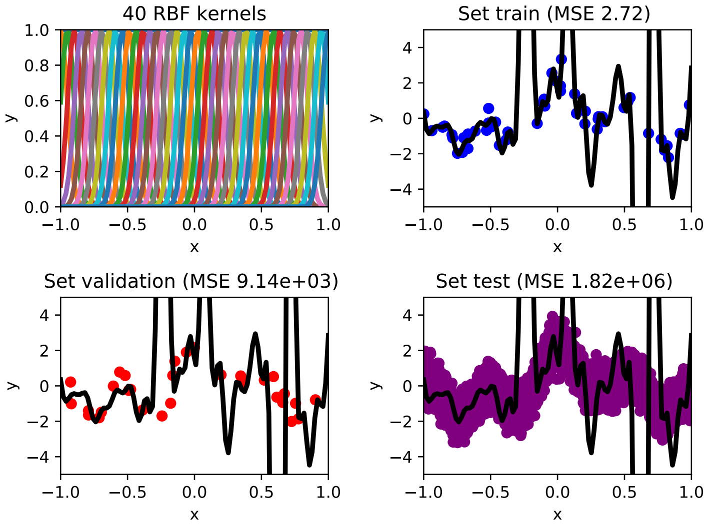


(c) $l = 10$

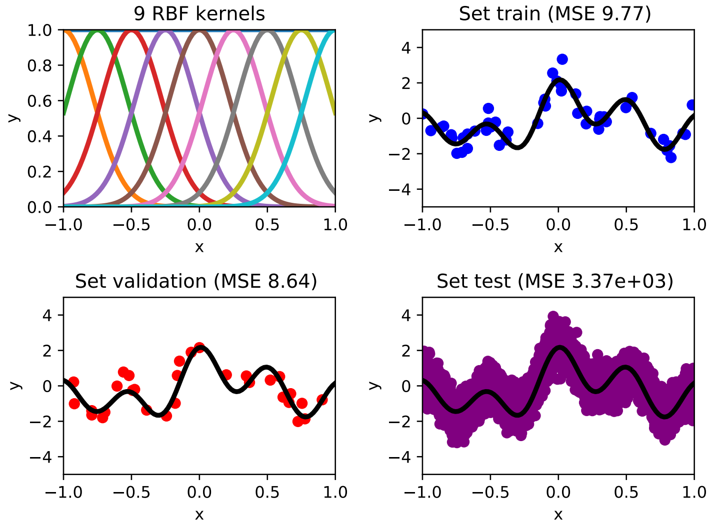


(d) $l = 22$

Figure 4: Results of Linear Regression for varying degree l.



(a) Training set, $l = 40$



(b) Validation set, $l = 9$

Figure 5: Results with lowest cost for training and validation set.

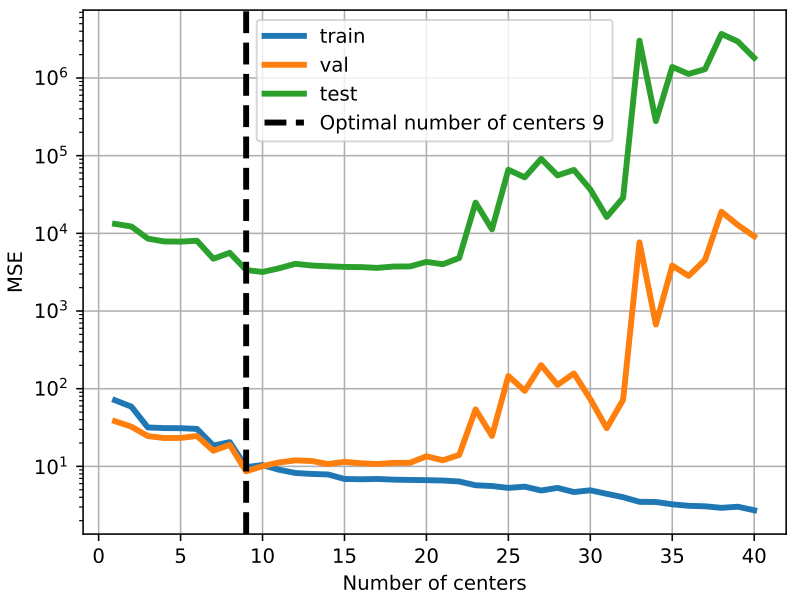


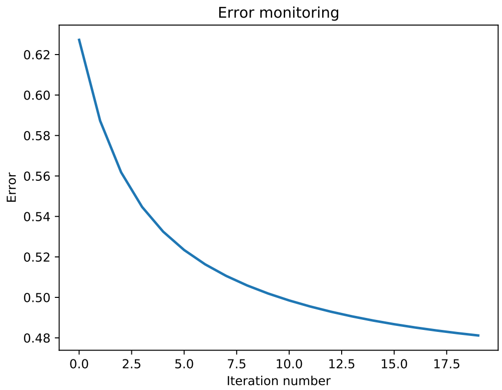
Figure 6: Training, validation and testing costs as a function of the degree l.

2 Logistic Regression

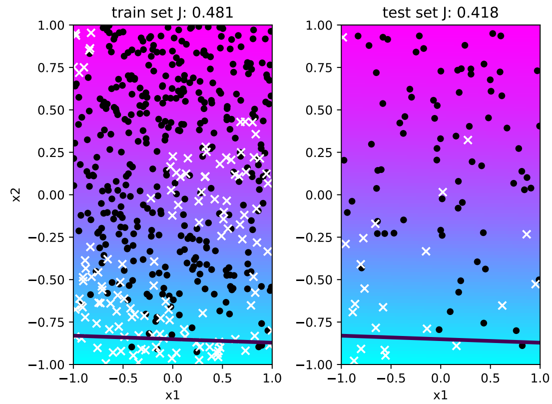
2.1 Derivation of Gradient

2.2 Logistic Regression training with gradient descent

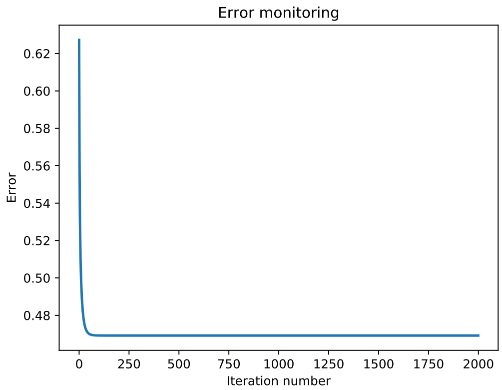
•



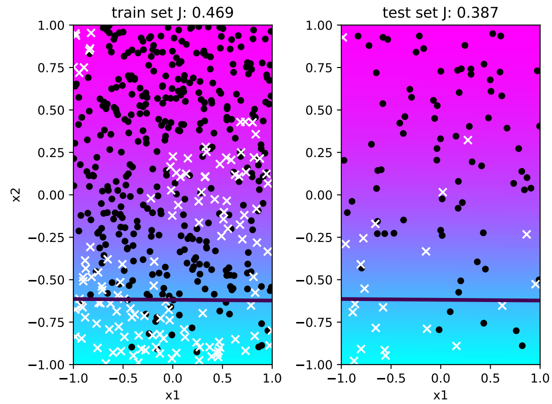
(a) Error, 20 iterations



(b) Decision boundary, 20 iterations



(c) Error, 2000 iterations



(d) Decision boundary, 2000 iterations

Figure 7: GD errors and decision boundaries for varying number of iterations, degree $l = 1$ and learning rate $\eta = 1$.

•

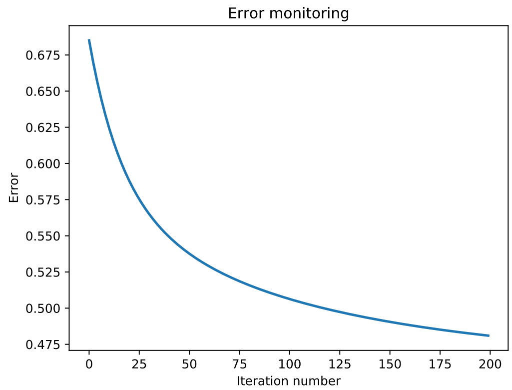
•

degree l	learning rate η	iterations	training cost	testing cost
1	1	60	0.470	0.391
2	5	75	0.450	0.354
7	10	1000	0.310	0.361
20	8	1500	0.297	0.398

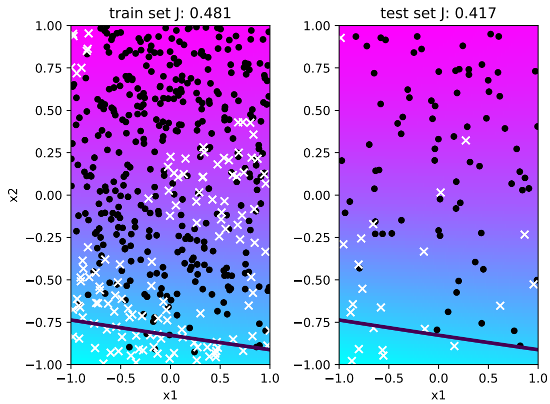
Table 1: Chosen values for learning rate η and number of iterations and obtained training and testing costs for degrees $l = 1, 2, 7, 20$.

•

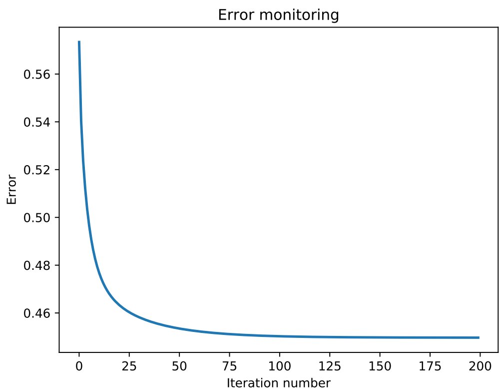
•



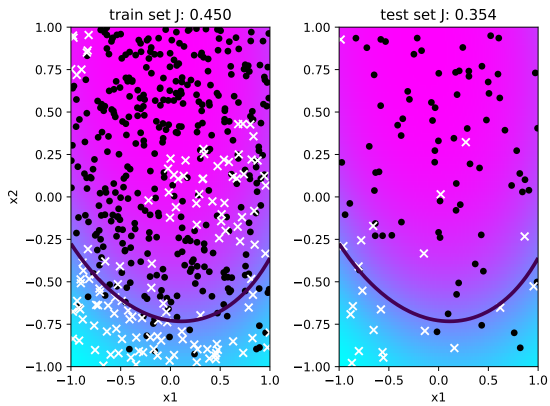
(a) Error, $\eta = 0.1$



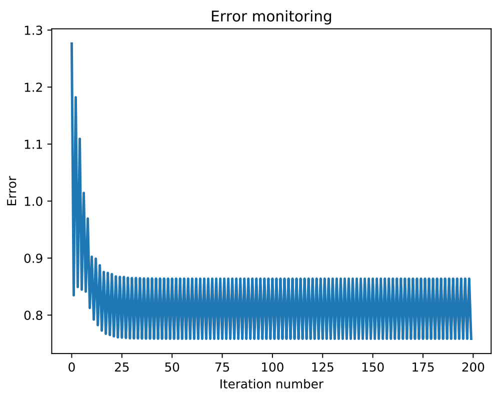
(b) Decision boundary, $\eta = 0.1$



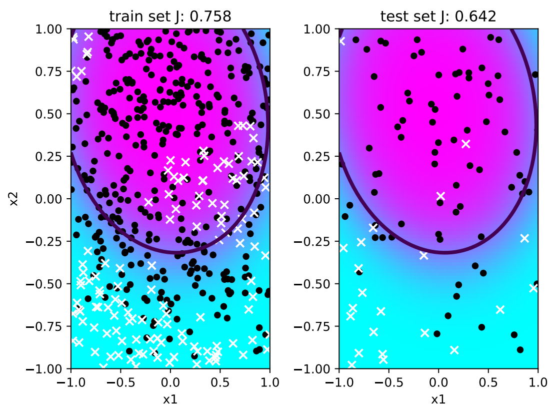
(c) Error, $\eta = 2$



(d) Decision boundary, $\eta = 2$

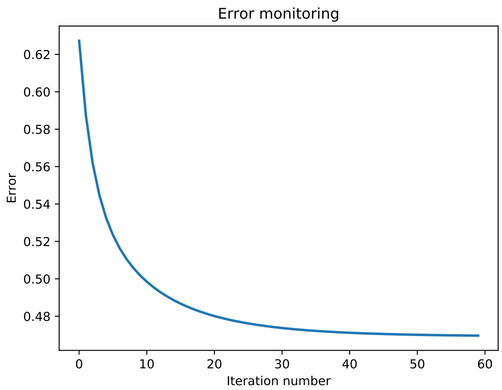


(e) Error, $\eta = 20$

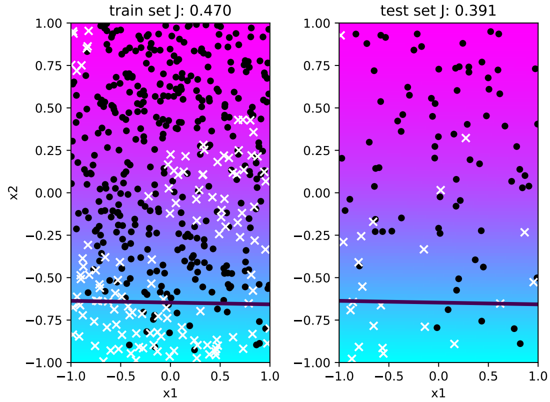


(f) Decision boundary, $\eta = 20$

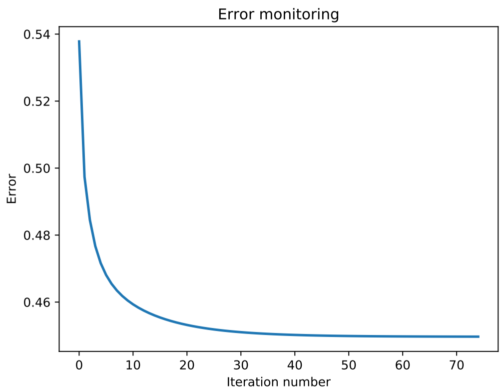
Figure 8: GD errors and decision boundaries for varying learning rate η, degree $l = 2$ and 200 iterations.



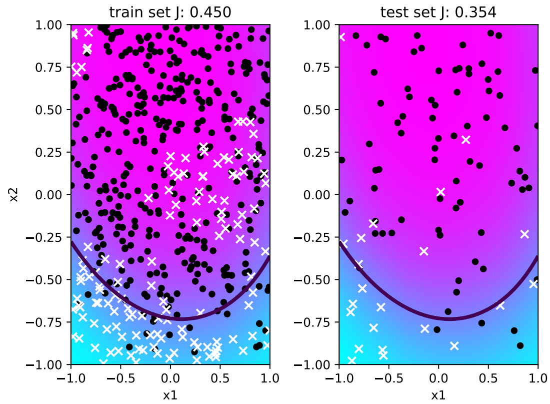
(a) Error, $l = 1$, $\eta = 1$, 60 iterations



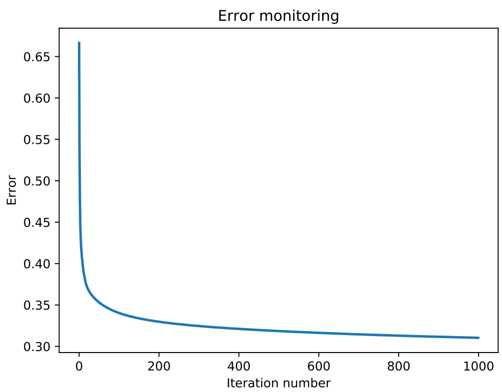
(b) Decision boundary, $l = 1$, $\eta = 1$, 60 iterations



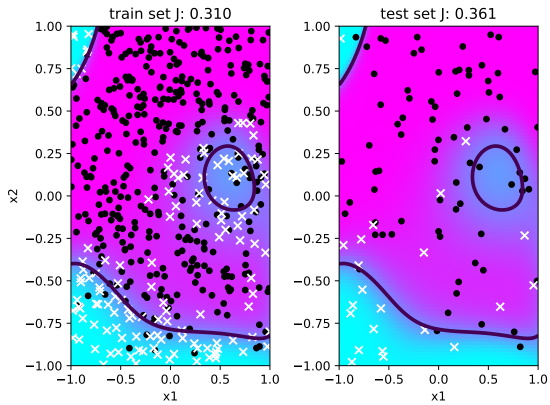
(c) Error, $l = 2$, $\eta = 5$, 75 iterations



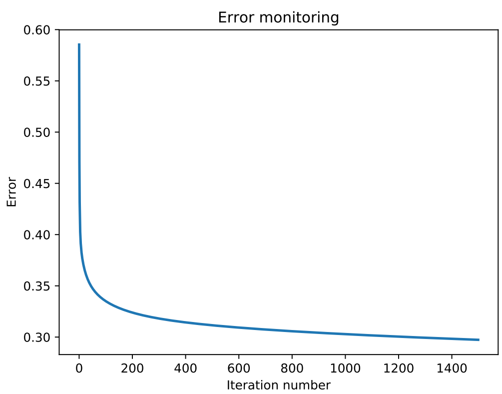
(d) Decision boundary, $l = 2$, $\eta = 5$, 75 iterations



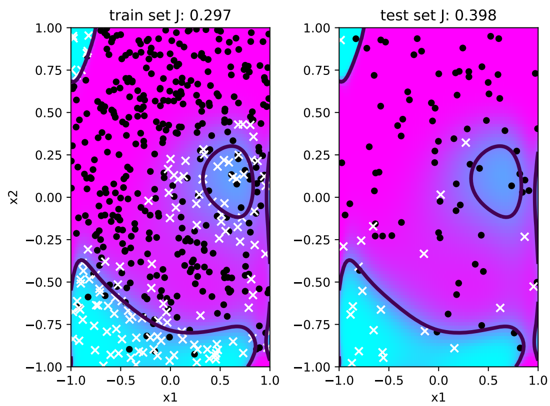
(e) Error, $l = 7$, $\eta = 10$, 1000 iterations



(f) Decision boundary, $l = 7$, $\eta = 10$, 1000 iterations



(g) Error, $l = 20$, $\eta = 8$, 1500 iterations



(h) Decision boundary, $l = 20$, $\eta = 8$, 1500 iterations

Figure 9: GD errors and decision boundaries for degrees $l = 1, 2, 7, 20$ and chosen parameter values.