DLE Homework 1
Double Descent

Vojtěch Michal, michavo3

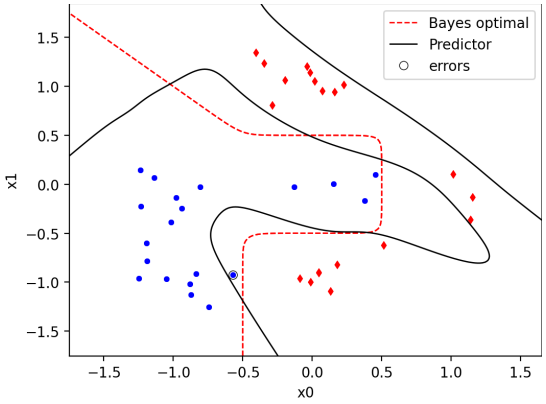

Fig. 1: $N = 40$, $D = 10$, test error 11.95 %.

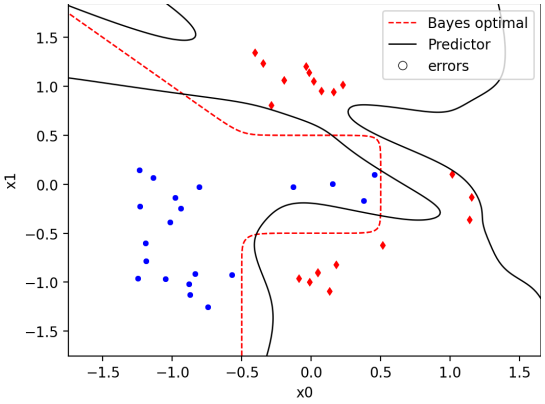

Fig. 2: $N = 40$, $D = 20$, test error 13.51 %.

***Abstract*—The assigned[1] homework strives to demonstrate the phenomenon of double descent. Classifier decision boundaries learned for various complexities of the hidden layer are investigated. Additionaly, the dependence of classification errors and losses on both the training set size and hidden layer complexity is observed.**

I. Classifier decision boundary

Figures 1 through 11 show decision boundaries for various values of hidden layer complexity D and fixed size N of the training set. It can be observed that with growing D, the learned decision boundary approaches the theoretically optimal Bayesian decision boundary corresponding to the given ground truth model. The shape of the boundary in Fig 11 matches the expected shape given in the assignment, hence verifying the correctness of implemented algorithms.

II. Fixed layer complexity

Results from experiments for the case of fixed $D = 40$ and varying training sample size N are shown in Fig. 12 – the classification error rate – and in Fig. 13 – the evaluated mean-square-error loss.

III. Fixed training set size

Results from experiments for the case of fixed $N = 40$ and varying hidden layer complexity D are shown in Fig. 14 – the classification error rate – and in Fig. 15 – the evaluated mean-square-error loss. For smoothness, all graphs are generated by averaging results of 200 iterations with same parameters.

[1]The homework assignment is available on https://cw.fel.cvut.cz/wiki/courses/bev033dle/labs/lab0_ddescent/start

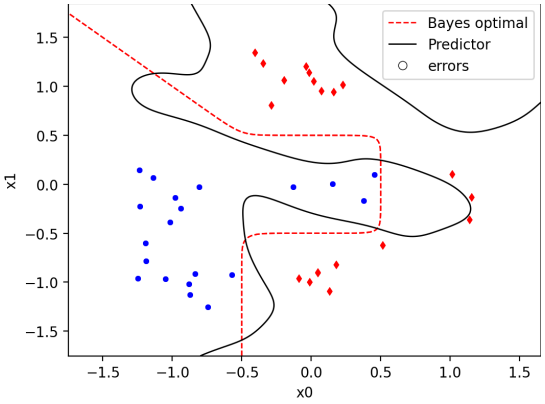

Fig. 3: $N = 40$, $D = 30$, test error 12.89 %

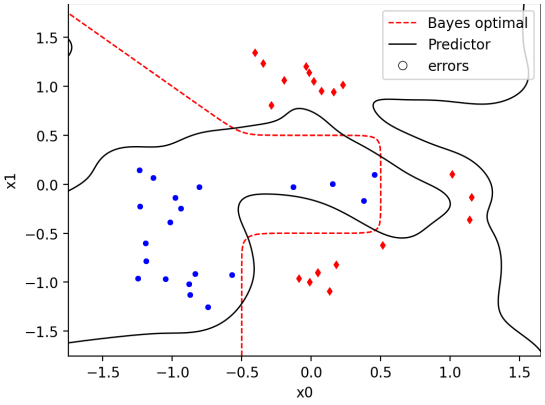

Fig. 4: $N = 40$, $D = 40$, test error 13.32 %

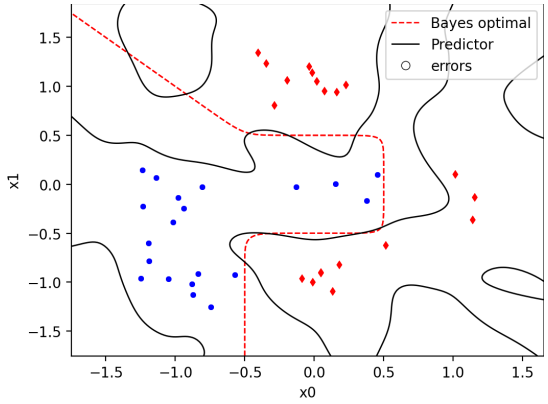

Fig. 5: $N = 40$, $D = 50$, test error 10.86 %

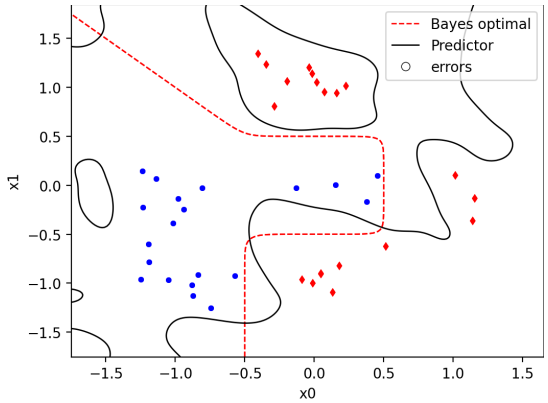

Fig. 6: $N = 40$, $D = 75$, test error 10.75 %

Results acquired by experiments confirm results presented in recommended papers. My intuitive interpretation is that when the neural net has many parameters, it may perform ”auxiliary optimizations”, e.g. regularization.

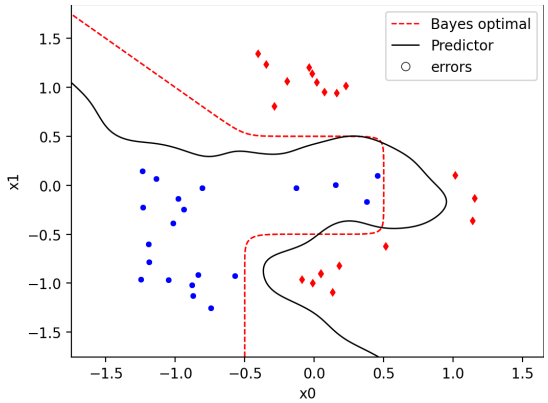

Fig. 7: $N = 40$, $D = 100$, test error 8.93 %

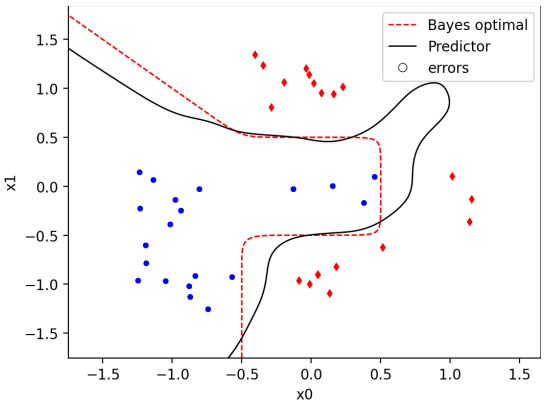

Fig. 8: $N = 40$, $D = 150$, test error 3.69 %

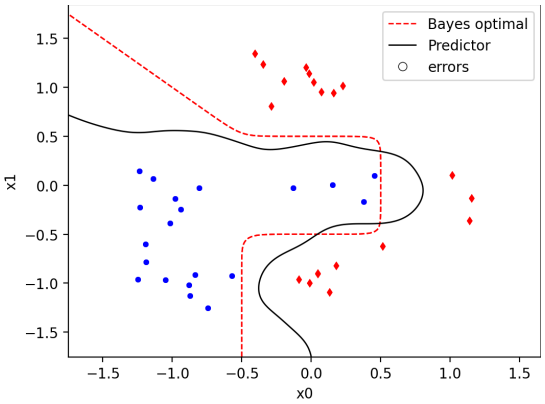

Fig. 9: $N = 40$, $D = 250$, test error 4.97 %

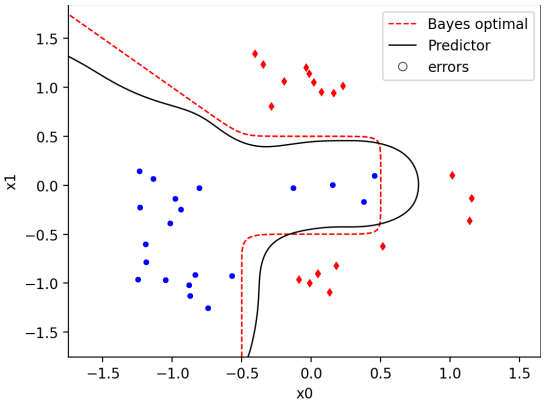

Fig. 10: $N = 40$, $D = 500$, test error 3.81 %

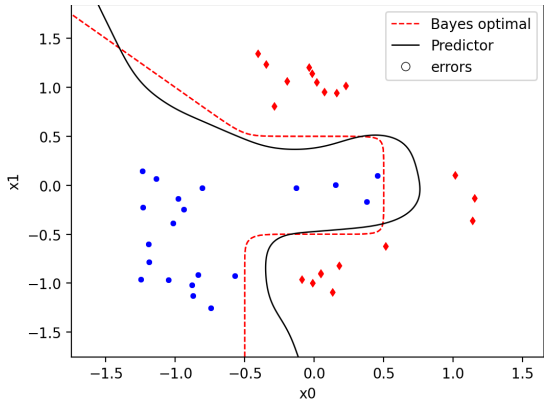

Fig. 11: $N = 40$, $D = 1000$, test error 4.25 %

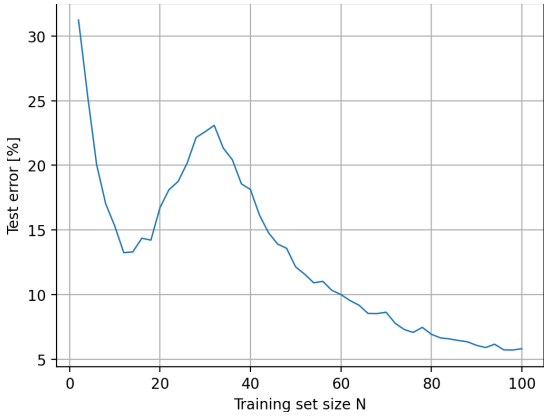

Fig. 12: Test error rate for fixed $D = 40$ and varying N.

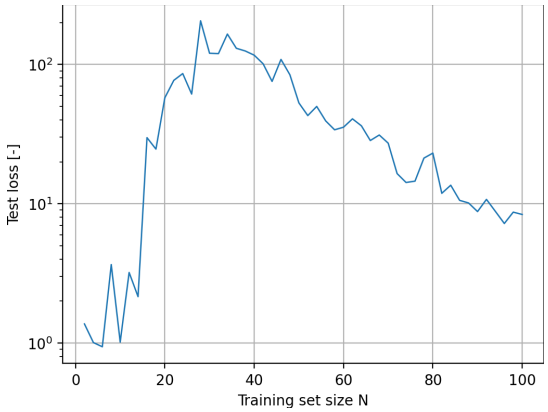

Fig. 13: Loss for fixed $D = 40$ and varying N.

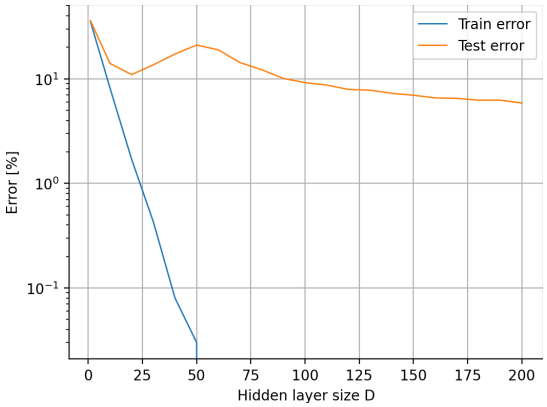

Fig. 14: Test error rate for fixed $N = 40$ and varying D.

Fig. 15: Loss for fixed $N = 40$ and varying D.