Hyperparameter tuning

Understanding validation set overfitting when tuning your hyperparameters

Team - P37

Mart Mägi
MSc student in Data Science

Musa Salamov
MA student in Innovation and Technology Management

Kaja Jakobson
MSc student in Conversion Master in IT

Project owner

Viacheslav Komisarenko
Ph.D. student and Junior Research Fellow in Machine Learning at UT

The problem

Hyperparameter tuning:

- goal is to achieve an optimal model
- involves dozens of evaluations on the validation set:
    - noise from random data gets included in model
    - resulting in the best hyperparameters being overfitted on the validation set

Explore the limitation of validation set usage for hyperparameter tuning

Initial expectations

Theoretical background

hyperparameter optimization can lead to overfitting;

Empirical study

Setting test environment; Performing hyperparameter optimization; Interpreting results;

Outcomes

Better understanding of the overfitting which is caused by hyperparameter tuning; possible strategies that mitigate/overcome the problem

Empirical study

The test environment

Artificially generated data

Number of futures: 10 vs 20;
Number of instances:
1000 vs 10000

Hyperparameter tuning

Automated:
GridSearchCV
Manually: for 2 hyperparameters

Classifier

Logistic regression

Hyperparameters for LR

1. C : float, default: 1.0 : Inverse of regularization strength; must be a positive float. Like in support vector machines, smaller values specify stronger regularization.
2. tol : float, default: 1e-4 : Tolerance for stopping criteria. This tells the algorithm to stop searching for a minimum (or maximum) once some tolerance is achieved, i.e. once it is close enough.
3. solver : {‘newton-cg’, ‘lbfgs’, ‘liblinear’, ‘sag’, ‘saga’}, default: ‘liblinear’ Algorithm to use in the optimization problem.
4. max_iter : int, default: 100 : Useful only for the newton-cg, sag and lbfgs solvers. Maximum number of iterations taken for the solvers to converge

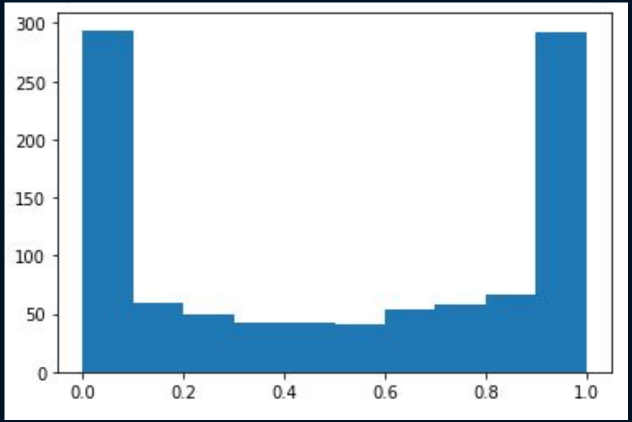
Solver selection

	Sl	C	Train_acc_liblinear	Val_acc_liblinear	Build_time_liblinear	Train_acc_newton-cg	Val_acc_newton-cg	Build_time_newton-cg	Train_acc_lbfgs	Val_acc_lbfgs	Build_time_lbfgs
1.0	1	0.001	82.835821	83.030303	0.001768	69.402985	69.090909	0.006571	69.402985	69.090909	0.004296
2.0	2	0.006	82.686567	83.636364	0.001376	76.119403	76.969697	0.006133	76.119403	76.969697	0.003330
3.0	3	0.011	82.686567	83.333333	0.001390	79.104478	79.696970	0.005732	79.104478	79.696970	0.002957
4.0	4	0.016	82.985075	83.636364	0.001373	80.895522	81.818182	0.006945	80.895522	81.818182	0.003080
5.0	5	0.021	83.283582	83.636364	0.001444	81.194030	83.030303	0.007570	81.194030	83.030303	0.003458
...	...	...	...	...	...	...	...	...	...	...	...
996.0	996	4.976	83.134328	84.242424	0.001697	82.985075	84.545455	0.007992	82.985075	84.545455	0.005040
997.0	997	4.981	83.134328	84.242424	0.001753	82.985075	84.545455	0.008276	82.985075	84.545455	0.004354
998.0	998	4.986	83.134328	84.242424	0.001744	82.985075	84.545455	0.007393	82.985075	84.545455	0.003890
999.0	999	4.991	83.134328	84.242424	0.001721	82.985075	84.545455	0.007218	82.985075	84.545455	0.003865
1000.0	1000	4.996	83.134328	84.242424	0.001749	82.985075	84.545455	0.007863	82.985075	84.545455	0.004247

1000 rows × 17 columns

- Best result: liblinear

Relatively small data



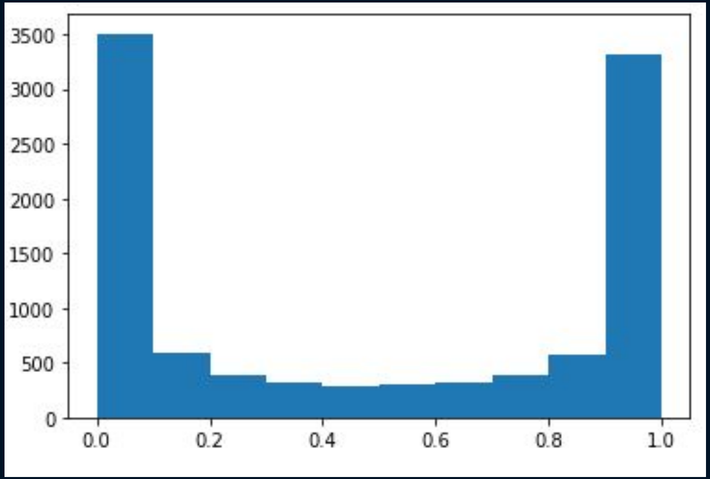
- Number of instances: 1000
- Number of features: 10

	Sl	C	Train_acc	Val_acc	Build_time
1.0	1	0.001	77.761194	73.333333	0.003977
2.0	2	0.006	86.268657	87.272727	0.004561
3.0	3	0.011	86.716418	87.272727	0.003087
4.0	4	0.016	86.865672	86.969697	0.003844
5.0	5	0.021	86.716418	86.969697	0.003019
...	...	...	...	...	...
996.0	996	4.976	87.014925	88.181818	0.004034
997.0	997	4.981	87.014925	88.181818	0.003875
998.0	998	4.986	87.014925	88.181818	0.003854
999.0	999	4.991	87.014925	88.181818	0.005348
1000.0	1000	4.996	87.014925	88.181818	0.005487

1000 rows × 5 columns

- Initial val_acc: 73.33%
- Max val_acc: 88.18%

Relatively big data



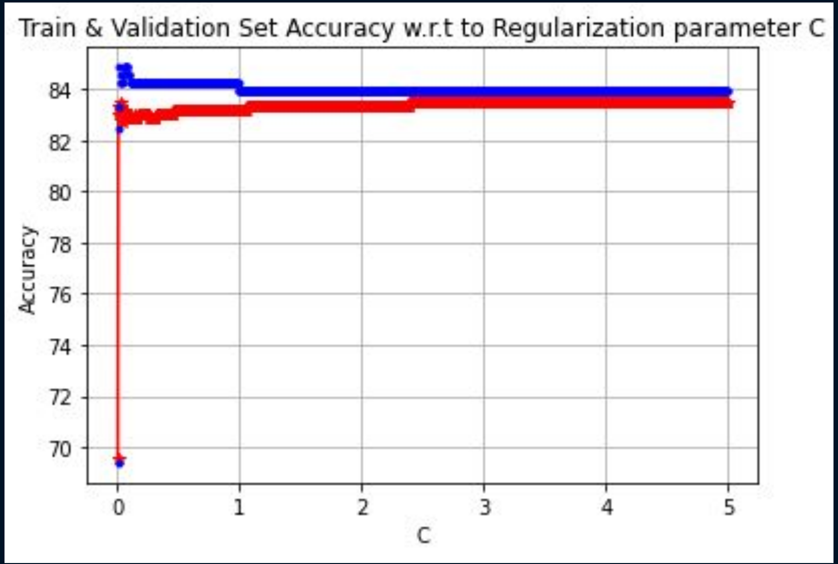
- Number of instances: 10000
- Number of features: 20

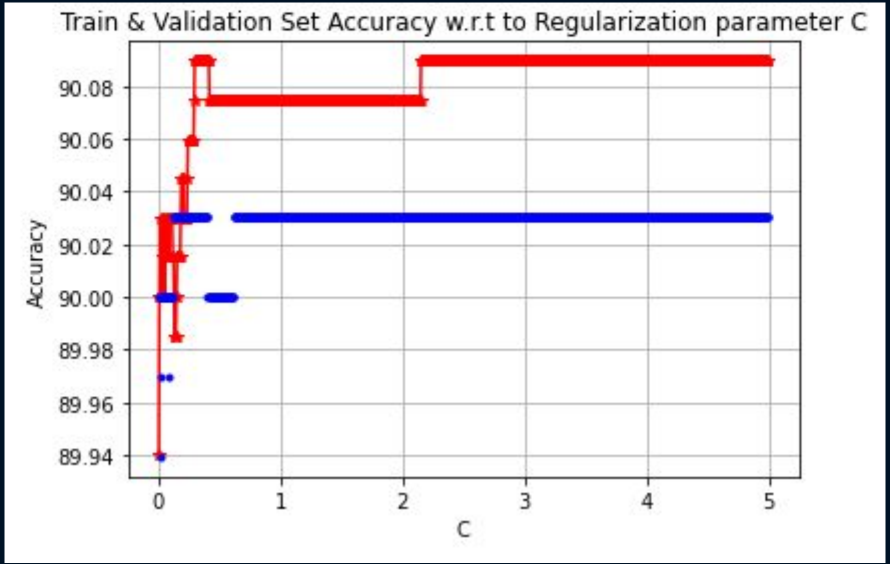
	Sl	C	Train_acc	Val_acc	Build_time
1.0	1	0.001	89.940299	90.000000	0.020243
2.0	2	0.006	90.000000	90.000000	0.023894
3.0	3	0.011	90.029851	89.939394	0.023087
4.0	4	0.016	90.000000	90.000000	0.015033
5.0	5	0.021	90.000000	89.969697	0.028266
...	...	...	...	...	...
996.0	996	4.976	90.089552	90.030303	0.019288
997.0	997	4.981	90.089552	90.030303	0.019595
998.0	998	4.986	90.089552	90.030303	0.019161
999.0	999	4.991	90.089552	90.030303	0.019998
1000.0	1000	4.996	90.089552	90.030303	0.018911

1000 rows × 5 columns

- Initial val_acc: 90%
- Max val_acc: 90.03%

Comparison of the results

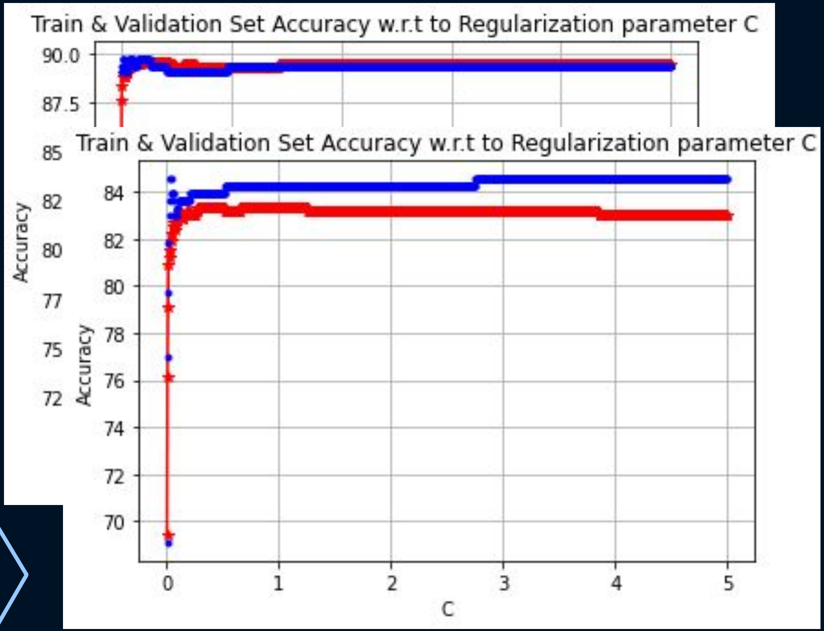


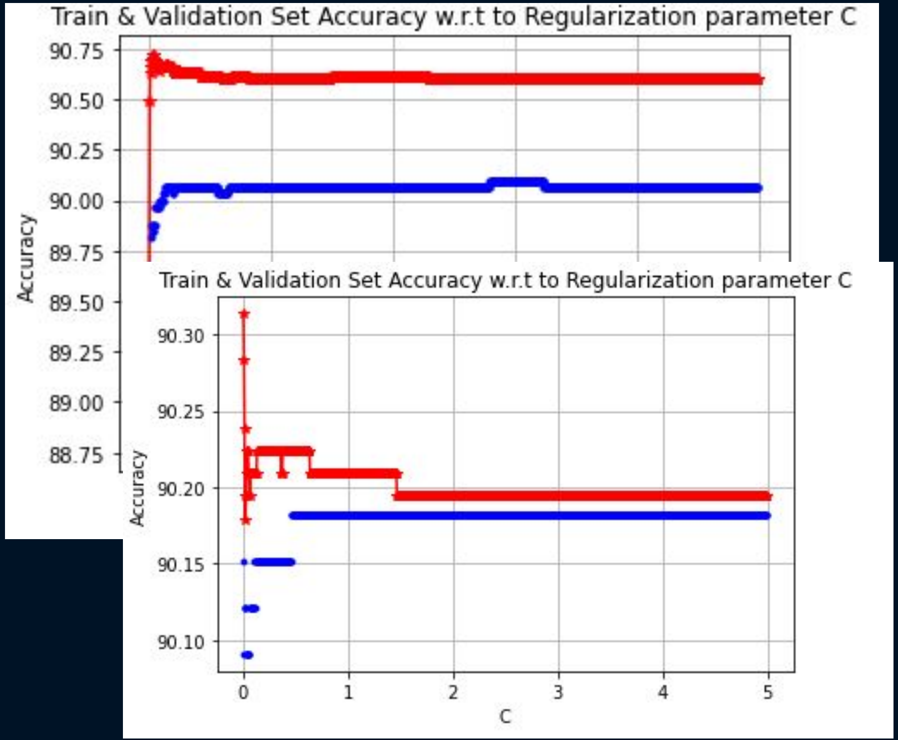


- Relatively small data
- Validation set accuracy is higher than train set accuracy over the iterations

- Relatively big data
- For almost every iteration, train set accuracy is higher than validation set accuracy

Other experiment results





Outcomes

Hyperparameter tuning may lead into overfitting on validation set; depends on:

- size of the data
- classifier
- hyperparameters of the classifier

Possible mitigation strategies:

- acquiring more data
- cross-validation
    - train-validation-**test**

Lessons learnt

- The importance of the data quality
- Complexity of computational time
    - Classifier
    - Data
- The importance of hyperparameters

Thanks!

Do you have any question?

References

1. Benner, J. , Cross-Validation and Hyperparameter Tuning: How to Optimise your Machine Learning Model, Aug 6, 2020 https://towardsdatascience.com/cross-validation-and-hyperparameter-tuning-how-to-optimise-your-machine-learning-model-13f005af9d7d
2. Bhadauriya, R., Cross-Validation(CV) and Hyper-Parameter Tuning, Sept 23 https://medium.com/@creatrohit9/cross-validation-cv-and-hyper-parameter-tuning-5db4d209820d
3. Prashant Bhardwaj Cross-Validation and Hyperparameter Tuning, Apr 8 https://medium.com/almabetter/cross-validation-and-hyperparameter-tuning-91626c757428
4. Gorodetski, M., Hyperparameter Tuning Methods - Grid, Random or Bayesian Search?, Apr 28 https://towardsdatascience.com/bayesian-optimization-for-hyperparameter-tuning-how-and-why-655b0ee0b399
5. Koehrsen, w., A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning, Jun 24, 2018 https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f
6. Dewancker, I., McCourt, M., Clark, S., Bayesian Optimization Primer https://static.sigopt.com/b/20a144d208ef255d3b981ce419667ec25d8412e2/static/pdf/SigOpt_Bayesian_Optimization_Primer.pdf
7. Neural Network Hyperparameter Tuning using Bayesian Optimization, 30/11/2021 https://analyticsindiamag.com/neural-network-hyperparameter-tuning-using-bayesian-optimization/
8. Strategy for Deep Learning Hyper-parameter Tuning: Bayesian Optimization, Oct 20 https://medium.com/@tzjy/strategy-for-deep-learning-hyper-parameter-tuning-bayesian-optimization-75a524bdaf0d