

Lasso

UNDERSTANDING: Can you explain the effect that Lasso has on the weights and which regularizer you would choose to get the best results possible?

ANSWER: Lasso leads to very sparse weights. Sparsity means that many of the coefficients of the weights w_i are zero. This causes less computation and makes the solution more interpretable. The best regularizer for Lasso is the L_1 -regularizer, since it is a regularizer that provides sparse weights and still remains convex.

TRANSFERRING:

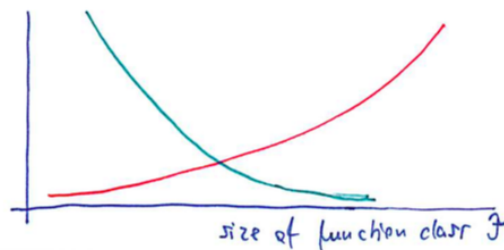
Ridge Regression

UNDERSTANDING: Assume you have a test dataset with small variances between the testpoints and you expect that the distribution of this data follows a complex polynomial. How would you choose the parameter λ ?

ANSWER: You would have to choose a small λ , since λ is a regularization parameter and therefore it punishes complex functions harder than less complex functions. But since you assume that your data is following a complex polynomial, you would want to include also complex polynomials and therefore you would need them to not be punished that hard.

ERM and RRM framework

UNDERSTANDING:



In the above illustration which curve represents the approximation error and which represents the estimation error? Explain why!

ANSWER: The approximation error is the discrepancy between your best function in a certain Function class and the Bayes classifier. If the Function class gets bigger than the approximation error gets smaller, because it is more likely the true best function, namely the Bayes classifier lies within your Function class. Therefore the green curve shows the approximation error. On the other side the estimation error is the discrepancy between the function learned by your algorithm and the best function within the Function class. Therefore it gets bigger if the Function class gets bigger, too. This is why the red curve shows the estimation error.