

References

- [1] A. Niculescu-Mizil and R. Caruana, “Predicting good probabilities with supervised learning.” in *Proceedings of the 22 nd International Conference on Machine Learning*, 2005, pp. 625 – 632, two questions arise: where does the sigmoid train set come from? and how to avoid overfitting to this training set? If we use the same data set that was used to train the model we want to calibrate, we introduce unwanted bias. For example, if the model learns to discriminate the train set perfectly and orders all the negative examples before the positive examples, then the sigmoid transformation will output just a 0,1 function. So we need to use an independent calibration set in order to get good posterior probabilities. This, however, is not a draw back, since the same set can be used for model and parameter selection. To avoid overfitting to the sigmoid train set, an out-ofsample model is used.