Evaluation

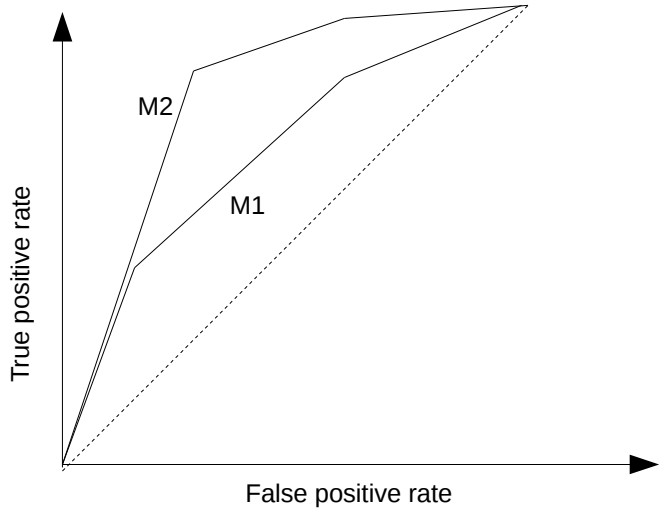
1. A classifier was trained and tuned on a training set until the error was 0. Then a new dataset, a test set, was used to evaluate the performance of the learned model, giving a significant error. How can the error on the test set be high, if the error on the training set was 0? What measures can be taken to prevent this phenomenon and improve generalization to unseen data?

2. What is the role of the validation set as opposed to the test set?

3. What is cross-validation? What is it used for?

4. A model can distinguish between points of 5 classes. What is the most appropriate single representation of its accuracy?

5. Given the two ROC curves below, which model is to be preferred? Justify your answer:



6. A dataset has 500 points of class A, and 25 points of class B, which metric is most appropriate to represent its accuracy? Justify your answer.