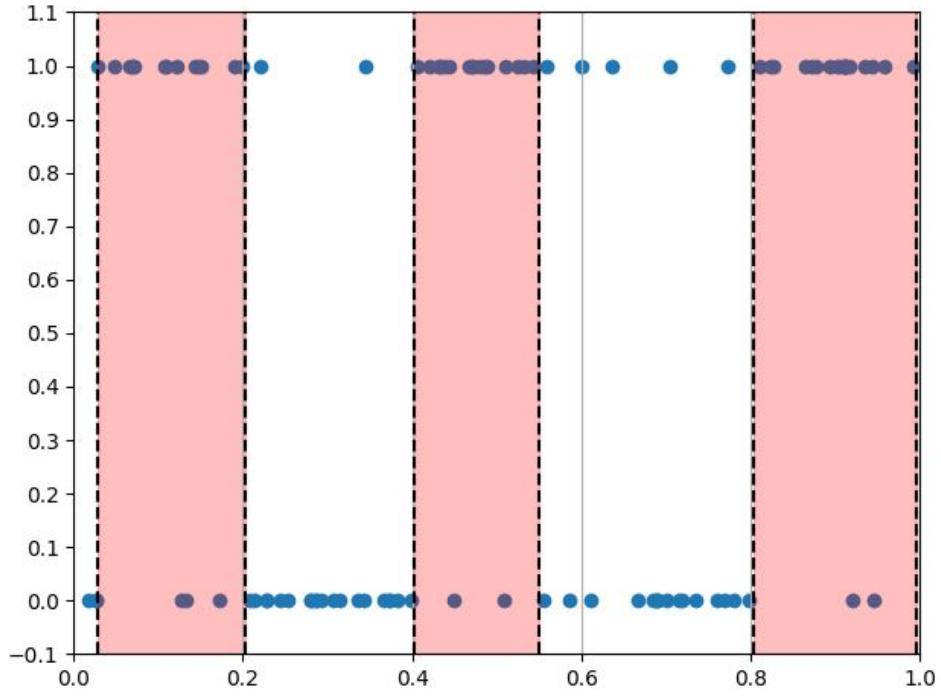
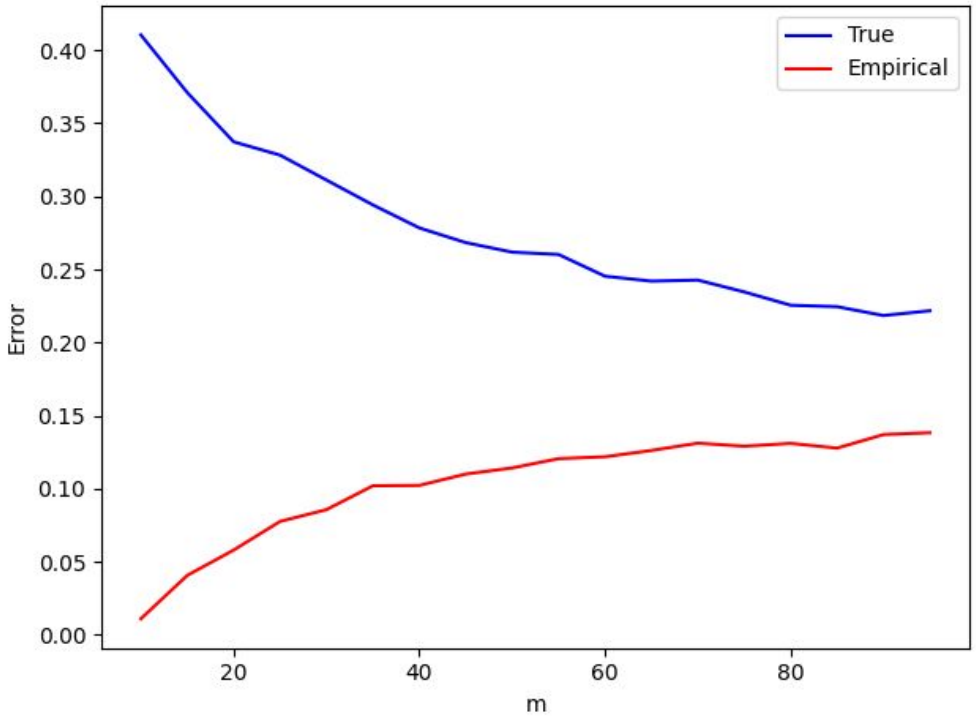
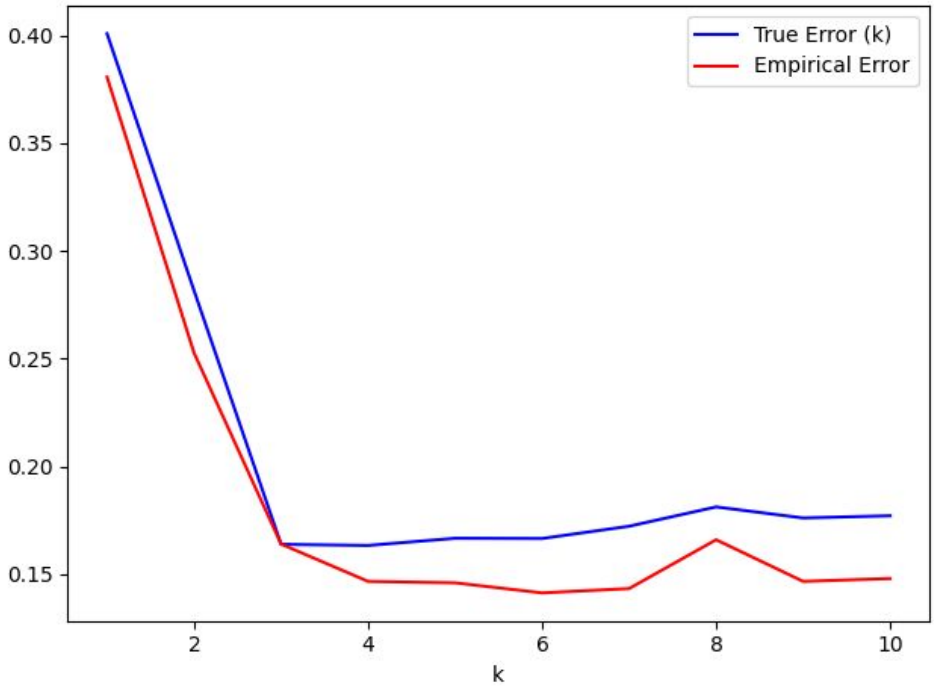
1) The red vertical stripes are the intervals provided by the find_best_interval:



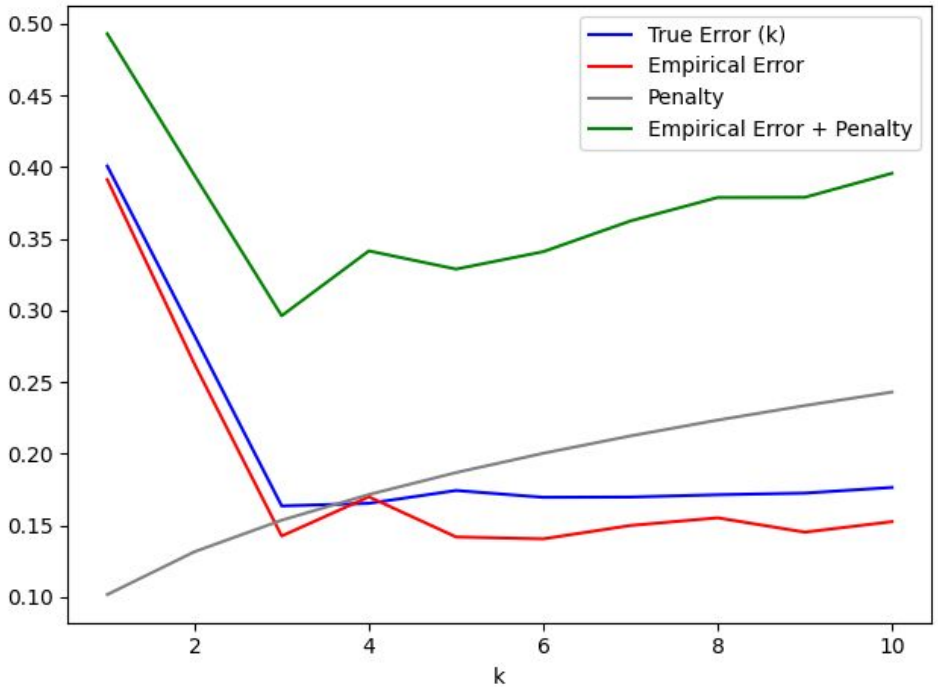
2) The best hypothesis is: h(x) = 1 if x is in one of the intervals {(0, 0.2), (0.4, 0.6), (0.8, 1)}, 0 else
3) As seen in the following chart the true error decreases as m increases which makes sense for two reasons:
    a) The true error should converge to its minimum as m goes to infinity, while the empirical error should converge to the true error
    b) Empirical error is smaller for few examples since it's easy to train a model based on very little examples (for example if m≈k)



4) k* intervals doesn't necessarily constitute a good hypothesis, it may cause overfitting. This explains why k* is 6 but the best k in terms of true error is 3



5) Here we can clearly see the expected result - adding the penalty function causes the best k to be 3. Without the penalty function it's still 6



6) In all 3 tests the best hypothesis was based on k=3. So to find the optimal hypothesis itself all we needed is to run find_best_intervals with k=3 and a big enough sample