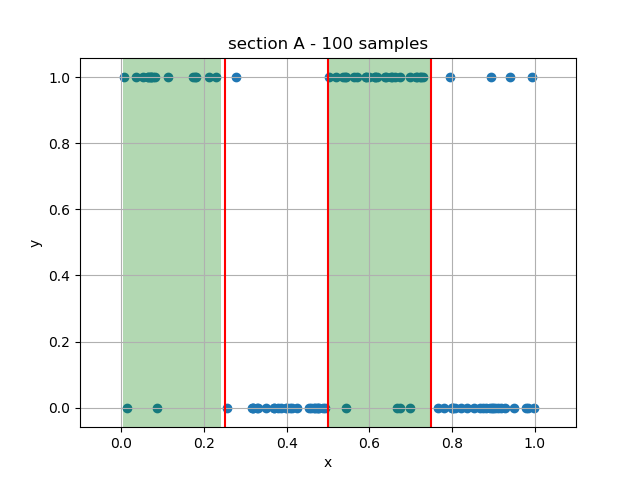
Assignment #2 – Programming Part

**Question A**



**Question B**

We’ve seen in Lecture 1 that the optimal classifier for Zero-One Loss is the **maximum-aposteriori (MAP)**.

That yields the hypothesis:

The true error in this case is:

**Question C**

****

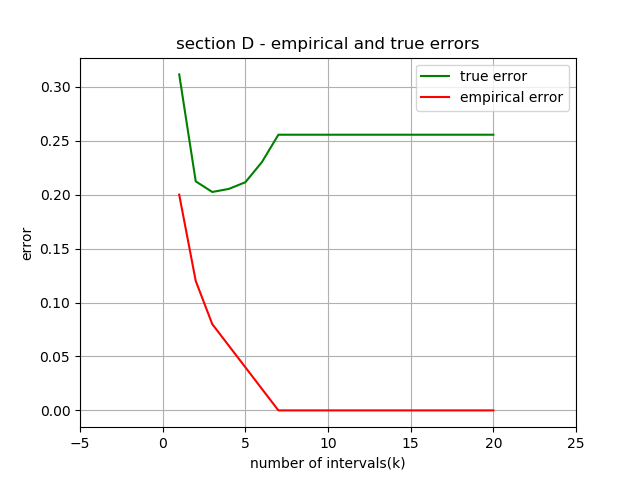
The plot implies that -

* The Empirical Error increases as the number of samples grows. This result stays true to the Law of Large Numbers: The true error represents an expectation value, while the empirical error is an average of losses. As we sum more and more samples, the average should converge to the expectation as we approach infinity. Since the true error is higher than the empirical error, the empirical error increases as it gets closer to the true error.

This result can also be attributed to the structure of the data distribution: the more samples we have, the larger the probability that we “hit an opposite label” becomes (recall: the hypothesis will always choose 1 or 0 for a given interval, but in reality there is a probability that this label is wrong as defined for the ground truth distribution). Considering the fact that the Loss function is a zero-one loss, we accumulate more errors (albeit on average) as we use more samples.

* The True Error decreases as the number of samples grows, as the ERM algorithm gives better hypothesizes that converge to the optimal MAP hypothesis as proved in Question B (notice how the green line approaches 0.15 as we add more samples, which is the best true error value we can get, for the MAP classifier). This is simply due to the fact that the more training data we have, the more accurate interval boundaries the hypothesis can determine: otherwise we have areas of “uncertainty” as we don’t have enough data to determine the precision of the intervals.

**Question D**

****

The plot implies that:

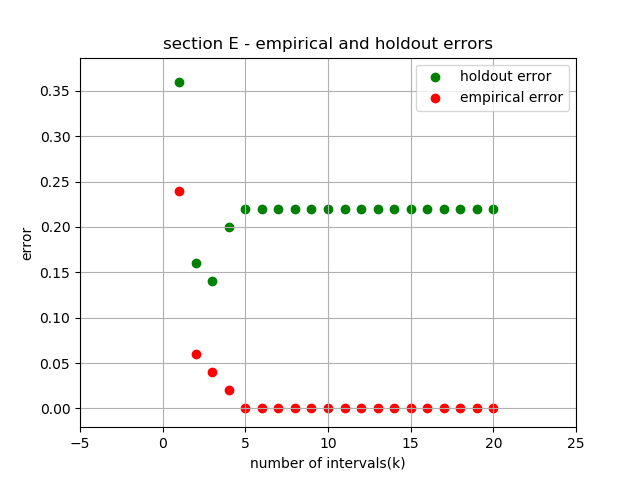
* The empirical error is continuously decreasing. This is due to the fact that the hypothesis class is more expressive and can describe the dataset more accurately.
* The true error decreases but at some point starts to increase. This is because at this point the variance component is increasing faster than the bias term (the empirical error) decreases:

describes the variance, describes the bias and both depend on |H| in an inverse manner).

All the results match the theory we learned about model selection.

In our case K\* is 7 (actually 7 to 20). As we can see from the graph, if we want to minimize the **true** error an hypothesis with K\* intervals is not the best choice (which is exactly the case of overfitting). It has a true error greater than the true error of an hypothesis with 3 intervals for example.

**Question E**



As we can see the holdout error behaves more like the true error.

K=3 has the lowest holdout error so we will choose it as best hypothesis class.

After this we can find an hypothesis from this class that minimizes the empirical error on both train and test datasets together. This is because, as we saw in subsection C, the true error decreases as we train on more samples.