What is the role of the number of training points to accuracy?

It can be easily seen that the amount of data used has an effect on the accuracy. Lower amounts of training points result in a lower accuracy rate, but as you approach larger amount these improvements become negligible and only add to the computational cost.

e.g.

Points of data	Accuracy (%)
100	67
500	83.1
1000	87.5
2000	90.94
4000	93.43
10000	95.44

What is the role of *k* to accuracy?

Building off our answer to the first question, I constructed this table to better display what happens as k changes:

k	Points of data	Accuracy (%)
1	100	69.23
	500	84.50
	1000	88.20
5	100	67.00
	500	83.10
	1000	87.50
10	100	55.58
	500	77.80
	1000	84.11

As k gets larger, the accuracy starkly drops which is understandable considering that a bigger selection of neighbours will dilute the accuracy. (Consider the demographics of a state vs. the demographic of a neighbourhood in a city of that state; you can more accurately guess the demographic of a neighbourhood as opposed to the wider demographic of the entire state by narrowing the scope of your guess to that specific neighbourhood)

What numbers get confused with each other most easily?

By looking at the confusion matrix, we look for the largest values outside of the diagonal (because of course 2 is going to look like 2 the most!) to spot the numbers that got confused the most. 9 and 4 got confused the most, followed by 7 and 2 (with half as many incidences), 7 and 9 and lastly, 5 and 3.