MP 4

a.2 Multiplying all probabilities together might result in underflow.

a.6
$k = 1 \times 10^{-10}$, because it results in higher accuracy
The result is reasonably good, because training on 5000 samples gives around 80% accuracy

b.1
three set of features:
1. aspect ratio of the number
2. number of empty spaces in the number
3. number of "+" and "#" in the number

Performance:

Sample size	Accuracy basic features	advanced features	final features
10%	0.722	0.629	0.755
20%	0.785	0.65	0.808
30%	0.802	0.65	0.819
40%	0.804	0.659	0.817
50%	0.812	0.658	0.837
60%	0.814	0.66	0.838
70%	0.811	0.666	0.841
80%	0.825	0.664	0.843
90%	0.823	0.658	0.843
100%	0.823	0.665	0.841