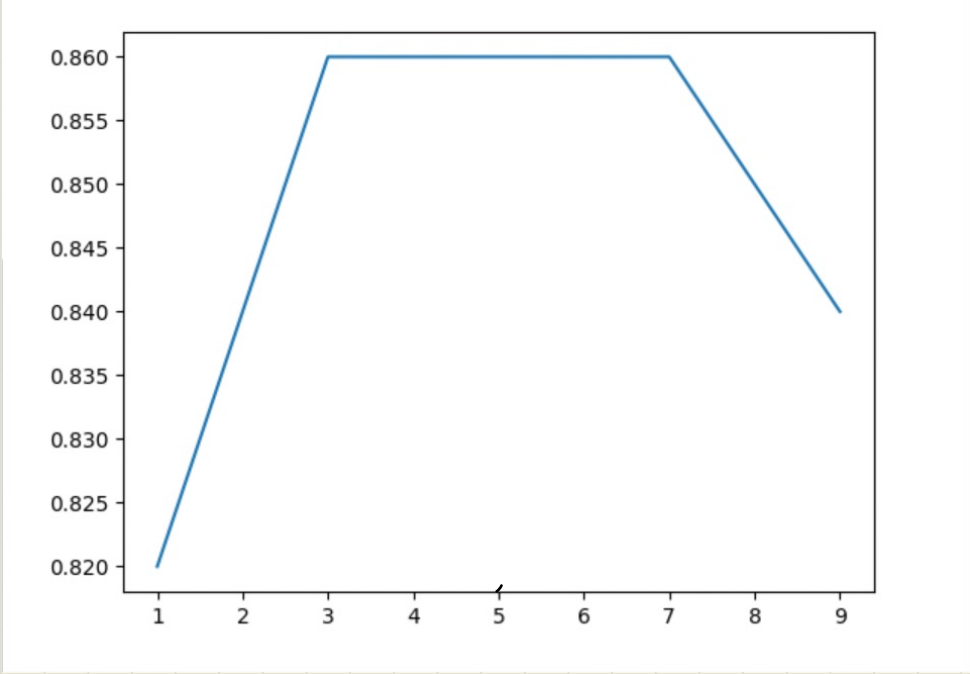
Q3



I would chose $k^* = 5$

rates for k^*-2, k^*+2. are both 86%

When k is too small, prediction is under-fit and not that accuary

when k is too large. prediction is over-fit

not that accuracy

(b) I found learning-rate = 0.2

iteration = 300 work best

(c) please see q2_2_entropy.jpg

I would choose less iteration since it's already converge