

Ensemble methods introduction*(find and work with a partner)*

1. Concern: will **bootstrap samples** all be basically the same? To address this, let's compute the following probability. For data set of size n , what is the probability that a given example will *not* be selected in a bootstrap sample?

- (a) probability that an example is not chosen the first time =
 (b) probability that the example is not chosen any of n times =
 (c) for large n

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n.$$

So as n grows large, the answer to (b) converges to =

- (d) Therefore on average, a bootstrap sample contains what percent of training examples?

Credit: Jessica Wu

2. **Bagging**: reducing the error rate. Say I'm running a bagging algorithm with $T = 3$ classifiers on a test example where the true label is $y_{\text{test}} = 1$. Each individual classifier has $r = \frac{1}{4}$ chance of being wrong (which means predicting a 0 in this case).

$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

- (a) The table above shows all possible ways the 3 classifiers could classify this test example. For which rows is the *overall* classification wrong? (i.e. more than $T/2$ classifiers incorrect)
- (b) For each row you just identified, what is the *probability* of such a configuration of votes? (Hint: use r .)
- (c) Add up all these probabilities to find the overall probability of an incorrect classification. Compare this number to r . Has bagging made an improvement?