1. (a)

**Expectation value of Z**

since X and Y are independent univariate random variables, we know that:

To calculate , using the *Law of unconscious statistician* we can tell that

where is the density function. For a uniform distribution in [0, 1],

Where [a, b] represents the interval for the distribution, which is [0, 1] in our case. Then we can continue with our calculation,

Since X and Y are two independent identical distributions, we know

Finally ,

**Variance of Z**:

Lastly,

1. (b) Given that is the dimension for , and using and we computed in 1. (a), we have:

**Expectation:**

**Variance:**

Since each is independent from all , , for all We have:

1. (c) The maximum Euclidean distance in d-dimensional space can be represented as the distance between point and , which will be:

=

Therefore the squared maximum Euclidean distance will simply be .

From 1. (b) we know that the variance of squared Euclidean distance between two points in unit cube with dimensions is , therefore which is proportional towhere as the expectational value is proportional to d. From this we can conclude that in higher dimensions, grows slower than , and is relatively smaller than , thus “most points are approximately the same distance”. On the contrast grows linearly with d and as d gets larger, gets relatively large, thus “most points are far away”.

1. (b): This is the performance from the function select\_model(), using max\_depth of (5, 10, 20, 50, 100) and split criteria of information gain and gini-coefficient. Performances of all models have accuracy rate of 70% to 80%, with the best performance 80% from the classifier with 100 layer and information-gain criteria.

A screenshot of a social media post

Description automatically generated

1. (c): This is the visualization from the first two layers of the decision tree, using the classifier with 100 decision-tree layer and information-gain split criteria.

A close up of a device

Description automatically generated

2. (d): The is the output from the compute\_information\_gain() function. The five different words chosen to calculate information gain are [“the”, “if”, “Clinton”, “changed”, “trade”], with “the” being the top-most word from the decision tree in 2 (c) , the the rest of the words randomly chosen from the data set.

A screenshot of a social media post

Description automatically generated