problem1

# Description

k-nearest-neighbor estimates는 Linear model보다 덜 stable하지만 Bias가 작은 예측이다. Training data가 충분히 클 경우, k-near-neighbor averaging을 통해 optimal conditional expectation을 구할 수 있다고 생각할 수 있다. 그러나 “차원의 저주”에서 알 수 있듯, 차원이 높은 경우 이 방법은 적절하지 않다.

개 데이터 관측치가 원점을 중심으로 하는 차원 단위구에 균등하게 분포한다고 가정하자. 원점으로부터 가장 가깝게 이웃하는 점을 추정하고자 한다. 원점으로부터 가장 가까운 관측치까지 median 거리는 다음과 같이 표현할 수 있다.

예를 들어 차원인 원의 인 sample을 가정하면, 원점으로부터 가장 가까운 관측치까지의 거리를 추정했을 때 median값이 약 로 비교적 가깝다고 할 수 있다. 하지만 차원 구를 가정한다면, 이 값은 약 이다. 즉, 원점에서 가장 가까운 점이 구의 반지름의 절반보다도 먼 거리에 위치한다는 것을 알 수 있다. 따라서 이 점은 더 이상 원점과 가깝다고 할 수 없으며, 원점과 이웃하는 k개의 점에 의한 평균으로 추정한 값은 원점에서의 추정치로 사용하기에 적절하지 않다. 이러한 현상은 작은 sample일수록 더 심하게 나타난다.

위에서 사용한 원점으로부터 가장 가까운 관측치까지 median 거리 공식을 유도해보자.

# result

원점으로부터 각 점의 거리는 확률변수이고, 각 거리를 라고 하자. 의 분포를 알면 median 거리를 추정할 수 있다. 의 분포를 찾기 위해 어떤 점의 원점으로부터 거리가 보다 작을 확률을 구해본다.

개의 점이 차원 단위구에 균등하게 분포한다는 사실을 거리와 관련하여 생각해보자. 단위구 내의 일정한 거리를 연결한 새로운 내부 구를 생각했을 때, 내부 구에 분포된 점의 개수는 내부 구의 부피가 차지하는 비율에 비례한다.

반지름이 인 차원 구의 부피는

이다.

따라서, 차원 단위구에서 어떤 점의 원점으로부터 거리가 보다 작을 확률은 다음과 같이 표현할 수 있다.

원점으로부터 거리가 가장 가까운 점은 중 값이 가장 작은 점, 즉 거리가 인 점이다. 각 점의 거리가 서로 독립이면, 순서통계량 개념을 이용하여 원점에서 가장 가까운 점의 원점으로부터의 거리에 대한 분포를 구할 수 있다.

의 median 값을 구하기 위해

식을 풀면

이고, 이것이 원점으로부터 가장 가까운 점의 median 거리에 대한 추정치이며, , 값에 의존하므로 로 쓸 수 있다.

problem2

# Description

차원 단위구 안에 균등하게 분포하는 개 점을 발생시키는 함수를 만들자.

먼저, <Choosing a Point from the Surface of a Sphere>의 method2를 이용해 단위구 표면에 균등하게 분포된 점의 좌표를 구한다.

다음으로, 표면에 분포된 점을 구 내부로 균등하게 분포시키기 위해 개 점에 대한 거리를 지정해주고자 한다. Problem1에서 구했듯이, 차원 단위구 안에 균등하게 분포하는 점들의 원점으로부터 거리에 대한 분포함수는

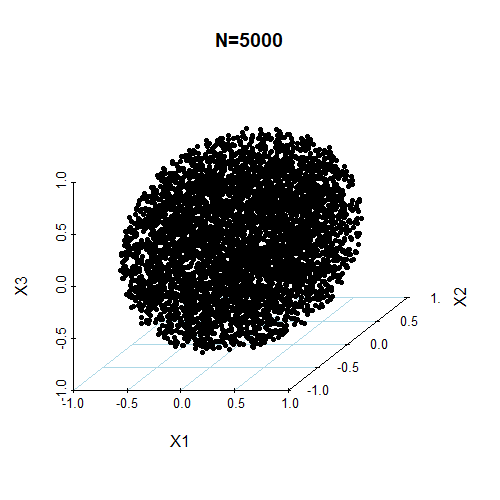
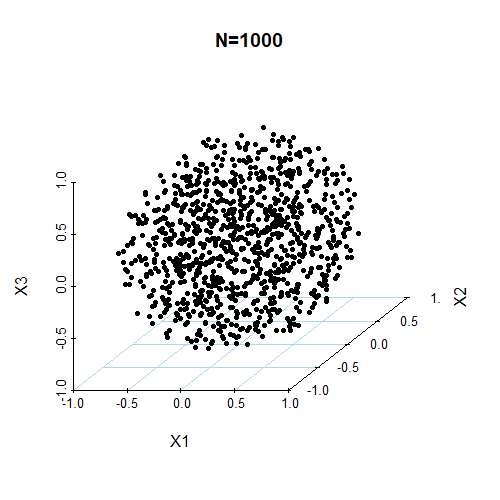
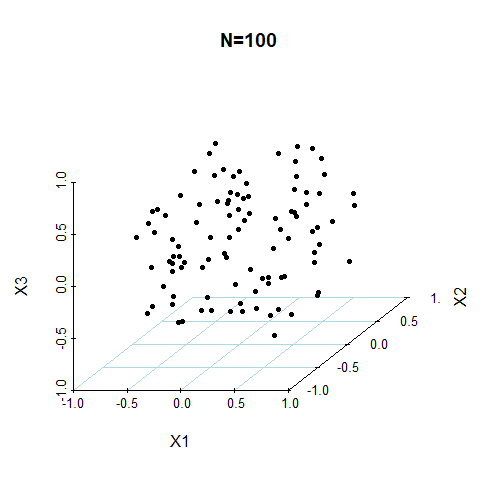
이며, closed form이고 strictly increasing이다. 이 때,

이므로, 의 분포는 을 따른다.

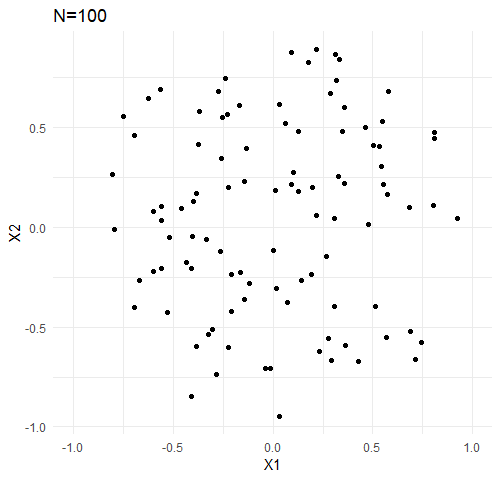
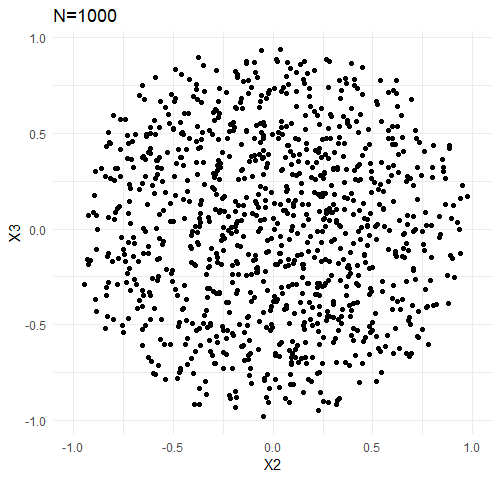
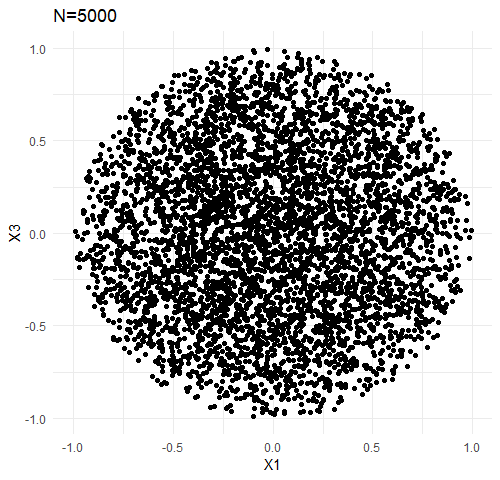
따라서, 을 따르는 확률변수 을 랜덤으로 발생시켜 을 통해 개의 거리를 만들 수 있다. 이렇게 구한 개의 값을 각각 위에서 구한 좌표에 곱하면, 원점으로부터 거리가 인 점을 생성할 수 있다.

# result

3차원 단위구에 균등하게 분포하는 점을 생성했다.



2차원 평면으로 정사영한 그래프는 다음과 같다.

problem3

# description

Problem2에서 샘플이 적절히 구해졌는지 Problem1과 비교해본다.

# result

[표에 행을 추가해야 하는 경우 마우스 포인터를 표의 왼쪽 가장자리에 올려놓으면 행 사이에 나타나는 더하기 기호를 클릭합니다.]

|  |  |  |
| --- | --- | --- |
| 문제 | 담당자 | 날짜 |
|  |  |  |
|  |  |  |
|  |  |  |

discussion

[표에 행을 추가해야 하는 경우 마우스 포인터를 표의 왼쪽 가장자리에 올려놓으면 행 사이에 나타나는 더하기 기호를 클릭합니다.]

appendix