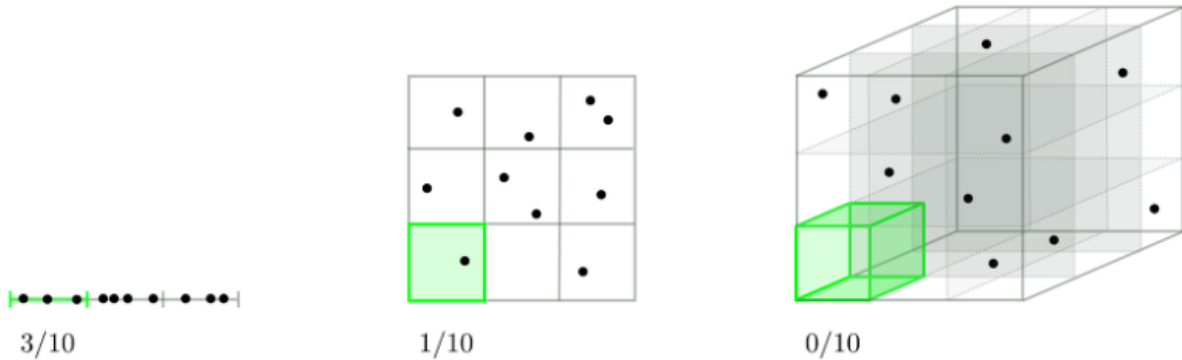


1. 차원의 저주

←〈차원이 작은 데이터들에 대한 추측은 쉽게 파악할 수 있지만,
차원이 큰 데이터들에 대한 추측은 쉽게 파악할 수 없다.〉

머신러닝에서 데이터 셋의 특성(feature)가 많아지면, 각 특성인 하나의 차원(dimension) 또한 증가하게 된다. 이렇게 데이터의 차원이 증가할 수록 데이터 공간의 부피가 기하 급수적으로 증가하기 때문에, 데이터의 밀도는 차원이 증가할 수록 희소(sparse)해진다. 아래의 그림은 데이터의 차원이 증가할수록 각 영역(region)에 포함되는 데이터의 개수를 나타낸 그림(출처: [jermwatt's blog](#))이다. 그림에서 알 수 있듯이 차원이 증가할 수록 데이터의 밀도가 희소해지는 것을 알 수 있다.



데이터의 차원이 증가할수록 데이터 포인트 간의 거리 또한 증가하게 되므로, 이러한 데이터를 이용해 머신러닝 알고리즘을 학습 하게 되면 모델이 복잡해지게 된다. 따라서, 오버피팅(overfitting) 위험이 커진다.

이러한 차원의 저주를 해결하기 위한 방법 중 하나는 데이터의 밀도가 높아질 때까지 학습 데이터셋의 크기를 늘리는 것이다. 하지만, 데이터셋의 크기에 비해 차원은 기하급수적으로 커지기 때문에 매우 힘든 일이라 할 수 있다.

2. 특징 가중치 값이 상대적으로 클 때

: 직선 회귀식에서 특징 가중치가 크면, 해당 가중치를 가진 독립변수에 의해서 예측값이 결정되는 꼴이 된다. 즉, 다른 독립변수들의 존재감이 무의미해진다.

(차라리, 애초에 큰 가중치를 가진 독립변수만 독립변수로 선택했어도 되는 상황이 된다.)