일반적으로 discriminative classifiers가 거의 항상 선호된다. 하지만 generative learning과 discriminative learning 두 알고리즘 사이에서 “어떤 알고리즘이 더 나은가?”라는 질문에 대한 답에 접근하기 위한 한 가지 요소에는 training set size에 있다. 두 알고리즘의 연산 복잡도와 error는 training set과 관련이 있으며 discriminative learning가 비록 generative learning보다 더 낮은 error를 보이지만, generative learning은 error가 높다는 단점에도 그 error에 도달하기까지 discriminative learning보다 더 빠르게 도달한다는 장점이 존재한다.

generative model은 empirical estimates of the probabilities 또는 Laplace smoothing of the probabilities를 이용하여 산출한 p(x|y)와 p(y)를 추정하여 p(y|x)를 계산하는데 집중되어 있는 반면, discriminative model은 p(y|x) 자체가 어떠한 linear combination의 형태인 함수로 되어 있다 여기고 p(y|x)를 직접적으로 추정하는 방식이다. 둘의 error에 관한 차이점을 보면 discriminative model의 error이 generative model의 error보다 작거나 같다는 것을 알 수 있다. 쉽게 생각하면 discriminative model의 error은 점진적으로 무한소에 수렴하기 때문에 당연히 generative model보다 안 좋은 성능을 내는 discriminative model은 없다. Discriminative model의 error boundary를 보면 Vapnik’s uniform convergence bounds에 의해 example data의 개수에 영향을 받고, 따라서 연산 복잡도 또한 order n으로 된다. Generative model의 경우, theorem 4에 의해 logscale로 convergence를 달성할 수 있으며, 이는 discriminative model의 n과는 다른 확연히 빠른 속도를 보인다. 즉, error(gen) >= error(dis) but O(gen) < O(dis)라고 할 수 있다.

UCI Machine Learning repository의 datasets을 이용하여 위의 이론적 결과를 실험적으로 보인다. Figure 1을 보았을 때, generative model이 discriminative model보다 바르게 flat하는 것을 볼 수 있다. 몇몇 dataset에서는 discriminative model이 generative model을 압도하여 더 낮은 error를 보이는데, 이는 곧 이론에서 설명한 것과 같이 discriminative model의 궁극적인 error가 generative model보다 작다는 것과 일맥상통하다. 하지만 그 외의 dataset에서는 discriminative model이 추월하지 못하는 경향이 보이는데, 이는 모델을 학습시키기 위한 충분한 data m의 크기가 작기 때문이다.