

이과대 경진대회 중간보고서

Manifold Learning

양승준

22/8/5

Introduction

Optimization Algorithm들에서의 사례

초기 ML에서는 NN의 가중치를 갱신하기 위해 Gradient Descent Method(GD)를 사용하였다. 이 GD의 최적화 속도를 개선하기 위하여 Momentum이라는 Method가 등장하였다. 이 Momentum은 기존의 GD에 Momentum 항을 추가하는 방법으로, 뉴턴역학에서 보존력이 작용하는 계에서 질량을 가진 입자를 기술하는 방식과 유사하다.¹ 그러나 Multidimension Space 상에서 최적화를 할 때 GD와 Momentum은 Gradient가 큰 방향으로의 학습이 잘되지만 작은 방향으로 학습이 되지 않는 문제가 존재했다. 이를 해결하기 위해 다시 Momentum이나 GD에서 항들을 추가하거나 변형하게 된다.

Manifold Learning

초기의 Classification Algorithm들은 유클리드 공간 상에서 유클리드 거리를 이용하여 이루어졌다. 그러나 샘플 데이터들이 Swiss Roll과 같이 3차원 상에 놓인 곡면 상에 놓여있다면 곡면상에서 거리가 먼 두 점이 가깝게 인식되는 상황이 발생한다. Manifold Learning은 샘플 데이터들이 어떤 manifold위에 존재할 것이라는 가정 하에 이루어진다. 이 Manifold를 유클리드 공간으로 Mapping시켜 기존의 알고리즘을 적용시킬 수 있도록 만드는 데에 목적이 있다. Manifold를 Mapping 시키는 방법들에는 Isomap, Local Linear Embedding, Laplacian Eigenmaps, Hessian Eigenmaps등이 있다.² 이번 연구 프로젝트에서는 이 방법들을 공부하고 더 나아가 새로운 방법을 만들어보려고 한다.

Methods of Manifold Learning

Isomap

Manifold Learning의 핵심은 manifold에서 유클리드 공간 위로의 mapping을 찾는 것이다. Isomap은 2가지 가정하에 이루어진다.

- Manifold상에서의 geodesic의 길이는 mapping에 대해 invariant하다.

¹Ning Qian. (1999). "On the momentum term in gradient descent learning algorithms"

²Izenman, A. J. (2008). "Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning"

- Manifold는 convex하다.

직관적으로는 둘둘 말려있는 종이를 다시 펴평하게 펴는 것을 떠올리면 된다. Mapping이 geodesic이라는 Global한 성질을 보존한다. 이 방법은 manifold에 구멍이 뚫려있다면 적용하기에 어려움이 있다.

Isomap의 구체적인 알고리즘은 다음과 같은 순서를 따른다.

1. K를 자연수라고 할 때, 각 점에서 1부터 K번째로 가까운 점을 연결하여 Graph를 만든다.
2. 각 점들 사이의 Geodesic distance들을 구한다.
3. MDS를 적용한다.

Local Linear Embedding(LLE)

LLE는 Isomap과 거의 동일한 가정을 사용한다. 다만 다른 점은 Isomap은 global한 성질을 보존시켰다면, LLE는 local한 성질을 보존시킨다.

1. K를 자연수라고 할 때, 각 점에서 1부터 K번째로 가까운 점을 연결하여 Graph를 만든다.
2. ...
3. Eigenproblem

Laplacian Eigenmaps(LE)

LE는 LLE와 첫 번째와 세 번째 단계가 동일하다.

Hessian Eigenmaps(HE)

기존의 Isomap은 convexity라는 강력한 조건을 가정하고 있었다. Hessian Eigenmap은 높은 차원상에 있는 데이터들에 대해 convexity가 만족되지 않더라도 mapping을 찾을 수 있도록 해준다.