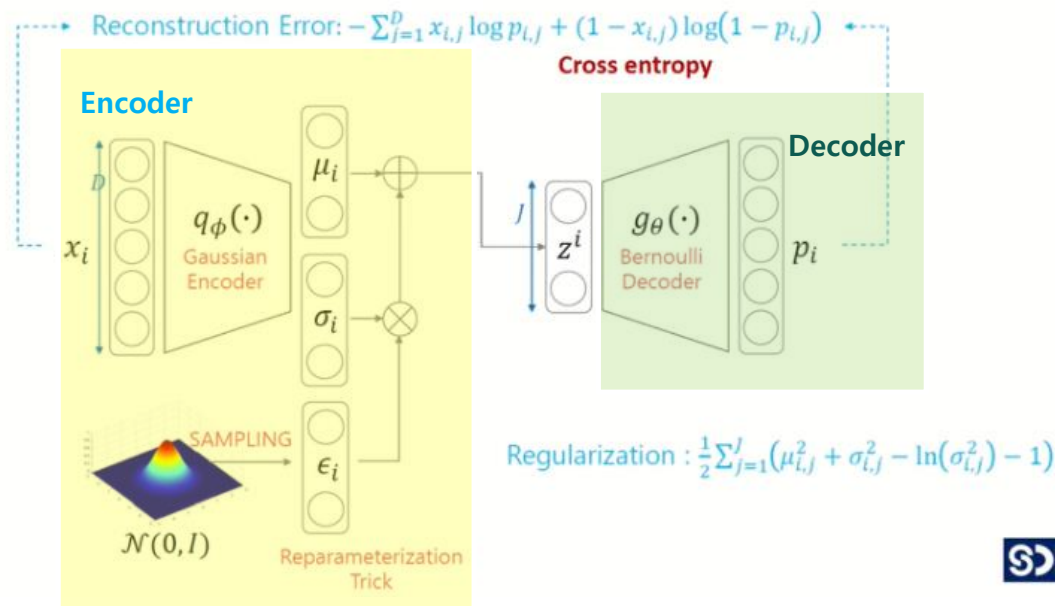


VAE 모델

VAE(Variational AutoEncoder)는 generative model 의 한 종류로 Input X 를 잘 설명하는 feature 를 추출하여 Latent vector Z 에 담고 이 latent vector Z 를 통해 X 와 유사하지만 완전히 새로운 데이터를 생성해 내는 것을 목표로 한다.



구조

Encoder : 입력 데이터 x 를 저차원의 잠재공간(latent space)으로 매핑하여 잠재 변수 z 의 평균과 표준편차로 출력하고 출력한 평균과 표준편차를 기반으로 한 정규분포에서 z 를 샘플링하는 과정을 거침

하지만 이런 방식으로 샘플링한다면 미분이 불가능해 역전파가 불가능하기 때문에 재매개변수화 트릭(Reparameterization Trick)을 사용하여 표준 정규분포에서 샘플링한 ϵ 값을 표준편차에 곱한 뒤 평균을 더하는 과정을 가짐

Decoder : 잠재 공간의 저차원 벡터 z 를 다시 복원하여 복원된 데이터를 출력

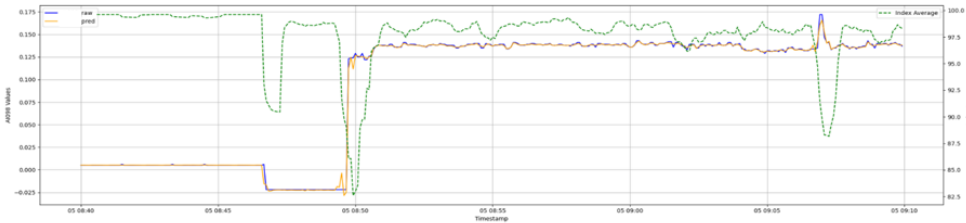
학습과정

VAE 의 학습과정은 복원된 데이터와 원본 데이터간의 차이를 최소화하기 위해 reconstruction loss 를 계산하고 잠재 변수의 사후 분포와 사전 분포 사이의 차이를 최소화하기 위해 regularization loss 를 계산하여 두 loss 의 합을 최소화하는 방향으로 모델의 가중치를 업데이트한다.

이상치 탐지과정

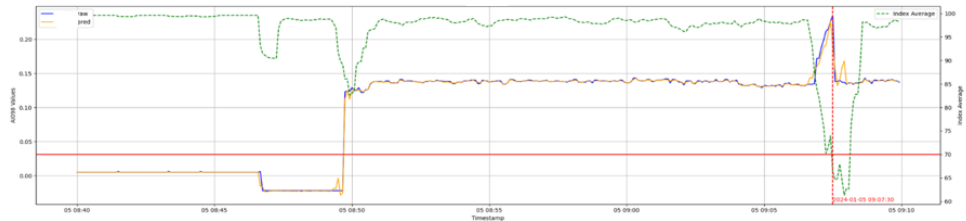
VAE 모델을 통한 학습과정에서 Train 데이터세트(정상구간)가 가지는 분포를 학습하여
Train 데이터세트가 가지는 분포내의 Input data 가 들어온다면 정상적으로 복원이 되겠지만
기존의 분포와 다른 분포를 가지는 Input data 가 들어온다면 정상적으로 복원하지 못해
실제값과 예측값사이의 잔차를 기준으로 이상탐지를 수행할 수 있음

정상데이터 예측



Train데이터세트가 가지는 분포내의
데이터가 입력되어 복원데이터와
원본데이터와의 차이가 크지 않음

이상데이터 예측



Train데이터세트가 가지는 분포와
다른 데이터가 입력되어 예측은 되지만
복원데이터와의 차이가 생겨 이상탐지

VAE 적용 환경

- 상관관계가 낮은 변수를 넣더라도 예측이 가능하여 모델에 다양한 태그를 사용이 가능하고 발전소의 정비 구간 등 비정상 구간의 데이터를 포함하여도 예측이 가능함
- 모델의 index 점수를 학습데이터로 사용하여 설비의 전반적인 상태를 감시 하는 목적으로 사용 가능할 것으로 기대