

1. 다양한 **Learning Rate Scheduler** 중 적어도 한 가지를 소개
과제에서 사용한 **Adam(Adaptive Moment Estimation)**은 딥러닝 최적화 기법 중 하나로써 학습의 방향과 크기를 모두 개선한 기법이다. 수식을 통해 알아보도록 하겠다.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla f(x_{t-1})$$

$$g_t = \beta_2 g_{t-1} + (1 - \beta_2) (\nabla f(x_{t-1}))^2$$

β_1 : Momentum의 지수이동평균 ≈ 0.9

β_2 : RMSProp의 지수이동평균 ≈ 0.999

Adam에서는 m 과 g 가 처음에 0으로 initialization이 되어있다. 따라서 학습 초반에는 m_t 와 g_t 가 0에 가깝게 bias되어 있을 것이라고 판단해 먼저 이를 unbiased하게 만들어주는 과정을 거친다.

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \widehat{g}_t = \frac{g_t}{1 - \beta_2^t}$$

학습이 계속 진행되다 보면 $(1 - \beta_1)$ 와 $(1 - \beta_2)$ 는 거의 1에 가까워지기 때문에, \widehat{m}_t 와 \widehat{g}_t 는 결국 m_t , g_t 와 같은 값이 된다.

$$x_t = x_{t-1} - \frac{\eta}{\sqrt{\widehat{g}_t + \epsilon}} \cdot \widehat{m}_t$$

ϵ : 분모가 0이 되는 것을 방지하기 위한 작은 값 $\approx 10^{-8}$

η : 학습률 ≈ 0.001

2. 오늘 수업에서 소개되지 않은, **Training Error**와 **Generalization Error** 사이 간극을 줄이는 방안

Data Augmentation

: 데이터에 대한 **Regularization**으로, 어떤 이미지에 대해 좌우 반전, **Random crops**(다양한 사이즈로 자르기), **Scales**(crop한 뒤 원본 크기로 키우기) 등 데이터를 조금씩 변경하여 다양한 데이터를 넣어주는 방식을 **Data Augmentation**이라고 한다.

DropConnect

: Dropout에서는 입력을 Drop 했다면, DropConnect에서는 **weight** 값에 대해서 Drop을 준다. **weight** 값에 대해서 임의로 0으로 설정해 주는 것. 임의로 **weight matrix**를 임의적으로 만들어주고 이를 곱해주는 것이다.