

## 1.1 Probability와 Likelihood

### 1. Probability:

- Probability는 특정 이벤트 A가 발생할 가능성을 나타내는 척도입니다.
- Probability는 0과 1 사이의 값을 가지며,  $P(A)$ 로 표현됩니다.
- 예를 들어, 동전을 던졌을 때 앞면이 나올 확률은  $P(\text{앞면}) = 0.5$ 입니다.

### 2. Likelihood:

- Likelihood는 관찰된 데이터 X가 주어졌을 때, 모델 파라미터  $\theta$ 의 확률을 나타냅니다.
- Likelihood는  $L(\theta|X)$ 로 표현되며, 데이터 X를 지원하는 정도를 나타냅니다.
- 예를 들어, 동전 던지기 실험에서 앞면이 400번 나왔다면, 동전의 확률 p에 대한 Likelihood는  $L(p|X) = p^{400} * (1-p)^{600}$ 입니다.

## 1.2 Prior와 Posterior distribution

### 1. Prior distribution:

- Prior distribution은 데이터를 관찰하기 전의 모델 파라미터  $\theta$ 에 대한 확률 분포입니다.
- Prior distribution은  $p(\theta)$ 로 표현됩니다.
- 예를 들어, 동전 던지기 실험에서 동전의 확률 p에 대한 Prior distribution은  $p \sim \text{Uniform}(0, 1)$ 일 수 있습니다.

### 2. Posterior distribution:

- Posterior distribution은 데이터 X를 관찰한 후의 모델 파라미터  $\theta$ 에 대한 확률 분포입니다.
- Posterior distribution은 Bayes' Theorem에 의해 계산되며,  $p(\theta|X)$ 로 표현됩니다.
- Bayes' Theorem:  $p(\theta|X) = (p(X|\theta) * p(\theta)) / p(X)$

## 1.3 Bayes' Theorem

Bayes' Theorem은 다음과 같이 표현됩니다:

$$p(\theta|X) = (p(X|\theta) * p(\theta)) / p(X)$$

여기서:

- $p(\theta|X)$ : Posterior distribution
- $p(X|\theta)$ : Likelihood
- $p(\theta)$ : Prior distribution
- $p(X)$ : Marginal likelihood

Bayes' Theorem은 Prior distribution과 Likelihood를 결합하여 Posterior distribution을 계산합니다. 이를 통해 데이터를 관찰한 후 모델 파라미터의 확률 분포를 업데이트할 수 있습니다.

## 2. Kullback-Leibler Divergence

Kullback-Leibler (KL) Divergence는 두 확률 분포 P와 Q 사이의 차이를 측정하는 지표입니다. KL Divergence는 다음과 같이 정의됩니다:

$$D_{KL}(P||Q) = \sum_x P(x) \log(P(x) / Q(x))$$

여기서:

- $P(x)$ 는 참 분포,  $Q(x)$ 는 근사 분포입니다.
- $D_{KL}(P||Q)$ 가 작을수록 두 분포가 유사하다고 볼 수 있습니다.

KL Divergence는 기계 학습 분야에서 다양하게 활용됩니다. 예를 들어, Variational Autoencoder (VAE)에서는 Encoder와 Decoder 사이의 KL Divergence를 최소화하여 의미 있는 Latent Space를 학습합니다.

## 3. Variational Autoencoder (VAE)

VAE는 Generative Model의 한 종류로, Encoder와 Decoder로 구성됩니다. VAE의 목표는 입력 데이터 X의 분포를 근사하는 것입니다.

VAE의 수식적 표현은 다음과 같습니다:

$$\text{Encoder: } q(z|x) = N(\mu(x), \sigma^2(x))$$

$$\text{Decoder: } p(x|z) = N(\mu_x(z), \sigma_x^2(z))$$

여기서:

- $q(z|x)$ 는 Encoder의 Posterior distribution
- $p(x|z)$ 는 Decoder의 Likelihood
- $\mu(x), \sigma^2(x)$ 는 Encoder의 출력
- $\mu_x(z), \sigma_x^2(z)$ 는 Decoder의 출력

VAE는 Encoder와 Decoder 사이의 KL Divergence를 최소화하여 의미 있는 Latent Space를 학습합니다.

## 4. K-fold Cross Validation

K-fold Cross Validation은 모델의 성능을 평가하는 방법 중 하나입니다. 이 방법은 다음과 같이 진행됩니다:

Test	Train	Train	Train
Train	Test	Train	Train
Train	Train	Test	Train
Train	Train	Train	Test

1. 데이터셋을 K개의 균등한 부분집합으로 나눕니다.
2. K-1개의 부분집합을 사용하여 모델을 학습합니다.
3. 남은 1개의 부분집합을 사용하여 모델의 성능을 평가합니다.
4. 이 과정을 K번 반복하여 각 부분집합에 대한 성능 평가 결과를 평균합니다.

K-fold Cross Validation을 통해 모델의 일반화 성능을 보다 정확하게 추정할 수 있습니다.

일반적으로 K=5 또는 K=10을 많이 사용합니다.