1. Sample Space and Events

S (Sample Space)	발생할 수 있는 모든 경우의 수의 집합
Event (사건)	Sample Space의 부분 집합
Event Space	Sample Space의 요소로 구성될 수 있는 모든 조합의 경우의 수 (2^S개로 구성됨) Event Space 내에 공집합과 전체 집합도 포함됨 / 공집합과 전체 집합도 사건 일부

2. 확률 (Probability)

- **불확실한 사건의 발생을 숫자로서 표현하는 것** **(※ 불확실성이 있다는 것은 특정 상황을 확률로서 표현할 수 있음)**

- **확률을 구하는 방법**

  ① **Classical Approach의 확률 :** 발생 가능한 사건이 각각 동일한 발생 가능성을 가진다고 가정함
  (※ 완벽한 동전 또는 주사위는 존재하지 않음 / 불확실성에 확실성을 부여하려하기에 모순적임)

  ② **Relative Frequency Approach (Frequentist)의 확률 :** 무수한 실험 후 확률 결정
  (※ 무한적인 실험을 요구하기에 비현실적이며, 무수한 실험의 결과가 수렴하지 않으면 문제가 발생함)

  ③ **Subjective Approach (Bayesian)의 확률 :** 상황 발생에 대한 믿음의 정도를 가지고 결정
  (※ 반복적인 실험에 대한 확률을 알 수 없을 경우 사용 / 주관적인 견해와 믿음이 반영됨)

- **확률로 인정받기 위한 조건 (Probability Measure)**

  ① Event Space 내에서 각각의 Event에 대한 확률값은 0이상이어야함

  ② 모든 경우의 수의 확률 합은 1이어야함

  ③ 교집합이 없는 두 상황에 대한 확률값은 각각의 상황에 대한 확률값을 더한 것과 같음

- Classical Approach, Relative Frequency Approach, Subjective Approach 중 1개를 선택하여 특정 상황에 대한 확률값을 구하고, 확률 인정 조건을 통해 확률이라는 것을 증명/인정함

3. 조건부 확률 (Conditional Probability)

- **$P(F|H)$: H라는 사건이 발생할 때, F 사건도 같이 발생할 확률** = $P(F \cap H) / P(H)$

- **$P(F|H) = P_H(F)$ →** H에 대한 Sample Space로 관점을 이동하여, H Sample Space에서 F 사건 발생을 표현함

- **만약 2개의 사건이 독립이라면 조건에 상관없이 각자의 상황은 발생해야함.**
  그러므로 F와 H가 독립이라면 서로에게 영향을 주지 않기에 서로에 대한 조건부 확률은 각 상황이 발생하는 확률과 같음 ($P(F|H) = P(F)$, $P(H|F) = P(H)$). 이를 통해서 F와 H가 독립일 때, $P(F \cap H) = P(F)P(H)$라 할 수 있음.

- **Rule of Total Probability ($P(E) = \sum P(B_i \cap E) = \sum P(B_i)P(E|B_i)$)**
  E라는 사건이 내부적으로 독립된 $B_1 \sim B_n$ 사건들로 구성 되어있을 시, E 사건이 발생할 수 있는 확률은 E사건과 B_i 사건이 같이 발생할 수 있는 모든 경우의 확률의 합과 같음. **조건부 확률로 바꿔서 표현하게 되면 B_i라는 각각의 사건에 대한 상태만 알아도 전체 상태인 E를 유추할 수 있음.**

- **Bayes' Rule**
  : 새로운 정보가 제공되면 불확실성에 대해 Update될 수 있음.
  : $P(B_i|E)$에서 사건에 대한 새로운 정보는 E에 추가되어 확률을 Update할 수 있음.
  : Bayes' Rule은 확률의 Updating Rule을 알려줄 수 있음.

4. 확률 변수 (Random Variables) / 확률 분포 (Probability Distribution)

- **확률 변수 X**
  : Sample Space 내에 속한 각각의 사건을 기호가 아닌 실수 형태의 Label로 표현하는 방법
  : 각 사건의 발생에 대해 대응되는 실수를 배치하며, 사건의 불확실성을 실수 영역에서 다룰 수 있게됨.
  → **이산확률변수 (Discrete Random Variable) :** 사건에 대한 수학적 표현 • 실수 형태 Label이 Discrete한 경우
  → **연속확률변수 (Continuous Random Variable) :** 사건에 대한 수학적 표현 • 실수 형태 Label이 Continuous한 경우

- **확률 분포 P(X)**

  : Sample Space 내에 속한 각각의 사건에 대응되는 실수인 확률변수에 따라 매칭된 확률값의 집합

  → 확률 변수 X에 대한 확률 분포 P(X)를 통해 **"기존 사건 대 실수 확률값"**의 표현을 **"사건의 실수 표현 대 실수 확률값" 관계로 재정의**하면서 **사건의 불확실성을 숫자의 집합으로 만들어서 수학적 분포 • 수학적 수식으로 정의할 수 있게됨.**

  → **이산확률분포 (Discrete Probability Distribution) :** 이산확률변수에 따른 확률값의 Discrete한 집합

  → **연속확률분포 (Continuous Probability Distribution) :** 연속확률변수에 따른 확률값의 Continuous한 집합

5. Mean, Variance, Covariance, Correlation

- **Mean (기댓값, 평균)**은 확률 분포 내 사건 표현값과 확률값의 곱의 합이며, **확률 분포의 무게 중심을 보여주는 척도임.** 특정 사건이 발생할 때 해당 사건에 대해 예상되는 확률을 지칭함.

- **Variance (분산)**은 확률 분포 내 각 사건이 Mean으로부터 오차에 대한 평균을 구한 값이며, **확률 분포 내에서 Mean을 중심으로 전반적으로 확률이 얼마나 넓게 퍼져있는지를 보여주는 척도임.** Variance가 높을수록 사건이 Mean을 중심으로 넓게 펴져 다양한 사건이 발생할 수 있음. Variance가 낮을수록 사건이 Mean을 중심으로 모이게 되어 Mean에 유사한 사건들이 발생할 수 있음.

- **Covariance (공분산)은 2개 변수 X, Y 간의 선형성(Linearity)을 파악하기 위한 척도임. Covariance는 2개 변수 사이에 선형성 유무와 선형성 방향을 알려줄 수 있음. 그러나 Covariance는 단위를 가지는 척도이기에 선형성의 강도를 표현할 수 없음.**

- **Correlation (상관관계)은 Covariance와 같이 2개 변수 X, Y간의 선형성을 나타내기 위한 척도임. Covariance와 달리 단위가 없는 무단위 척도이기에 2개 변수 사이의 선형성 유무, 선형성 방향, 선형성 강도 모두 표현할 수 있음.**

  (※ Covariance, Correlation은 2개 변수 사이의 선형성을 알려줄 뿐 '인과 관계'를 알려주지 않음.)

6. Statistics

- **Random Sample**

  : Random Sample은 모집단에서 구성원이 Sampling에 의해 선택될 확률이 동등하다는 원칙하에 모집단으로부터 선택될 Sample (표본)을 의미함.

  : 모집단 내에서 선택될 확률만 같은 상황이기에 Random Sample은 아직 모집단에서 실제로 Sampling • 실험 행위가 수행되기 전의 표본임. 모집단에서 어떤 요소를 뽑은 것이 아니라 앞으로 뽑을 예정인 것임.

  : Random Sample은 실제 Sampling • 실험을 진행하기 전까지는 어떤 값이 나올지 모르기에 Sample 간에는 독립적이며 (Indepdent), 모집단 내에서 어떤 것이든 동등한 확률로 선택될 수 있기에 (Identically Distributed) 모집단과 동일한 확률변수 특성을 가지게 됨.

- **Statistic (통계량)**

  : Statistic은 Random Sample로 구성된 값을 의미함. Sample을 뽑아야 값을 알 수 있음.

  : Statistic은 모집단에서 앞으로 선정될 요소인 Random Sample로 구성된 값이기 때문에 Sampling • 실험을 진행하면서 Statistic (통계량)에 대한 값을 점차적으로 알아갈 수 있음.

  : 그러므로 Statistic은 Sampling • 실험하기 전까지는 불확실성을 지니기 때문에 '확률변수'라 정의할 수 있으며, Sampling • 실험을 수행에 따라 사건이 발생함에 따라 실수 확률 분포로 표현될 수 있음.

- **Estimator**

  : Estimator는 Sample을 가지고 모집단의 특성을 추정하기 위해 사용하는 Statistic • 수학적 도구

  : Estimator는 Sample의 평균인 Sample Mean과 분산인 Sample Variance를 사용함. 각각은 Sampling • 실험을 통해 **실제로 Sample을 모으기 전까지는 Sample Mean과 Sample Variance를 구할 수 없음.**

  : Sampling • 실험을 진행할수록 Sample Mean과 Sample Variance 값이 업데이트됨. **Sample Mean과 Sample Variance는 고정된 값이 아닌 Sampling • 실험 횟수에 따라 변동되기에 불확실성을 지니게되며, 확률로 취급되어 확률 분포로서 표현될 수 있음.**

  : 만약 모집단에서 **Sample을 계속 무한히 뽑을수록 Sample이 모집단과 유사해지면서 Sample Mean과 Sample Variance가 모집단의 확률 분포와 유사해짐. 그러므로 Sample Mean과 Variance를 통해 모집단의 Mean과 Variance를 추정할 수 있음.**

- **최적의 Estimator 조건**

  ① **Unbiasedness :** 모집단 분포의 중심 Mean이 Estimator (Sample Mean)와 최대한 일치해야 모집단 추정이 정확해짐

  ② **Efficiency :** 모집단을 추정 정확도의 분포가 짧아야함

  ③ **Consistency :** Sample을 점점 많이 뽑을수록 모집단에 유사해짐

  ④ **Sufficiency :** Estimator가 모집단에 대해서 정보를 충분히 제공해야함

- **Bias-Variance Decomposition을 통해 모집단 vs Sample간의 Error가 Bias와 Variance의 조합으로 구성되는 것**을 볼 수 있으며, **Bias와 Variance의 조합에 따라 Error가 모집단과 Sample 사이의 Error가 최소화**될 수 있음.

  → **Bias가 생겨서 모집단의 Mean과 Sample의 Mean이 서로 조금 어긋나더라도 Variance가 좁으면 모집단과 가까운 위치에 있기에 모집단 vs Sample간의 Error가 적을 수 있음.**

- **Machine Learning의 학습 ≒ 최적의 Estimator를 찾아내는 문제**

  : 모집단에서 Sample을 추출하면서 Estimator를 통해 모집단을 추정함

  : Machine Learning에서 내가 원하는 Dataset에 대해 Prediction을 잘해주기 위해 Sample 데이터 (Train Dataset)을 제공해서 학습시켜서, 그에 따라 모집단에도 동일하게 정확하게 Prediction이 작동되게함.

  : Machine Learning에서 학습을 통해 최적 성능의 Network를 설계하는 문제는 모집단에 대해 Unbiased하고 Efficient한 최상의 Estimator를 구하는 것과 같음.

7. Derivative • Differentiation / Gradient

- **Differentiation (미분)**

  : 특정 변수 X의 변화에 대한 결과값 Y의 변화율을 관찰함.

  : 그래프 상 특정 지점의 변화율인 기울기를 의미함.

- **Partial Differentiation (편미분)**

  : 다변수로 구성된 공식이나 관계 W = f(X, Y, Z)에서 특정 변수 X의 변화에 대해서만 결과값 W의 변화율을 관찰함.

  : 미분 대상인 X를 제외한 나머지 변수는 상수로 취급하여 X에 대해서만 W의 미분을 계산함.

- **Gradient**

  : W = f(X, Y, Z)와 같은 다변수 함수의 기울기를 정의하기 위해 각 독립변수 X, Y, Z에 대한 W의 편미분 모음

8. Vector / Dot Product / Norm

- **Vector** : 크기와 방향을 가지는 물리량
- **Dot Product** : Vector와 Vector가 곱해서 실수가 나오게 하는 곱셈 연산
- **Norm**

  ① L^1 Norm ($||a||_2 = \sqrt{\sum a_i^2}$) : Vector의 모든 요소의 절대값의 합

  ② L^2 Norm ($||a||_1 = \sum |a_i^2|$) : Vector의 각 성분의 제곱을 더해서 루트를 적용한 값