(9/15)엔트로피 구하기

1. A, B, C, D 네개의 심볼이 있을 때 엔트로피 구하기

다른 정보가 없음으로, 각각의 심볼이 나타날 확률은 1/4이다.

$$P_A = \frac{1}{4}, P_B = \frac{1}{4}, P_C = \frac{1}{4}, P_D = \frac{1}{4}$$

이 때의 엔트로피는

$$H = -\sum P_i log_2(P_i)$$
$$H = -(P_A log_2(P_A) + P_B log_2(P_B) + P_C log_2(P_C) + P_D log_2(P_D))$$
$$H = -(\frac{1}{4} log_2(\frac{1}{4}) + \frac{1}{4} log_2(\frac{1}{4}) + \frac{1}{4} log_2(\frac{1}{4}) + \frac{1}{4} log_2(\frac{1}{4}))$$
$$H = 2$$

2 bits이다.

2. 전체 100개의 발생중에서 각각 60, 20,10,10 번 발생했을 때의 엔트로피

각각의 심볼이 나타날 확률은

$$P_A = \frac{60}{100}, P_B = \frac{20}{100}, P_C = \frac{10}{100}, P_D = \frac{10}{100}$$

이 때의 엔트로피는

$$H = -\sum P_i log_2(P_i)$$
$$H = -(P_A log_2(P_A) + P_B log_2(P_B) + P_C log_2(P_C) + P_D log_2(P_D))$$
$$H = -(\frac{60}{100} log_2(\frac{60}{100}) + \frac{20}{100} log_2(\frac{20}{100}) + \frac{10}{100} log_2(\frac{10}{100}) + \frac{10}{100} log_2(\frac{10}{100}))$$
$$H \approx 1.571$$

약 1.571 bit이다.

3. 1번과 2번을 비교하시오

아무런 정보도 없이, 모든 확률이 1/n으로써 동일하다고 가정할 경우, 엔트로피는 2 bits이고,

정보를 획득하여 엔트로피를 계산하는 경우, 약 1.571 bits이다.

정보가 있는 2번이 정보가 없는 1번에 비하여 약 0.429(bits) 정도 엔트로피가 낮다.

이처럼 각각의 발생 빈도에 대한 정보를 이용하여 엔트로피를 줄이는 방향으로 계산하는 것이 가능하다.