

Probability and Statistics in ML (ML에서 확률과 통계)

- 확률 기호
 - Permutation : 순서가 있는 경우 (순열)

$$nPr = \frac{n!}{(n-r)!}$$
 - Combination : 순서가 없는 경우 (조합)

$$nCr = \frac{n!}{r!(n-r)!}$$
 - 확률 분포
 - 확률 벡터

- 베이즈 정리 : $P(y, x) = P(x|y)P(y) = P(x, y) = P(y|x)P(x)$

$$\rightarrow P(y|x) = \frac{\overbrace{P(x|y)P(y)}^{\text{우도 : 가 능 도}}}{\underbrace{P(x)}_{\text{사전 확률}}}$$

\uparrow
 사후 확률

- 기계학습에 적용 (Maximum A Posteriori) MAP

$\boxed{?}$ $\xrightarrow{\text{특성들}}$ $x = (7.0, 3.2, 4.7, 1.4)^T \xrightarrow{\text{사전 확률 추정}}$

$P(\text{setosa} x) = 0.18$	$\xrightarrow{\text{argmax}}$	versicolor
$P(\text{versicolor} x) = 0.72$		
$P(\text{virginica} x) = 0.10$		

- 최대 우도 : 목적 함수에 likelihood를 이용, 매개변수 θ 를 모르는 상황에서 θ 를 추정하는 문제

- 평균과 분산 : 데이터의 요약 정보

- 유용한 확률 분포

- 가우시안 분포 : 평균을 기준으로 대칭성을 띈다.
- 베르누이 분포 : 1, 0 두가지로만 나타내는 확률 분포
- 이항 분포 : 베르누이 실험 시 행을 여러번 하는 것

Information Theory

- 확률이 작을수록 ~~많은~~ 많은 정보

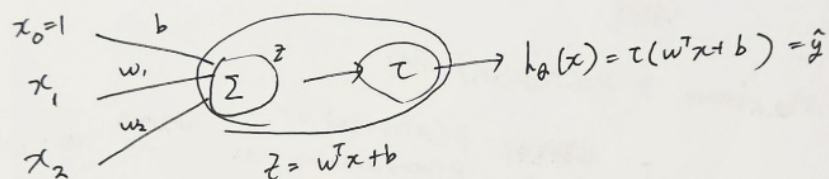
- 자기정보 : 특정 사건 e_i 의 정보량

$$h(e_i) = -\log_2 p(e_i)$$

- Cross Entropy : 하나의 변수가 가질 수 있는 서로 다른 분포를 가진 경우, 해당 분포들의 차이를 의미
; Entropy : 변수의 불확실성을 나타내는 지표

- KL divergence : 두 확률분포의 차이를 계산하는데 사용하는 함수

Objective Function : 입력신호의 총합을 신호로 변환하는 함수



- MSE (평균제곱오차)

- 오차가 클수록 e 값이 커진다.

- sigmoid와 결합시, 에러가 클 때 gradient가 0에 가까워져 학습이 느려질 수 있다. (느린 학습)

- CrossEntropy

- 분류 문제에서 사용

- one-hot 인코딩된 vector를 사용

- 클래스별 확률을 따지면 /

② MSE vs Cross Entropy

- 데이터가 연속적인 분포인 gaussian 분포에 가까울 때 → MSE Loss

- 데이터가 범주형 베르누이 분포에 가까울 때 → Cross Entropy Loss