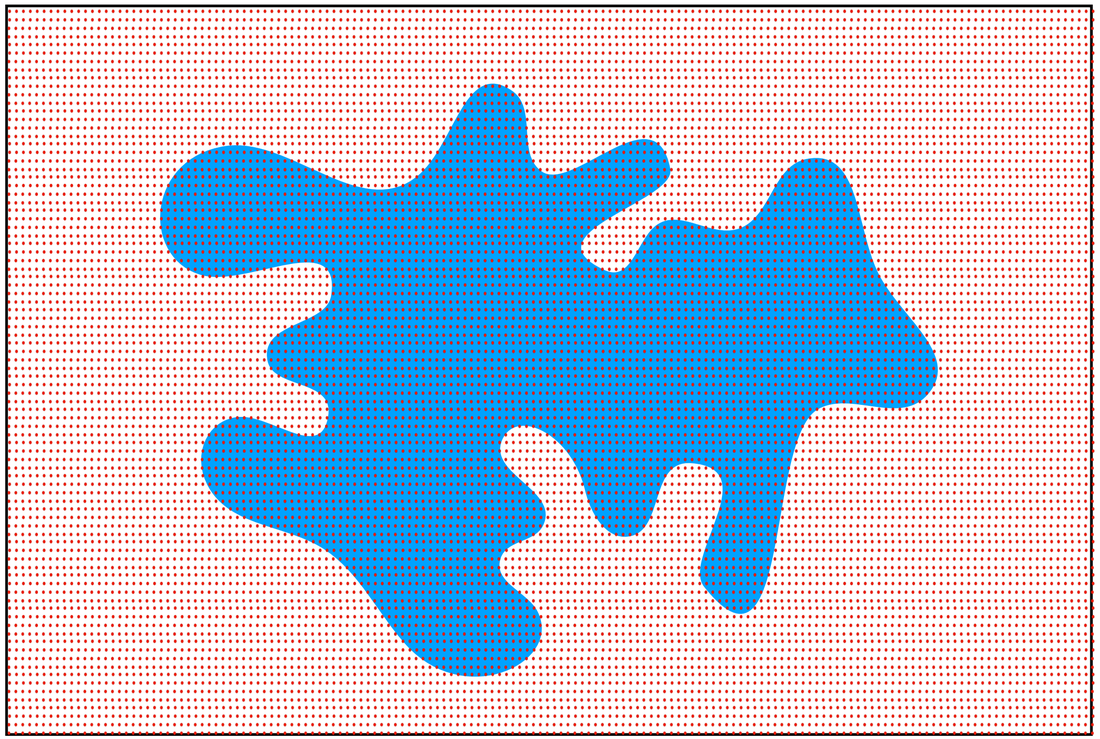
ICT이노베이션스퀘어 AI복합교육 고급 언어과정

자연어처리를 위한 Softmax

현청천

2021.04.19

Monte Carlo method



샘플 모델

복잡한 함수를 난수를 이용해 함숫값을 확률적으로 근사화하는 방법

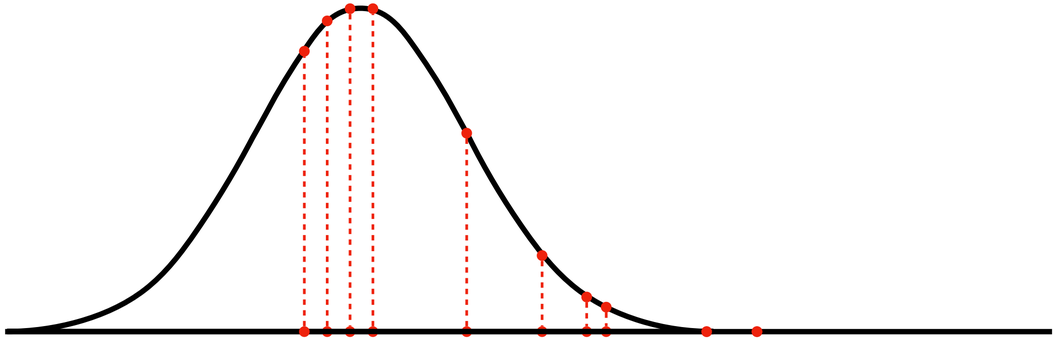
Maximum Likelihood Estimation

키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63

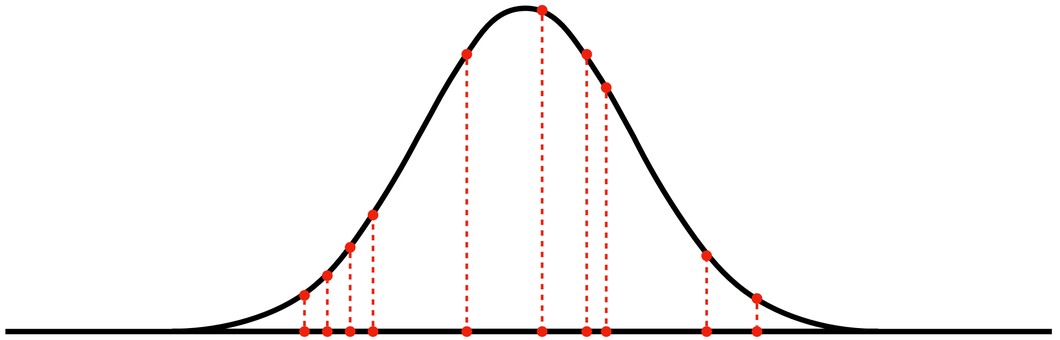
사건으로부터 확률분포를 예측

Maximum Likelihood Estimation

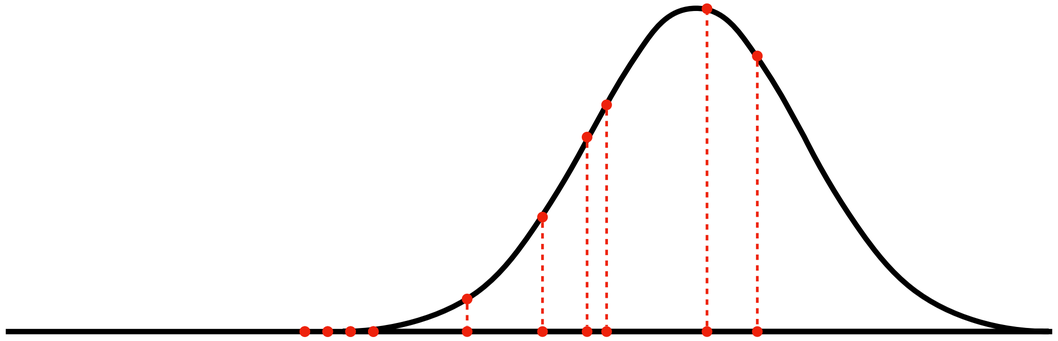
$p(X;\theta_1)$



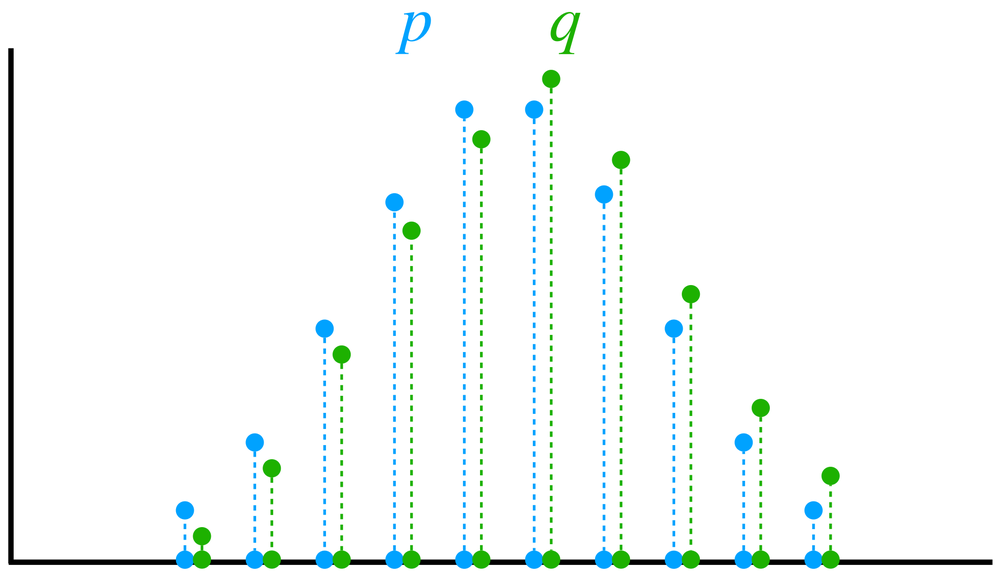
$p(X:\theta_2)$



$p(X;\theta_3)$



Cross Entropy (이산확률분포)



$$H(p,q) = \sum_{x} p(x) \log \frac{1}{q(x)}$$

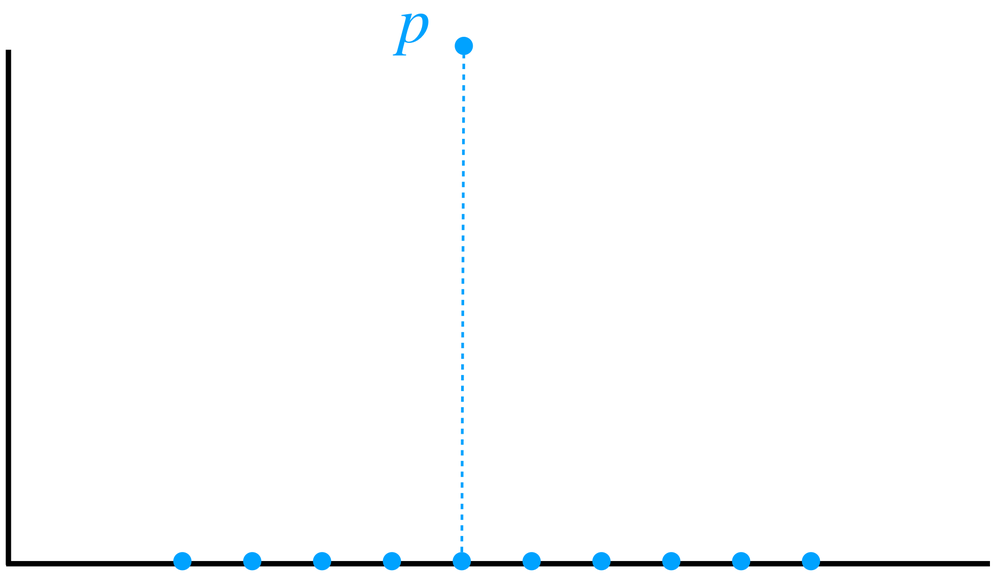
확률분포 p에 대한 확률분포 q의 정보량의 기댓값

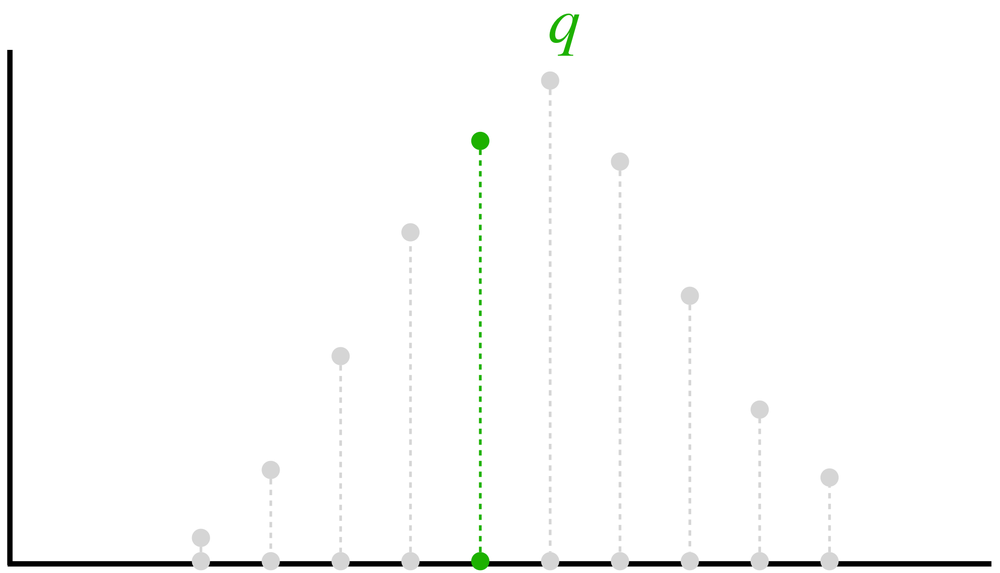
Cross Entropy Loss

Cross Entropy loss $$CE = \frac{1}{N}\sum_{i=n}^{N}\sum_{j=1}^{C} y_{ij}\ \log\frac{1}{\hat{y}_{ij}} = -\frac{1}{N}\sum_{i=n}^{N}\sum_{j=1}^{C} y_{ij}\ \log\hat{y}_{ij}$$

여러 샘플의 정답확률분포와 예측확률분포의 Cross Entropy의 평균

Cross Entropy Loss





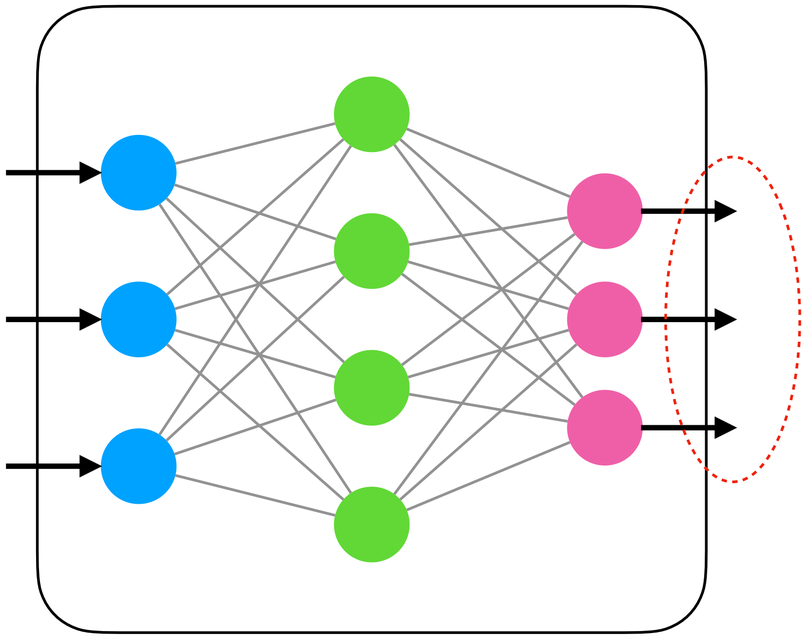
$$CE = -\frac{1}{N}\sum_{i=n}^{N} \log \hat{y}_i$$

$$NLL = -\sum_{i=1}^{n} \log \hat{y}_i$$

Minimize cross entropy loss

Minimize negative log likelihood

What is Softmax

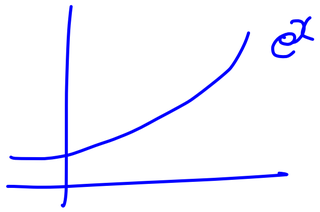


$$\hat{y}_i = \frac{e^{x_i}}{\sum_{j=1}^{C} e^{x_j}}$$

입력을 확률분포로 변환하는 함수

What is Softmax

$$\hat{y}_i = \frac{e^{x_i}}{\sum_{j=1}^{C} e^{x_j}}$$



x	e^x	$\hat{y}$
0.9	2.46	0.38
1.2	3.32	0.52
-0.4	0.67	0.10

$$\sum_{j=1}^{C} e^{x_j} = 6.45$$

$$\sum_{j=1}^{C} \hat{y}_j = 1$$

Softmax and Cross Entropy loss

$$CE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} \log \hat{y}_{ij}$$

$\hat{y}$
0.38
0.52
0.10

y
0
1
0

CE
0
0.66
0

$= \lg \frac{1}{0.52}$

$\lg y - \lg \hat{y}$

$= \lg \frac{y}{\hat{y}} = \lg\left(\frac{1}{0.52}\right)$

감사합니다.