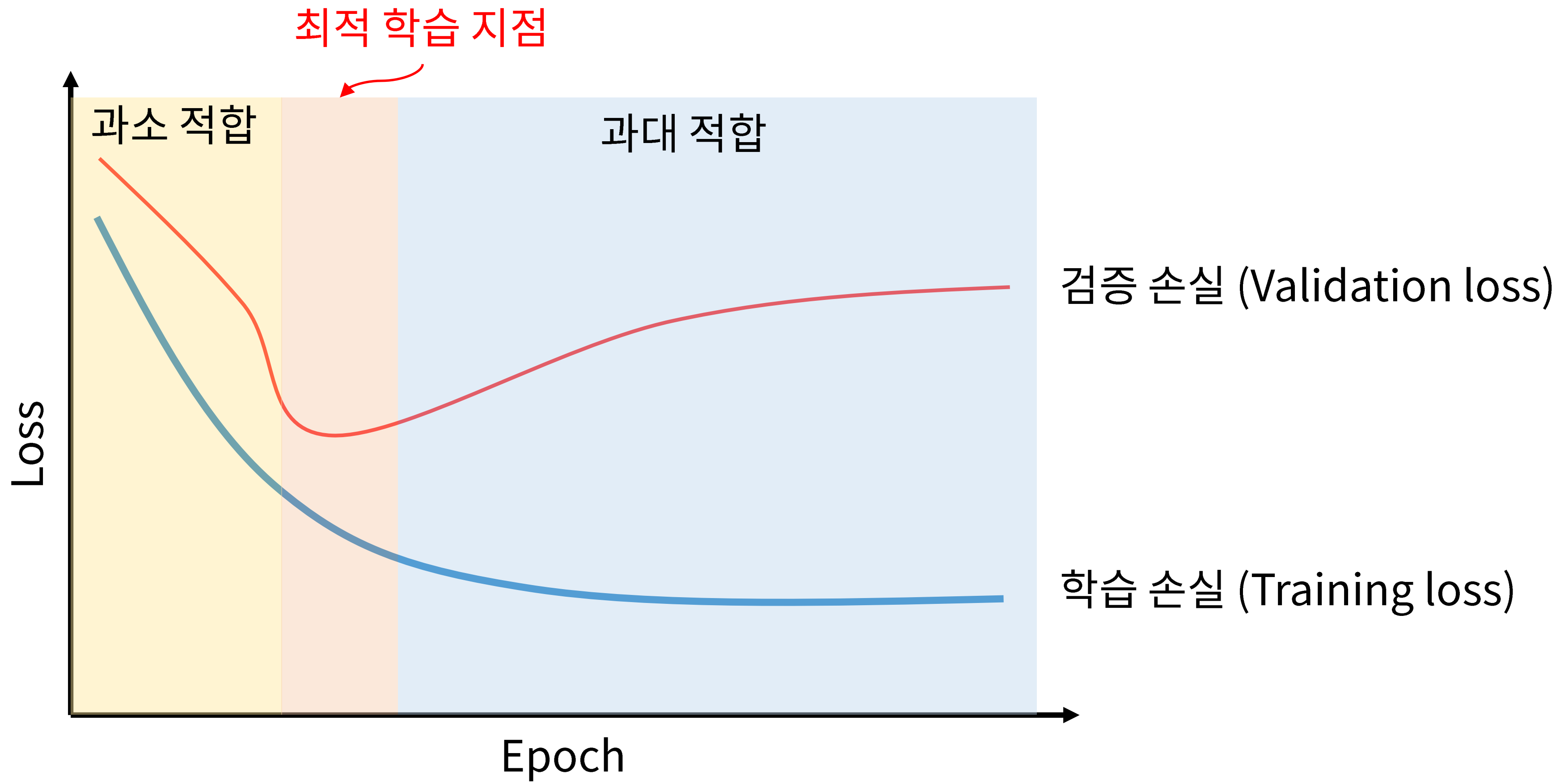
Fast Campus **Online** | Life Changing **Education**

Chapter 08. 효과적이면서도 쉽게 쓸 수 있는 기법들

STEP2.
정규화 기법의 이해

과적합 현상



실제로 학습을 진행하다 보면, 학습 데이터에 과적합(Overfitting)되는 현상이 발생한다.

Loss vs. Complexity

$$Cost = Loss(Data|Model) + \lambda Complexity(Model)$$

Loss에 집중 : 학습 데이터에 대한 신뢰도가 높음. 학습 데이터에 속하지 않은 입력에 취약.

Complexity를 낮춤 : 모델의 복잡도가 지나치게 높아지지 않도록 제약. 데이터 학습보다 일반화에 투자.

L-2 정규화 L-2 Regularization

$$Cost = Loss(Data|Model) + \lambda Complexity(Model)$$

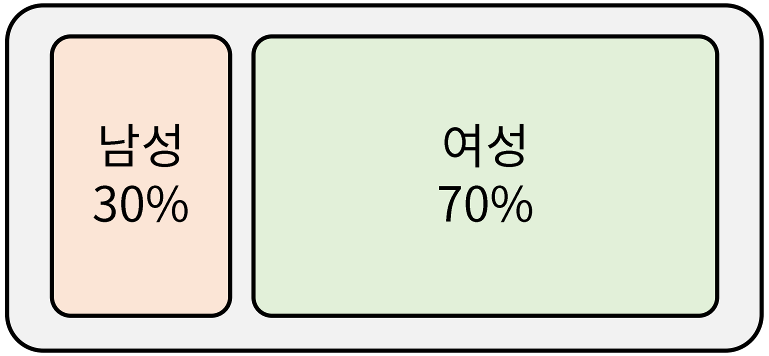
$$\boxed{Complexity(Model) = \frac{1}{N}\sum_{i}\frac{1}{2}w_i^2 = \frac{1}{2}\|\boldsymbol{w}\|^2}$$

- 아주 큰 가중치에 대해 패널티 부여
- 더 구불구불한 것 보다, 더 평평한 형태를 선호
- 베이지언 사전 확률 분포 (정규 분포)

L-2 Regularization (Ridge)는 가중치의 L-2 Norm을 최소화 하는 방법이다.

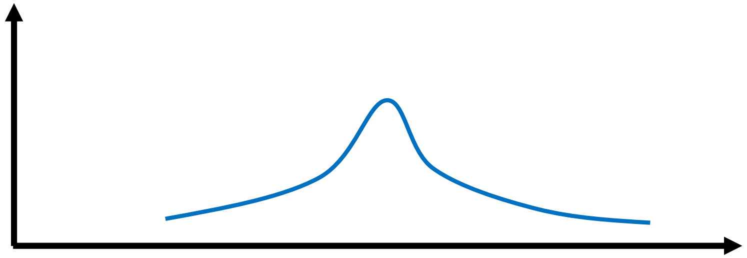
가중치가 정규분포의 형태를 이루도록 한다.

사전 확률 분포 Prior Probability Distribution



우리 학급에 어떤 학생이…

"이 학생은 0.7의 확신도로 여성이군!"



평균 70.0, 분산 5.2

영어 시험의 반 평균이 …

"반 평균은 70점 근방일 확률이 가장 높겠군!"

사전 확률

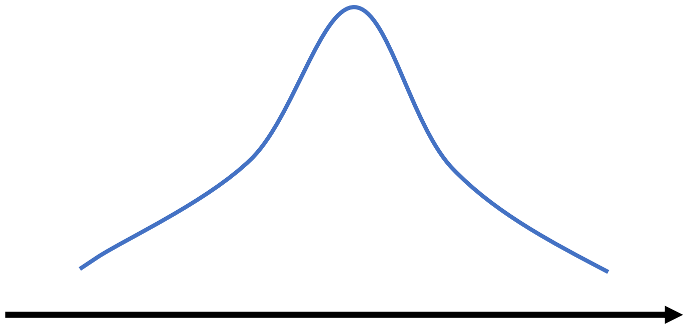
사전 확률 기반 예측

사전 확률 분포는 *a priori*라고도 하며, 데이터를 보기 전에 확률 분포를 예측하는 것을 말한다.

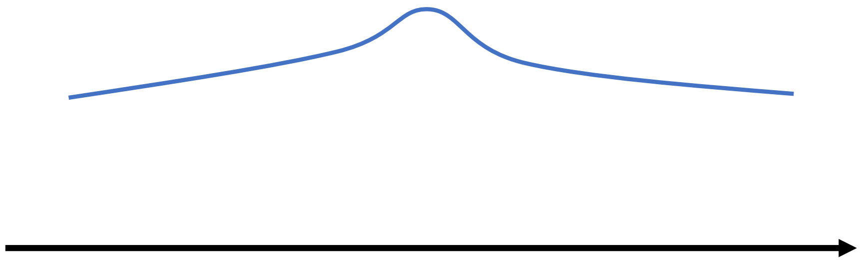
베이지언 통계학에서 많이 사용하는 방법이며, 여기서는 개념만 짚고 넘어가자.

L-2 정규화와 람다 값

$$Cost = Loss(Data|Model) + \lambda \frac{1}{2}\|\boldsymbol{w}\|^2$$



높은 λ일 때의 가중치 분포



낮은 λ일 때의 가중치 분포

람다 값이 크면 가중치는 정규 분포에 가깝게 나타난다.

람다 값이 0에 가까울 수록 정규화가 이루어지지 않으며, 가중치는 평평한 분포를 지향한다.

L-1 정규화 L-1 Regularization

$$Cost = Loss(Data|Model) + \lambda Complexity(Model)$$

$$Complexity(Model) = \frac{1}{N}\sum_{i}|w_i| = \|\boldsymbol{w}\|_1$$

- 가중치의 절대값에 패널티를 줌
- 값이 양수 또는 음수로 존재하면 줄이려 함
- 값이 희소(Sparse)해 지는 특성이 있음
- 베이지언 사전 확률 분포 (라플라스 분포)

L-1 Regularization (Lasso)는 가중치의 L-1 Norm을 최소화 하는 방법이다.
가중치가 라플라스 분포의 형태를 이루도록 한다.