

# 오버피팅

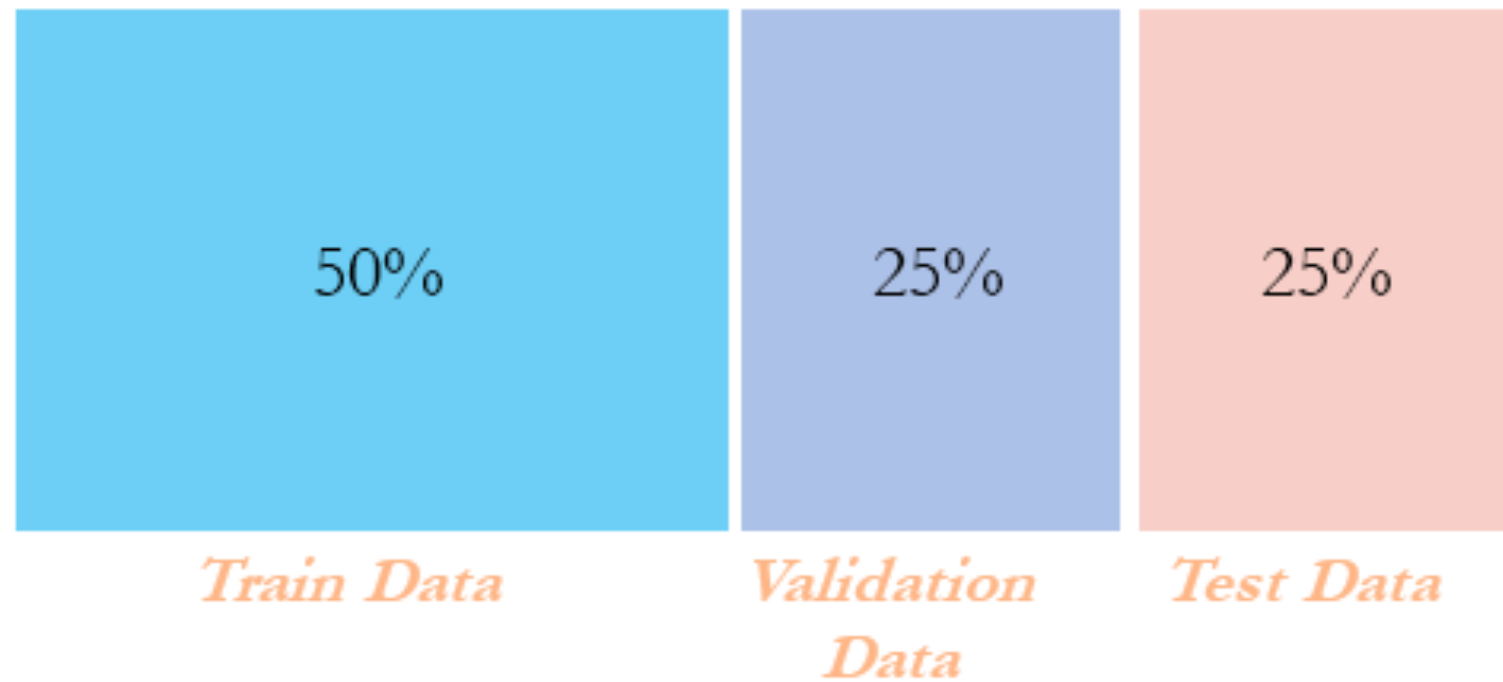
Overfitting 이 일어나는 이유와 방지 방법

DAY 1

JARAM DeepLearning Study

박상원

## *Machine Learning Modeling Methodology*

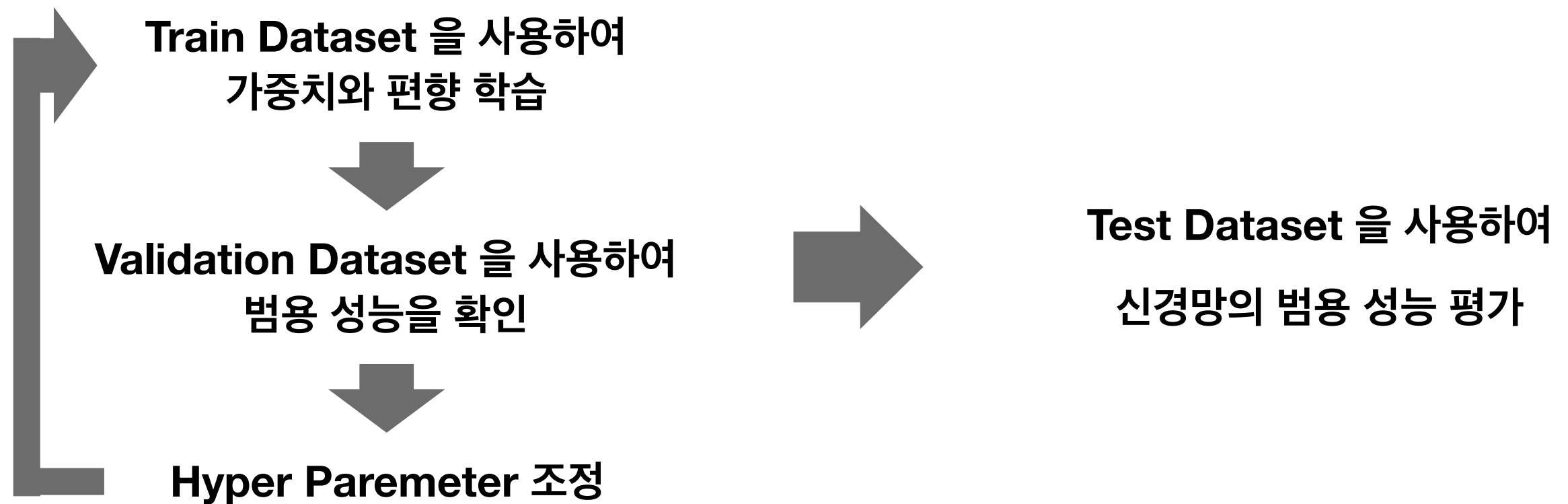


**Train Data**      학습을 위한 데이터

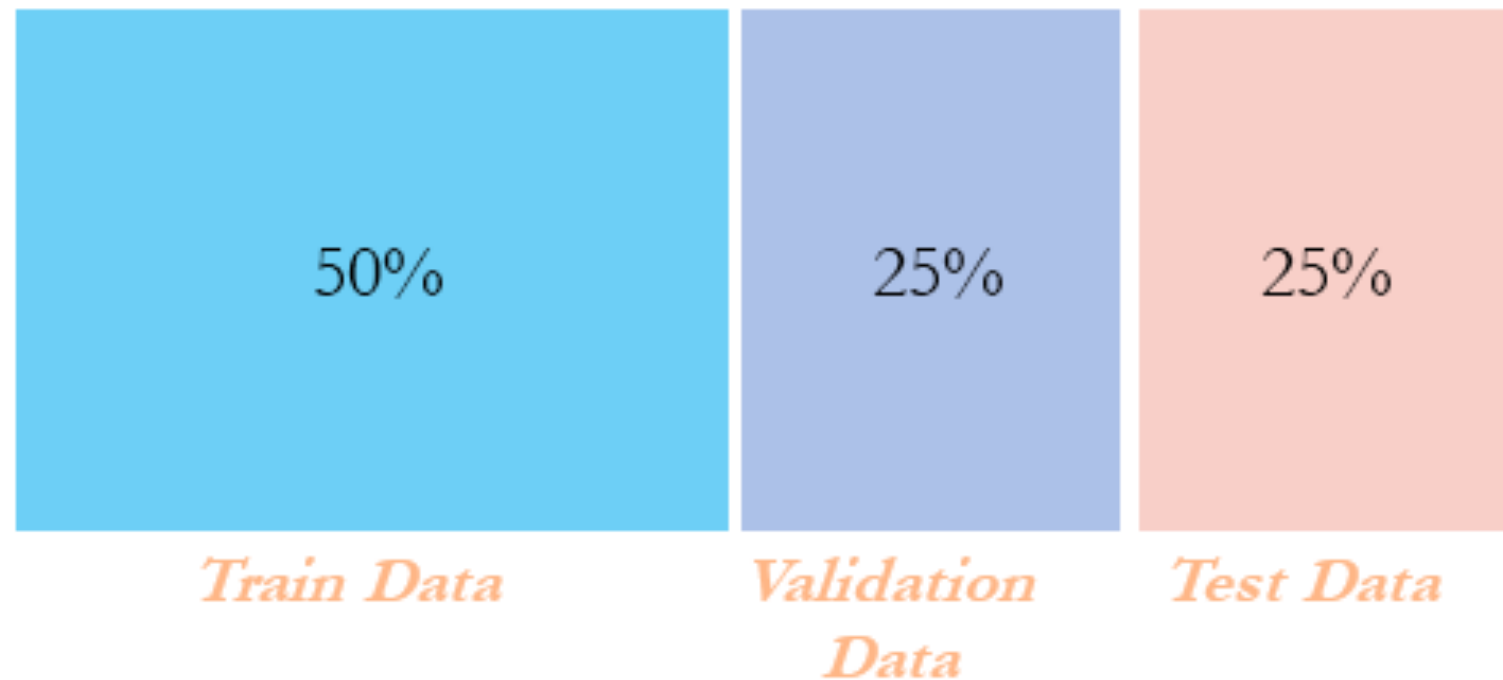
**Validation Data**      하이퍼 파라미터의 성능 평가를 위한 데이터

**Test Data**      학습된 신경망의 범용 성능 평가를 위한 데이터

# Test Dataset 과 Validation Dataset 의 차이



## *Machine Learning Modeling Methodology*



**Train Data**      학습을 위한 데이터

**Validation Data**      하이퍼 파라미터의 성능 평가를 위한 데이터

**Test Data**      학습된 신경망의 범용 성능 평가를 위한 데이터

} 학습과정에 관여

# 오버피팅

높은 Variance 를 띄는 상태

신경망이 훈련 데이터(Training Data)에만 지나치게 적응하여  
그 외의 데이터에는 제대로 대응하지 못하는 상태

## 오버피팅이 발생하는 경우

- 매개변수가 많고 표현력이 높은 모델을 사용할 경우
- 훈련 데이터가 적을 경우

# Bias - Variance Tradeoff

$$\text{Total Error} = \text{Bias}^2 + \text{Variance} + \text{Noise}$$

<b>Bias</b> <sub>(편향)</sub>	예측한 값의 평균과 실제 값의 차이
<b>Variance</b> <sub>(분산)</sub>	예측한 값들이 예측한 값의 평균을 중심으로 퍼져있는 정도
<b>Noise</b>	데이터 자체가 가지는 본질적인 한계

모델 복잡도를 높이면 Variance 가 커져서 오버피팅이 일어나고

모델 복잡도를 낮추면 Bias 가 커져서 언더피팅(Under fitting)이 일어나는데

이 때의 Bias 와 Variance 는 서로 상반되어서 B-V Tradeoff 라고 함



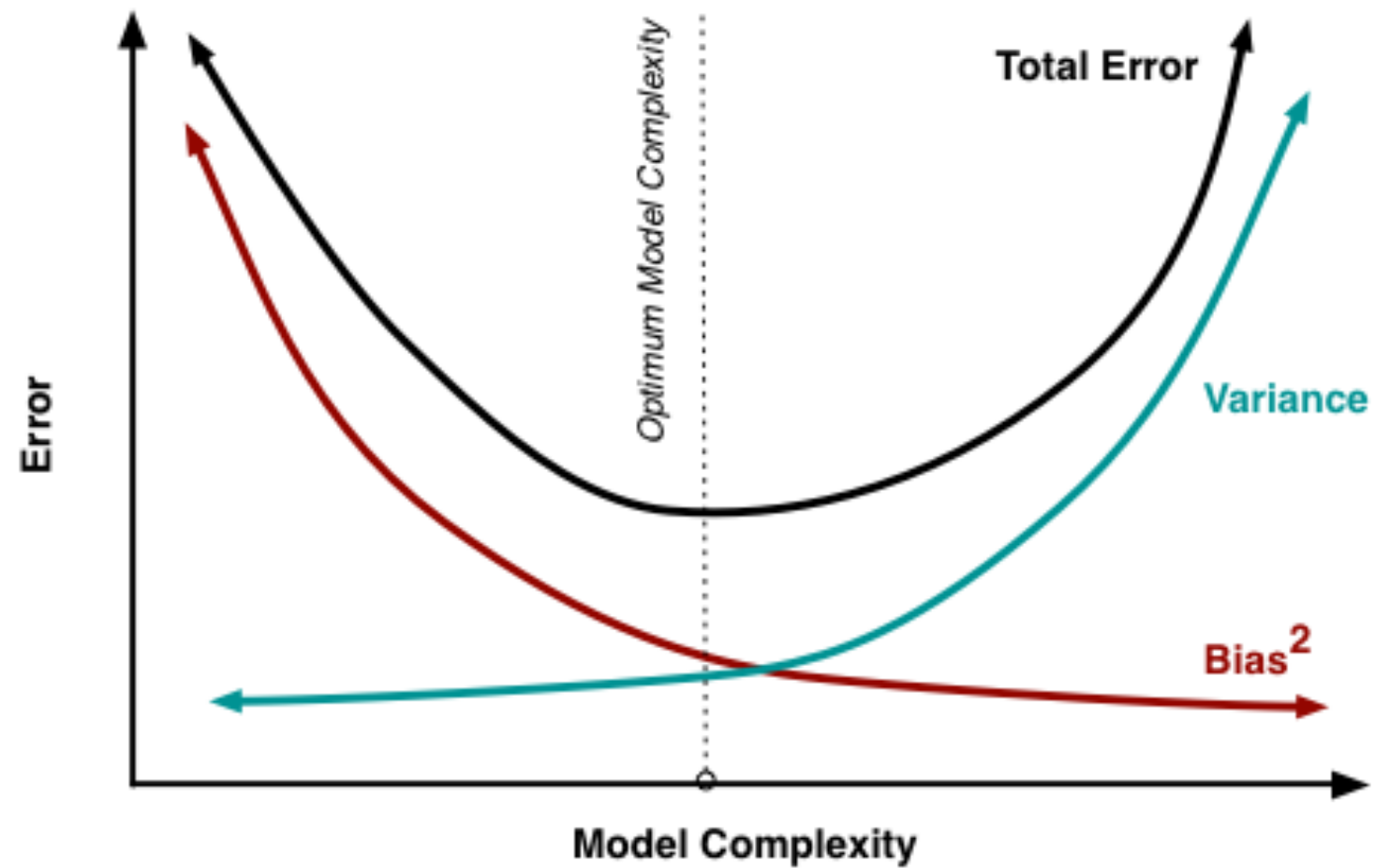
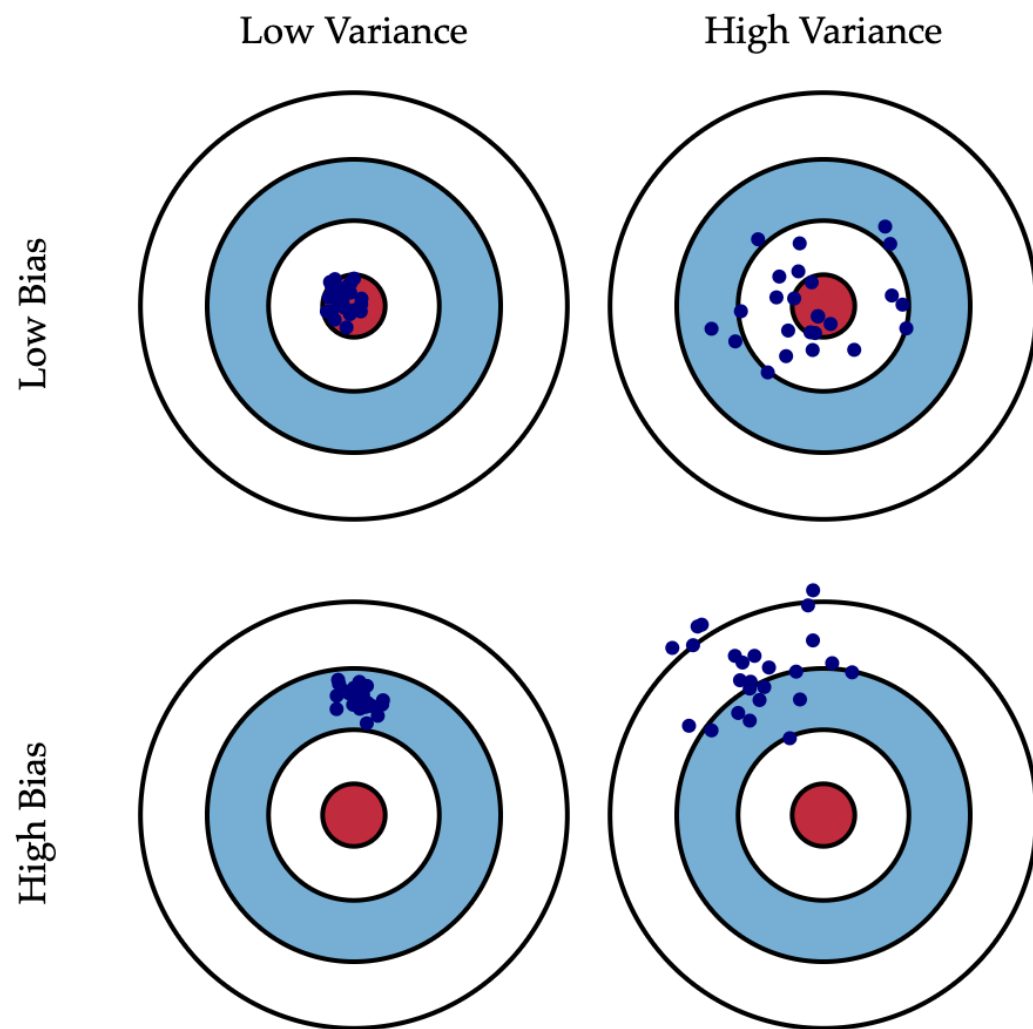


그림 출처 : <http://scott.fortmann-roe.com/docs/BiasVariance.html>

그림 출처 : <http://scott.fortmann-roe.com/docs/BiasVariance.html>

# 오버피팅

높은 Variance 를 띄는 상태

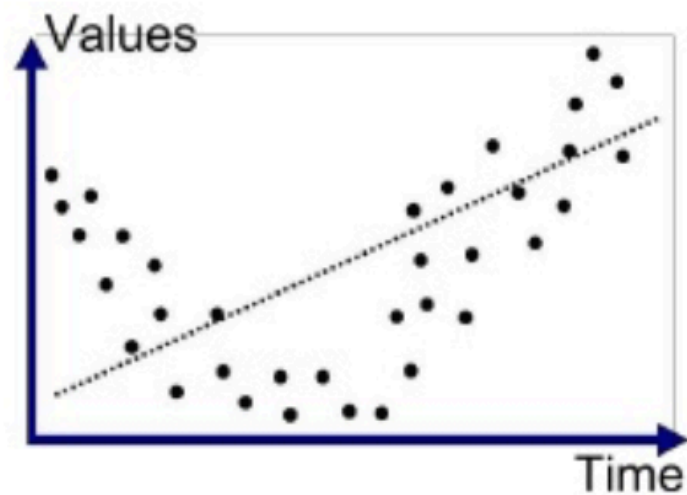
신경망이 훈련 데이터(Training Data)에만 지나치게 적응되어  
그 외의 데이터에는 제대로 대응하지 못하는 상태

## 오버피팅이 발생하는 경우

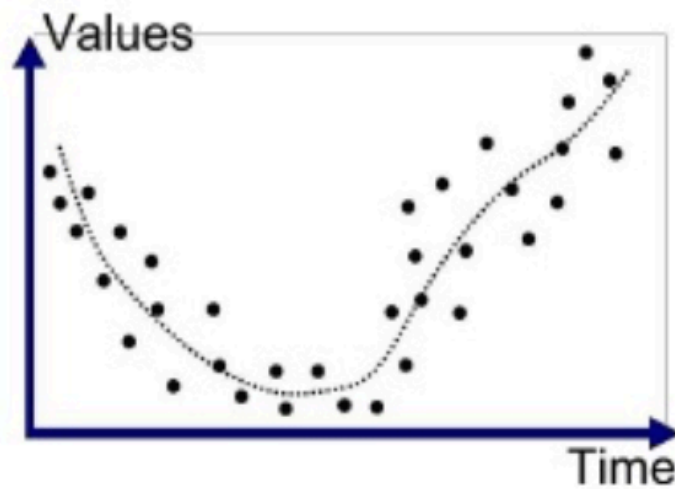
- 매개변수가 많고 표현력이 높은 모델을 사용할 경우
- 훈련 데이터가 적을 경우



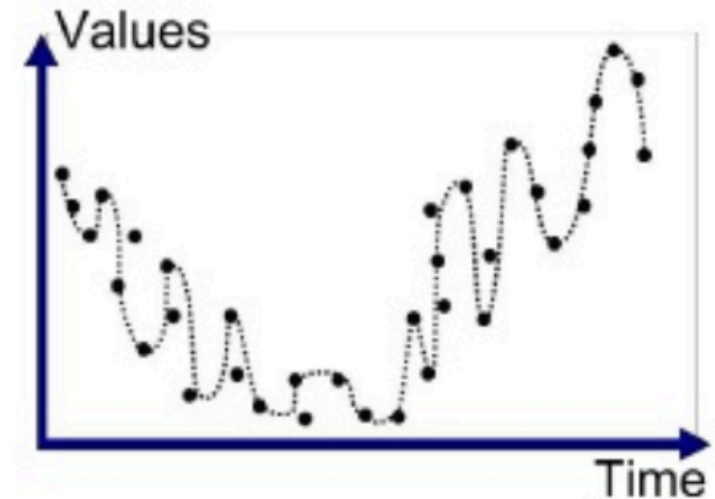
# Example



high bias  
low variance



medium bias  
medium variance



low bias  
high variance

위의 세가지 그림은 서로 다른 모델에서 값을 예측하는 그래프이며 점을 정답으로, 선을 모델의 예측값으로 표현하였다. **Bias** 는 선과 점 사이의 간격이 얼마나 넓은지, **Variance** 는 선의 변동(fluctuation)이 얼마나 많은지 살펴보면 된다.

# 오버피팅을 방지하는 방법

Variance 를 낮추고, Bias 를 높이면 된다!

1. Regularization (일반화)
2. Dropout
3. Hyperparameter Optimization (하이퍼파라미터 최적화)
4. Data Augmentation (데이터 확장)
5. Transfer Learning (전이 학습)
6. Early Stopping (조기 종료)

# Regularization

L2 regularization 을 이용한 Weight Decay

$$\begin{aligned} L_2 &= \sqrt{\sum_i^n x_i^2} \\ &= \sqrt{x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2} \end{aligned}$$

오버피팅은 보통 가중치 매개변수의 값이 커서 발생한다. 따라서 이 방법은 학습 과정에서 큰 가중치에 대해서는 그에 상응하는 큰 페널티를 부과하여 오버피팅을 억제하는 방법이다.

L2 Regularization 을 사용함으로써 이상치(Outlier)의 영향을 적게 받을 수 있다.

L1 Regularization 도 있는데 이 일반화는 상수 값을 빼주도록 되어 있기에 작은 가중치들은 0에 수렴하게 되어 몇개의 중요한 가중치만 남는다.

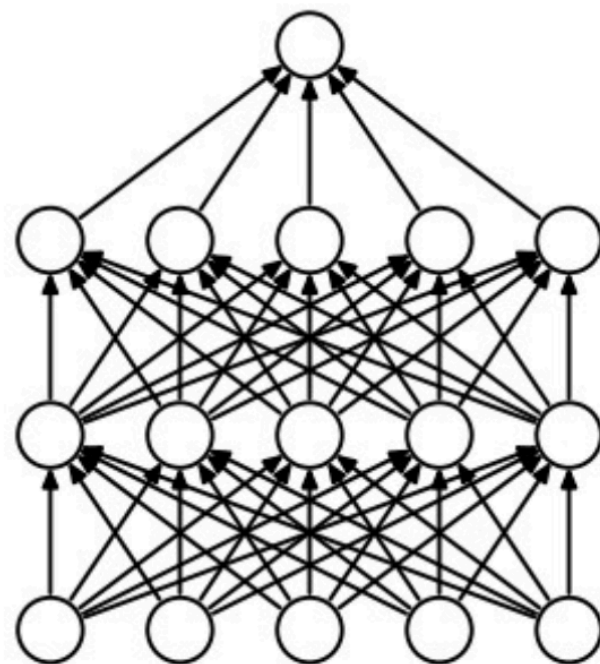
따라서 몇 개의 의미있는 가중치를 원할때에는 L1 Regularization 을 사용하지만, 일반적으로 L2 Regularization 을 사용한다.

# Dropout

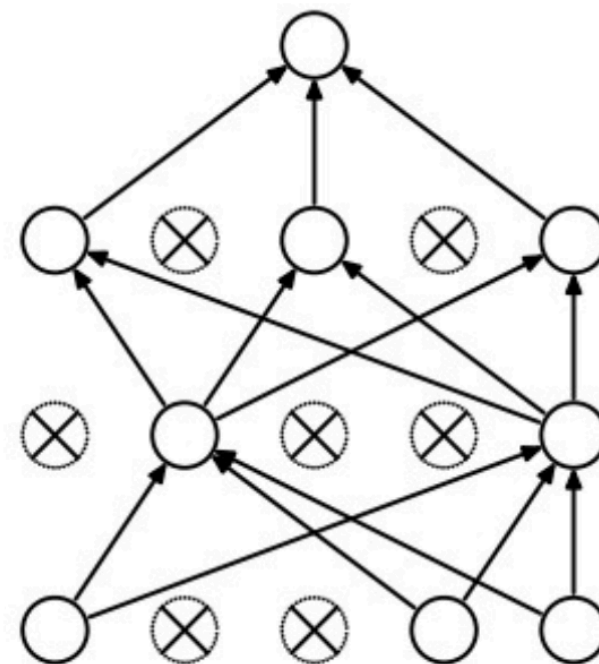
각 계층마다 뉴런을 임의로 Drop 하여 나머지 뉴런만 학습하는 방법

신경망 모델이 복잡해져 **Regularization** 만으로 대응하기 어려워질 경우 사용하는 방법이다.

Dropout은 뉴런을 임의로 삭제하면서 학습하는 방법이다. 훈련 때는 임의의 뉴런을 골라 삭제하여 신호를 전달하게 하지 않는다. 단, 시험 때는 모든 뉴런에 신호를 전달한다.



(a) Standard Neural Net



(b) After applying dropout.

# Hyperparameter Optimization

Validation Dataset 을 이용한 최적화

대략적인 범위를 설정하고, 그 범위에서 무작위로 값을 선택해 정확도를 평가한 다음  
좋은 정확도를 내는 곳으로 범위를 축소하는 방식을 반복하는 방법이다.

어느 정도 범위가 좁아지면 그 범위 내에서 값 하나를 골라낸다.

범위는 보통 10의 거듭제곱 단위로 지정하며, 이를 로그 스케일(Log Scale)로 지정한다고 한다.

딥러닝 학습에는 시간이 오래 걸리기 때문에 학습을 위한 에폭을 작게 하여  
1회 평가에 걸리는 시간을 단축하는 것이 효과적이다.

# Data Augmentation

훈련 데이터가 부족할 때 사용하는 방법

훈련 이미지가 개수가 적을 때, 이미지를 회전하거나, 밝기 변화, 자르기, 뒤집기 등의 약간의 변화를 주어 훈련 데이터의 개수를 늘리는 방법이다.

# Transfer Learning

미리 학습된 가중치를 이용하는 방법

미리 학습된 가중치를 초기값으로 설정하여 새로운 데이터를 대상으로 재학습하는 방법이다.

# Early Stopping

Validation Accuracy 가 더 이상 올라가지 않을 때 학습을 멈추는 방법

반복해서 학습시킬 때 Train Dataset 의 정확도는 올라가나

Validation Dataset 의 정확도가 멈추거나 낮아지는 지점이 오는데 이 때 학습을 멈추는 방법이다.

**Q & A**



# 자료 출처

1. <http://taewan.kim/post/norm/>
2. <https://nittaku.tistory.com/272>
3. <https://nittaku.tistory.com/289>
4. <https://umbum.tistory.com/222>
5. <https://brunch.co.kr/@chris-song/32>
6. <https://www.opentutorials.org/module/3653/22071>
7. [https://modulabs-biomedical.github.io/Bias\\_vs\\_Variance](https://modulabs-biomedical.github.io/Bias_vs_Variance)
8. <http://blog.naver.com/PostView.nhn?blogId=laonple&logNo=220527647084>
9. <https://medium.com/mighty-data-science-bootcamp/%EC%B5%9C%EC%84%A0%EC%9D%98-%EB%AA%A8%EB%8D%B8%EC%9D%84-%EC%B0%BE%EC%95%84%EC%84%9C-%EB%B6%80%EC%A0%9C-bias%EC%99%80-variance-%EB%AC%B8%EC%A0%9C-%EC%9D%B4%ED%95%B4%ED%95%98%EA%B8%B0-eccbaa9e0f50>
10. <https://bywords.tistory.com/entry/%EB%B2%88%EC%97%AD-%EC%9C%A0%EC%B9%98%EC%9B%90%EC%83%9D%EB%8F%84-%EC%9D%B4%ED%95%B4%ED%95%A0-%EC%88%98-%EC%9E%88%EB%8A%94-biasvariance-tradeoff>

The background of the slide features a dark, textured surface, likely a wooden desk. On the left, an open notebook with lined pages is visible, with a yellowish page on the left and a white page on the right. A brown bookmark is tucked into the bottom of the notebook. On the right side, a white ceramic cup filled with dark coffee is partially visible, with a small amount of steam or a reflection on the surface.

# Thank you!

DAY 1

JARAM DeepLearning Study

박상원