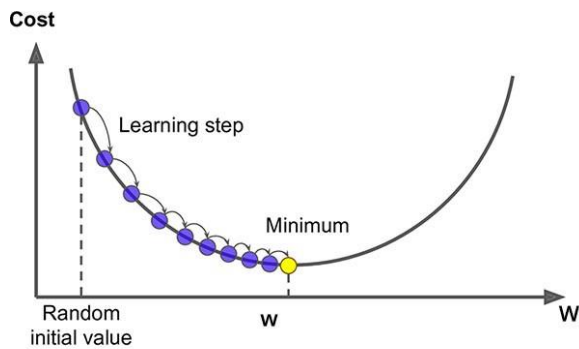


## Gradient Descent (경사하강법)

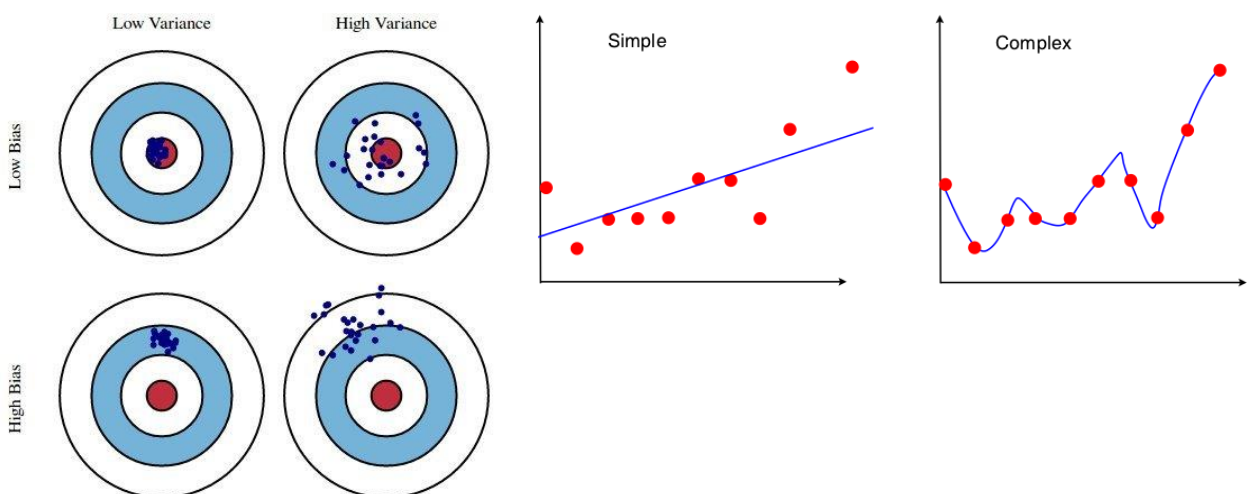


- 비용함수를 최소화하기 위한 알고리즘
- 초기값을 무작위로 설정
- 적당한 Learning rate을 설정하는 것이 중요
- Local minimum, Global minimum
- Scale이 다를 경우 학습이 오래 걸림

-> GD기반 모델들의 조건

이름	내용
BGD	<ul style="list-style-type: none"> <li>-전체 훈련 세트를 사용해 G를 계산.</li> <li>-큰 데이터셋에는 학습이 느림</li> <li>-최적의 G값에 수렴함</li> </ul>
SGD	<ul style="list-style-type: none"> <li>-한 개의 샘플을 무작위로 선택하여 G를 계산</li> <li>-큰 데이터셋에 사용하기 적합함</li> <li>-무작위성을 갖기 때문에 최적의 G값 주변에서 요동침</li> <li>-최적의 G값에 수렴시키기 위해서는 Learning schedule 설정이 필요</li> </ul>
Mini-BGD	<ul style="list-style-type: none"> <li>-Batch단위로 G를 계산</li> <li>-전체데이터를 사용하여 G를 갱신하는 것이 아니기 때문에 G값 주변에서 요동침</li> </ul>

## Bias Variance trade-off



➔ 모델의 복잡도가 커지면 분산이 늘어나고, 편향은 줄어든다. 반대로 모델의 복잡도가 줄어들면 편향이 커지고 분산이 작아진다.

## 규제가 있는 선형모델

이름	내용
Ridge	-선형모델의 비용함수인 MSE 에 <b>L2 규제항</b> 을 더해준 식을 비용함수로 가짐 - SGD 에 penalty 를 'l2'로 준 것으로 사용 가능
Lasso	-선형모델의 비용함수인 MSE 에 <b>L1 규제항</b> 을 더해준 식을 비용함수로 가짐 -덜 중요한 가중치를 제거하려고 함 (0 으로 만들) -SGD 에 penalty 를 'l1'로 준 것으로 사용 가능
Elastic Net	-Ridge 와 Lasso 의 조합 -l1_ratio 로 ridge 와 lasso 정도를 조정가능

➔ **파라미터로 알파를 가짐(값을 늘릴수록 자유도가 낮은 모델 생성)**