Supervised Learning의 종류

(know the answer)

- Regression (회귀)

  연속적인 숫자들을 예측. (차이가 있다)

- Classification (분류)

  범주적인 숫자들 (맞거나 틀리거나)

- UNSupervised Learning. : Don't know Answer.

  PCA.

- Re enforcement Learning.

  상황을 프로그램 환경과 주고 선물을 받게된다.

Supervised Learning의 종류

- 예측하려는 변수의 종류에 따라
- Regression (회귀): 연속 변수를 예측
- Classification (분류): 이산 변수를 예측

회귀 vs. 분류

- 예측하는 변수가 다름(연속 vs. 이산)
- 오류의 형태가 다름

- 회귀: 예측과 실제의 거리
  - 예: 3.84로 예측했는데 4.28 오차를 수치로 산출한다.

- 분류: 예측과 실제의 차이
  - 예: 고릴라로 예측했는데 판다

평가지표의 종류

- 크로스 엔트로피 모델들을 fitting 시킬때 중요.

예측이 90% 의 확률로 A라고 했는데 B인경우 큰패널티를 적용.
49% 51% 인경우는 패널티↓.

틀린정도에따라 패널티강도가 달라진다.

$$H(p,q) = -\sum p(x)\log q(x)$$

- 분류에서 확률로 예측할 때 로그-우도

분류에서 나올 수 있는 경우

		예측	
		양성 Positive	음성 Negative
실제	양성 Positive	진양성 True Positive	위음성 False Negative
	음성 Negative	위양성 False Positive	진음성 True Negative

정확도 Accuracy

		예측	
		양성 Positive	음성 Negative
실제	양성 Positive	진양성 True Positive	위음성 False Negative
	음성 Negative	위양성 False Positive	진음성 True Negative

정밀도 Precision

		예측	
		양성 Positive	음성 Negative
실제	양성 Positive	진양성 True Positive	위음성 False Negative
	음성 Negative	위양성 False Positive	진음성 True Negative

실제 양성 중에 양을 맞춘 확률

재현율 Recall

정밀도 : 양성으로 예측해야 상승

↕

· 양성으로 예측하면 내려간다.

↳ · 정밀도는 같이 높아지기 어렵다.

		예측	
		양성 Positive	음성 Negative
실제	양성 Positive	진양성 True Positive	위음성 False Negative
	음성 Negative	위양성 False Positive	진음성 True Negative

· 정밀도 : 예측 양성 중에 맞춘 확률.

· 재현율 : 실제 양성 중에 맞춘 확률.

Kappa

$$Kappa = \frac{O - E}{1 - E}$$

O : observed accuracy, E : Expected Accuracy

F1 score

- 정밀도와 재현율의 조화평균

$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

LM의 정규화

$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}{n}}$

- RMSE만 최소화하는 대신

평균 제곱근 편차 (Root Mean Square Deviation)

- RMSE + (w의 크기)를 동시에 최적화

Cost function.

(정규화)
모델의 복잡도를 최적화시킬 수 있는 부분

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q$$

$y - (ax+b)$ 최소화

모델이 복잡해지면 (over fitting) ⇒ cross validation으로 완전히 예방X.
새로운 데이터에는 맞지 않을 확률↑.

Lasso

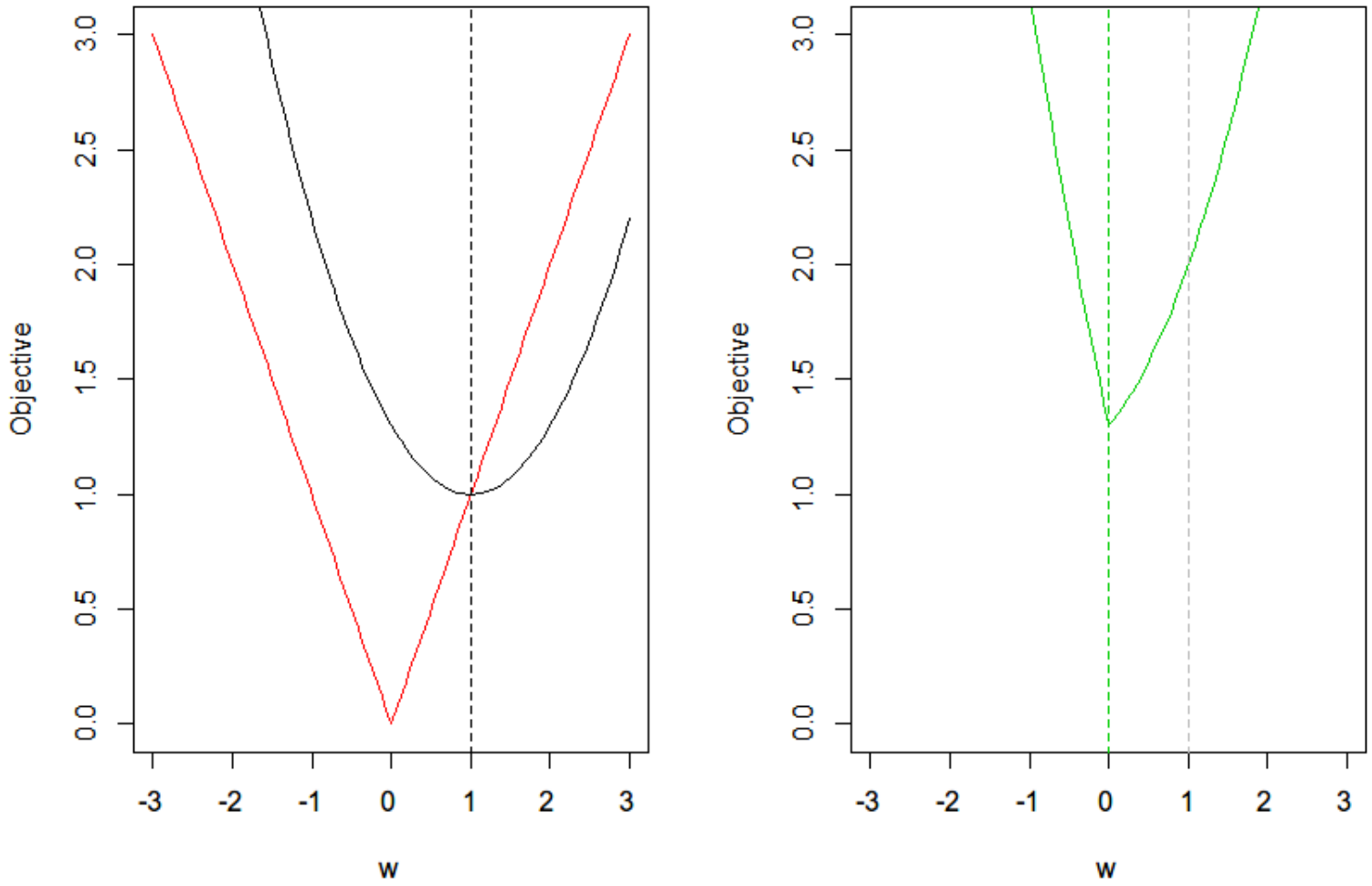
summary : 자잘한 변수들은 W를 0으로

- q = 1
- w의 절대값의 합도 함께 최소화
- w를 0으로 만드는 경향이 있음
- 변수 선택의 기능

y=ax+b
마치 y=0x+b
x. 없게.

가능한 변수를 다 넣고 LASSO를 넣으면. 필요없는 변수들을 제거해 준다.

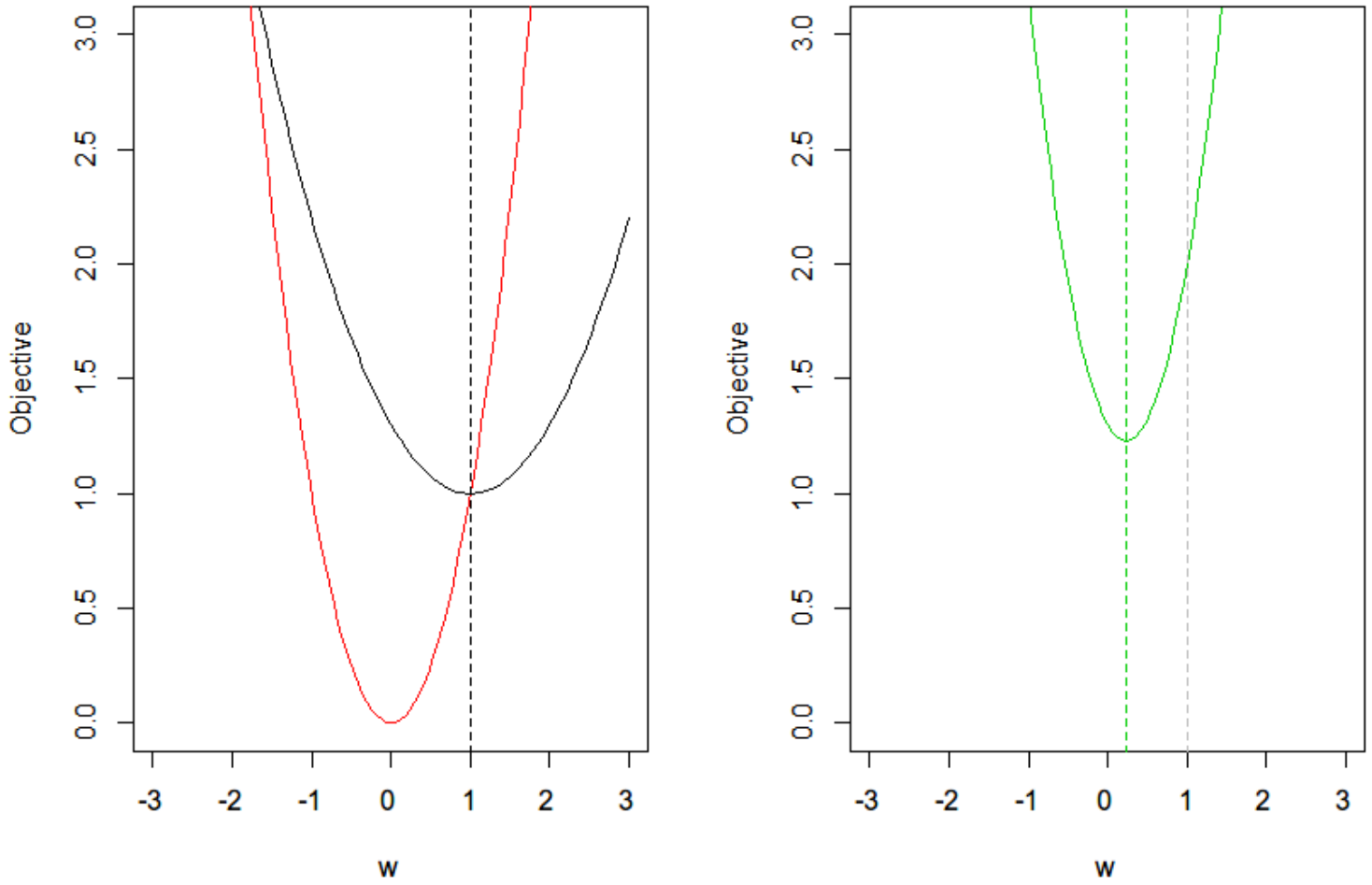
Lasso


Objective
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-3
-2
-1
0
1
2
3
w
Objective
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-3
-2
-1
0
1
2
3
w

Ridge

- q = 2
- w의 제곱의 합도 함께 최소화
- 대체로 Lasso에 비해 예측력이 좋음
- 변수 선택 X

Ridge


Objective
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-3
-2
-1
0
1
2
3
w
Objective
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-3
-2
-1
0
1
2
3
w

Elastic Net

- RMSE + Lasso + Ridge
- Lambda: 정규화항의 가중치
- Alpha: 정규화항에서 Lasso의 비중

→ CV로 결정

Logistic Regression

- Linear Model for Classification