

Supervised Learning의 종류

- Regression
- Classification

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
- 분류: 예측과 실제의 차이
 - 예: 고릴라로 예측했는데 판다

평가지표의 종류

- 크로스 엔트로피

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

- 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

재현율 Recall

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

Kappa

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

O : observed accuracy, E : Expected Accuracy

F1 score

- 정밀도와 재현율의 조화평균

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

LM의 정규화

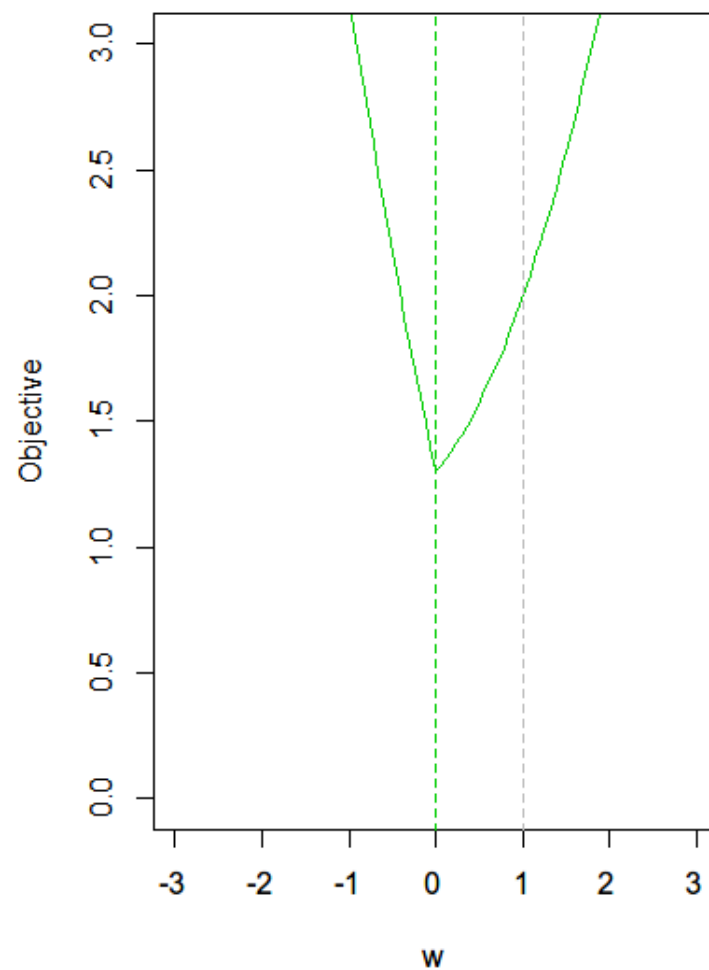
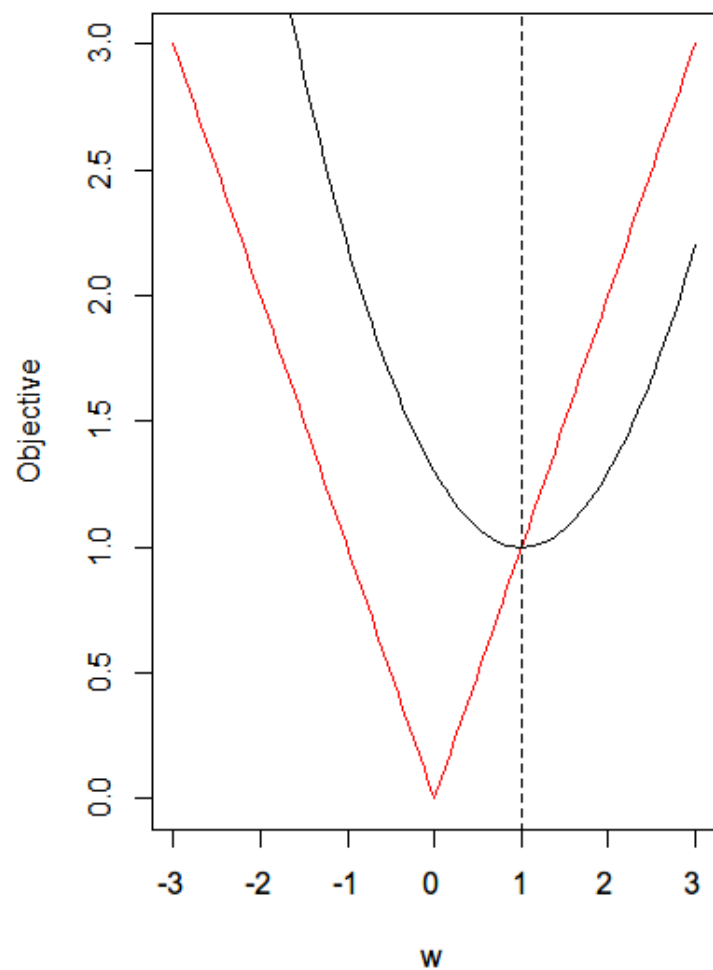
- RMSE만 최소화하는 대신
- RMSE + (w의 크기)를 동시에 최적화

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

Lasso

- $q = 1$
- w 의 절대값의 합도 함께 최소화
- w 를 0으로 만드는 경향이 있음
- 변수 선택의 기능

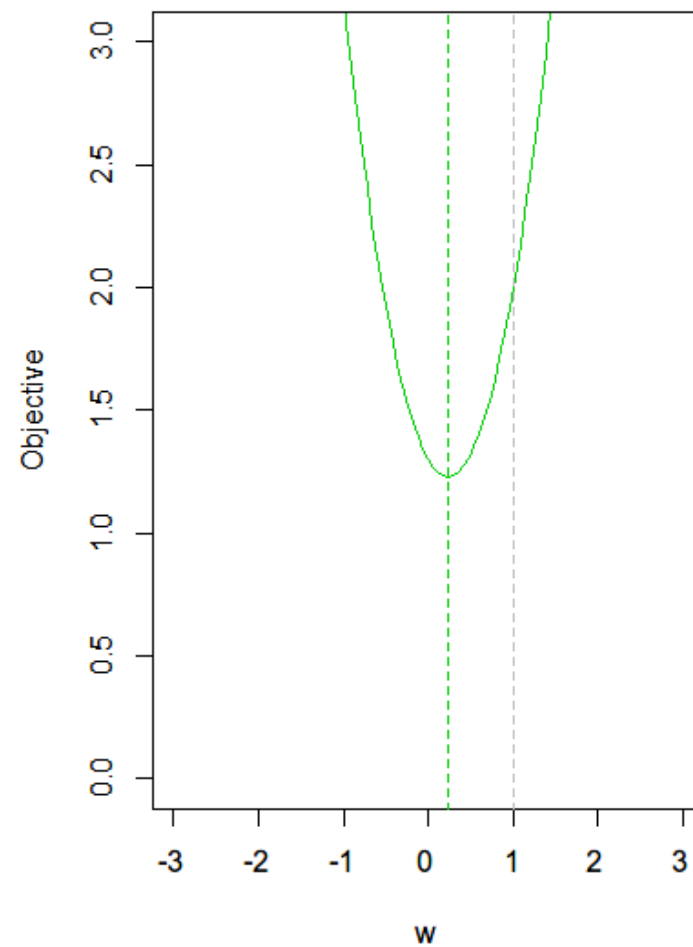
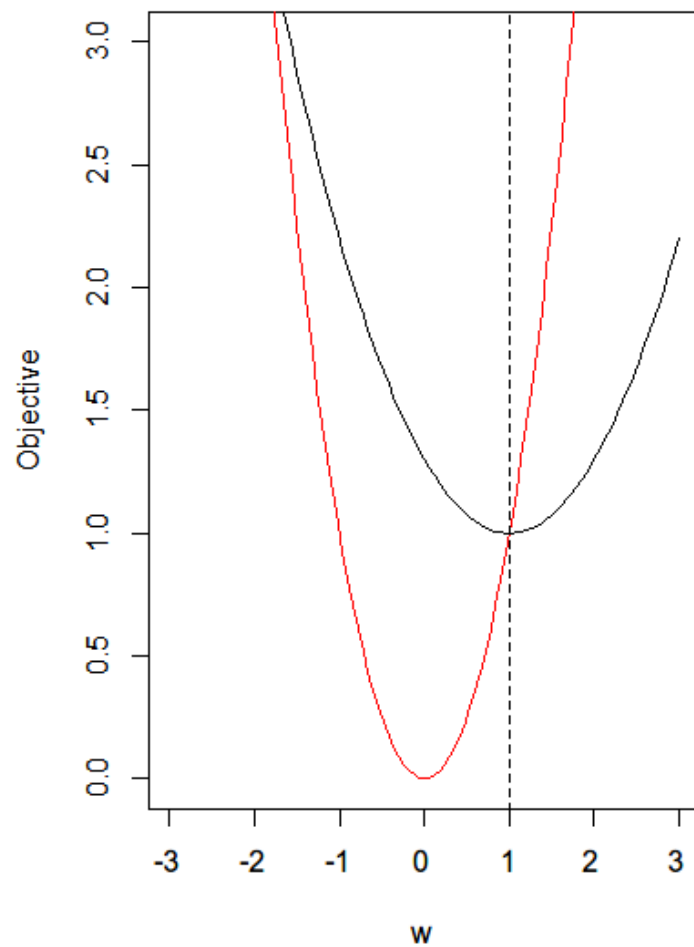
Lasso



Ridge

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

Ridge



Elastic Net

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

Logistic Regression

- Linear Model for Classification

