

Supervised Learning의 종류

(know the answer)

- Regression (회귀)

연속적인 값을 예측. (차이가 있다)

- Classification (분류)

이산적인 값 (맞거나 틀리거나)

- **UNSupervised Learning**. : Don't know Answer.
PCA.

- **Reinforcement Learning**.

어떤 프로그램을 평가가 주어 시행을 하게 된다.

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 산수한다.
- 분류: 예측과 실제의 차이
 - 예: 고릴라로 예측했는데 판다

평가지표의 종류

- 크로스 엔트로피 모델 성능을 Affinity 시킬때 중요.

예를 들어 90% 정확도를 A라고 했을 때 B안정우 큰편차의 경우
49% 51% 안정우는 무시할 수 ↓.

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

- 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

진양성, 진음성이라 한 것만
주게는 정확도

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

정밀도 Precision

ex) 회사에서도 음성인식률 파악 \Rightarrow 완성 \Rightarrow 신제품바나지
 ex) 병원에서도 환자 음성인식률 파악 \Rightarrow 완성 \Rightarrow 음성인식률 파악
 오류 X.

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

\rightarrow 최대한 음성으로 바뀌어리면된다.

ex) 회사 신임직원 \Rightarrow 최대한 파악

재현율 Recall

→ . 정답은 손이 높아진다.

. 실수로 많이 넣어 버린다.

		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

ex) 10km ↑
→ 90km ↓
[평균 80km X]

비율, 속도, 등을 평균내야.

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

LM의 정규화

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

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

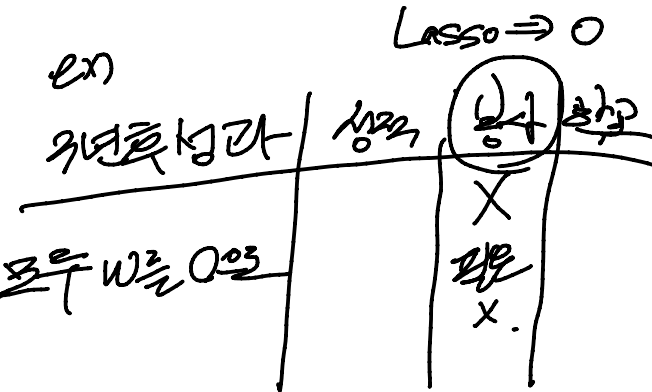
$y - (ax+b)$ 최소화

모델이 복잡해지면 (over fitting) \Rightarrow cross validation으로
 체크해보아야 하는 것이지 ∞ 로 해줄 필요는
 없다.

(정규화)
 모델의 복잡도를
 낮추는 것
 (과적합 방지)

Lasso

Summary : 자칫한것을 모두 w 를 0으로



- $q = 1$

- w 의 절대값의 합도 함께 최소화

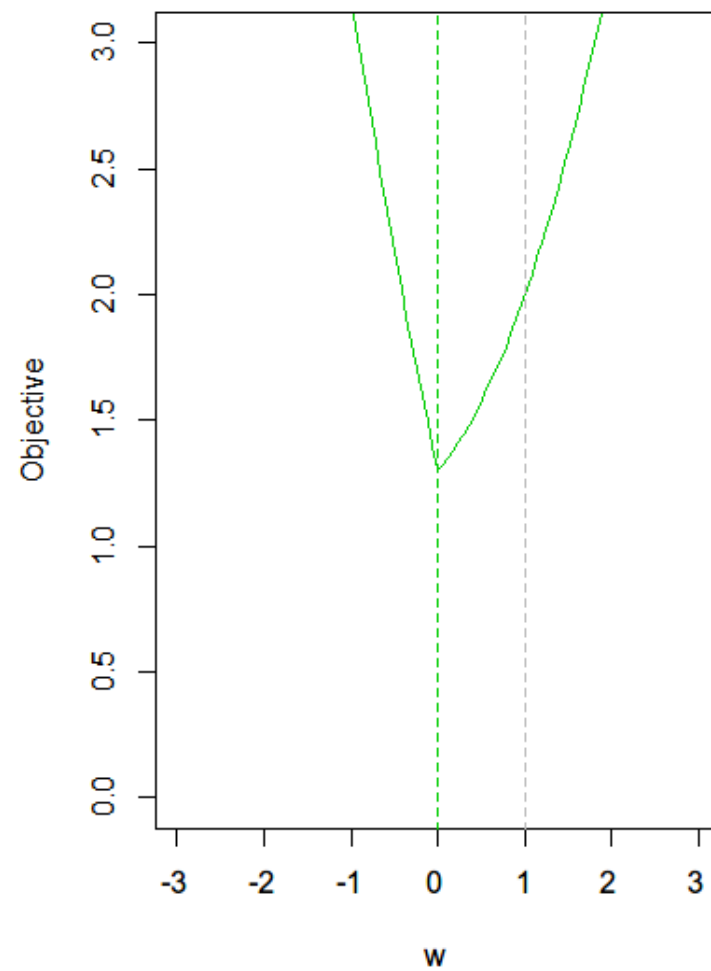
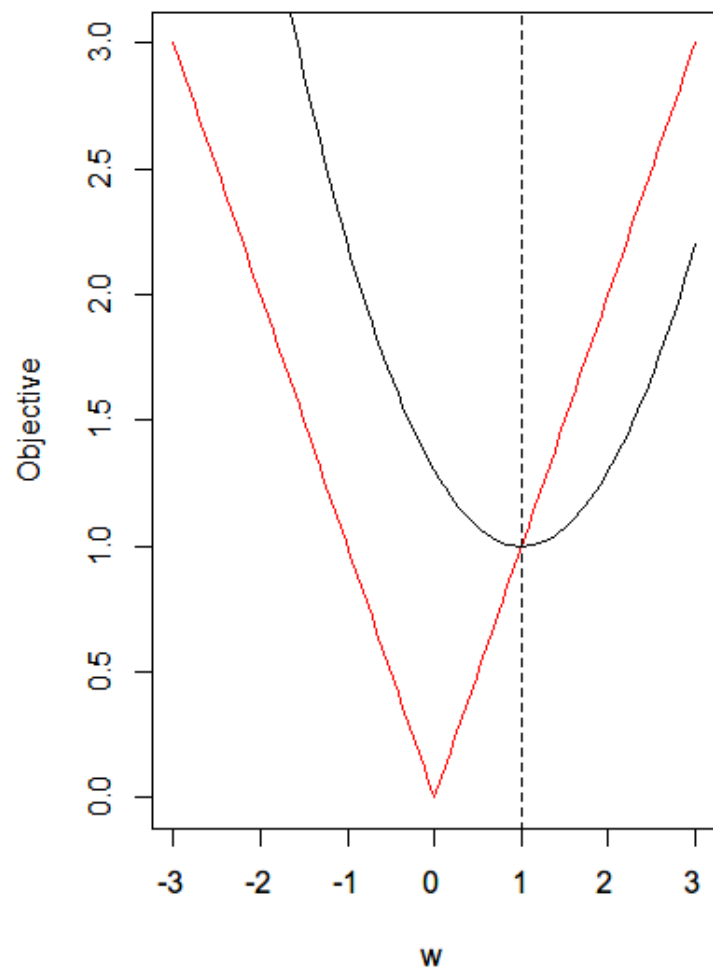
- w 를 0으로 만드는 경향이 있음

$y = ax + b$
 여기서 $y = 0ax + b$
 X. 사례.

- 변수 선택의 기능

가장한변수를 제외한 LASSO를
 넣으면. 최상의 변수를 찾아
 있다.

Lasso

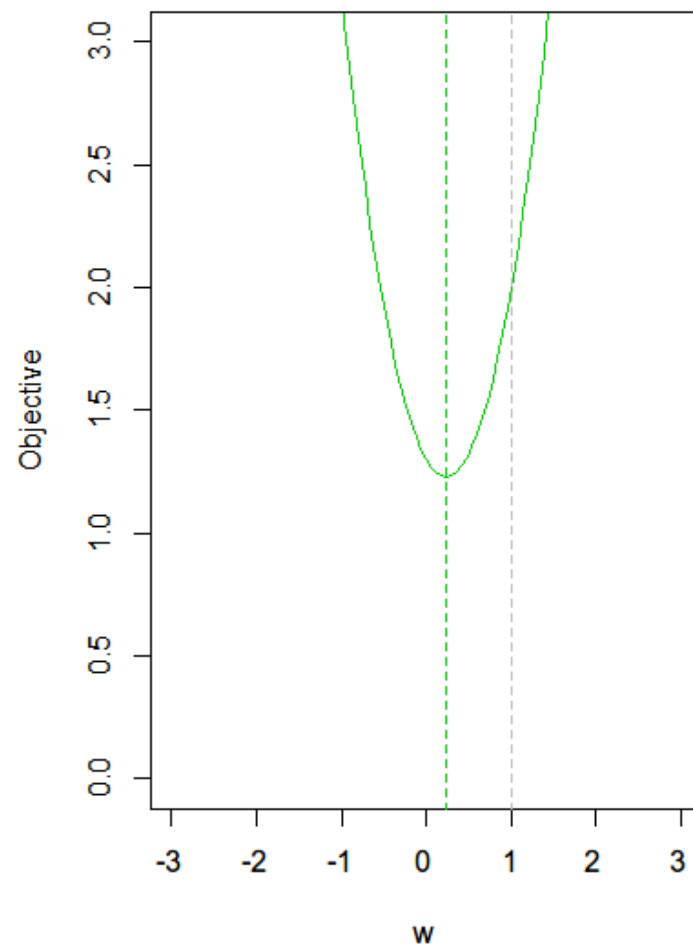
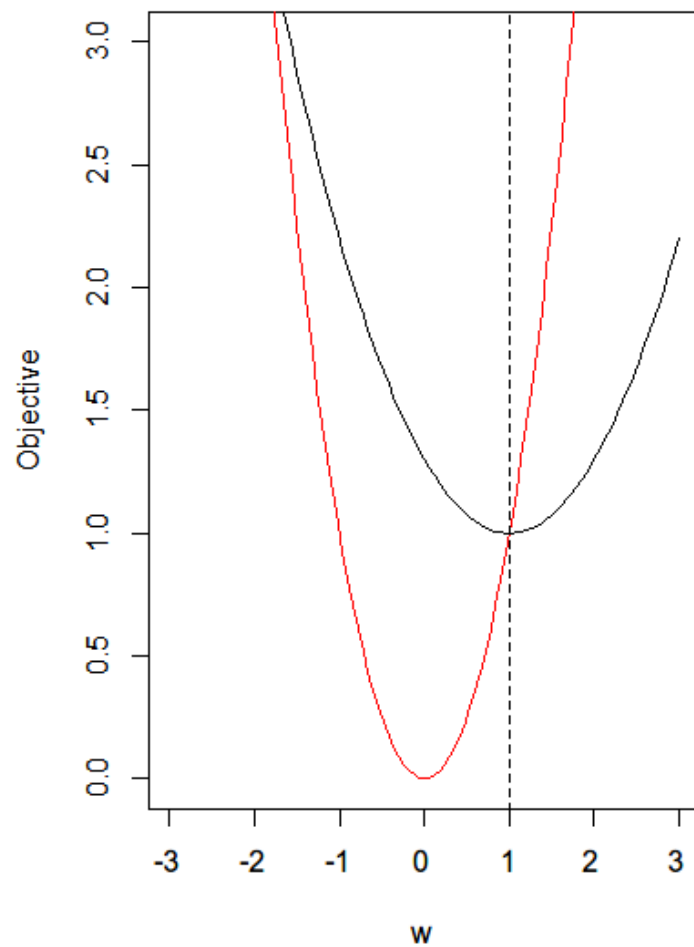


Ridge

일반적으로 Lasso 보다 overfitting
예방능력이 좋다.

- $q = 2$
- 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

