

Prioritizing what to work on

스팸 메일 분류기

지도 학습 :

x = features of email
 y = Spam (1) or not spam (0)
ex. Choose 100 words indicative of spam/not spam
(신체론은 훔치신 만이)

for low error

- 많은 데이터 수집
- 복잡한 features
- 복잡한 algorithm

gut feeling < evidence !!

Error analysis - manually examine

- 간단한 알고리즘 (quick) 으로 검증세트에 적용
- learning curve 로 보완 방향 잡기
- 검증세트 error 분석