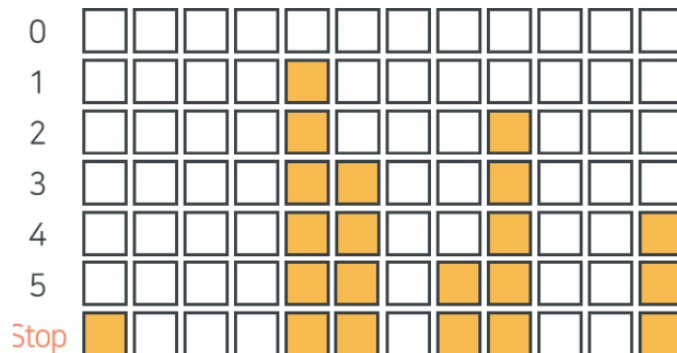
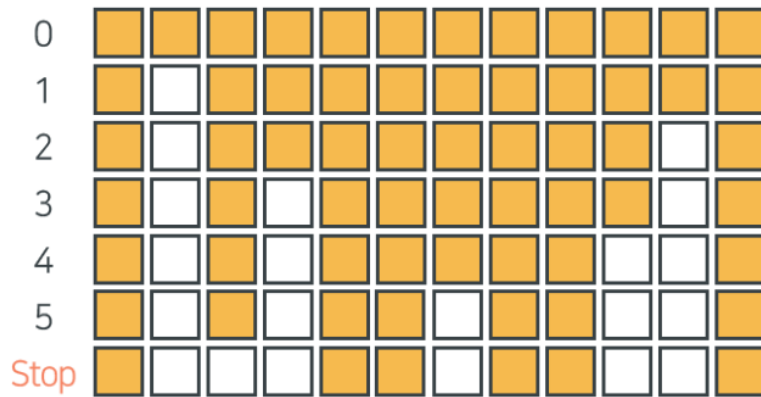


다중회귀분석(변수선택)

- 변수선택의 이유
 - 독립변수의 갯수가 많으면 결과해석에 여러 애로사항이 발생
 - 독립변수들끼리 중복된 정보가 많은 경우,
독립변수들에 대한 회귀계수들 추정에 여러가지 부정적인 영향을 미침
- 변수선택법
 - 전진선택법(foward selection)
 - 절편만있는모델에서 출발하여 중요한 변수를 하나씩 추가하는 방식.
 - 변수를 추가할때는 적절한 검정을 통해 선별되어짐
$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$
$$H_0 : \beta_2 = 0 \Rightarrow \text{축소}$$
$$H_1 : \beta_2 \neq 0 \Rightarrow \text{완전}$$
완전모형과 축소모형을 비교하는 형태로 검정을 실시
사전에 정의된 유의수준을 기준으로 변수 선별

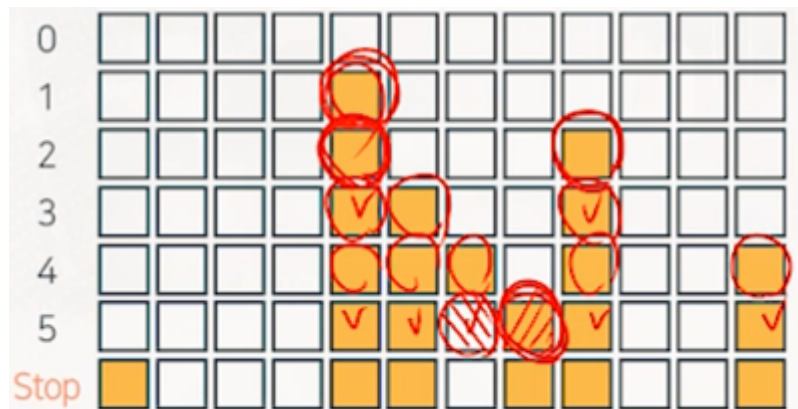


- 한번 선택된 변수는 제거되지 않는 단점이 있음.
- 후진제거법(backward elimination)
 - 모든 변수가 포함된 모델에서 가장 중요하지 않은 변수부터 하나씩제거.
 - 한번 제거된 변수는 선택되지 않는 단점이 있음.



◦ 단계선택법(stepwise method)

- 절편만 포함된 모델에서 출발해 가장 중요한 변수부터 추가하고, 모델에 포함되어 있는 변수 중에서 중요하지 않은 변수를 제거함.
- 더 이상 새롭게 추가되는 변수가 없을때 까지 매 단계마다 변수의 추가 또는 삭제를 반복함



- 모든가능한조합의회귀분석: 모든가능한 독립변수들의 조합에 대한 회귀모형을 생성한 뒤 가장 적합한 회귀모형을 선택
- 모형 선택의 기준
 - R-squared는 변수 선택을 고민하는 문제에선 사용하지 않는 것이 좋음
 - 불필요한 독립변수가 추가되더라도 R-squared값이 증가하게 되므로 무조건 변수를 추가하는 방법은 좋지 않음
 - 수정된 결정계수(Adjusted R^2)
 - 결정계수 R^2 는 새로운 독립변수가 추가되면 항상 증가함
 - 이를 보완한 수정결정계수 Adjusted R^2 는 추가된 독립변수가 종속변수를 설명하는데 기여하는 바가 큰 경우에만 증가함

$$Adjusted\ R^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

- 그 밖에 AIC, Mallow's Cp 등의 다양한 적합도 지표를 이용할 수 있음