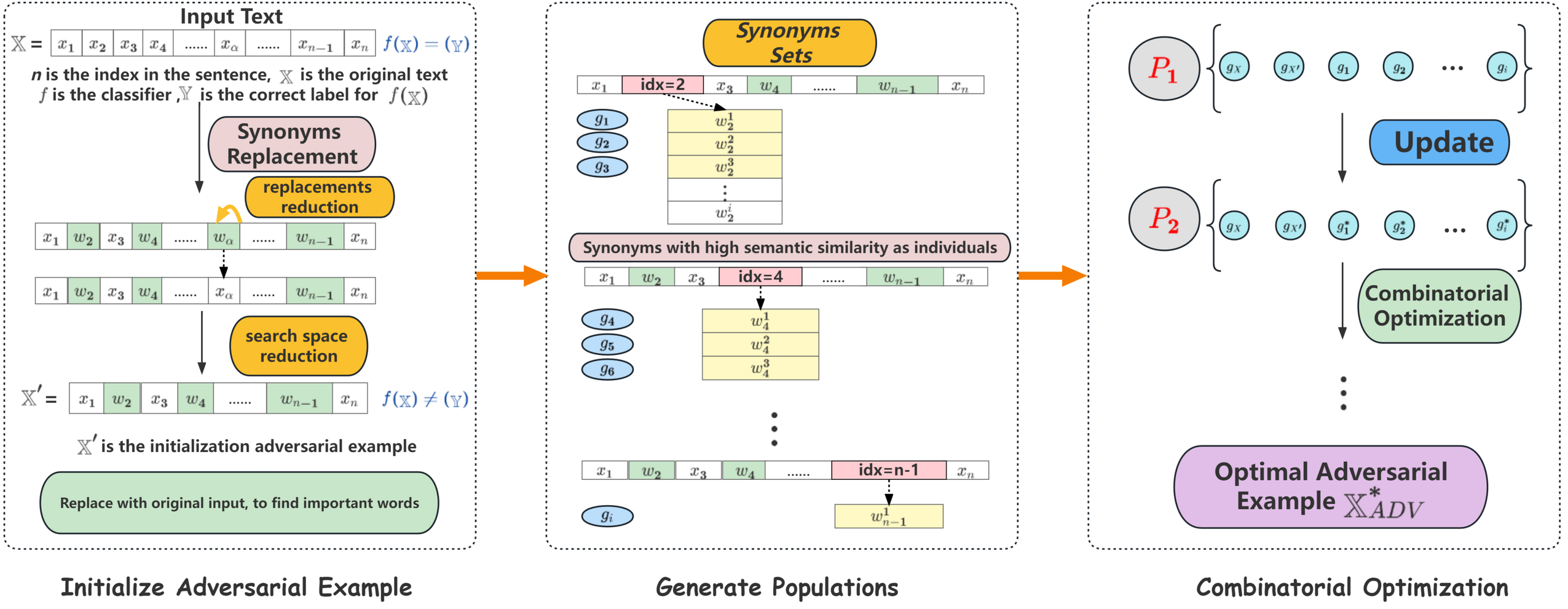
Input Text
$\mathbb{X}$ = x_1 x_2 x_3 x_4 ..... x_α ..... x_{n-1} x_n $f(\mathbb{X}) = (\mathbb{Y})$
n is the index in the sentence, $\mathbb{X}$ is the original text
f is the classifier ,$\mathbb{Y}$ is the correct label for $f(\mathbb{X})$
Synonyms Replacement
replacements reduction
search space reduction
$\mathbb{X}'$ = x_1 w_2 x_3 w_4 ..... w_{n-1} x_n $f(\mathbb{X}) \neq (\mathbb{Y})$
$\mathbb{X}'$ is the initialization adversarial example
Replace with original input, to find important words
Synonyms Sets
idx=2
g_1 g_2 g_3
w_2^1 w_2^2 w_2^3 w_2^i
Synonyms with high semantic similarity as individuals
idx=4
g_4 g_5 g_6
w_4^1 w_4^2 w_4^3
idx=n-1
g_i
w_{n-1}^1
P_1 g_X $g_{X'}$ g_1 g_2 ... g_i
Update
P_2 g_X $g_{X'}$ g_1^* g_2^* ... g_i^*
Combinatorial Optimization
Optimal Adversarial Example $\mathbb{X}_{ADV}^*$
Initialize Adversarial Example
Generate Populations
Combinatorial Optimization