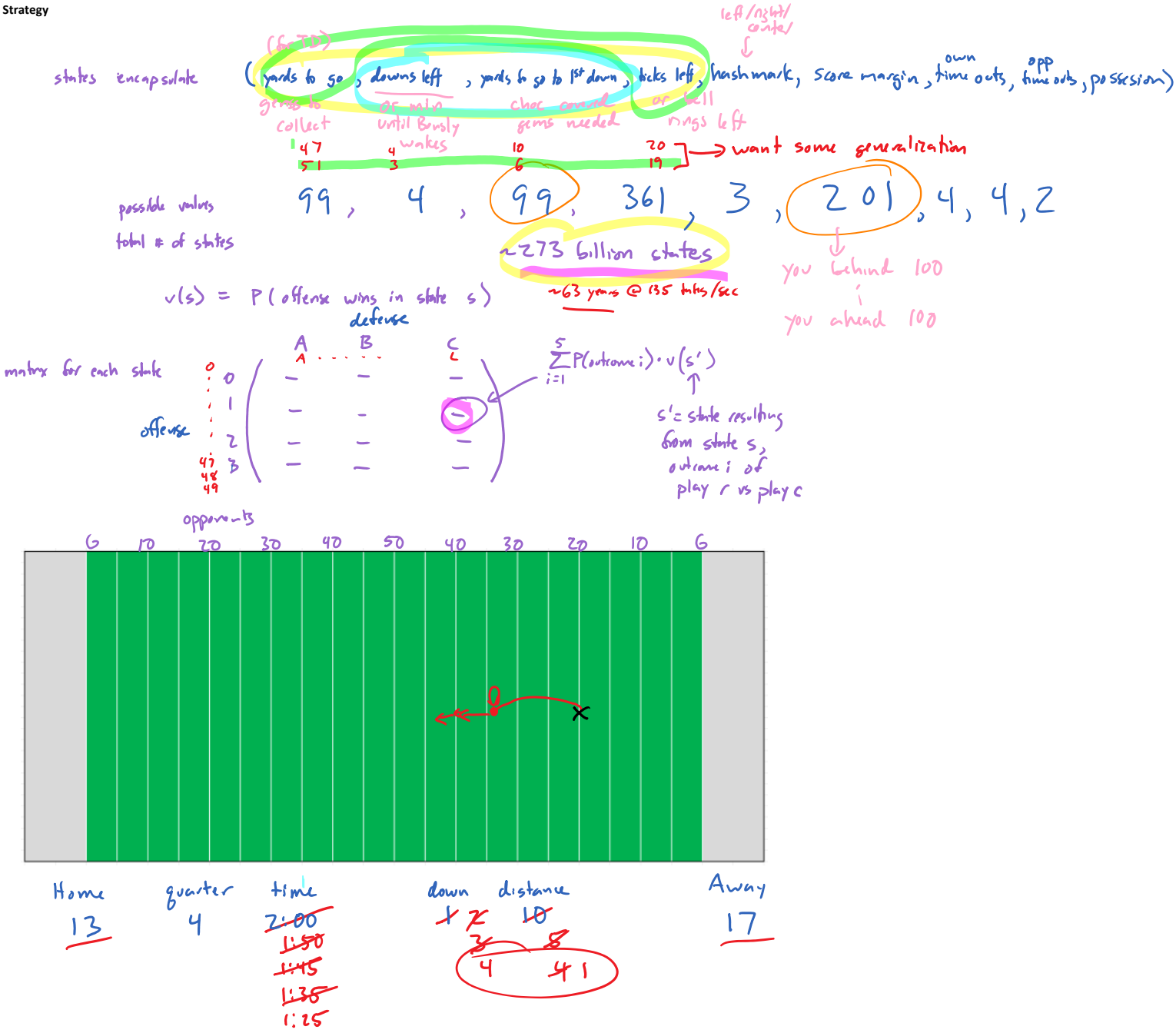
NFL Strategy



Function Approximators — input: state/action output: $\hat{q}(s,a)$ (approximation of exact $q(s,a)$)

Linear Approximator

Define features of state/action pairs

$f_1(s,a) = \begin{cases} 1 & \text{if action } a \text{ moves ball to closest sideline} \\ 0 & \text{otherwise} \end{cases}$

$f_2(s,a) = \begin{cases} 1 & \text{if short yardage } (\leq 2 \text{ yards to get } 1^{st} \text{ down}) \\ 0 & \text{otherwise} \end{cases}$

$f_3(s,a)$

$\vdots$

$f_4(s,a)$

fxn of only s (not a)

↓

make a copy $f_{i,a}$ for each action a

$f_{2,a'}(s,a) = \begin{cases} f_2(s,a) & \text{if } a' = a \\ 0 & \text{otherwise} \end{cases}$

$$\hat{q}(s,a) = w_1 \cdot f_1(s,a) + \cdots + w_n f_n(s,a)$$

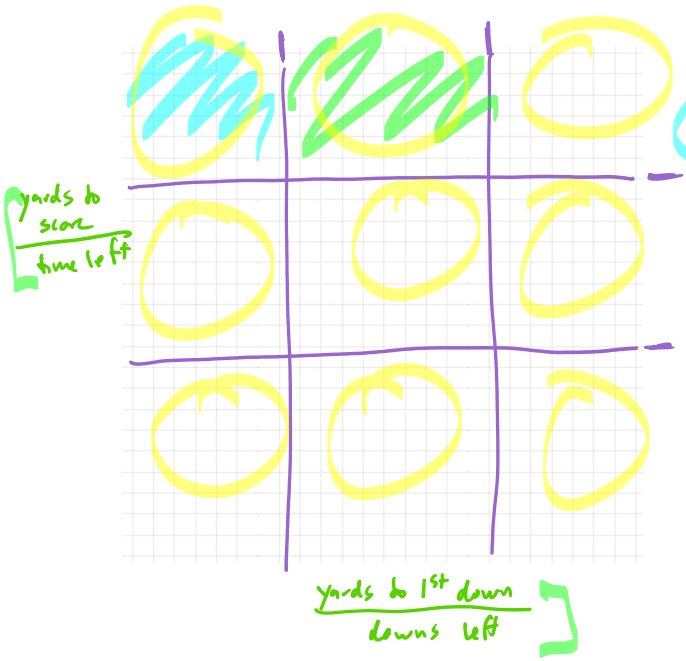
for nonterminal s

↑ learn weights w/ q-learning

for a bunch of episodes
  $S \leftarrow S_0$
  while s not terminal
    In state s
    Choose action a (ε-greedy)
    Observe transition (s, a, r, s')
    Update for each feature i $w_i \leftarrow w_i + \alpha \cdot f_i(s,a) \cdot \left(r + \gamma \max_{a'} \hat{q}(s',a') - \hat{q}(s,a) \right)$

$\hat{q}(s',a_1) = \hat{q}(s',a_2) = \cdots = \hat{q}(s',a_n)$

State Aggregation (Coarse Coding/Buckets/Grid)



$f_{11}(s,a) = \begin{cases} 1 & \text{if state } s \text{ belongs in } 1,1 \\ 0 & \text{otherwise} \end{cases}$

$f_{12}(s,a) = \begin{cases} 1 & \text{if state } s \text{ belongs in } 1,2 \\ 0 & \text{otherwise} \end{cases}$

$\vdots$

$f_{11,1}(s,a) = \begin{cases} 1 & \text{if state } s \text{ belongs in } 1,1 \text{ and } a=1 \\ 0 & \text{otherwise} \end{cases}$

$f_{11,2}(s,a) = \begin{cases} 1 & \text{if state } s \text{ belongs in } 1,1 \text{ and } a=2 \\ 0 & \text{otherwise} \end{cases}$

$\vdots$

while not done

  $s \leftarrow s_0$

  while s not terminal

    choose action a

    observe transition (s,a,r,s')

    update $q(s,a) \leftarrow q(s,a) + \alpha\left(r + \gamma \cdot \max_{a'} q(s',a') - q(s,a)\right)$

    $s \leftarrow s'$

$n=36$ (9 grid cells in partition · 4 actions)

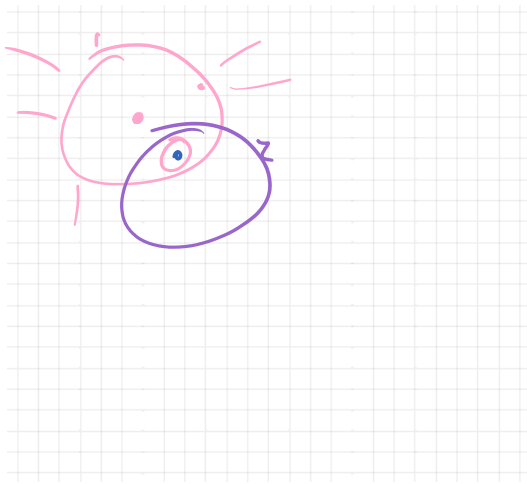
$\hat{q}(s,a) = w_1 \cdot f_1(s,a) + \cdots + w_n f_n(s,a)$

exactly 1 is 1, rest are 0

so only updating $f_{r,c,a'}$ for s goes in cell r,c and $a=a'$

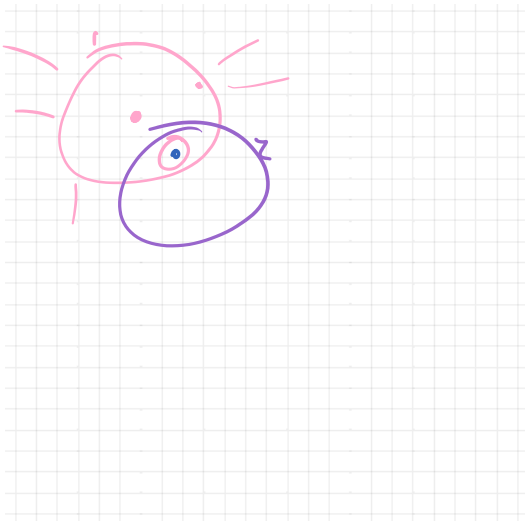
$w_i \leftarrow w_i + \alpha \cdot f_i(s,a) \cdot \left(r + \gamma \max_{a'} \hat{q}(s',a') - \hat{q}(s,a)\right)$

$w_{r,c,a}$ is the only one updated $\rightarrow$ call that $\hat{q}(\hat{s},a)$



$f_1 = \begin{cases} 1 & \text{if } s \text{ in region } 1 \\ 0 & \text{otherwise} \end{cases}$

$f_2 = \begin{cases} 1 & \text{if } s \text{ in region } 2 \\ 0 & \text{otherwise} \end{cases}$



$f_1 = \begin{cases} 1 & \text{if } s \text{ in region 1} \\ 0 & \text{otherwise} \end{cases}$

$f_2 = \begin{cases} 1 & \text{if } s \text{ in region 2} \\ 0 & \text{otherwise} \end{cases}$