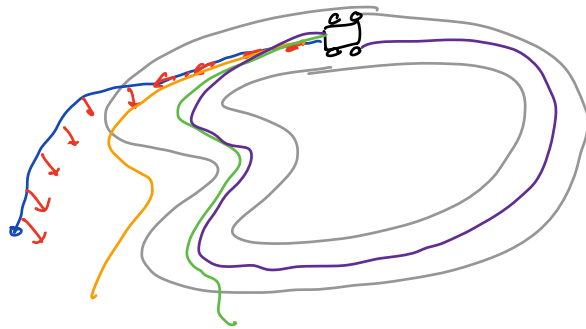


TRAIN-TEST MISMATCH "DISTRIBUTION"

- ① BIAS IN ENVIRONMENTS IN WHICH ROBOT
(INTENTIONAL)
- ② EMBODIMENT PROBLEM
- ③ BACKGROUND CHANGES (PREFERENCES, WEATHER)

$$P(\underset{\text{input}}{X}, \underset{\text{output}}{Y}) = \underbrace{P(X)}_{\text{CORRELATED SHIFT}} \underbrace{P(Y|X)}_{\text{LABEL SHIFT}}$$



$$\pi: X \rightarrow Y$$

ALGORITHM

(1) COLLECT DATA FROM HUMAN EXPERT

$$D = \{(x_1, y_1^*), (x_2, y_2^*), \dots\}$$

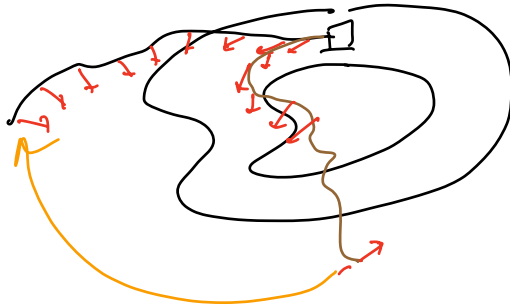
(4) CALL INTERACTIVE EXPERT ON ALL STATES π_0

$$D = \{(x_1, y_1^*), (x_2, y_2^*), (x_3, y_3^*)\}$$

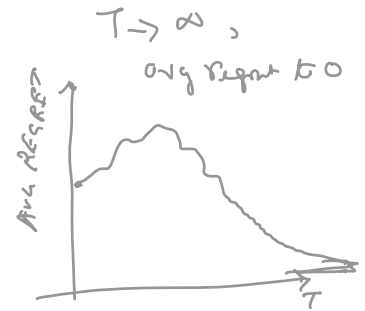
→ (2) TRAIN π_0 ON D.

(3) EXECUTE π_0

c



	l_1	l_2	l_3	l_4	...	l_T
π^1	0.1	π_2 1.0	1.0	-	-	-
π^2	π_1 1.0	0.1	0.1	-	-	-
...	0.3	...	π_3 0.5	-	-	-
...	-	-	-
π^N	1.0	0.5	1.0	-	-	-



NO REGRET

$$\text{Avg REGRET} = \frac{1}{T} \left(\underbrace{\sum_{t=1}^T l_t(\pi_t)}_{\text{(LEARNER)}} - \min_{\pi^* \in \Pi} \underbrace{\sum_{t=1}^T l_t(\pi^*)}_{\text{(BEST POLICY IN HINDSIGHT)}} \right)$$