Driving a car with low dimensional input features
Jean-Claude Manoli, jcma@stanford.edu

PROBLEM
The motivation for this project is a robotics competition in which autonomous cars race against each other around a small circuit. The challenge is to keep the car within the track boundaries knowing only the distance driven, the heading and the color of the track under the car.

MODEL
This problem can be posed as a partially observable Markov decision process (POMDP).

State Space
\[ s = \{ x_m, y_m, \theta, \kappa, p, v, c_b \} \in S \]
Observation Space
\[ o = \{ d, \hat{\theta}, \kappa, \hat{v}, c_b \} \in \Omega \]
Action Space
\[ A = A_s \times A_p \]
\[ A_s = \{ \text{nothing, steer left, steer right} \} \]
\[ A_p = \{ \text{nothing, throttle up, throttle down} \} \]

Reward
\[ R(s) = v \cdot \Delta t \cdot \cos(\theta - h(x_m)) \]

Belief State Space
\[ b = \{ \hat{x}_m, \hat{y}_m, \hat{\theta}, \kappa, \hat{v} \} \in B \]
\[ \hat{x}_m^{(t_2)} = \int_{t=t_1}^{t_2} \hat{v} \cos(\hat{\theta} - h(\hat{x}_m)) \, dt \]
\[ \hat{y}_m^{(t_2)} = \int_{t=t_1}^{t_2} \hat{v} \sin(\hat{\theta} - h(\hat{x}_m)) \, dt \]
When \( c_b > 0 \)
\[ \hat{y}_m^{(t_2)} = \text{sign} \hat{y}_m \left( \frac{1}{2} w_a + c_b w_b \right) \]

Lap or checkpoint:
set \( \hat{x}_m \) to 0 or \(-1\).

Q-LEARNING
\[ \hat{Q}^*(b, a) \leftarrow r(s) + \gamma \max_{a' \in A} \hat{Q}^*(b', a') \]
\[ a = \arg\max_a \hat{Q}^*(b, a) \]

BASELINE
A simple PD controller on belief \( \hat{y}_m \)
\[ f(s) = \kappa + k_p y_m + k_d \hat{y}_m \]
\[ a = \begin{cases} \text{steer left} & \text{if } f(s) < -\epsilon \\ \text{steer right} & \text{if } f(s) > \epsilon \\ \text{do nothing} & \text{otherwise} \end{cases} \]

PIPELINE

<table>
<thead>
<tr>
<th>100 episodes</th>
<th>Score</th>
<th>Crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 75% throttle</td>
<td>16.0</td>
<td>0</td>
</tr>
<tr>
<td>Baseline 100% throttle</td>
<td>12.9</td>
<td>35</td>
</tr>
<tr>
<td>DQN</td>
<td>22.1</td>
<td>0</td>
</tr>
</tbody>
</table>

Baseline

DQN

Lap & Threshold Detection
Observation checkpoint
Position Tracking
Belief
Agent