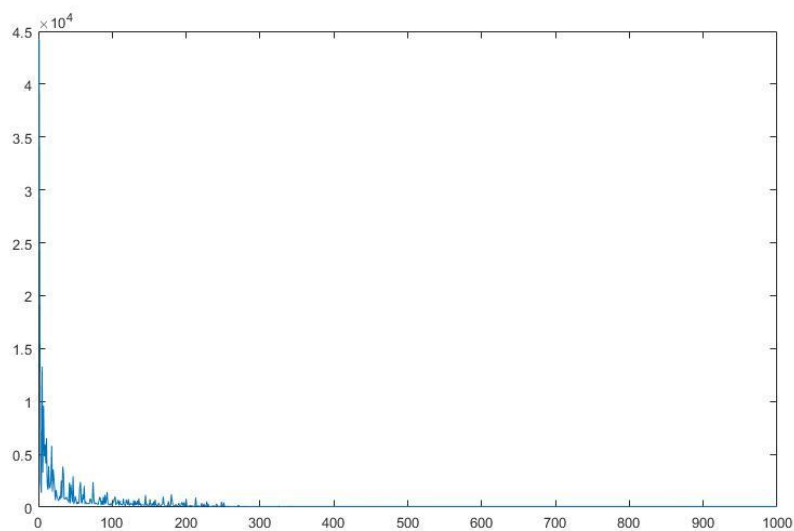


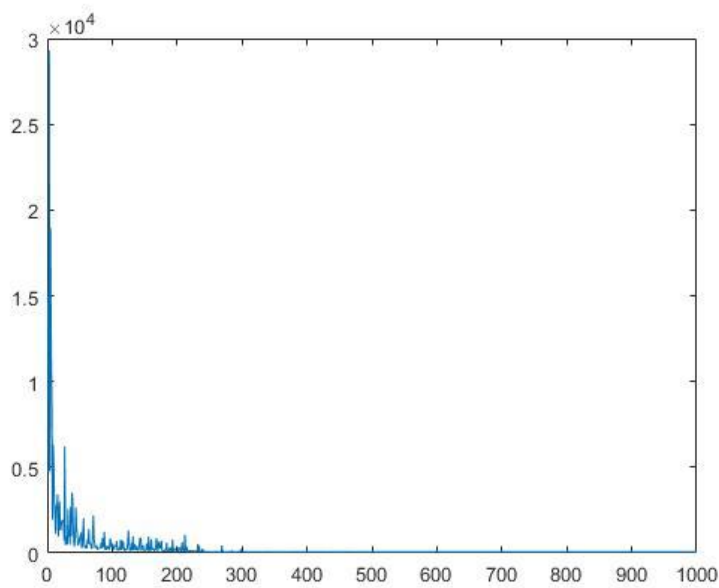
طبق صحبتی که با حل تمرین های درس (آقای زارع) داشتم قرار شد که ارائه نمودار تعداد حرکت در هر اپیزود به جای نمودار میانگین پاداش ارائه شود:

یادگیری کیو:

$Q=0$

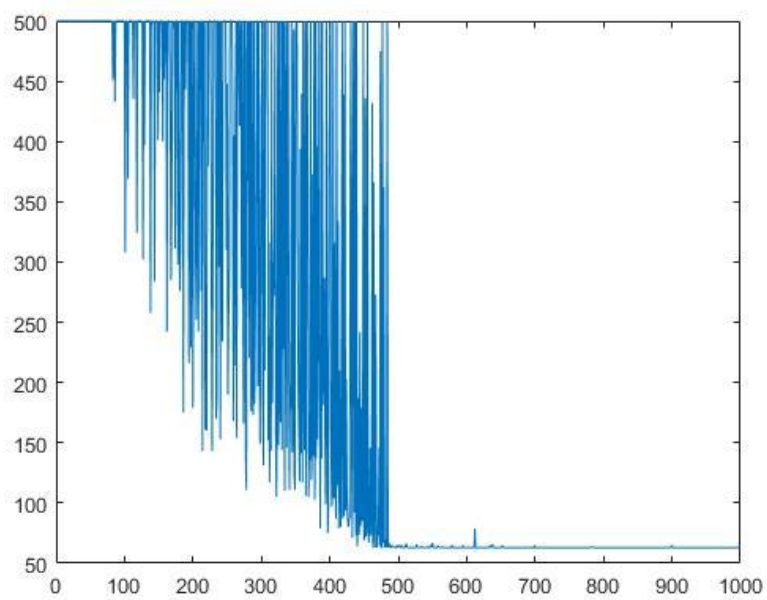


$Q = \text{random}$

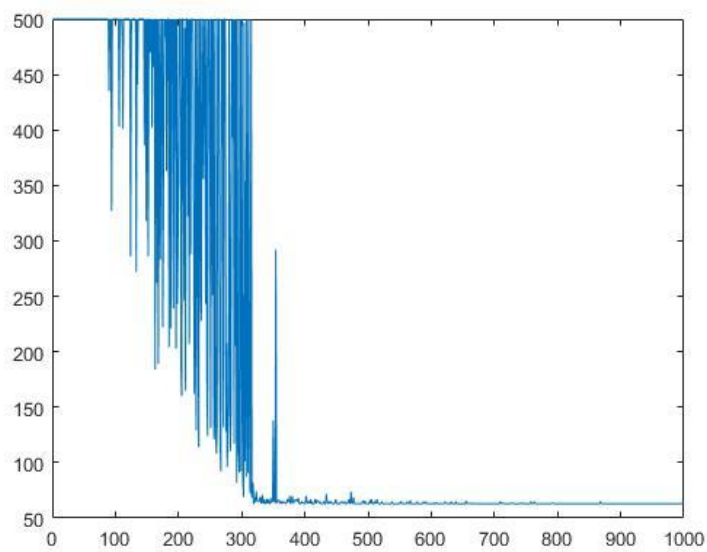


یادگیری سارسا لاند:

$0 = \lambda$



$0.5 = \lambda$



$1 = \lambda$

