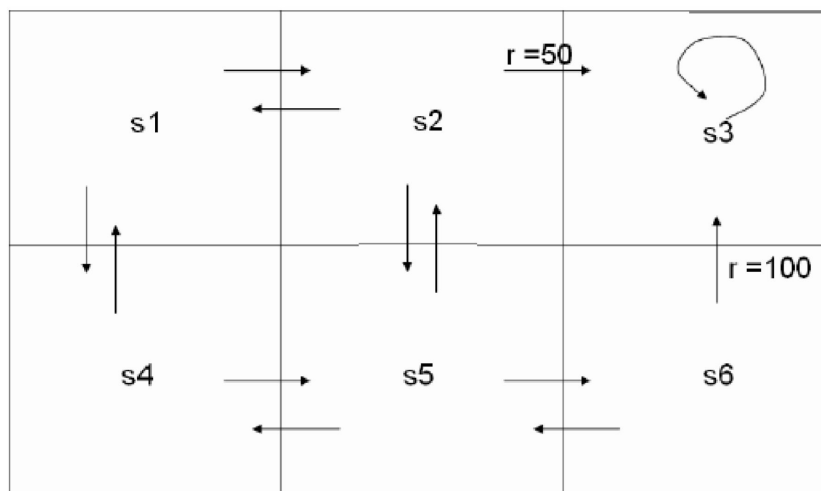




- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

## Reinforcement learning

فرآیند تصمیم گیری مارکوفی را برای رباتی در فضای شکل زیر در نظر بگیرید. مقدار reward ها در کنار فلش های انتقال نوشته شده اند. reward های فلش های بدون مقدار را صفر در نظر بگیرید.  $\gamma = 0.8$



۱. برای هر state مقدار  $V^*(s)$  را مشخص کنید.
۲. فلش های مربوط به policy optimal خود را انتخاب کنید.
۳. مقدار متفاوتی برای  $\gamma$  ارائه دهید که منجر به optimal policy متفاوت می شود و تعداد action های تغییر یافته باید حداقل باشد.
- (آ) مقدار جدید خود را برای  $\gamma$  بدهید.
- (ب) توضیح دهید کدام action تغییر می کند.
۴. آیا میتوان reward را تغییر داد به گونه ای که  $V^*$  تغییر کند اما optimal policy ( $\pi^*$ ) تغییر نکند؟

(موفق باشید:)