گزارش تمرین سوم

یادگیری تقویتی

پاییز ۱۴۰۰

بخش اول

علی ساعی زاده ۸۱۰۱۹۶۴۷۷

1- فروش خرچنگ

در این مسئله، پایه تصمیم ما تعداد خرچنگ‌های موجود و ماه جاری (جمعیت در حال افزایش است یا کاهش با توجه مقطع زمانی سال) خواهد بود بنابراین استیت ما برابر تعداد خرچنگ های موجود در یک اسکیل معقول، ماه و بودجه باقی مانده است. برای مثال اگر بودجه را ثابت درنظر بگیریم و بازه های خود را 1000 تایی در نظر بگیریم و 1000000 خرچنگ داشته باشیم 12000 استیت خواهیم داشت (تعداد ماه در نظر گرفته شده) که اگر بودجه نیز متغیر بود استیت ها با توجه به مقدار بودجه افزایش پیدا می‌کرد.

اعمال ما در واقع فروش یا عدم فروش یا جبران جمعیت خرچنگ‌ها یا اعلام ورشکستگی است که باتوجه به استیتی که در آن داریم تعیین می شود. برای مثال زمانی که 1000 خرچنگ باقی مانده اقدام به فروش نمی‌کنیم و با توجه به بودجه تصمیم میگریم که اعلام ورشکستگی کنیم با جمعیت را جبران کنیم.

احتمال انتقال بین استیت ها نیز می توان با شرایط جوی و محیطی تصمیم گرفته شود برای مثال رشد توسط جمعیت طی سال ها محاسبه شود و هر ماه با توجه به افزایش جمعیت و فروش استیت بعدی بدست می آید اما احتمالی برای عدم فروش و یا عدم رشد جمعیت با توجه به شرایط منطقه باید درنظر گرفته شود. برای مثال اگر زمانی شریاط جوی به صورت غیرمتقربه تغییر کند احتمال وقوع این شرایط باید در رشد جمعیت خرچنگ ها در نظر گرفته شود که در واقع احتمال خطای ما در مورد پیش‌بینی ما در مورد جمعیت خواهد بود.

پاداش باید بصورت متناسب و ترکیبی از میزان فروش، ماه، بودجه باقی مانده و خرچنگ باقی مانده باشد. به این صورت که اگر جمعیت زیاد باشد اما میزان فروش کم باشد باید تنبیه درنظر گرفته شود. در مثالی دیگر، اگر جمعیت کم باشد و فروش زیاد باشد و بودجه نیز کم باشد باید تنبیه بسیار شدید در نظر گرفته شود.

2- فروش سهام

برای هر دوره با توجه به سود و زیان تصمیم های گذشته و جیب خود تصمیم نهایی را برای فروش سهام در پایان دوره می‌گیریم. بنابراین تصمیم ما براساس نتیجه خرید و فروش های قبلی است. بنابراین استیت ما سود و زیان ما از خرید های قبلی است که هم میتوان فضایی گسسته برای آن درنظر گرفت هم پیوسته. همچنین برای تصمیم بهتر می توان نتیجه هر تصمیم را نیز در استیت ها اضافه کرد. برای مثال اگر در کل 1000 تومان سود کرده باشیم اما این سود حاصل 6 نگهداری سهام و 1 فروش سهام باشد باشد تصمیم ما بر نگهداری سهام خواهد بود. بنابراین استیت ما شامل تعداد تصمیم موفق از فروش و یا نگهداری سهام است و سود و زیان هر تصمیم و پس انداز ما در یک دوره است.

اکشن ما در انتهای هر دوره خرید یا فروش سهام خواهد بود که مارا به استیت بعدی (پس انداز جدید و دوره جدید) خواهد برد.

انتقال استیت ما بصورت قطعی نخواهد بود با توجه به تخمینی که از هزینه های زندگی خود داریم پس اندازه جدید بدست می آید که با توجه به شریط غیرقابل پیش‌بینی زندگی ممکن است اتفاقی در زندگی ما بیفتد که پس انداز ما را بطور کلی عوض کند برای مثال یک تصادف و یا بردن یک جایزه می تواند از این دست اتفاقات باشد. همچنین اتفاق های نامعمول بازار های مالی در انتهای دوره می تواند انتقال بین استیت های مارا تغییر دهد (تغییر ناگهانی سود و زیان سهام ها).

پاداش ها با توجه به پس انداز و سود و زیان بدست آمده در پایان دوره باید تعیین شود. برای مثال اگر پس انداز ما 1000000 باشد و ما سودی برابر 1000 تومان بدست آورده باشیم باید پاداش کمتری نسبت به پس انداز 1000000 تومان و سود 10000 داشته باشد. که البته این پاداش ها با توجه به ریسک پذیری فرد و شریاط بازار نیز تعیین شود برای مثال اگر بازار در دوره رکود باشد هر سودی باید پاداش بسیار بالایی دریافت کند اما اگر بازار در رونق باشد اگر سود پایین باشد باید تنبیه درنظر گرفته شود.

3- تولید کارخانه

استیت ما شامل تقاضای سال گذشته، تجهیزات کارخانه، نیروی کار و بودجه برای پیش‌بینی مقدار تولید در اول سال برای سال آینده خواهد بود. برای مثال اگر تجهیزات قابلیت تولید 1000 عدد محصول را نداشته باشد نباید میزان تولید 10000 تعین شود همچنین اگر نیروی انسانی و تجهزیات قابلیت تولید 10000 محصول را داشته باشد اما تقاضا 1000 باشد باید مقدار کمتری محصول تولید شود.

اکشن ما تعداد محصول تعیین شده در اول سال است و با توجه به اسکیل ما تعداد آن متفاوت است.

استیت ما بصورت قطعی با اعمال ما تغییر نمی‌کند برای مثال اگر آتش سوزی در کارخانه رخ دهد (اتفاقی نادر اما امکان پذیر) استیت ما کاملا تغییر پیدا می‌کند. بنابراین برای رخ دادن حوادث باید احتمال در نظر گرفته شود. که مارا به استیت متفاوتی از استیت پیش بینی شده خواهد برد.

پاداش ما باید با توجه به استیت و عرضه و تقاضا تعیین شود. برای مثال اگر 1000 محصول تولید کردیم و 800 خرید داشتیم بودجه ما نیز تحت تاثیر قرار می‌گیرد باید تنبیه درنظر گرفته شود و گار اختلاف تولید و فروش کم باشد باید پاداش تعیین شود. همچنین اگر تولید ما کمتر از تقاضا باشد باید تنبیه شدیدی درنظر گرفته شود.

4- آتش نشانی

استیت ما با توجه به منطقه آتش سوزی، جمعیت ساکن، میزان صدای آژیر (نوع آن با توجه به میزان آتش سوزی) و ارزش آن منطقه و تعداد ماشین های موجود باید تعیین شود برای مثال در شرایطی که دو ماشین موجود است با شرایطی که 100 ماشین موجود است کاملا متفاوت خواهد بود.

اکشن ما شامل تعداد ارسال ماشین به منطقه خواهد بود.

احتمال انتقال بین استیت ها قطعی نیست برای مثال با توجه به خرابی غیر متقربه یک ماشین، استیت می تواند تغییر کند.

پاداش باید با توجه به هزینه آتش سوزی و تعداد ماشین های باقی مانده تعیین شود و نجات جان انسان باید بیشترین ارزش را داشته باشد. برای مثال اگر 10 انسان کشته شود باید تنبیه بسیار شدیدی در نظر گرفته شود همچنین در صورت آتش سوزی کامل بانک و عدم رسیدگی به آن تنبیه شدید تری نسبت به زمین زراعی باید درنظر گرفته شود.