Assignment NO.5 Solutions

Neural Networks | Fall 1400 | Dr.Mozayani

Teacher Assistants:

Amirali Molaei

Samin Heydarian

Student name : **Amin Fathi**

Student id : **400722102**

Problem 1.a

In this section, you need to provide an MDP (Markov Decision Process) model. It should be noted that
you need to determine states, actions, state transition probabilities, and rewards for your model. (35pts)

(a)

In a village, we want to make a decision at the beginning of each month whether the sale of shrimps is allowed or not. Every time we decide to sell shrimps, the number of shrimps will be reduced and we gain a profit from the sale of them. It should be noted that if the population of shrimps is reduced too much, it costs us a lot of money to compensate for their population, otherwise, the whole shrimp industry in this village will go broke

برای حل این سوال ابتدا باید در نظر بگیریم که برای حل مسئله MDP به تعریف این ۴ مورد نیاز داریم :

۱ - **حالات** محیط یا همان states

۲ - **اعمالی** که عامل می تواند انجام بدهد یا همان actions

۳ - **پاداش** به دست آمده پس از انجام عمل توسط عامل دریک حالت یا همان reward

۴ - **انتقال حالات** یا همان state transition

اعمال (action) فقط به حالت فعلی عامل بستگی دارد .

پاداش (reward) فقط به حالت فعلی و action فعلی بستگی دارد .

در این مسئله که بسیار شبیه به مسئله ای موجود در لینک انتهای پاسخنامه است ، ابتدا state ها را به صورت زیر تعریف می کنیم که در واقع بیانگر تعداد ماهی های موجود می باشد :

empty, low, medium, high.

Empty => هیچ ماهی ای موجود نمی باشد

low => مقدار ماهی های موجود کمتر از مقدار ترشولد t1 می باشد .

medium => مقدار ماهی های موجود بیشتر از مقدار ترشولد t1 و کمتر از مقدار ترشولد t2 است.

high => مقدار ماهی های موجود بیشتر از مقدار ترشولد t2 می باشد.

اعمال مورد نیاز در این مساله عبارتند از :

Sale_allowed

Sale_not_allowed

نکته : برای حالت empty تنها اکشن مجاز re-breeding به معنای پرورش مجدد میگو است ، چرا که صنعت فروش میگو در خطر است ! و نیاز به پرورش میگو داریم

پاداش :

پاداش ها برای این مساله عبارتند از :

چنانچه در حالت low بودیم ، به طور فرض ۱۰ ملیون تومان از فروش میگو سود حاصل میشود

چنانچه در حالت medium بودیم ، ۵۰ ملیون تومان سود حاصل از فروش میگو حاصل میشود

چنانچه در حالت high بودیم ، ۱۰۰ ملیون تومان سود حاصل از فروش میگو حاصل میشود

چنانچه در حالت empty باشیم ف مقدار reward برابر منفی دویست ملیون می باشد که این هزینه برای پرورش میگو های جدید محاسبه می شود.

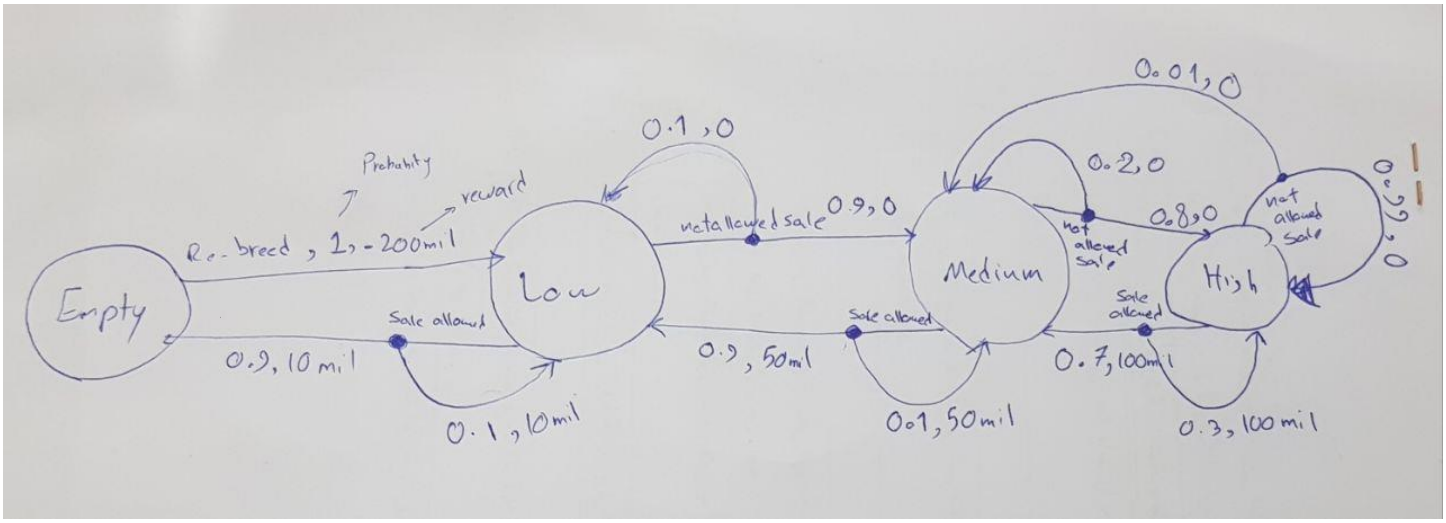
تابع گذار حالات :

مشخصا با فروش میگو ها با احتمال بیشتری(مثلا0.9) به به حالت با مقدار میگوی کمتری منتقل می شویم و با احتمال کمتری در همان حالت می مانیم (مثلا0.1) همچنین مقدار پاداش (reward) با توجه به حالتی که در آن هستیم تعیین می شود

و با عدم فروش میگو با احتمال بیشتری (مثلا 0.9) به حالت با مقدار میگوی بیشتر منتقل میشویم و با احتمال کمتری در همان حالت می مانیم (مثلا 0.1) ، مشخصا عدم فروش میگو reward ای به دنبال ندارد .

چنانچه در حالت empty باشیم ، تنها عمل ممکن re-breed یا بازیبی و افزایش میگو های در معرض فروش است که مسلما این کار نیاز به سرمایه و زمان بسیاری دارد و در نتیجه مقدار reward آن منفی و به طور فرض برابر منفی دویست میلیون تومان در نظر گرفته شده است ، طبیعی است از انجا که این اقدام تنها عمل ممکن می باشد مقدار احتمال آن برابر ۱ می باشد .

چنانچه در حالت high باشیم ، با عدم فروش میگو reward ای نمیگیریم و همچنین با احتمال بسیار بالایی (مثلا 0.99) در حالت فعلی می مانیم و با احتمالی کمتر (فاسد شدن میگو ها) (0.01) به حالت medium منتقل می شویم.



مدل MDP برای فروش میگو

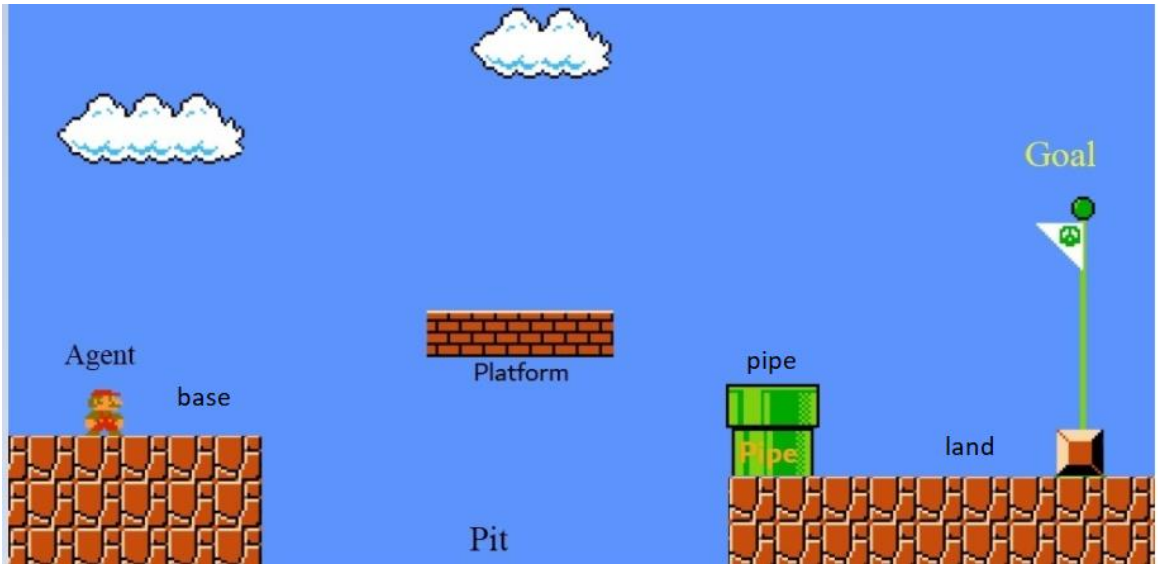
(تابع گذار به صورت دوتایی (پاداش ، احتمال) در نظر گرفته شده است)

منبع :

[Real World Applications of Markov Decision Process | by Somnath Banerjee | Towards Data Science]

Problem1.b

In the Mario game, the goal is to reach the end flag without falling in the pit or dying by enemies. Assume that Mario (our agent) can either jump or move forward. The speed of Mario affects its jump distance, for instance, if he jumps at high speed, he may slip off the edge of the platform and fall (either in the pit or on the green pipe), or if the speed is too low, he can't reach the platform after jumping. A piranha plant will also come out of the pipe stochastically and kill Mario if it hits him. The game ends whether Mario gets killed or reaches the flag



در این سوال state ها همانطور که درشکل مشخص کرده ایم برابر است با :

۱ - base :

در واقع قسمت آغازین بازی است که عامل ما در آنجا در حالت سکون بازی را شروع میکند.

۲ – platform :

قسمت اجری وسط شکل میباشد

۳ – pit :

گودال است و سقوط آن مرگ و پایان بازی را به همراه دارد

۴ – pipe :

لوله است که اگر عامل روی ان بیوفتد ممکن است به صورت تصادفی توسط گیاه پیرانا کشته شود و بازی پایان یابد .

۵ – land : قسمت سمت راست بازی است که در نهایت منجر به رسیدن به goal می شود

۶ – goal :

پایان بازی است و عامل برنده است و بازی پایان یافته

۷ – dead :

عامل در اینجا مرده است و بازی پایان یافته

برای تعیین action ها:

میتوانیم بپریم یا حرکت کنیم که برای هر دو این ها هم سرعت بالا و پایین در نظر میگیریم.

Low jump : پریدن با سرعت پایین

Low move : حرکت کردن با سرعت پایین

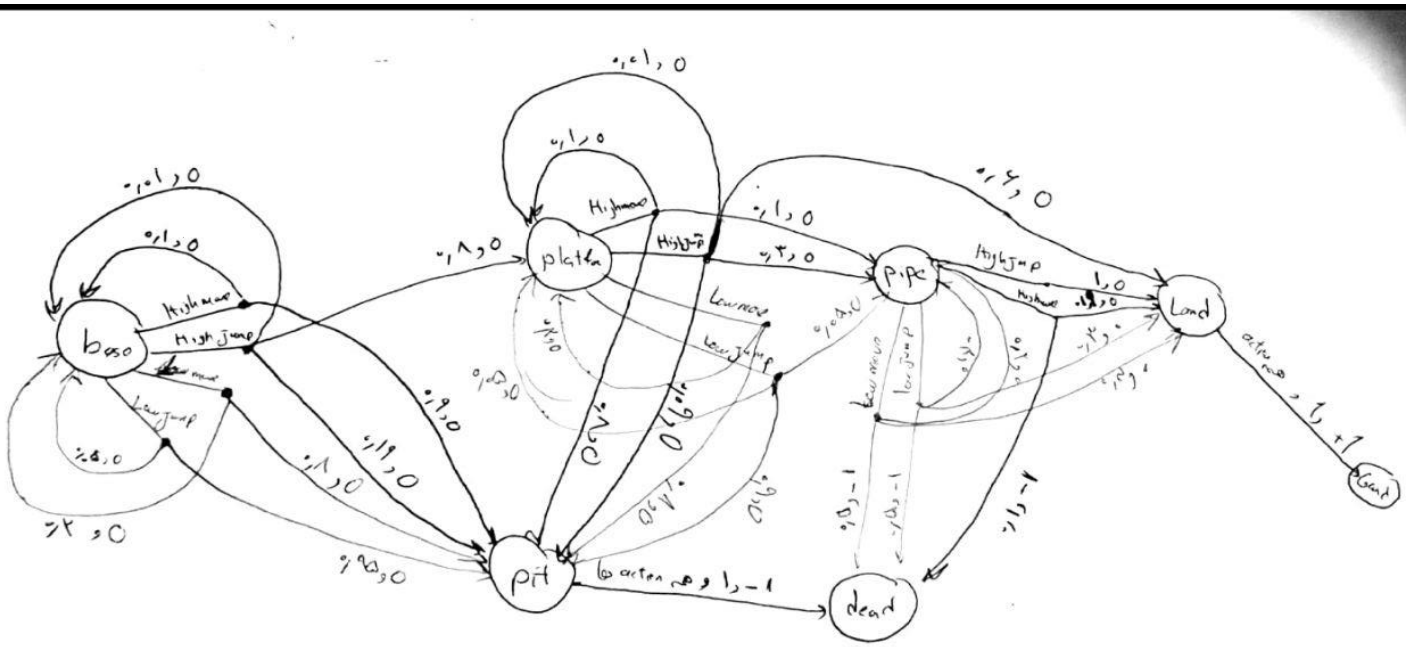
High jump : پریدن با سرعت بالا

High move : حرکت با سرعت بالا

برای تعیین reward ها :

میتوان برای رسیدن به goal ، پاداش مثبت در نظر گرفت (۱+) و برای dead هم پاداش منفی (۱-)

: MDF



تابع احتمالات :

۱ - : اگر در حالت BASE باشیم و قصد حرکت با سرعت زیاد داشته باشیم با احتمال 0.1 همچنان در این State خواهیم بود و با احتمال 0.9 به pit یا پرتگاه منتقل میشویم .

اگر که با سرعت کم حرکت کنیم با احتمال 0.2 در حالت فعلی خواهیم بود و با احتمال 0.8 به دره خواهیم افتاد .

اگر با سرعت بالا بپریم ، با احتمال 0.01 در حالت فعلی خواهیم بود و با احتمال 0.19 به دره می افتیم و با احتمال 0.8 به سکوی platform خواهیم رسید .

اگر هم که با سرعت کم بپریم ، با احتمال 0.05 در استیت فعلی خواهیم ماند و با احتمال 0.05 به استیت pipe منتقل می شویم و با احتمال 0.9 به دره سقوط می کینم.

۲ – اگر در pit باشیم :

به ازای همه اکشن ها خواهیم مرد و پاداش -1 خواهیم گرفت

۳- اگر در platform باشیم :

اگر با سرعت بالا حرکت کنیم با احتمال 0.1 به استیت pipe میرسیم و با احتمال 0.1 هم در حالت فعلی می مانیم و با احتمال 0.8 به دره خواهیم افتاد

اگر که با سرعت کم حرکت کنیم با احتمال 0.2 در حالت فعلی خواهیم بود و با احتمال 0.8 به دره خواهیم افتاد .

اگر با سرعت بالا بپریم ، با احتمال 0.01 در حالت فعلی خواهیم بود و با احتمال 0.09 به دره می افتیم و با احتمال 0.6 به استیت land خواهیم رسید و با احتمال 0.3 به استیت pipe منتقل می شویم .

اگر هم که با سرعت کم بپریم ، با احتمال 0.05 در استیت فعلی خواهیم ماند و با احتمال 0.95 به دره خواهیم افتاد .

۴ _ اگر در pipe باشیم :

اگر با سرعت بالا حرکت کنیم با احتمال 0.1 به استیت dead میرسیم و پاداش 1- میگیریم و با احتمال 0.9 هم به حالت land منتقل میشویم .

اگر که با سرعت کم حرکت کنیم با احتمال 0.5 به استیت dead میرسیم و پاداش 1- میگیریم و با احتمال 0.3 به استیت land منتقل میشویم و با احتمال 0.2 در حالت فعلی می مانیم.

اگر با سرعت بالا بپریم ، با احتمال 1 به استیتland منتقل می شویم .

اگر هم که با سرعت کم بپریم ، با احتمال 0.5 به استیت dead میرسیم و پاداش 1- میگیریم و با احتمال 0.3 به استیت land منتقل میشویم و با احتمال 0.2 در حالت فعلی می مانیم.

۵ – اگر در land باشیم :

به ازای همه action ها فقط و فقط به استیت goal میرسیم و پاداش 1+ میگیریم.

Problem 2

Imagine our agent wants to go from state S to T. State T has a reward of +120 and states with red color have a reward of -90. Taking each step has a -1 reward. Run each episode with the following actions and update values by Q-Learning algorithm:

episode 1: Right, Down, Down, Down, Down, Down, Left.

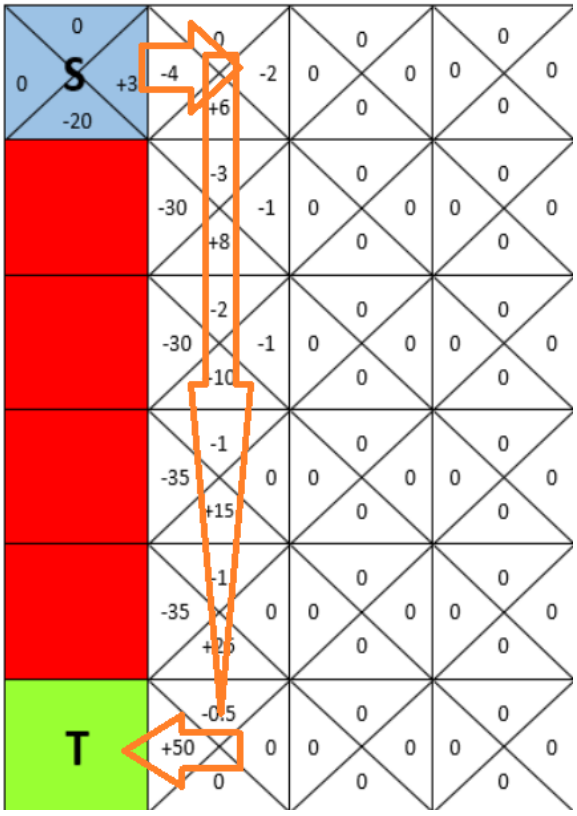
episode 2: Right, Down, Down, Left.

Note that If the agent goes into a red-colored state or state T, the episode terminates. Set α = 0.9 and λ = 0.8

الگوریتم q-learning :

$$Q(a,i) \leftarrow Q(a,i) + \alpha(R(i) + \gamma \max_{a'} Q(a',j) - Q(a,i))$$

حال برای اپیزود اول الگوریتم را اجرا میکنیم.



New(S ,R) = 3 + 0.9 * (-1 + 0.8 *6 – 3) = 3.72

New (S1 , D) = 6 + 0.9 * (-1 + 0.8 * 8 -6) = 5.46

New (S2 , D) = 8 + 0.9 * (-1 + 0.8 * 10 -8) = 7.1

New (S3 , D) = 10 + 0.9 * (-1 + 0.8 * 15 -10) = 10.9

New (S4 , D) = 15 + 0.9 * (-1 + 0.8 * 25 -15) = 18.6

New (S5 , D) = 25 + 0.9 * (-1 + 0.8 * 50 -25) = 37.6

New (S6 , L) = 50 + 0.9 * (120 + 0.8 * 0 - 50) = 113

حال برای اپیزود دوم الگوریتم را اجرا میکنیم (در این مرحله از موارد اپدیت شده در مرحله قبل استفاده میکنیم) :

New (S ,R) = 3.72 + 0.9 * (-1 + 0.8 * 5.46 – 3.72) = 3.4

New (S1 ,D) = 5.46 + 0.9 * (-1 + 0.8 * 7.1– 5.46) = 3.4

New (S2 ,D) = 7.1 + 0.9 * (-1 + 0.8 * 10.9 – 7.1) = 3.4

New (S 3,L) = -30 + 0.9 * (-90 + 0.8 * 0 +30) =-84

Problem 4

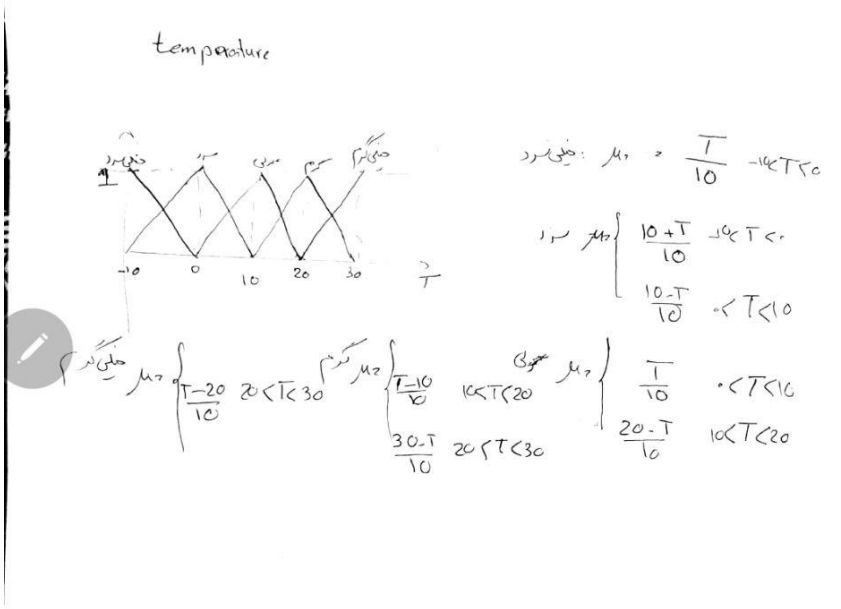
We want to model a fuzzy controller in this part, the fuzzy controller will be for a steam turbine.

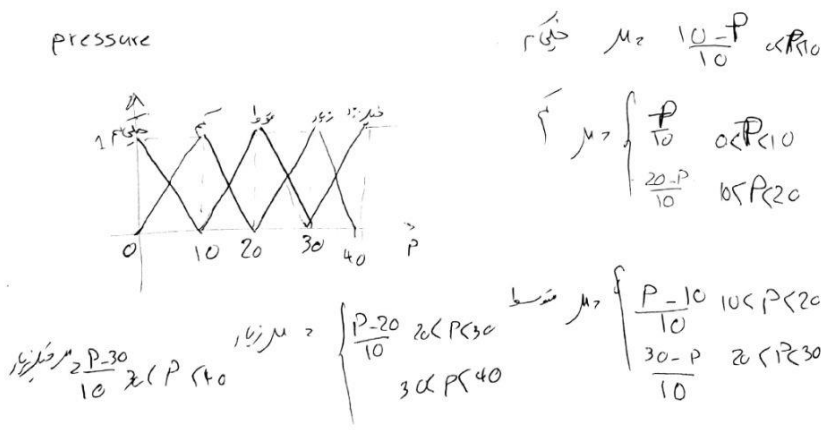
- **Inputs: temperature and pressure (5 descriptors each)**
- **Output: throttle setting (7 descriptors)**

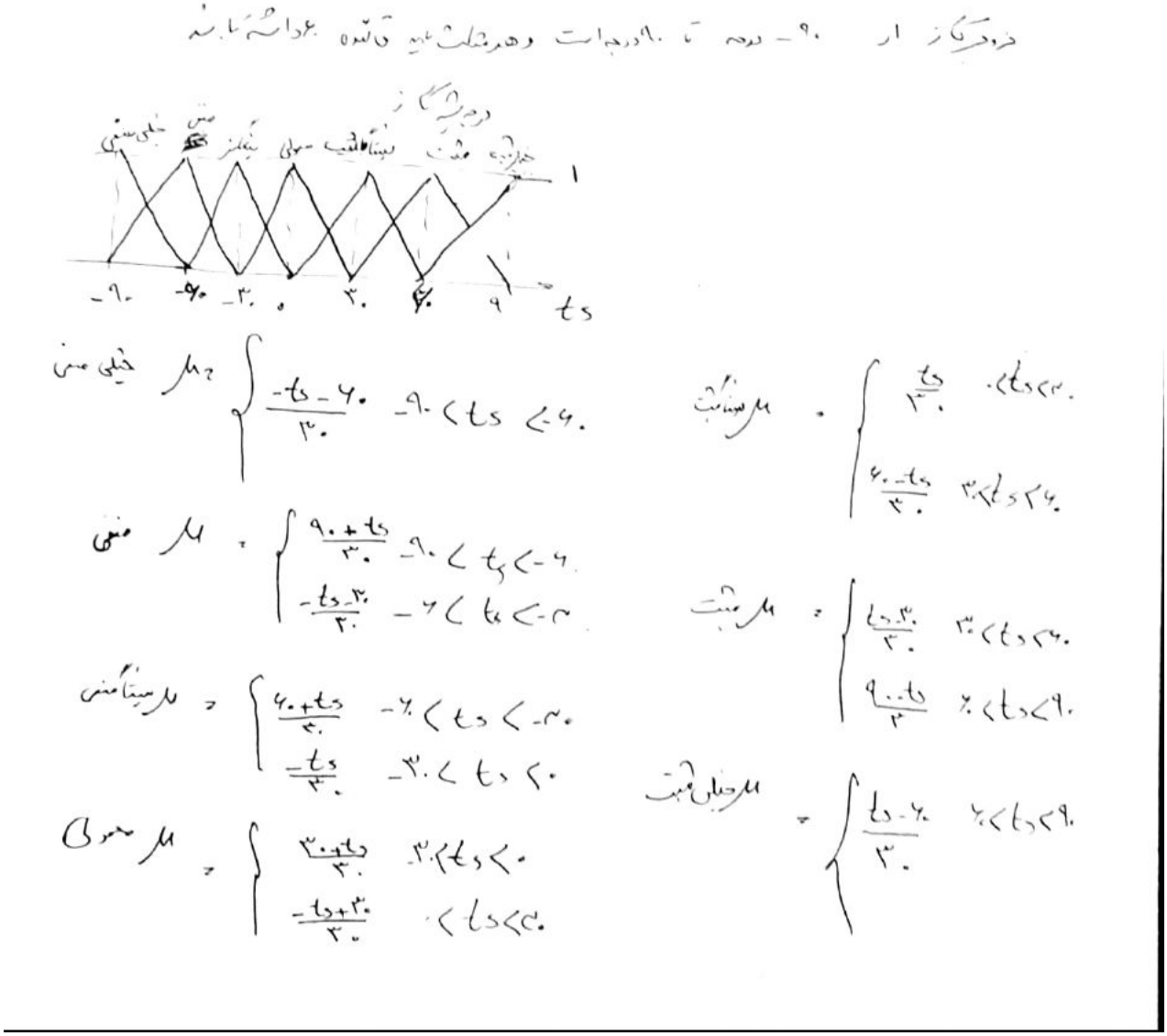
After modeling the fuzzy controller answer this question.

“If for inputs temperature is 70% and pressure is 30% determine the throttle position.”

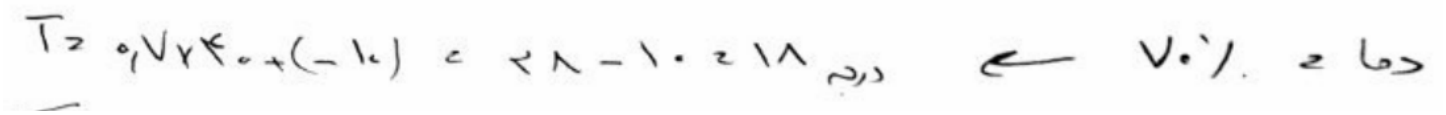
ابتدا توابع فازی را رسم و فرم توابع را برای هر ۳ متغیر دما – فشار و شیر گاز مینویسم



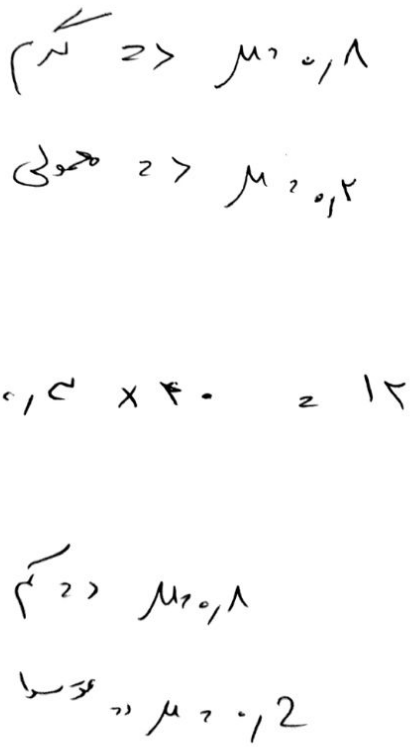




حال دمای ۷۰ درصد و فشار ۳۰ درصد را حساب میکنیم:

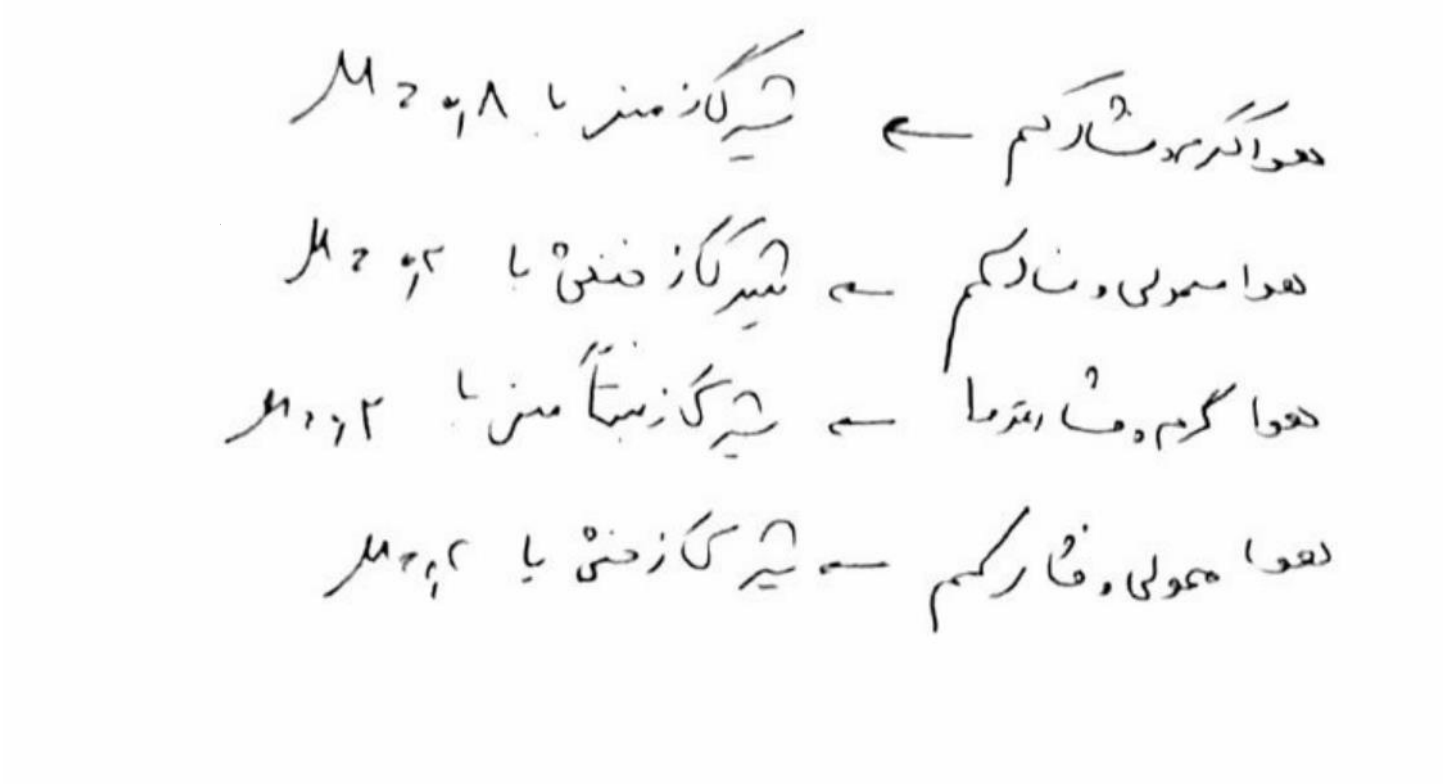


که با این دما دو میو در دمای گرم و معمولی داریم ، سپس فشار را حساب کرده و میو های آن را هم حساب میکنم.

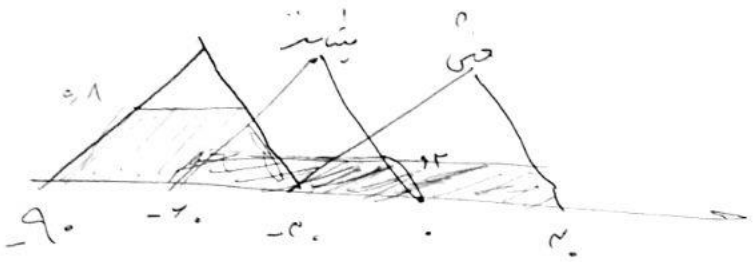


با این اعداد و با توجه به جدول زیر که با اطلاعات متخصص ساخته ایم ۴ حالت پی می آید

خیلی گرم	گرم	معمولی	سرد	خیلی سرد	دما/فشار
خیلی منفی	خیلی منفی	خنثی	خیلی مثبت	خیلی مثبت	خیلی کم
خیلی منفی	منفی	خنثی	مثبت	خیلی مثبت	کم
منفی	نسبتا منفی	خنثی	نسبتا مثبت	مثبت	متوسط
نسبتا منفی	نسبتا منفی	خنثی	نسبتا مثبت	نسبتا مثبت	زیاد
نسبتا منفی	نسبتا منفی	خنثی	نسبتا مثبت	نسبتا مثبت	خیلی زیاد



نتایج حاصل را در نمودار درجه شیر رسم کرده و با روش میانگین وزن دار حاصل نهایی را حساب می کنیم .



$$\bar{x} = \frac{x_1\mu_1 + x_2\mu_2 + x_3\mu_3}{\mu_1 + \mu_2 + \mu_3}$$

$$\bar{x} = \frac{0.1 \times (-90) + 0.2 \times (-30) + 0.7 \times 10}{0.1 + 0.2 + 0.7}$$

$$= \frac{-9 - 6 + 7}{1} = -8$$

دوم: $\frac{-48}{12} = -4$