Machine Learning

HW #1

فاطمه نورزاد
۸۱۰۱۹۸۲۷۱

سوال ۱:

(به قسمت الف و ب یک جا پاسخ داده شده است.)

انسان ها معمولا lost aversion دارند.. به این معنی که همواره تصمیمی (هر چند غیر منطقی) میگیرند که بیشترین سود و کمترین ضرر را برایشان داشته باشد. علاوه بر این انسان ها همواره به دنبال تصمیم هایی هستند که کمترین ابهام را داشته باشد.(Ambiguity Averse)

علاوه بر این موضوعات انسان ها به دو دسته تقسیم میشوند :

۱) punishment averse : این دسته از انسان ها تصمیمی را میگیرند که کمترین ضرر را داشته باشد فارغ از این که این تصمیم باعث چه سودی شود.

۲) reward seeker : این دسته از انسان ها برعکس دسته قبلی به دنبال بیشترین سود و جایزه هستند و ضرر یک تصمیم برایشان به اندازه پاداش احتمالی ارزش ندارد.

من بین این دو قرص، قرص اول را که قطعا سود بخش است را انتخاب میکنم.

به دلیل این انتخاب میتوان گفت من punishment averse هستم. چراکه به دنبال ریسک کردن نبودم و به این علت که قرص دوم punishment هم دارد، قرصی را انتخاب کردم که هیچ ضرری نداشته باشد. علاوه بر این موضوع باید دقت شود که باتوجه به احتمالات بسیاری که برای قرص دوم وجود دارد، گریز از آن نشانه علاقه من برای دوری از ابهام است.

برای فردی که reward seeker است به احتمال زیاد قرص دوم مورد علاقه است. به این علت که برای این افراد punishment تعریف متفاوتی نسبت به punishment averse دارد.

سوال ۲:

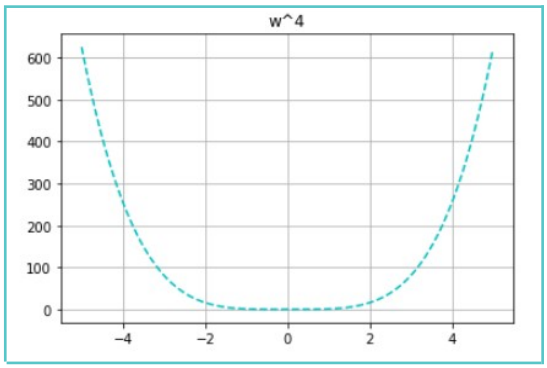
ریسک پذیری را به این صورت تعریف میکنیم که افراد ریسک پذیر در واقع reward seeker هستند، به این معنی که به دنبال پاداش بیشتری هستند بدون توجه به ضرر احتمالی. یعنی تصمیمی را میگیرند که بیشترین سود را داشته باشد. اگر تابع ارزش این افراد را مشاهده کنیم، مقـداری که برای

در مقابل آن ها افراد ریسک گریز قرار دارند که punishment averse نامیده میشوند. این افراد به دنبال این هستند که کمترین ضرر متوجه آن ها شوند و به پاداش توجه چندانی ندارند. اگر تـابع ارزش این افراد را مشاهده کنیم، و مقداری که برای مجازات در نظر میگیرند عددی منفی و از نظر اندازه بزرگتر از مقدار در نظر گرفته شده برای پاداش است.

مجازات ها مقادیر منفی w را نشان داده و مقادیر مثبت آن نمایش دهنده پاداش است.

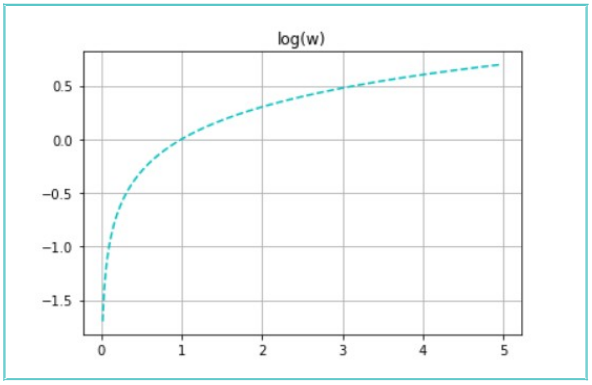
اگر نمودار های هر تابع ارزش داده شده را در بازه [۵ , ۵-] نمایش دهیم، به ترتیب زیر برای هر نمودار توضیحاتی داده و در ادامه نتیجه گیری میکنیم. (نمودار ها در jupiter notebook رسم شده اند.)

۱) w^4 :



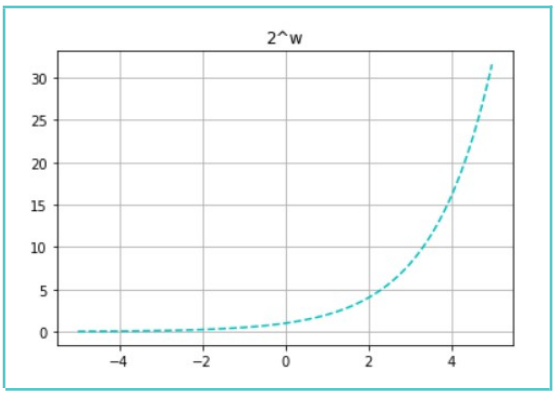
همان طور که دیده میشود تابع برای مقادیر منفی که به عنوان مجازات تلقی میشود، مقادیر بسیار بالای مثبتی میدهد. بنابراین فردی که ریسک گریز اسـت علاقـه ای بـه این گزینـه نـدارد چراکـه مقدار مجازات آن زیاد است. اما به طور هم زمان دیده میشود که برای قسمت پاداش هـا این تـابع ارزش مقدار زیادی دارد. بنابراین این تابع ارزش میتواند برای یک فرد ریسک پذیر باشد. چراکه برای این افـراد مقدار پاداش ها مهم است و ضرر ها را در نظر نمیگیرند.

۲) $\log(w)$:



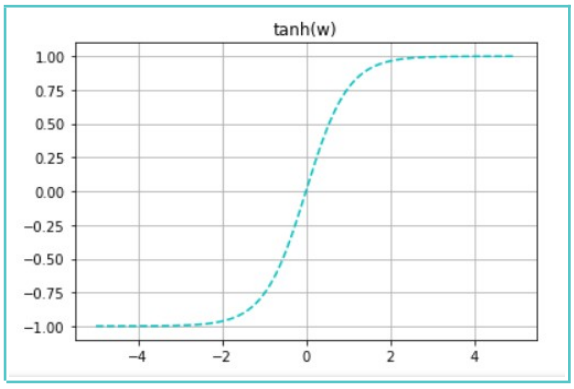
برای این تابع با توجه به این که ورودی تابع لگاریتم عددی منفی نمیتواند باشد، اطلاعاتی در رابطه با چگونگی رفتار این فرد در رابطه با مجازات ها نداریم. پس میتوان گفت مجازات ها برای فرد مهم نبوده و آنها را در نظر نمیگیرد (یعنی مهم نیست که ب مجازات ها چه مقداری اختصاص دهـد) کـه در این صورت با توجه به این که فقط به پاداش ها توجه دارد میتواند مربوط به یک فرد ریسک پذیر باشد.

۳) 2^w :



همان طور که در شکل دیده میشود برای مجازات ها مقادیر بسیار کمی(در مقایسه با پاداش ها) در نظر گرفته شده است به همین علت میتواند مورد مناسبی برای یک فرد ریسک گریز باشد.

۴) $\tanh(w)$:



برای این تابع هم همین طور که دیده میشود برای مقادیر منفی که نمایش دهنده مجازات هستند، مقادیر کم و منفی در نظر گرفته شده است. با توجه به تعاریف بالا برای یک فرد ریسک گریز، این تابع میتواند با تابع ارزش مناسب باشد. مقادیر پاداش برای این تابع زیاد نیستند اما این موضوع برای یک فرد ریسک گریز اهمیت چندانی ندارد. تا وقتی که مقادیر ارزش برای مجازات ها کم باشد، تابع برای یک فرد ریسک گریز مورد مناسبی است.

بنابراین برای جمع بندی میتوان گفت دو تابع اول برای یک فرد ریسک پذیر و دو تابع دوم برای یک فرد ریسک گریز است.

در میان دسته اول w^4 با توجه به این که مقدار پاداش و به همان نسبت مقدار مجازات زیاد است بنابراین شخص ریسک زیادی کرده است و میتوان گفت این تابع ریسک پذیری را به مقدار بیشتری نمایش میدهد. در نتیجه به تغییرات واکنش بیشتری نشان میدهد.

در میان دسته دوم که مربوط به فرد ریسک گریز است، با توجه به این که در تابع $\tanh(w)$ مقادیر منفی دارای ارزش منفی هستند این تابع نشان دهنده ریسک گریزی بیشتری است. در قسمت مربوط به پاداش ها دیده میشود که که تابع 2^w نسبت به تغییرات واکنش زیادی دارد که التبه این موضوع همان طور که بیان شد برای یک فرد ریسک گریز اهمیت چندانی ندارد.

سوال ۳:

همان طور که در سوال گفته شده هدف اکبر افزایش تعداد برد های خود است. با توجه به این که برد به معنی گرفتن پاداش یا مجازات و به بیان بهتر گرفتن امتیاز مثبت و یا امتیاز منفی است، میتوان گفت تنها موضوعی که برای اکبر اهمیت دارد علامت امتیاز است و نه مقدار آن. بنابراین تابع ارزش وی تابعی از علامت امتیازات خواهد بود. علامت منفی برای وی مجازات زیادی به همراه داشته و علامت مثبت هم برای وی پاداش زیاد.

این موضوع تا حدودی برای کبری هم صادق است با این تفاوت که برای کبری بیشتر شدن امتیازات در کل اهمیت دارد. به این معنی که علاوه بر این که علامت امتیاز مهم است مقدار آن هم اهمیت دارد. علامت برای کبری مهم است چراکه نمیخواهد از امتیازاتی که تا به حال بدست آورده مقداری کم شود. اندازه برای وی مهم است چراکه میخواهد در صورتی که امیتازی منفی گرفته مقدار آن کمترین مقدار ممکنه باشد تا از مجموع امتیازاتش، مقدار کمتری کم شود و به همین صورت اگر امتیاز مثبتی گرفته، مقدار آن بیشترین مقدار ممکنه باشد. بنابراین تابع ارزش کبری تابعی از علامت و اندازه امتیاز است. به این ترتیب که برای علامت منفی باید ارزش برای اندازه کم، کم و برای اندازه بزرگ زیاد باشد. چراکه در صورتی که علامت منفی و اندازه کم باشد به مقدار کمی از مجموع امتیازات کم میشود اما اگر اندازه یک امتیاز منفی بزرگ باشد میتواند مجموع امتیازات کبری را بسیار کم کند. بنابراین از نظر کبری علامت منفی با اندازه کوچک مجازات کمتری نسبت به امتیاز با علامت منفی و اندازه بزرگ دارد.(یعنی کبری باید نسبت به مقادیر بزرگ آن ارزش منفی زیادی قایل شود به این معنی که از آن گریزان بوده و در نتیجه مجازات زیادی برای این حالت قایل باشد.) و برای علامت مثبت ارزش برای اندازه کم نسبت به اندازه زیاد کمتر است. به این علت که اندازه کم، در مجموع امتیازات وی تاثیر کمی دارد. بنابراین کبری در ذهن خود برای علامت مثبت با اندازه کمتر پاداش کمتری نسبت ب امتیاز با عدد بزرگتر در نظر میگیرد. (چراکه علامت مثبت و اندازه زیاد به معنی افزایش امتیاز به مقدار زیاد است که به مفهوم پاداش زیاد برای کبری است.).

سوال ۴:

برای حل قسمت الف نیاز به یک فرض اولیه داریم. به این ترتیب که زندانی میداند که نگهبان در یک بازه زمانی مشخص به چند بار به سلول ها سر میزند. در این صورت میتواند با توجه به این که چند بار به سلول او آمده است متوجه شود که چند بار به سلول بقیه رفته است. اما نکته ای که وجود دارد این است که در نمیتواند متوجه شود که دقیقا چند بار به هرکدام از سلول ها سر زده است. به این معنی که میتواند در مجموع بگویید که زندان بان به دو سلول دیگر چند بار رفته است، اما این که این تعداد برای هر سلول دیگر چه مقدار است برای وی قابل تعیین کردن نیست. اما اگر اطلاعات دیگری را به فرضیات وی اضافه کنیم میتوان این احتمالات را به صورت دقیق تر بدست آورد. مثلا اگر بداند که نگهبان به تعداد مساوی به دو سلول دیگر سر میزند. به این ترتیب احتمالات نصف شده و میتوان تا حدودی دقیق تر شد.

اگر این زندانی حتی از تعداد کل رفت و آمد های نگهبان خبر نداشته باشد، هیچ احتمالی را نمیتواند بدست آورد. به این معنی که او فقط میبیند که زندان بان به سلول او آمده و خارج میشود و هیچگاه نخواهد دانست که این تعداد بار ها چه تعداد از کل بار ها است.

برای حل قسمت ب با دو فرض متفاوت به حل پرداخته شده است.

فرض ۱) در این قسمت فرض میکنیم زندانی هیچ اطلاعی از گراف داده شده در سوال ندارد. به این معنی که این زندانی در ابتدا به صورت کاملا رندوم احتمال انتخاب هر سلول توسط نگهبان را در نظر میگیرد. بعد از گذشت چند لوپ و مشاهده به آپدیت کردن احتمالات انتخاب کردن هرسلول میپردازد. به این ترتیب این زندانی احتمالات رفتن نگهبان به هر سلول و در واقع احتمال یال های گراف را با شمارش کل بارهایی که حرکت کرده میتواند بدست آورد. اما اشکالی که این روش دارد این است که زندانی با فرض اولیه غلطی شروع به آپدیت کردن باور خود کرده است. در این حالت قطعا نمیتواند به احتمالاتی که روی یال های گراف بیان شده برسد.

برای کد زدن این قسمت ابتدا به صورت کاملا رندوم (یعنی با احتمال های یکسان ۱/۳) یکی از سلول ها انتخاب میشود. سپس سلول بعدی با احتمال ۱/۲ انتخاب میگردد و این کار برای ۴ بار تکرار

میشود. سپس با توجه به انتخاب های این قسمت، احتمال برای انتخاب سلول بعدی محاسبه میگردد. این کار به تعداد زیادی انجام میشود تا به حدودا به احتمالاتی برای یال های گراف ها برسیم.

فرض ۲) در این قسمت برخلاف قسمت قبلی احتمال انتخاب هر سلول براساس گراف داده شده در صورت سوال است. به این معنی که نگهبان برای رفتن از هر سلول به سلول دیگر با احتمالی برابر عدد های روی یال های گراف حرکت میکند. زندانی هم براساس این که نگهبان جند بار هر یک از یال ها را پیموده و کل حرکت های نگهبان احتمال هر یال را انتخاب میکند. در این حالت باتوجه به این که زندانی با توجه به حرکت نگهبان با توجه به احتمالات روی یال ها به محاسبه پرداخته است، به همان احتمالات روی یال ها همگرا میشویم. (به صورت تقریبی البته)

به این ترتیب همان طور که مشاهده میشود در قسمت اول دانشی در رابطه با احتمال روی یال ها نخواهیم داشت و اما در قسمت دوم به احتمال یال ها رسیدیم. پس در حالت اول به خطا میرسیم.

(کد های فرض ۱ و ۲ در قسمت کد های ضمیمه شده موجود است.)

سوال ۵:

در این قسمت برای محاسبه بتا باید باتوجه به صورت سوال مقادیر را تعیین کنیم :

$P_s(o) = 0.5 \ , \ P_o(o = win) = 0.1 \rightarrow$

$$0.5 < \frac{e^{\beta(0.1)}}{e^{\beta(0.1)} + e^{\beta(0.9)}} \rightarrow e^{\beta(0.9)} < e^{\beta(0.1)} \rightarrow \beta(0.9) < \beta(0.1) \rightarrow \beta < 0$$

سوال ۶:

الف)

فضای حالت در واقع بیانگر این است که مار پس از انجام یک عمل و گرفتن یک پاداش یا مجازات، از یک حالت به حالت دیگری برود. بنابراین فضای حالت این بازی مختصات مار در صفحه بازی خواهد بود. به این ترتیب که با انجام هر عمل و گرفتن هر پاداش یا مجازات به مختصات دیگری در صفحه میرود. پس فضای حالت شامل زوج مرتب های (x,y) است.

فضای اعمال هم با توجه به این که مار ۴ جهت اصلی را طی میکند، شامل حرکت به بالا، پایین، چپ و راست خواهد بود. به این معنی که اگر عمل بالا رفتن اجرا شود در زوج مرتب حالت مار x بدون تغییر مانده و به y یک واحد اضافه میشود. بنابراین به استیت دیگری با شرایط بیان شده میورد. برای چپ، راست و پایین هم با توجه به توضیحات استیت بعدی مشخص میشود.

ب)

برای این که عامل یادبگیرد که باید بسته به حالتی که قرار دارد، (زوج مرتب مختصات در صفحه بازی) چه عملی را باید انتخاب کند و همچنین محل دیوار ها هم ثابت هستند، باید تابع ارزشی که تعیین میشود به این صورت باشد که در صورتی که مار با حرکتی به میوه ای برسد پاداشی میگیرد و در صورتی که میوه را تمام کند پاداشی بسیار بزرگ بگیرد. به این ترتیب عامل می آموزد که برای گرفتن پاداش باید به میوه رسیده و آن را تمام کند. اما برای این که یادبگیرد که باید از دیوار ها دوری کند، لازم است برای خوردن به دیوار مجازات زیادی در نظر بگیریم. به این ترتیب در نظر عامل تابع ارزش برای خوردن به دیوار مقدار مجازات زیادی در نظر گرفته است.

بنابراین رسیدن به میوه پاداش دارد و اتمام آن پاداشی بزرگتر و خوردن به دیوار مجازات زیادی.

برای این که عامل بیشتر از دیوار ها دوری کند که مخصوصا این موضوع برای زمانی که طـول مـار زیـاد است، میتوان مقدار مجازات تعریف شده را بسیار بیشتر از پاداش گرفتن تعریف کرد (از نظر اندازه) و به این ترتیب عامل کاملا punishment averse خواهد شد.

ج)

در این قسمت بهتر است که فضای حالات و اعمال را تغییر ندهیم و با عوض کردن تـابع ارزش و به بیان دیگر آپدیت کردن تعاریف پاداش یا مجازات عامل را آموزش دهیم. به این ترتیب که برای پاداش علاوه بر موارد اشاره شده در ب، برخورد به سر مار دیگر پاداش بسیار کمی در مقایسه با دو مـورد دیگـر خواهد داشت. برای مجازات هم علاوه بر مجازات های شامل برخورد با دیوار برخورد با بدنه مـار دیگـر را هم میتوان به مجازات ها در نظر گرفت. این مجازات چون مانند برخورد به دیـوار بـاعث بـاختن میشـود، مجازاتی به همان اندازه زیاد به همراه خواهد داشت.

اما اگر بخواهیم حتما فضای حالت یا فضـای اعمـال را تغیـیر دهیم، بهـتر اسـت این تغیـیر را در فضای حالت ایجاد کنیم. در واقع با اضافه کردن مار فضای اطراف از نظـر عامـل دارای یـک عامل non-deterministic دیگر شد. بنایراین برای از بین بردن این عامل، اضافه کردن مختصات مار دیگر به فضای حالت به ما کمک میکند. به این معنی که اگر مار در آن نقطه حضـور دارد بـه آن نقطـه نـرود. در واقع به این شکل ما سعی میکنیم این احتمالاتی بودن محیط را در نظر عامل کم کنیم که البته این کـار هزینه در بردارد. چراکه برای اضافه کردن مختصات مار دیگر نیاز بـه داشـتن اطلاعـات در رابطـه بـا وی داریم که این بدست آوردن خود باعث هزینه میشود.

بنابراین برای جمع بندی آپدیت کردن تابع ارزش و یـا همـان تغیـیر تعریفـات از پـاداش و جـزا میتواند تصمیم بهتری باشد چراکه هزینه کمتری در بردارد. گرچه که باعث میشود تا حدودی یـادگیری دیرتر همگرا شود. اگر هزینه کردن مسیله مهمی نباشد، میتوان مختصات مار دیگر را اضافه کرد.

د)

باتوجه به این که رفتن به حالات (در این جا مختصات) مختلف probabilistic است و به عبارت بهتر رفتن از یک حالت به حالت دیگر دارای احتمال است و در واقع مسیله non-deterministic میباشد، خیر همواره به تساوی نمیرسیم. تمامی این تصمیم ها که چه عملی انجام شود به صورت احتمالی بررسی میشوند. بنابراین تساوی به معانی ای که در دنیای deterministic با آن روبرو هستیم رخ نمیدهد. درواقع ممکن است در چند بازی سر مار ها برخورد کرده و ما به تساوی برسیم ولی حکمی در رابطه به تمامی بازی ها نمیتوان داد. باید به صورت آماری با این مسیله برخورد کرد. به این ترتیب که باتوجه به این که احتمال برخورد سر مارها به یکدیگر چقدر است گفت که در کل به صورت میانگین به آن احتمال بازی به تساوی کشیده میشود. برتری مار ها هم به این صورت است که ممکن است ماری با تصمیم هایی که در رابطه با حرکت های بعدی میگیرد سریع تر به تصمیم بهینه در هر شرایط برسد و بنابراین برنده شود.