

دو NeuronGroup از کتابخانه Brian2 تعریف میکنیم و به هر کدام یک معادله  $dv/dt$  میدهیم. معادله یک نورون رو سینوسی میدهیم و یک نورون را خطی زیر با این روش می توانیم فاصله زمانی های اسپایک زدن متفاوتی را به دست بیاوریم. (نمودار ۱)

دو نورون را با یک سیناپس به هم وصل میکنیم و معادله ی  $stdp$  را برای وزن این سیناپس به آن نسبت میدهیم.

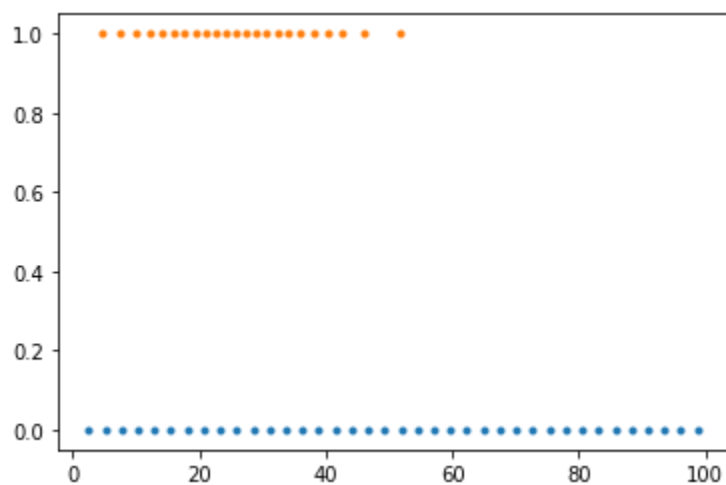
سپس هر دو نورون را به هم وصل میکنیم و جریان ثابتی را به هر دو نورون میدهیم. نورون اول جریان ۴ آمپر و نورون دوم جریان ۱.۵ آمپر

سپس بعد از تمام شدن زمان مشخص نتایج را بررسی میکنیم.

نمودار ۱ :

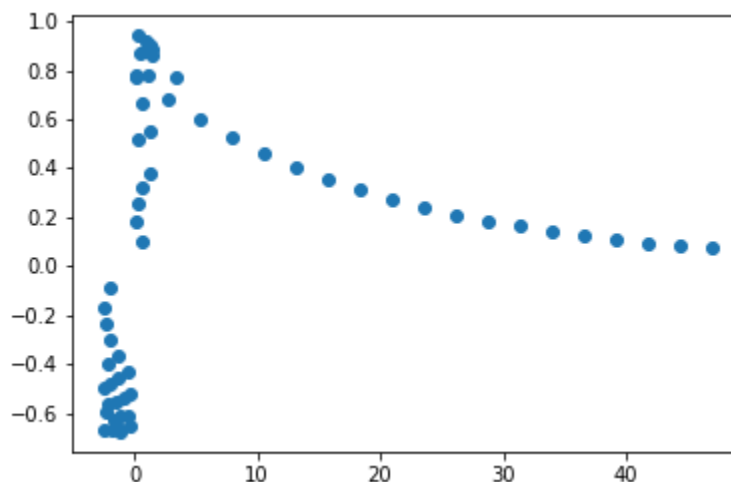
نمودار اسپایک زدن دو نورون (raster plot)

نورون با ایندکس ۱ نورون خروجی و دیگری نورون ورودی است.

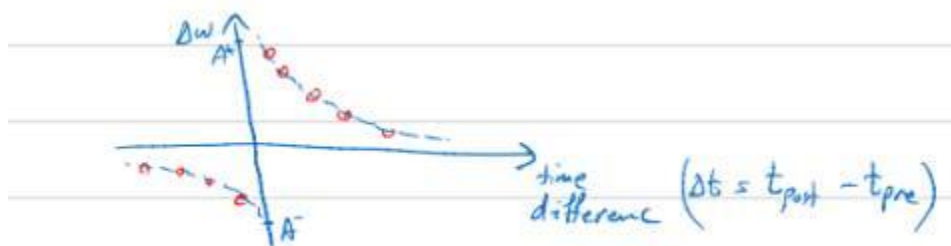


حالا اختلاف زمانی هر اسپایک و میزان تغییر هر اسپایک در آن فاصله زمانی را بررسی میکنیم. و به نمودار زیر میرسیم.

مشخص است که در دلتا تی های موجود ما که معمولا دلتا تی های مثبت هستند مدل ما طبق



انتظار ما که نمودار زیر است رفتار میکند

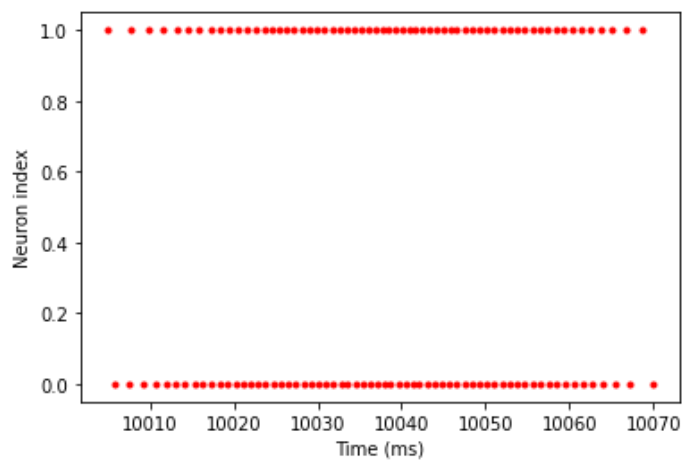


۲

حالا ما ۲۰ نورون در لایه اول تعریف میکنیم و طبق انکود اسپایک ریت ما جریان متناسب با پترن را به آن ورودی میدهیم و در نهایت اسپایک های دو نورون خروجی را بررسی میکنیم.

اولین پترن:

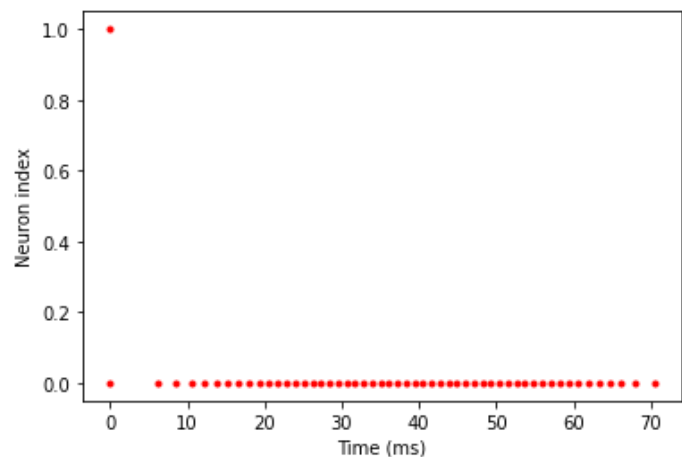
[0,0,0,0,0,0,0,0,0,0,0,0,2,1,2,1,2,1,2,1,2]



اسپایک دو نورون خروجی در فرایند یادگیری:

پترن دوم:

[1,2,1,2,1,2,2,1,2,1,2,0,0,0,0,0,0,0,0,0,0]

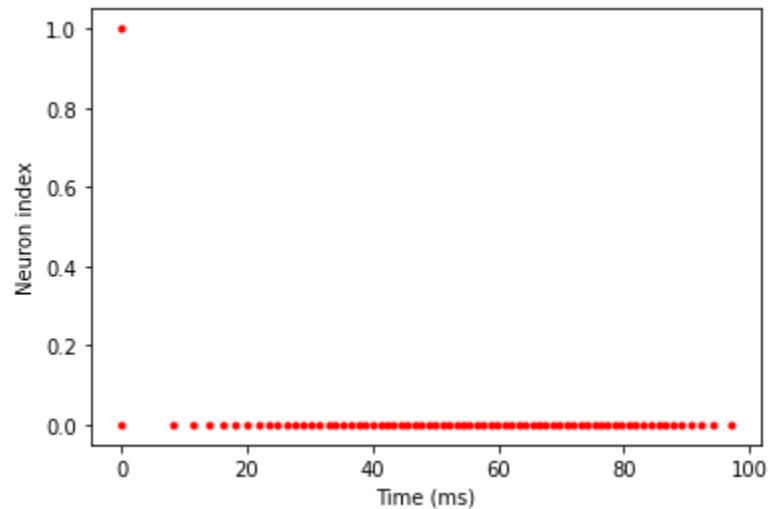


اسپایک دو نورون خروجی در فرایند یادگیری:

حالا یک پترن شبیه پترن اول می‌دهیم تا مدل را تست کنیم:

پترن:

[0,0,0,0,0,0,0,0,0,1,0,1,1,2,2,1,2,1,2]



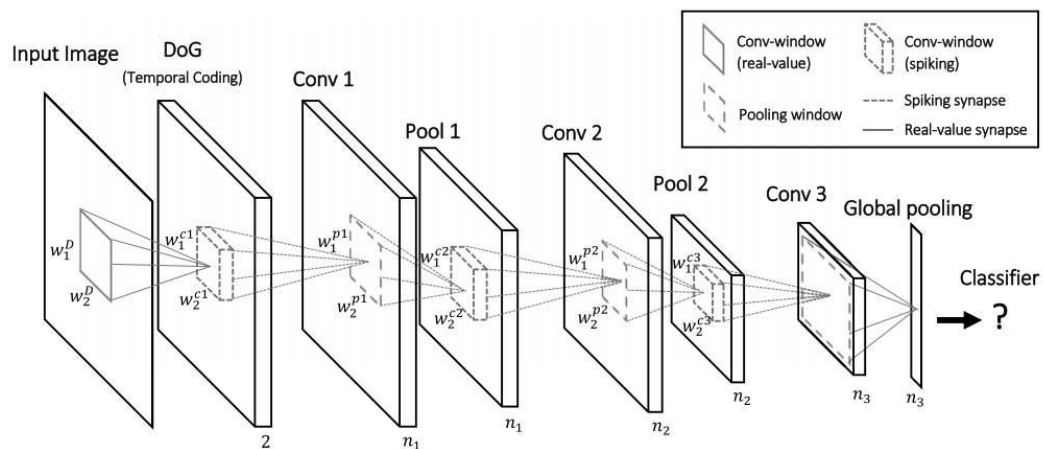
مشخص است که مدل پترن اول را یاد گرفته است.

۳

در این مقاله شبکه های کانولوژنی ای ساخته میشود که به صورت unsupervised و با کمک روش یادگیری stdp بتواند دیتا های عکسی را کلسیفای کند.

شبکه های کانولوژنی معمولا به صورت سوپروایزد و از تکنیک back propagation استفاده میکند اما شبکه های unsupervised مزایای بیشتری نسبت به شبکه ها سوپروایزد می توانند داشته باشند همچنین این شبکه ها به دلیل تعداد کم اسپایک ها بسیار اسپارس هستند و حجم محاسبات کمتری دارند همچنین در مورد مزایای این شبکه نسبت به متود های دیر آن سوپروایزد مانند random corpdسند یا auto encoder ها بحث می شود.

این شبکه از چندین لایه تشکیل شده است.



لایه اول که لایه اینپوت است و عکس را دریافت میکند

بعد از آن یک لایه DoG داریم که مرز ها و تفاوت ها را مشخص میکند و اطلاعات را از طریق تفاوت اسپایک ها منتقل می‌کند. و بعد از آن دو جفت لایه کانولوژن و پولینگ داریم که لایه کانولوژنی از با فیلتر آن از لایه های عقب تر تصاویر کلی تر مانند خطوط و اشکال را تشخیص میدهد و در لایه های جلویی اشکال پیچیده تر مانند جسم ها را تشخیص میدهد و بعد از آن به لایه کانولوژن دیگری می‌رسیم که اشکال بسیار پیچیده مانند چهره را تشخیص میدهد و بعد از یک لایه گلوبال پولینگ که کلسیفای می‌کند میرسیم. لایه پولینگ صرفاً کارش فشرده سازی اطلاعات است و ابعاد را کوچکتر میکنیم که محاسبات کمتر و دید کلی تری در لایه های بعدی داشته باشیم همانطور که قبلاً هم گفته شد یکی از مزیت های بزرگ این شبکه مقدار بسیار کمتر حجم محاسبات است.

از آنجا که ما در فرایند یادگیری از متود stdp استفاده میکنیم میتوانیم به صورت unsupervised الگو ها را یاد بگیریم و بنا به تعداد بعد لایه خروجی میتوانیم تصاویر را برحسب شباهت به n کلاس دسته بندی کنیم. یعنی هر ورودی که ما به شبکه میدهم وزن های آن به روش stdp که توضیح آن در این مقاله تمیگنجد جوری تغییر میکند و که هر الگویی که در عکس آورده می‌شود را یاد می‌گیرد.

این مدل را روی سه دیتا ست face motorbike dataset و eth-80 dataset و mnist اعمال کرده و به نتایج نسبتاً خوبی هم به آن میرسند.

۴

**learning Reinforcement:** یادگیری تقویتی نه supervised است نه Unsupervised.

به این دلیل که به واسطه عملی که agent انجام میدهد یک فیدبک از محیط میگردد Supervised نیست چون به agent نمیگوییم چی درست یا چی غلط است صرفاً سیگنالهای جایزه یا تنبیه می‌فرستیم. یک agent یادگیرنده داریم که در یک محیط تعامل انجام میدهد. این محیط حالتهای مختلفی دارد.

به عنوان agent ابتدا باید محیط را درک کنیم و بدانیم در چه state ای وجود داریم. امکان دارد observable fully یا observable partial باشد. Agent تصمیم بگیرد که عملیاتی را در قبال محیط انجام دهد این اکشن محیط را تحت تأثیر قرار میدهد.

بر مبنای عملی که انجام داده محیط مجازات میکند یا جایزه میدهد. تمام حسهای مثبت و منفی که ادراک می‌کنیم از اعمال و رفتار همان نتیجه فرایند یادگیری تقویتی است. مثلاً سیگنال reward در مغز ترشح میشود. هر حالت مثبت و منفی که احساس میکنیم به خاطر تغییر حالت شیمیایی در مغزمان است. در محیط Agent سعی می‌کند جایزه هایش را بیشینه کند.

جایزه خیلی فاصله ای با عمل ندارد • اگر زمان خیلی زیادی بین عمل و جایزه باشد ممکن است Action های دیگری هم انجام داده باشیم و ممکن است assignment credit سخت باشد که این جایزه یا تنبیه به خاطر کدام عمل بوده که در ازای آن تغییر ایجاد کنیم. • باید تشخیص دهیم چه بخش هایی از شبکه عصبی درگیر این عمل بوده‌اند. که بتوانیم تضعیف یا تقویت انجام دهیم. • کدام اسپایکها بوده که در ازاش credit گرفتیم. Dopamine : یکی از فرایندهای یادگیری تقویتی شناخته شده در مغز مبتنی بر یک modulator-neuro یه اسم dopamine است.

احساس رضایت در نتیجه دوپامین ایجاد میشود.

**STDP** دوفاز دارد :

**LTP** : موقعی که سیناپس ها تقویت میشود؛

**LTD**: موقعی که سیناپس ها تضعیف شود.

وقتی دوپامین ترشح میشود غلظتش می‌تواند نحوه انجام STDP را تغییر دهد. دوپامین میتواند یک سری از پترنهای فایر شدن را reinforce کند و عموماً تأثیر دوپامین روی STDP با تأخیر یک تا دو ثانیه- ای همراه است. دوپامین همیشه در مغز ترشح می‌شود اگر مقدار آن از یک حدی پایین تر بیاید punishment داریم و اگر از یک مقدار بیشتر شود. reward خاصیتی که دوپامین دارد که بخواهد الگوهای firing که با افزایش دوپامین همراه هستند را تقویت بکند و بقیه تضعیف شوند. بر مبنای یک فرایند سیناپتیکی به اسم traces eligibility synaptic یا tag synaptic است. این فرایند می‌گوید وقتی نورون ها فایر میشوند تا مدتی بعد از فایر اثر آن در نورون باقی می‌ماند موقعی که دوپامین بعد از تأخیر می‌آید سیناپسهایی که trace eligibility شان بیشتر شده باشد تقویت میشوند. اگر دوپامین کم شود، تضعیف میشود.

معادله RSTDP بر اساس ترشح:

The dynamics of these variables follow:

$$\begin{aligned}\frac{dc}{dt} &= -\frac{c}{\tau_c} + STDP(\tau)\delta(t - t_{pre/post}), \\ \frac{ds}{dt} &= cd,\end{aligned}$$

where  $\tau = t_{post} - t_{pre}$ ,  $d$  describes the extracellular concentration of DA and  $\delta(t)$  is the Dirac delta function.

با این معادله سیستم ما به یادگیری تقویتی میپردازد