



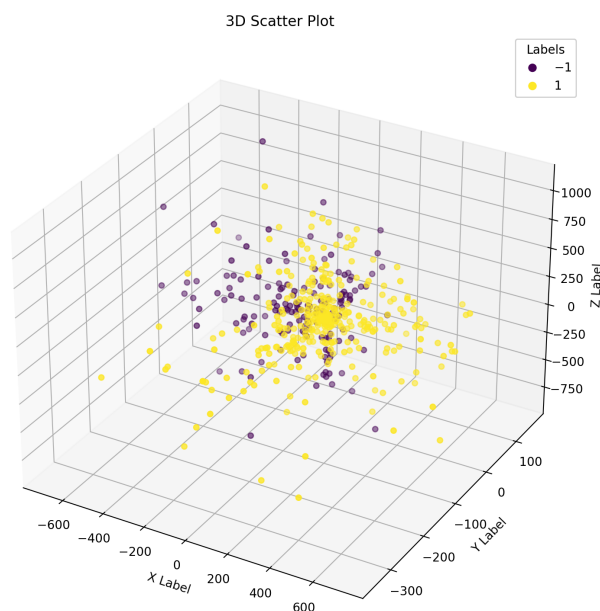
دانشکده علوم
دانشکده ریاضی، آمار و علوم کامپیوتر
تمرین سری چهارم حسابگری زیستی

داده های متناظر با سوالات ۱ و ۲ را میتوانید از این لینک دانلود کنید: [دانلود دیتاست](#)

مسئله ۱. در این سوال، هدف پیاده سازی الگوریتم پرسپترون (Perceptron) از ابتدا به صورت Scratch می باشد. دیتاست داده شده متشکل از ۵۰۰ نقطه می باشد و شامل سه ستون x ، y و z بوده که نشان دهنده موقعیت نقاط در صفحه مختصات هستند. ستون چهارم این دیتاست (label) نشان دهنده برچسب هر نقطه در فضای مختصات است که می تواند ۱ یا -۱ باشد.

حال باید با استفاده از الگوریتم پرسپترون، یک صفحه ای را پیدا کنید که نقاط را به دو بخش تقسیم کند؛ به طوری که نقاطی که در یک طرف صفحه قرار دارند برچسب ۱ داشته باشند و نقاط سمت دیگر برچسب -۱ را داشته باشند. (تضمین می شود که چنین صفحه ای وجود دارد.)

- الگوریتم را به ازای مقادیر مختلف η (نرخ یادگیری) اجرا و نتایج را گزارش کنید.

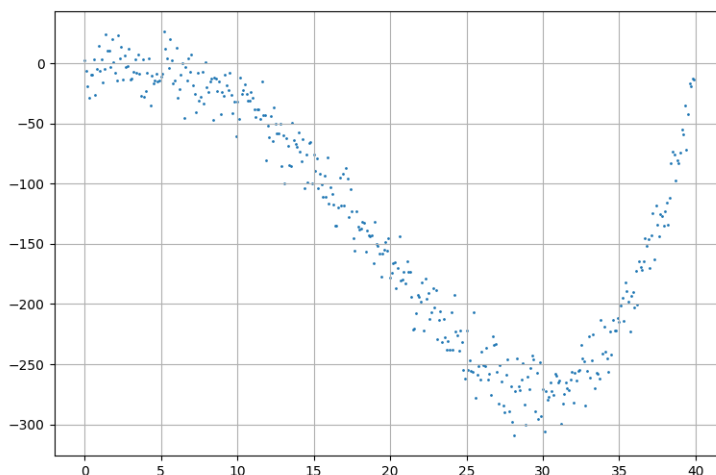


مسئله ۲. دیتاستی متشکل از ۴۰۰ سطر دو ستون x و y به شما داده شده است. هدف پیاده سازی یک شبکه عصبی تک لایه می باشد که بتواند با ورودی گرفتن x ، مقدار y را پیش بینی کند. برای بررسی دقت از معیار Mean-Squared-Error (MSE) استفاده کنید:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

مقادیر مختلفی برای تعداد نوروں های ورودی در نظر گرفته (چند جمله هایی از درجات مختلف ۱ تا ۴ برای برازش در نظر بگیرید.) و اثر Overfitting را با افزایش تعداد نوروں های ورودی بررسی کنید. انتظار می رود مدل ارائه شده، بر روی داده های train دارای شرایط زیر باشد:

- برای چند جمله ای $ax + b$: $MSE \leq 4500$
 - برای چند جمله ای $ax^2 + bx + c$: $MSE \leq 3750$
 - برای چند جمله ای $ax^3 + bx^2 + cx + d$: $MSE \leq 2400$
 - برای چند جمله ای $ax^4 + bx^3 + cx^2 + dx + e$: $MSE \leq 1300$
- دقت شود برای برازش توابع با درجات بالاتر، برای بررسی Overfitting میتوان از توابع آماده استفاده کرد.

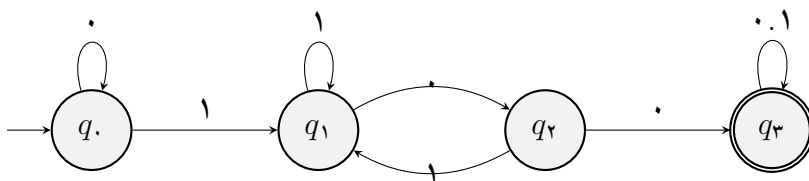


- تقسیم بندی داده های train و test را ابتدا با نسبت ۹۰ به ۱۰ انجام داده، و سپس سائز داده های test را در بازه های ۱۰ درصدی افزایش داده (۱۰, ۲۰, ۳۰, ..., ۹۰) و مجدداً مدل را ران کرده و نتایج را گزارش کنید.
- پس از پیاده سازی شبکه عصبی و تحلیل نتایج، به مدل Regularization را اضافه کرده و نتایج را با یکدیگر مقایسه کنید.

مسئله ۳. در این پرسش ابتدا با ماشین متناهی قطعی آشنا خواهید شد، سپس با پاسخ دادن به چند پرسش، شبکه های عصبی برای آن طراحی خواهید کرد.

ماشین متناهی قطعی (DFA)

در بیانی ساده می توان ماشین متناهی قطعی را جعبه سیاهی فرض کرد که ورودی را دریافت می کند و اگر متوجه الگویی خاص در ورودی ها شود، آن را در خروجی اعلام می کند. برای این کار از مجموعه ای از حالت ها استفاده می کند تا الگوهای مشاهده شده را بتواند ذخیره کند. یک ماشین متناهی قطعی در نظر بگیرید با الفبای $\{0, 1\}$ که بتواند با حداقل یکبار مشاهده ۱۰۰ وارد حالت پذیرش بشود.



شکل ۱: شماره داخل دایره، شماره حالت است و عدد روی یال، ورودی ای است که حالت کنونی توسط آن به حالت بعدی می رود. اگر ورودی ها تمام شوند و در حالتی باشیم که حالت دو خط دارد ورودی مورد نظر توسط دستگاه تشخیص داده شده است. (حالت سه)

به عنوان مثال، ورودی ۰۱۱۰۰۱ را در نظر بگیرید، جدول انتقال حالت برای این ورودی و ماشین قطعی فوق به صورت زیر خواهد بود:

جدول ۱: جدول انتقال حالت

حالت کنونی	ورودی	حالت بعدی	پذیرفتن
۰	۰	۰	۰
۰	۱	۱	۰
۱	۰	۲	۰
۱	۱	۱	۰
۲	۰	۳	۱
۲	۱	۱	۰
۳	۰	۳	۱
۳	۱	۳	۱

به کمک نورون McCulloch-Pitts توسعه یافته، DFA بیان شده را شبیه سازی کنید به این صورت که حالت فعلی و ورودی DFA هر دو به عنوان ورودی شبکه نورونی محسوب شده و حالت بعدی و اینکه خروجی پذیرفته شده یا خیر (۱ برای پذیرش و ۰ برای عدم پذیرش)، به عنوان خروجی شبکه محسوب میشوند. (بنابراین ۳ نورون ورودی و ۳ نورون خروجی خواهیم داشت).

توجه کنید که شماره حالت ها، ورودی و پذیرش شدن یا نشدن حالتها، همگی دودویی هستند. همچنین ترتیب زمانی انجام عملیات در این سوال مهم نیست. بنابراین نیازی به در نظر گرفتن تاخیر برای انجام عملیات نیست.

۱. جدول انتقال حالت DFA را به جدول حالت توضیح داده شده متناسب با شبکه نوروں ها تبدیل کنید.
۲. شبکه هر خروجی را به صورت جداگانه و به همراه توضیحات مختصری، رسم نمایید. برای این بخش دقت داشته باشید:
 - نیازی به کد نویسی در این بخش نیست.
 - حتما سه شبکه جدا از هم رسم شوند.
 - شبکه ای که برای هر خروجی رسم می کنید تا حد ممکن دارای کم ترین تعداد نوروں و کم ترین threshold باشد.
 - تعداد نوروں کم تر دارای اهمیت بالاتری نسبت به threshold کم تر است.
 - Threshold ها اعداد صحیح باشند.
 - تمام شبکه برای یک خروجی دارای threshold یکسان باشد.
۳. سه شبکه رسم شده در بخش قبلی را به صورت بهینه (با کمترین threshold و تعداد نوروں) و با threshold یکسان با هم ادغام کرده و رسم کنید.
۴. با استفاده از زبان پایتون شبکه های طراحی شده در بخش قبل پیاده سازی کرده و خروجی تمامی حالت ها به ازای تمامی ورودی ها را نمایش دهید.

برای تحویل تکلیف به نکات زیر توجه فرمایید:

- برای هر دو مسئله کد شما باید به زبان پایتون و در Jupyter Notebook پیاده سازی شود.
- خوانایی و دقت گزارش نهایی از اهمیت ویژه ای برخوردار است.
- همه ی کدهای پیوست گزارش باید قابلیت اجرای مجدد داشته باشند.