



۱۳۰۷

دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشکده مهندسی برق و کامپیوتر

نام و نام خانوادگی :

امیر اسماعیل زاده نوبری

شماره دانشجویی

40101924

درس یادگیری ماشین

امتحان میان ترم

استاد درس:

دکتر علیاری

لینک [GIT](#)

لینک [Drive](#)

بهار 1403

الحمد لله الذي
خلقنا من
الحمم

Contents

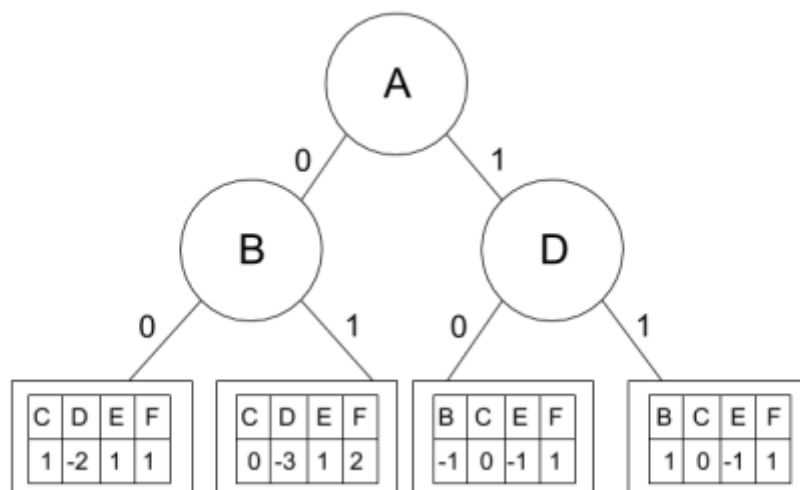
سوال 1	4
Error! Bookmark not defined.	1
Error! Bookmark not defined.	2
Error! Bookmark not defined.	3
Error! Bookmark not defined. : Logistic regression (الف	
Error! Bookmark not defined. : Perceptron (ب	
Error! Bookmark not defined.	4
Error! Bookmark not defined. Logistic Regression (الف	
Error! Bookmark not defined. : perceptron (ب	
Error! Bookmark not defined.	5
سوال 2	5
Error! Bookmark not defined.	1
Error! Bookmark not defined.	2
Error! Bookmark not defined. (آ	
Error! Bookmark not defined. (ب	
Error! Bookmark not defined. (ج	
Error! Bookmark not defined. (د	
Error! Bookmark not defined.	3
Error! Bookmark not defined.	4
سوال 3	6
Error! Bookmark not defined.	2
Error! Bookmark not defined.	3

1) خیر، این جمله به صورت کلی درست نیست. طبقه‌بند بیز یکی از روش‌های طبقه‌بندی است که در برخی موارد بسیار موثر و مفید است، اما نمی‌توان گفت که همیشه بهترین روش برای جداسازی یک مسأله دوکلاسه است. انتخاب بهترین طبقه‌بند به عوامل مختلفی از جمله: ماهیت داده‌ها، توزیع داده‌ها، پیچیدگی محاسباتی و بستگی دارد.

2) این جمله به طور کلی درست است، اما نیاز به توضیح و تحلیل بیشتری دارد. استفاده از رویکرد بیز برای تخمین پارامترهای توزیع می‌تواند به کاهش بیش‌برازش (Overfitting) کمک کند، به ویژه به دلیل ویژگی‌های ذاتی این رویکرد.

3) بله، این جمله به طور کلی درست است. استفاده از معیار **Information Gain** (کسب اطلاعات) برای ساخت درخت در شرایطی که بعضی از ویژگی‌ها حالات زیادی دارند، می‌تواند مناسب نباشد. از دلیل آن، مشکل بایاس که **Information Gain** نسبت به ویژگی‌هایی با تعداد حالات بیشتر دارد و همچنین تقسیم بیش از حد داده‌ها.

4) بله، هر شبکه عصبی چندلایه با توابع فعال‌ساز خطی قابل تقلیل به یک شبکه عصبی تک‌لایه با یک تابع خطی معادل است. این به این دلیل است که توابع خطی تجمعی می‌توانند به یک تابع خطی ساده‌تر تبدیل شوند.



(الف)

۱. برای نمونه $x = [1, 1, 0, 1, 0, 1]$ ، درخت پرسپترون داده شده چه برجسب خروجی ای را پیش بینی می کند؟

با توجه به نمونه داده شده $A=1$ و $D=1$ در leafnode سمت راست هستیم.

در نتیجه وزن های پرسپترون ما $W = \{w_b = 1, w_c = 0, w_e = -1, w_f = 1\}$

و بایاس ما 1 و ورودی پرسپترون $\{1, 0, 0, 1\}$ است.

داریم.

$$Y = \text{sign}(w_b B + w_c C + w_e E + w_f F + 1) = \text{sign}(1 + 1 + 1) = 1$$

خروجی ما عدد 1 را نشان میدهد

(ب)

۲. آیا مرزتصمیم درخت پرسپترون همواره خطی است؟ برای مقادیر کوچک حداکثر عمق، کیفیت آموزش درخت تصمیم و درخت پرسپترون را با ذکر دلیل مقایسه کنید. آیا تفاوتی دارند؟

خیر، مرز تصمیم درخت پرسپترون همواره خطی نیست. هر پرسپترون در گره های برگ مرزی خطی تعریف می کند، اما ساختار کلی درخت تصمیم می تواند مرزهای غیرخطی ایجاد کند. ساختار

سلسله‌مراتبی درخت تصمیم فضای ویژگی‌ها را به مناطق مختلف تقسیم می‌کند و در هر منطقه پرسپترون عمل می‌کند. این ترکیب از تقسیمات درخت تصمیم و مرزهای خطی از پرسپترون‌ها، منجر به مرزهای تصمیم تکه‌ای-خطی می‌شود که به صورت کامل خطی نیستند.

برای مقادیر کوچک حداکثر عمق، درخت‌های پرسپترون می‌توانند کیفیت آموزش بهتری نسبت به درخت‌های تصمیم ساده ارائه دهند زیرا توانایی شناسایی الگوهای پیچیده‌تر از طریق مرزهای تکه‌ای-خطی را دارند. با این حال، این مسئله با افزایش پیچیدگی و زمان آموزش همراه است. انتخاب بین استفاده از درخت تصمیم یا درخت پرسپترون بستگی به مسئله خاص، طبیعت داده‌ها و موازنه بین قابلیت تفسیر و بیانگری دارد.

سوال 3

سوال سه

بخش یک:

روابط لایه پنهان:

$$h1 = w1 * 1 + w3 * x1 + w5 * x2$$

$$h2 = w2 * 1 + w4 * x1 + w6 * x2$$

روابط لایه خروجی:

$$z = w7 * 1 + w8 * h1 + w9 * h2$$

خروجی شبکه عصبی:

$$P(Y = 1 | X, W) = \frac{1}{(1 + e^{-cz})}$$

مقدار z را با توجه به روابط لایه پنهان، به شکل زیر به دست می‌اریم:

$$z = w7 + w8 * (w1 + w3 * x1 + w5 * x2) + w9 * (w2 + w4 * x1 + w6 * x2)$$

بنابراین:

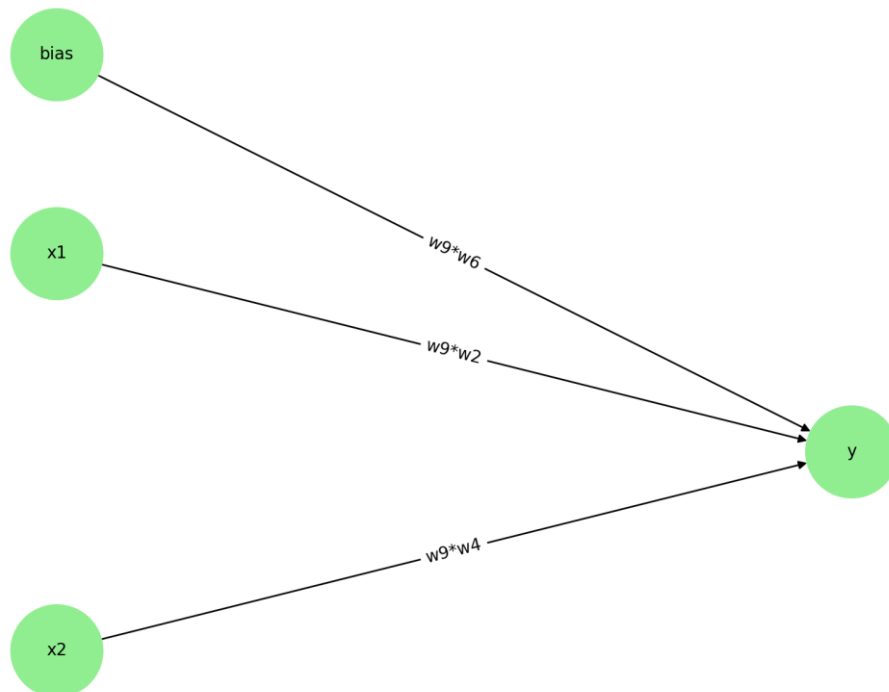
$$P(Y = 1 | X, W)$$

$$= \frac{1}{1 + e^{-c(w7 + w8 * (w1 + w3 * x1 + w5 * x2) + w9 * (w2 + w4 * x1 + w6 * x2))}}$$

بخش دوم:

با توجه به این که ترکیب چندین تابع خطی همچنان یک تابع خطی است، هر شبکه عصبی چندلایه با توابع فعال ساز خطی در لایه های پنهان می تواند به عنوان یک شبکه عصبی بدون هیچ لایه پنهانی نمایش داده شود.

بنابراین شبکه عصبی مورد سوال، بدون لایه پنهان به شکل زیر در می آید:



اگر روابط این شبکه را بنویسیم مشاهده میشود که با شبکه اصلی برابر است:

$$P(Y = 1 | X, W)$$

$$= \frac{1}{1 + e^{-c(w7 + w8 * (w1 + w3 * x1 + w5 * x2) + w9 * (w2 + w4 * x1 + w6 * x2))}}$$

سوال 4

Column	Actuator	Variable Symbol	Variable Description	Range	Units
1	1	-	Time stamp	0 – 86399	s
2	1	P51_05	P1 - juice pressure (valve inlet)	0 – 1000	kPa
3	1	P51_06	P2 - juice pressure (valve outlet)	0 – 1000	kPa
4	1	T51_01	T - juice temperature (valve outlet)	50 – 150	°C
5	1	F51_01	F - juice flow (1 st evaporator inlet)	0 – 500	m ³ /h
6	1	LC51_03CV	CV - control value (controller output)	0 – 100 ²⁾	%
7	1	LC51_03X	X - servomotor rod displacement	0 – 100 ²⁾	%
8	1	LC51_03PV	PV - process value (juice level in 1 st evaporator)	0 – 100 ³⁾	%
9	1	TC51_05	Juice temperature (1 st evaporator inlet)	50 – 150	°C
10	1	T51_08	Juice temperature (1 st evaporator outlet)	50 – 150	°C
11	1	D51_01	Juice density (1 st evaporator inlet)	0 – 25	Bx
12	1	D51_02	Juice density (1 1 st evaporator outlet)	13 – 41	Bx
13	1	F51_02	Steam flow	1 – 100	t/h
14	1	PC51_01	Steam pressure	100 – 300	kPa
15	1	T51_06	Steam temperature	50 – 150	°C
16	1	P51_03	Vapour pressure	0 – 250	kPa
17	1	T51_07	Vapour temperature	50 – 150	°C

داده ها در 17 بعد داده شده اند که بعد اول آن ها شماره سَمپل هاست که قابل حذف است.
داده های 9 و 17 نوامبر با توجه به شکل زیر آورده شده اند:

Item	Fault tag	Sample	Date	Fault description
1	f18	58800 – 59800	October 30, 2001	Partly opened bypass valve
2	f16	57275 – 57550	November 9, 2001	Positioner supply pressure drop
3	f18	58830 – 58930	November 9, 2001	Partly opened bypass valve
4	f18	58520 – 58625	November 9, 2001	Partly opened bypass valve
5	f18	54600 – 54700	November 17, 2001	Partly opened bypass valve
6	f16	56670 – 56770	November 17, 2001	Positioner supply pressure drop

طبق خواسته ی سوال از داده های 9 نوامبر برای train و validation و از داده های 17 نوامبر برای تست استفاده می کنیم.

با دستور های مختلف python می توان اطلاعات زیادی راجع به این دیتا دریافت کرد برای مثال با info(). داریم:

```
0 0      86400 non-null float64
1 1      86400 non-null float64
2 2      86400 non-null float64
3 3      86400 non-null float64
4 4      86400 non-null float64
5 5      86400 non-null float64
6 6      86400 non-null float64
7 7      86400 non-null float64
8 8      86400 non-null float64
9 9      86400 non-null float64
10 10    86400 non-null float64
11 11    86400 non-null float64
12 12    86400 non-null float64
13 13    86400 non-null float64
14 14    86400 non-null float64
15 15    86400 non-null float64
16 16    86400 non-null float64
```


که نشان از این است که داده های null نداریم . dtype از جنس float است.

با توجه به رنج سمپل داده شده های نرمال (فقط 5000) و دو فالت f16 و f18 را استخراج می کنیم. میبینیم که تعداد داده ی فالت بسیار کم است پس برای رفع این unbalance با روش Smote داده ها balance می کنیم:

```
[('NOV17', (86400, 17), 'double'), ('NOV9', (86400, 17), 'double')]
Shapes before balancing:
Normal: (50000, 16)
Fault 16: (276, 16)
Fault 18: (207, 16)
Shapes after balancing:
label
0      50000
1      50000
2      50000
Name: count, dtype: int64
Balanced shapes:
Normal: (50000, 16)
Fault 16: (50000, 16)
Fault 18: (50000, 16)
```

با تعریف تابعی ماتریس $M*N$ با مقادیر $100*200$ برای استخراج ویژگی تشکیل می دهیم
برای داده های نرمال و دو فال گفته شده:

```
(100, 200) (100, 200) (100, 200)
```

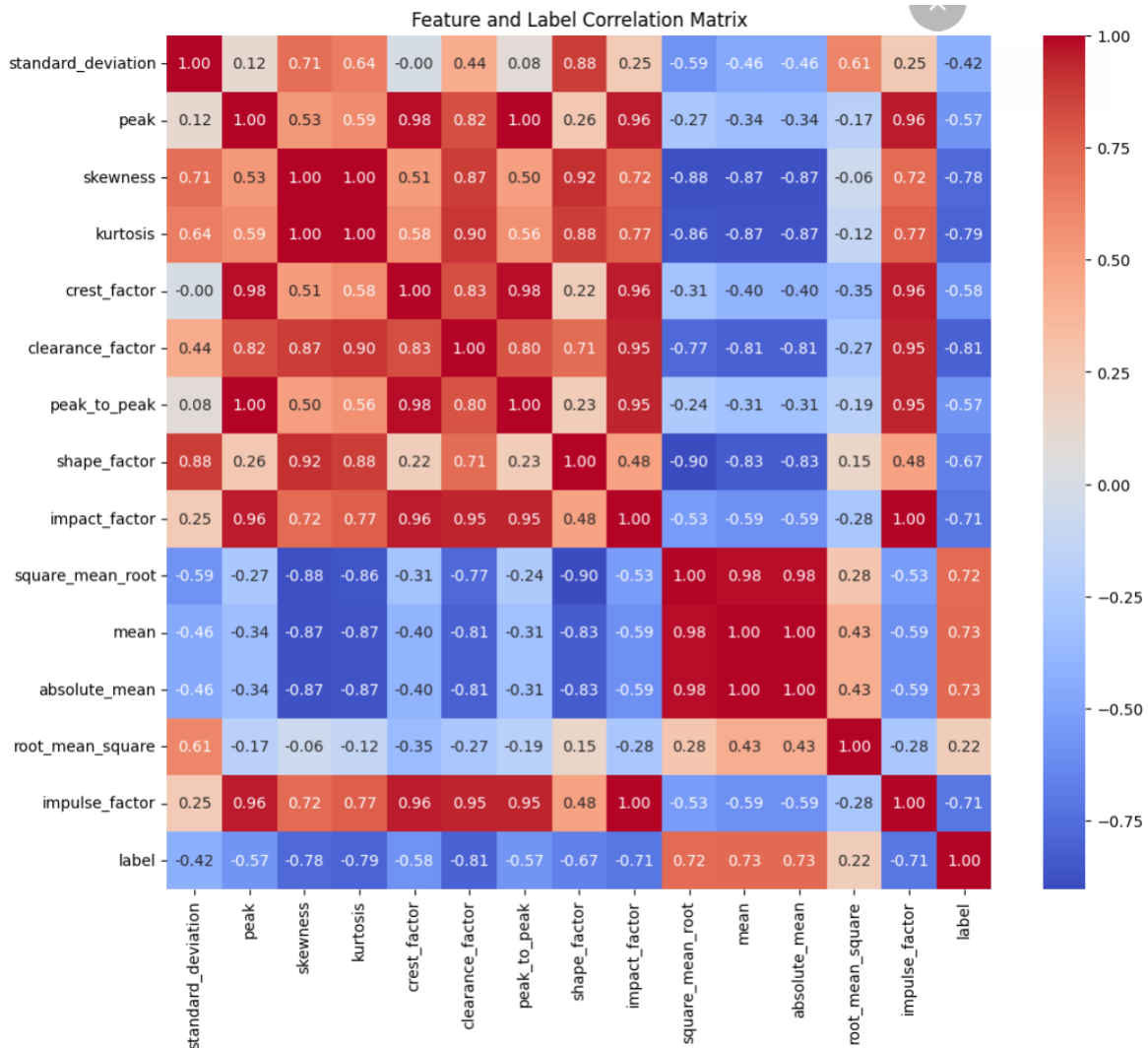
سپس برای هر سطر (هر 200 داده) 14 ویژگی زیر را که در تمرین هم زدیم با تعریف کلاسی در پایتون استخراج می کنیم:

Feature	Formula	Feature	Formula
Standard Deviation	$x_{std} = \sqrt{\frac{\sum_{i=1}^N (x(i) - \bar{x})^2}{N}}$	Shape Factor	$SF = \frac{x_{rms}}{\frac{1}{N} \sum_{i=1}^N x(i) }$
Peak	$x_p = \max x(i) $	Impact Factor	$IF1 = \frac{x_p}{\frac{1}{N} \sum_{i=1}^N x(i) }$
Skewness	$x_{ske} = \frac{\frac{1}{N} \sum_{i=1}^N (x(i) - \bar{x})^3}{x_{std}^3}$	Square Mean Root	$x_{smr} = \left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x(i) } \right)^2$
Kurtosis	$x_{kur} = \frac{\frac{1}{N} \sum_{i=1}^N (x(i) - \bar{x})^4}{x_{std}^4}$	Mean	$Mean = \frac{1}{n} \sum_{i=1}^n x_i$
Crest Factor	$CF = \frac{x_p}{x_{rms}}$	Absolute Mean	$Abs\ Mean = \frac{1}{n} \sum_{i=1}^n x_i $
Clearance Factor	$CLF = \frac{x_p}{x_{smr}}$	Root Mean Square	$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$
Peak to Peak	Maximum - Minimum	Impulse Factor	$IF2 = \frac{AbsMax}{\frac{1}{n} \sum_{i=1}^n x_i }$

و سپس آنها را به با اضافه کردن لیبل مربوطه به هم می چسبانیم:

(300, 15)

نتیجه 100 سمپل برای هر کلاس و 15 ویژگی (با لیبل)



نمودار همبستگی بین ویژگی ها و لیبل را مشاهده می کنیم :

مطلوب است corr بین هدف و ویژگی زیاد و بین ویژگی ها با هم کم باشد.

در بالا مشاهده می شود که ['standard_deviation', 'peak', 'crest_factor', '']
با هدف همبستگی بالایی ندارند پس حذفشان می کنیم.

(300, 9) (300,)

همچنان میتوانستیم ویژگی ها با CORR بالا باهم را (redundancy بالا) نیز حذف کنیم. که نکردیم.

داده ها را به train و validation با درصد تقسیم 80% برای 9 نوامبر تقسیم می کنیم.

همچنین آرگمان shuffle را نیز برای بر زدن فعال می کنیم:

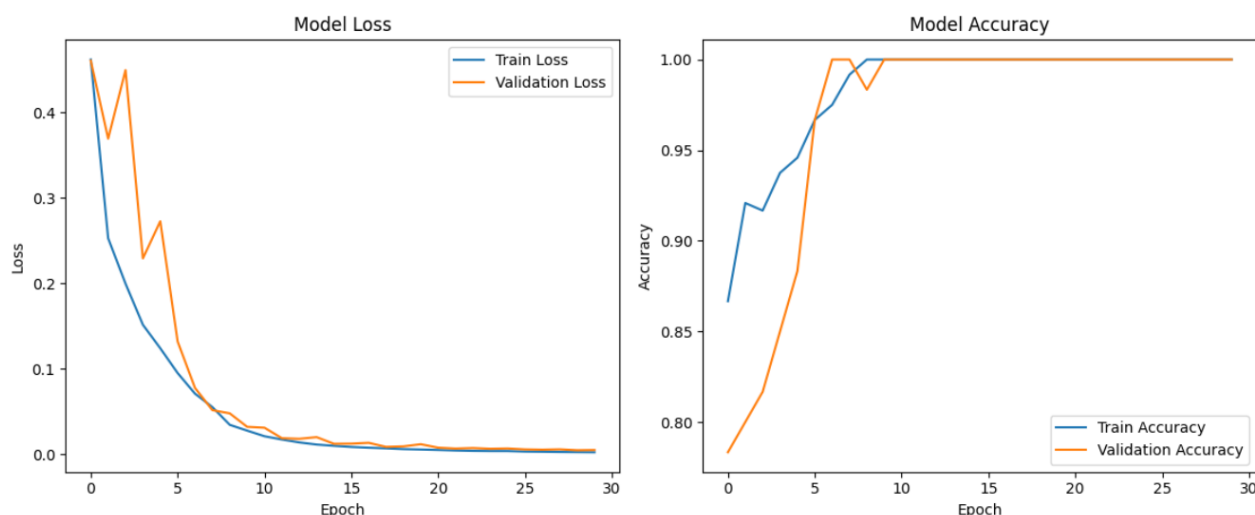
```
X_train_raw.shape, X_valid_raw.shape, y_train_raw.shape, y_valid_raw.shape  
  
((240, 9), (60, 9), (240,), (60,))
```

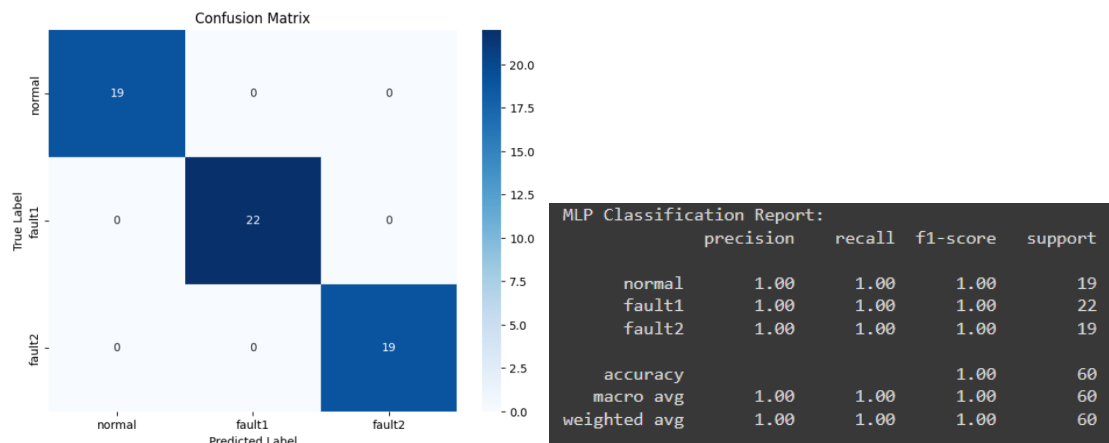
حال آنها را استاندارد سازی می کنیم.

برای استاندارد سازی آنها باید به داده های آموزش فیت و به تست اعمال بشه متاسفانه وقت نشد داده های تست را هم آماده کنم برای این کار.

(لیبل ها را one-hot کردیم)

MLP با 1 لایه پنهان Relu با 20 نرون و لایه نهایی softmax زدم. Learning rate را 0.01 و تعداد epoch را 30 با batchsize 10 گذاشتم. همچنین تابع اتلاف categorical crossentropy (برای one-hot) و optimizer را adam گذاشتیم.

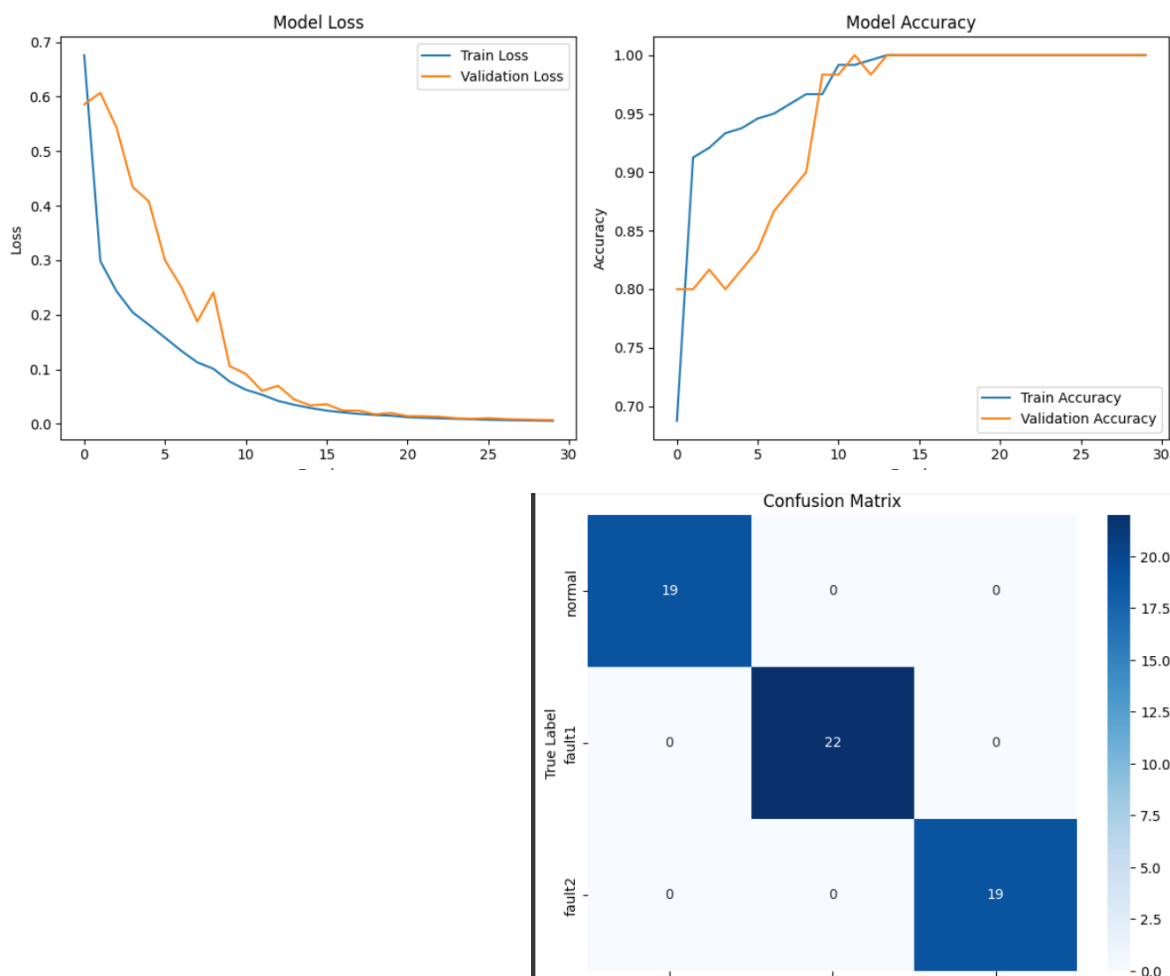




در بالا نمودار خطا accuracy ، classification report و confusion matrix آورده شده است

واضح است که دقت و خطا عالی بوده و داده missclassified نداریم

برای 10 نرون:

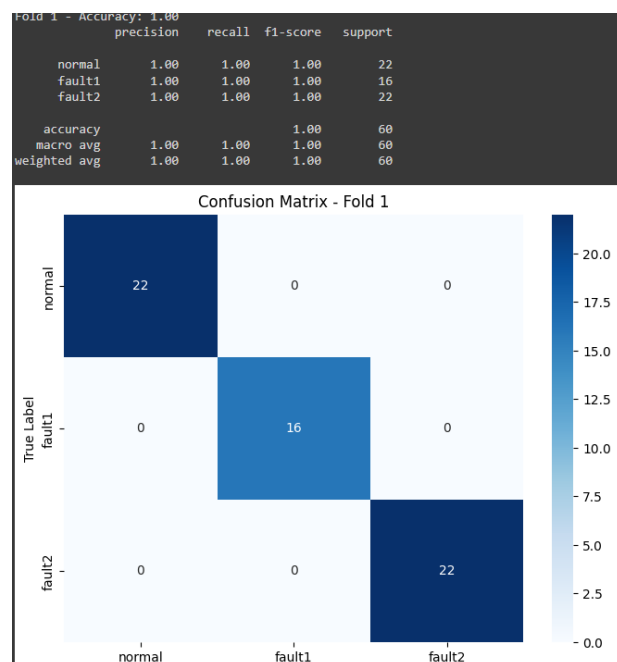


normal	1.00	1.00	1.00	19
fault1	1.00	1.00	1.00	22
fault2	1.00	1.00	1.00	19
accuracy			1.00	60
macro avg	1.00	1.00	1.00	60
weighted avg	1.00	1.00	1.00	60

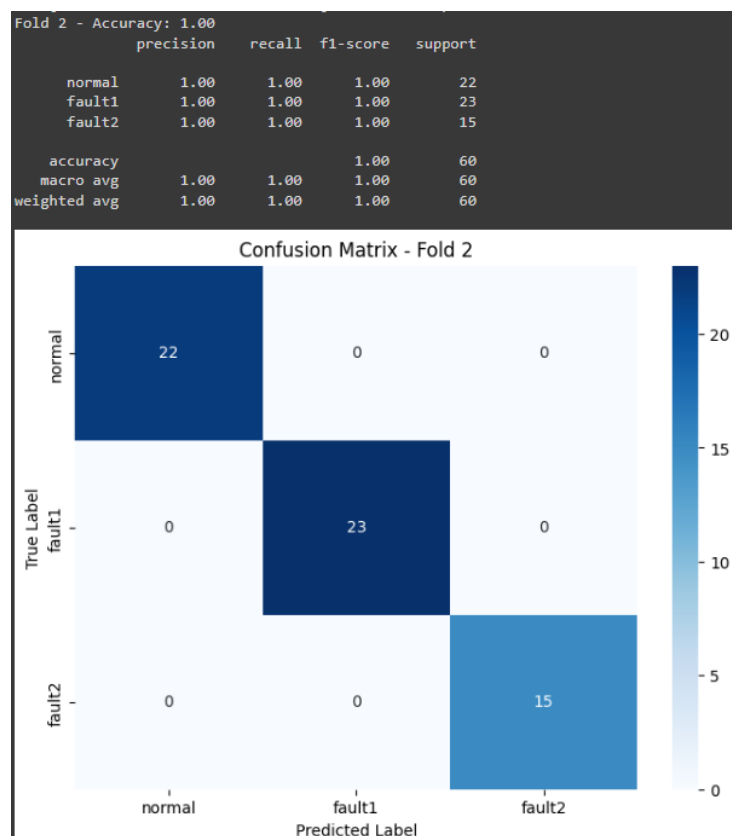
متأسفانه در preprocess داده های تست دچار مشکل شدم و نتوانستم این قسمت سوال را انجام بدم

K-fold

برای fold1:

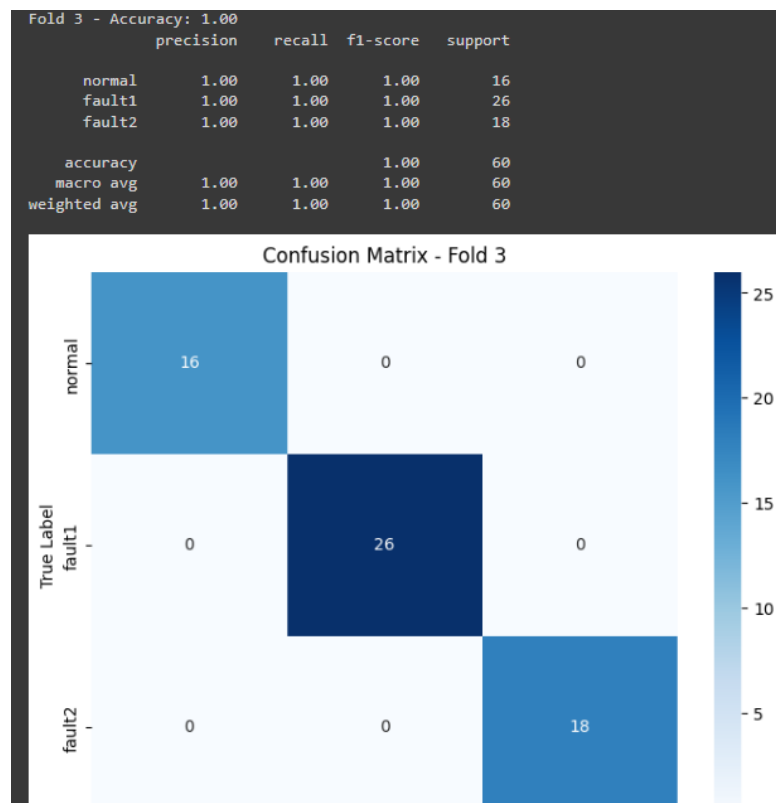


برای fold2:

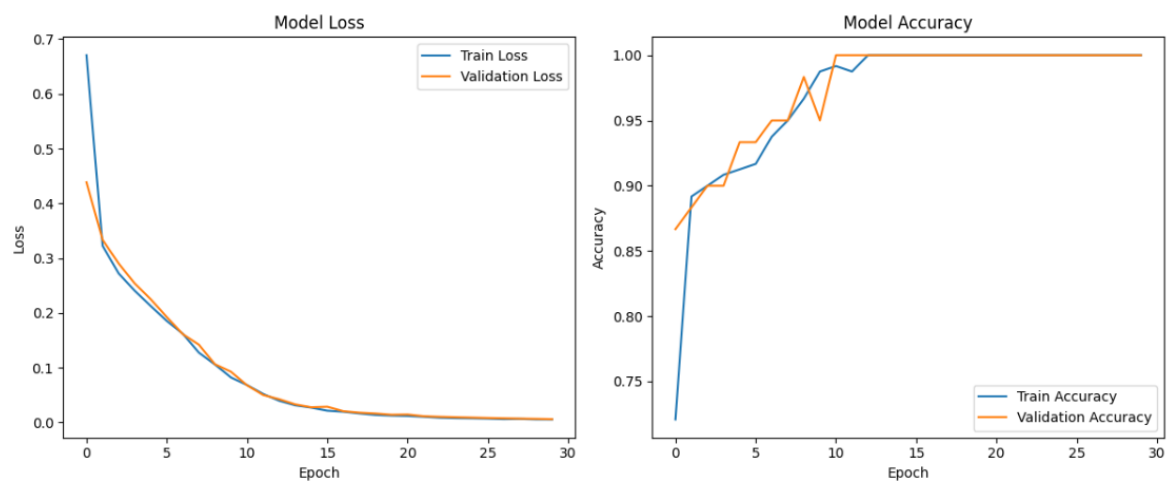


برای

fold3:



برای بقیه fold ها هم به همینطور بود.



Mean loss و mean accuracy را مشاهده می کنیم .