



HW3 Solutions

Machine learning | spring 1400 | Dr.Abdi

Teacher Assistant:

Zahra Dehghani

Student name : **Amin Fathi**

Student id : **400722102**

پیش از مراجعه به حل سوالاتا لازم به ذکر است برای حل سوال ۴ تئوری از لینک زیر :

[homework2-solution\(illinois.edu\)](https://homework2-solution(illinois.edu))

و برای حل سوالات ۱ تا ۷ تئوری از لینک زیر :

<https://dokumen.tips/documents/homework-2-ai.html?page=2>

برای حل قسمت عملی از لینک های زیر استفاده شده است :

[https://github.com/Archiboeuf/Facial-recognition-via-sparse-representation/blob/master/Code/FaceRecognition SparseRepresentation.ipynb](https://github.com/Archiboeuf/Facial-recognition-via-sparse-representation/blob/master/Code/FaceRecognition%20SparseRepresentation.ipynb)

Problem1

A

از سمت چپ شروع میکنیم به باز کردن معادله احتمال شرطی به جهت رسیدن به قسمت راست.

$$P(a|b,c) = P(b|a,c)$$

$$\Rightarrow \frac{P(a,b,c)}{P(b,c)} = \frac{P(a,b,c)}{P(a,c)}$$

$$\text{if } P(a,b,c) \neq 0 \Rightarrow \frac{1}{P(b,c)} = \frac{1}{P(a,c)}$$

$$\Rightarrow P(a,c) = P(b,c)$$

$$\text{if } P(c) \neq 0 \Rightarrow \frac{P(a,c)}{P(c)} = \frac{P(b,c)}{P(c)}$$

$$\Rightarrow P(b|c) = P(a|c)$$

B

از قسمت if می توان فهمید که a نسبت به (b,c) مستقل است :

$$P(a|b,c) = P(a)$$

$$\Rightarrow \frac{P(a,b,c)}{P(b,c)} = P(a)$$

$$\Rightarrow P(a,b,c) = P(a)P(b,c)$$

از قسمت then هم میتوان فهمید که b نسبت به c مستقل است و وابسته نیست.

$$P(b|c) = P(b)$$

$$\frac{P(b,c)}{P(c)} = P(b)$$

$$P(b,c) = P(b)P(c)$$

این دو لزوماً به هم مربوط نیستند، به طوری که میتوان بیشمار مثال احتمالاتی در نظر گرفت که عکس موضوع را اثبات کند به طور مثال اعداد زیر در این قسمت صادق نیست.

$$P(a) = \frac{1}{2} \quad P(b|c) = \frac{1}{3}$$

$$P(b) = \frac{1}{4}$$

C

در قسمت if داریم :

$$\begin{aligned} P(a|b) &= P(a) \\ \Rightarrow \frac{P(a,b)}{P(b)} &= P(a) \\ \Rightarrow P(a,b) &= P(a)P(b) \end{aligned}$$

که نشان میدهد a و b نسبت به هم مستقل هستند بنابراین به طور شهودی (و همانطور که در کلاس اثبات شد) نیز قابل درک است که گزاره then صحیح باشد.

Problem2

سوال غلط دارد و مقدار احتمال باید در یک ۱- ضرب شود تا بتوان حل کرد و در ادامه نیز با این فرض مسئله حل شده است.

از آنجا که در واقع گیرنده می تواند حداکثر تا یک بیت را تصحیح کند ، احتمال ارسال پیام درست در واقع برابر است با احتمال ارسال پیام بدون خطا + احتمال ارسال پیام با یک بیت خطا :

$$P(\text{true}) = P(0) + P(1);$$

$P(0)$: probability of 0 bit error

$P(1)$: probability of 1 bit error

در مورد $P(0)$ داریم :

انتخاب 0 بیت از $n+1$ بیت * احتمال مخدوش بودن بیت های غلط * احتمال درست بودن بیت های درست .

$$P(0) = \binom{n+1}{0} \times \epsilon^0 \times (1-\epsilon)^{n+1} = (1-\epsilon)^{n+1}$$

$$P(1) = \binom{n+1}{1} \times \epsilon^1 \times (1-\epsilon)^n = (n+1) \times \epsilon \times (1-\epsilon)^n$$

$$P(\text{true}) = P(0) + P(1) = (1+n\epsilon)(1-\epsilon)^n$$

$$(1+n\epsilon)(1-\epsilon)^n \approx 1 - 8$$

$$(1 + 0.001n)(1 - 0.001)^n \approx 1 - 0.01$$

$$(1 + 0.001n) \times 0.999^n \approx 0.99$$

مقدار مناسب n به طور تقریبی ۱۴۸ میباشد.

Problem3

A

از آنجا که سکه ها unbiased هستند، احتمال سکه معمولی و سکه جعلی به شرح زیر است :

$$P(\text{جعلی}) = \frac{1}{n} \quad P(\text{معمولی}) = \frac{n-1}{n}$$

در مورد سکه های معمولی می توان گفت احتمال شیر یا خط افتادن به صورت زیر است :

$$P(\text{معمولی} | \text{خط}) = \frac{1}{2} \quad P(\text{معمولی} | \text{شیر}) = \frac{1}{2}$$

این احتمال در مورد سکه ی جعلی که هر دو سوی آن خط هست به صورت زیر است :

$$P(\text{جعلی} | \text{خط}) = 1 \quad P(\text{جعلی} | \text{شیر}) = 0$$

احتمال اینکه سکه خط برداشته شده جعلی باشد برابر است با :

$$P(\text{خط} | \text{جعلی}) = \frac{P(\text{خط} | \text{جعلی}) \times P(\text{جعلی})}{P(\text{خط})}$$

حال به سراغ محاسبه مخرج عبارت روبه رو یعنی احتمال خط بودن می رویم، این احتمال حاصل جمع دو احتمال اینکه سکه جعلی باشد و خط بیاید و یا سکه معمولی باشد و خط بیاید است.

$$\begin{aligned} P(\text{خط}) &= P(\text{خط} | \text{معمولی}) P(\text{معمولی}) + P(\text{خط} | \text{جعلی}) P(\text{جعلی}) \\ &= \frac{1}{2} \times \frac{n-1}{n} + 1 \times \frac{1}{n} = \frac{0.5n + 0.5}{n} \end{aligned}$$

بنابراین پاسخ قسمت a برابر است با :

$$P(\text{فدا | حبلی}) = \frac{1 \times \frac{1}{n}}{\frac{0.5n + 0.5}{n}} = \frac{2}{n+1}$$

B

با استفاده از قوانین Bayesian احتمال مد نظر سوال را محاسبه می کنیم :

$$P(\text{فدا}^1, \dots, \text{فدا}^k | \text{حبلی}) = \frac{P(\text{فدا}^1, \dots, \text{فدا}^k | \text{حبلی})}{P(\text{فدا}^1, \dots, \text{فدا}^k)}$$

$$= \frac{\prod_{i=1}^k P(\text{حبلی} | \text{فدا}^i)}{P(\text{حبلی} | \text{فدا}^1, \dots, \text{فدا}^k) + P(\text{سری} | \text{فدا}^1, \dots, \text{فدا}^k)}$$

$$= \frac{P(\text{حبلی})}{P(\text{حبلی}) + P(\text{سری})}$$

$$= \frac{\frac{1}{n}}{\frac{1}{n} + \frac{0.5^k}{2^k + n - 1}}$$

$$= \frac{1}{0.5^k \frac{n-1}{n} + 1} = \frac{2^k}{2^k + n - 1}$$

C

احتمالی که به دنبال آن هستیم در واقع احتمال زیر است :

$$P(\text{اصولی})^k \cdot \underbrace{\text{خط و ... خط}}_{\text{خط و ... خط}}$$

که با کمک گرفتن از قسمت قبل به صورت زیر محاسبه می شود :

$$1 - P(\text{اصولی})^k$$

$$1 - \frac{2^k}{2^{k+n-1}} = \frac{2^{n-1}}{2^{k+n-1}}$$

Problem4

صورت سوال احتمال اینکه ما واقعا مبتلا به بیماری باشیم و تستمان مثبت باشد را می خواهد اما پیش از به دست آوردن آن باید اطلاعات داده شده را نوشت:

احتمال مبتلا بودن از تقسیم ۱ بیمار به ۱۰۰۰۰ نفر به دست می آید ، احتمال اینکه تست مثبت باشد در حالیکه ما واقعا مبتلا شده باشیم برابر است با 0.99 و احتمال اینکه تست منفی در بیاید در حالیکه ما واقعا مبتلا نشده باشیم هم برابر است با 0.99

$$\begin{aligned}P(T|D) &= 0.99 & P(D) &= 0.0001 = \frac{1}{10000} \\P(\neg T|\neg D) &= 0.99\end{aligned}$$

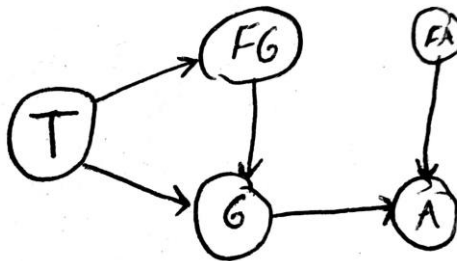
حال احتمال مد نظر سوال را محاسبه میکنیم:

$$\begin{aligned}P(D|T) &= \frac{P(T|D) \times P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\neg D)P(\neg D)} \\&= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} = 0.009904\end{aligned}$$

مشاهده میشود اینکه تست ما مثبت در آمده باشد و واقعا مبتلا شده باشیم بسیار کم است و این را مدیون پایین بودن مقدار $P(D)$ هستیم که در واقع مدیون نادر بودن بیماری هستیم .

Problem5

A



B

خیر چون یک لوپ تشکیل یافته توسط نود های T , FG , G داریم و اساسا با درخت طرف نیستیم.

C

برای حل سوال فرض می کنیم دمای صحیح همان دمای معمولی است و دمای ناصحیح همان دمای بالا (این فرض در جمله اول صورت سوال به صورت گنگی اشاره شده بود)

فرض کنید T دمای معمولی است و $\neg T$ آن به معنای دمای بالا

در این صورت :

$$\begin{aligned} P(G | T, FG) &= y \\ P(G | \neg T, FG) &= 1-y \\ P(G | T, \neg FG) &= x \\ P(G | \neg T, \neg FG) &= 1-x \end{aligned}$$

D

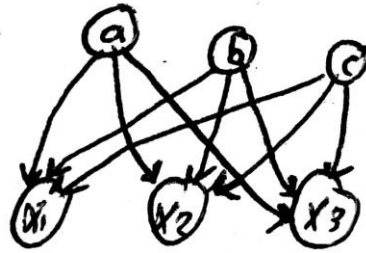
	G		$\neg G$	
	FA	$\neg FA$	FA	$\neg FA$
A	0	0	0	1
$\neg A$	1	1	1	0

تنها G و FG بر روی T تاثیر دارند پس :

با توجه به صورت سوال به دنبال احتمالات زیر برای دمای بالا هستیم .

$$P(\neg T | \neg FG, G) = \frac{P(G | \neg T, \neg FG) P(\neg FG | \neg T) P(\neg T)}{P(\neg FG, G)}$$

Problem6



CPT را برای یکی از X ها به دست می آوریم چون بقیه هم یکسان هستند و به همین شکل هستند.

Possibility	Value
$P(a)$	$\frac{1}{3}$
$P(b)$	$\frac{1}{3}$
$P(c)$	$\frac{1}{3}$
$P(X_1 = \text{ط} a)$	0.2
$P(X_1 = \text{سیر} a)$	0.8
$P(X_1 = \text{ط} b)$	0.6
$P(X_1 = \text{سیر} b)$	0.4
$P(X_1 = \text{ط} c)$	0.8
$P(X_1 = \text{سیر} c)$	0.2

B

هدف ما پیدا کردن احتمال زیر است :

$$P(\text{نوع سکه} | \text{سیرد خط, خط}) = \frac{P(\text{نوع سکه} | \text{سیرد خط, خط}) \cdot P(\text{سیرد خط, خط})}{P(\text{سیرد خط, خط})}$$

$$= \frac{P(\text{نوع سکه}) \cdot P(\text{نوع سکه} | \text{خط}) \cdot P(\text{نوع سکه} | \text{خط})}{P(\text{سیرد خط, خط})}$$

برای هر کدام از سکه ها پیدا میکنیم :

$$P(a | \text{سیرد خط, خط}) \propto P(a | \text{خط}) \cdot P(a | \text{خط}) \cdot P(a | \text{خط}) \cdot P(a) \\ = 0.2 \times 0.2 \times 0.2 \times \frac{1}{3} = 0.011$$

$$P(b | \text{سیرد خط, خط}) \propto P(b | \text{خط}) \cdot P(b | \text{خط}) \cdot P(b | \text{خط}) \cdot P(b) \\ = 0.6 \times 0.6 \times 0.4 \times \frac{1}{3} = 0.048$$

$$P(c | \text{سیرد خط, خط}) \propto P(c | \text{خط}) \cdot P(c | \text{خط}) \cdot P(c | \text{خط}) \cdot P(c) \\ = 0.8 \times 0.8 \times 0.2 \times \frac{1}{3} = 0.043$$

مشاهده می شود که برای سکه b این مقدار بیش از بقیه است بنابراین بیشترین احتمال دنباله صورت سوال مربوط به سکه b می باشد.

Problem7

A

مورد های دوم و سوم به راحتی با توجه به ساختار زنجیره ای قابل ادراک هستند اما در مورد اول نود ها B و M دارای تاثیر یکسانی همانند نود A هستند

$$P(B, \bar{I}, M) = P(B) P(I|B, M) P(M)$$

B

$$\begin{aligned} P(b, i, \neg m, \neg g) &= P(b) P(\neg m) P(i|b, \neg m) P(g|b, i, \neg m) P(\neg g) \\ &= P(b) P(g|b, i, \neg m) P(\neg g) \\ &= P(i) \times 0.8 \times 0.9 = 0.72 \times P(i) \end{aligned}$$

حال به سراغ محاسبه $P(i)$ می رویم :

$$\begin{aligned} P(i) &= P(i|b, m) P(b) P(m) + P(i|\neg b, m) P(\neg b) P(m) \\ &\quad + P(i|b, \neg m) P(b) P(\neg m) + P(i|\neg b, \neg m) P(\neg b) P(\neg m) \\ &= 0.9 \times 0.9 \times 0.1 + 0.5 \times 0.1 \times 0.1 + 0.5 \times 0.9 \times 0.9 \\ &\quad + 0.1 \times 0.1 \times 0.9 \\ &= 0.081 + 0.005 + 0.405 + 0.009 = 0.5 \end{aligned}$$

حاصل نهایی برابر است با :

$$P(b, i, \neg m, \neg g) = 0.72 \times 0.5 = 0.36$$

C

$$P(I) = P(\neg I) = 0.5 \quad P(G) = 0.37$$

$$P(G|B, I, M) = 0.9 \quad P(J) = 0.396$$

$$P(J|B, I, M) = \frac{P(J|B, I, M) P(G|J)}{P(G|J)}$$

$$= \frac{P(G, J|B, I, M)}{P(G|J)}$$

$$= \frac{P(J|B, I, M, G) P(G|B, I, M)}{P(G|J)}$$

$$= \frac{P(J|G) P(G|B, I, M)}{P(G|J)} \quad (B, I, M \perp\!\!\!\perp J|G)$$

$$= \frac{P(J) P(G|B, I, M)}{P(G)} = \frac{0.396 \times 0.9}{0.37}$$

$$= 0.963$$

D

$J \perp\!\!\!\perp B|G$

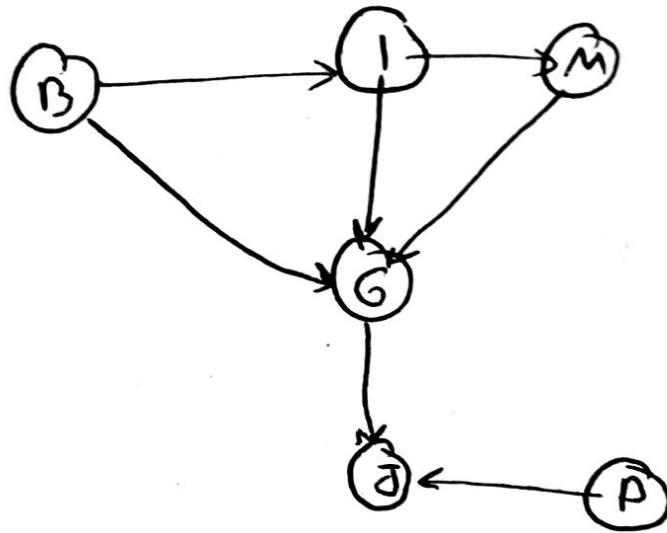
$J \perp\!\!\!\perp I|G$

$J \perp\!\!\!\perp M|G$

$B \perp\!\!\!\perp M$

E

عفو ریاست جمهوری فرمانی فراتر از همه فرمان هاست ! و فقط روی زندانی شدن تاثیر می گذارد یعنی J

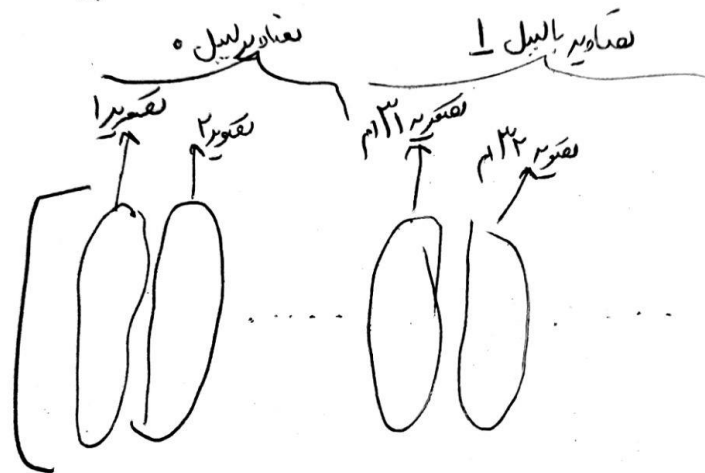


Problem 8

برای این سوال از دیتا ست MNIST استفاده کرده ام که در صورت سوال هم پیشنهاد شده بود.

دیتاستی که تصاویر اعداد دست خط ۲۸ در ۲۸ پیکسل و ۶۰۰۰۰ تایی برای train و ۱۰۰۰۰ تایی برای تست هستند .

برای اینکه دیتاست مناسب استفاده توسط SRC شود و با توجه به حجم بالای دیتاست من تصمیم گرفتم از هر کلاس (۰ و ۱ و ۲ و ...) ۳۰ تصویر اول را انتخاب کرده و به صورت ماتریس دو بعدی (تصاویر فلتن شده) در بیاورم، برای فهم این قضیه به مثال زیر توجه کنید:



همانطور در تصویر مشاهده می شود ، سی ستون اول دارای لیبل ۰ هستند و هر یک در واقع یک عکس از دادگان Train ما هستند که flatten شده اند و هر ستون ۷۸۴ تایی است .

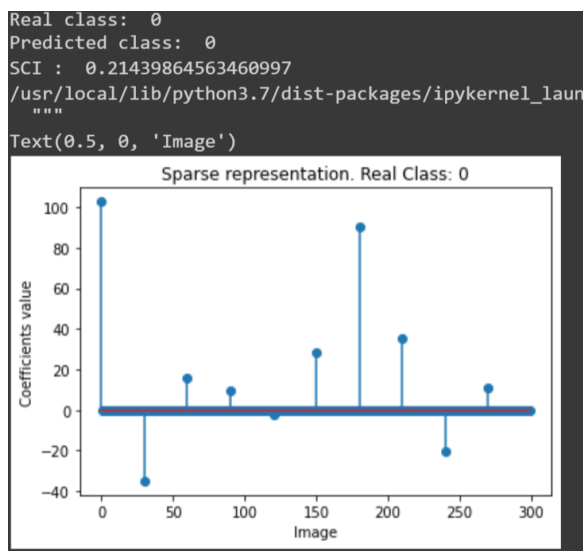
لازم به ذکر است بعد از آپلود دادگان، ابتدا آن ها نرمالایز میکنیم سپس از هر برچسپ ۳۰ تصویر را انتخاب کرده و به شکل تصویر بالا تبدیل میکنیم، که در واقع یک ماتریس با ۷۸۴ سطر و ۳۰۰ ستون خواهیم داشت.

البته خروجی حلقه for های به کار رفته به صورت ۳۰۰ سطر و ۷۸۴ ستون می باشد ، ابتدا آن را به numpy تبدیل کرده و ماتریس ترانهاد آن را حساب می کنیم . سپس حاصل را به صورت list که یک ماتریس دو بعدی است برای ادامه کار در متغیر xtrain ذخیره می کنیم .

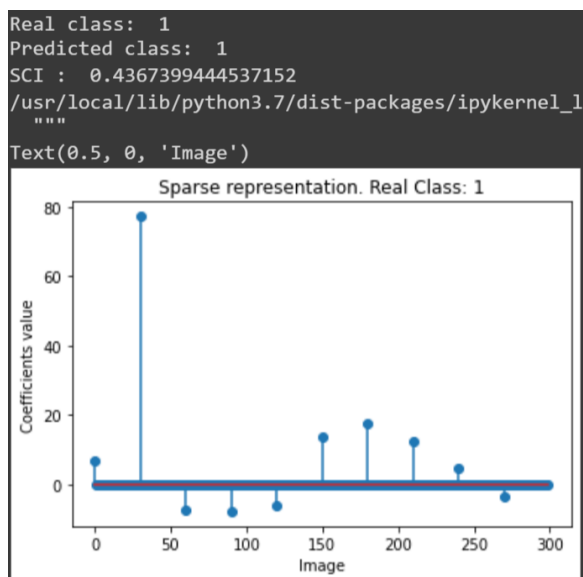
برچسپ این دادگان را نیز در یک list یک بعدی با ۳۰۰ درایه ذخیره می کنیم؛ لازم به ذکر است ۳۰ داده اول برچسپشان ۰ است و ۳۰ داده دوم برچسپشان ۱ و قص علی هذا.

این کار را بر روی دادگان test نیز انجام می دهیم.

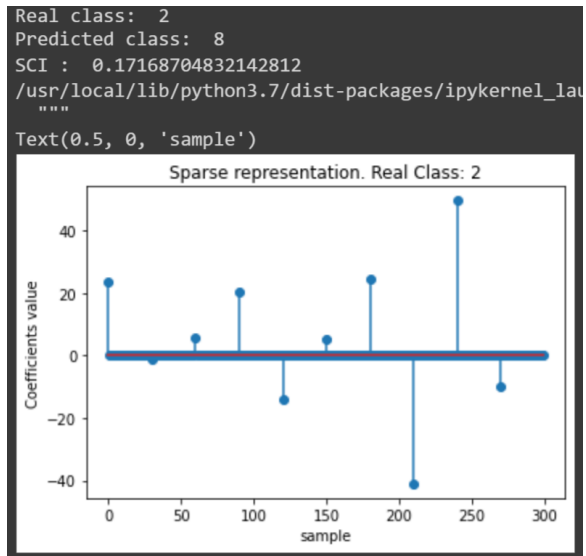
برای دادگان تست از هر کلاس ۱۰ نمونه و در نتیجه جمعا ۷۰ نمونه خواهیم داشت.



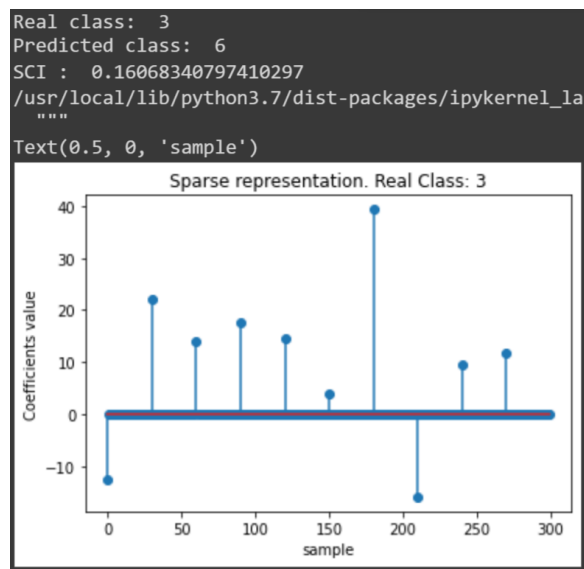
در مورد کلاس + ، به درستی همه موارد دادگان تست را + پیشبینی کرده ، مشاهده میشود بعد از + ، ۶ بالاترین امتیاز را داراست.



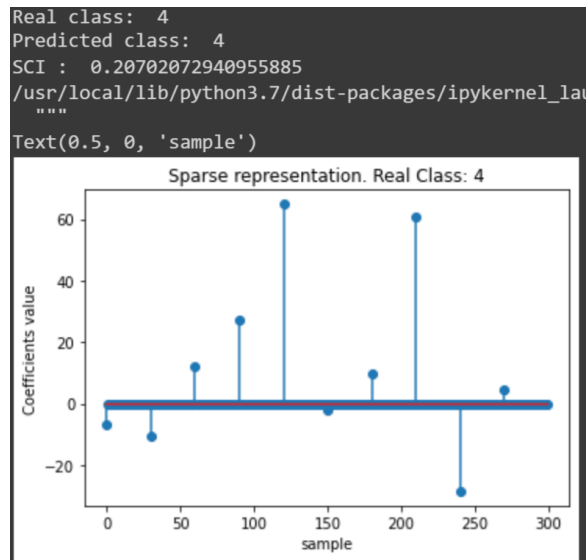
در مورد کلاس ۱، به درستی همه موارد دادگان تست را ۱ پیشبینی کرده ، مشاهده میشود که امتیاز کلاس ۱ با اختلاف از دیگر کلاس ها بیشتر است.



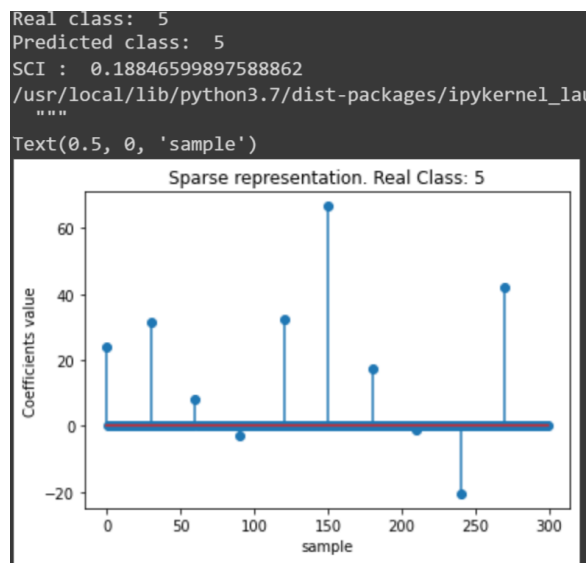
در مورد کلاس ۲، به اشتباه همه موارد دادگان تست را ۸ پیشبینی کرده ، امتیاز کلاس های مختلف در تصویر بالا مشاهده می شود.



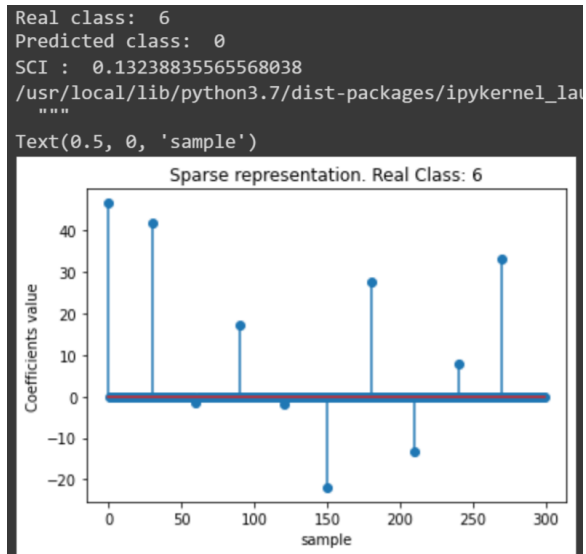
همچنین این روش، کلاس مربوط به دادگان با برچسب ۳ را اشتباها ۶ پیش بینی کرده ، امتیاز سایر کلاس ها را در تصویر بالا مشاهده می کنید.



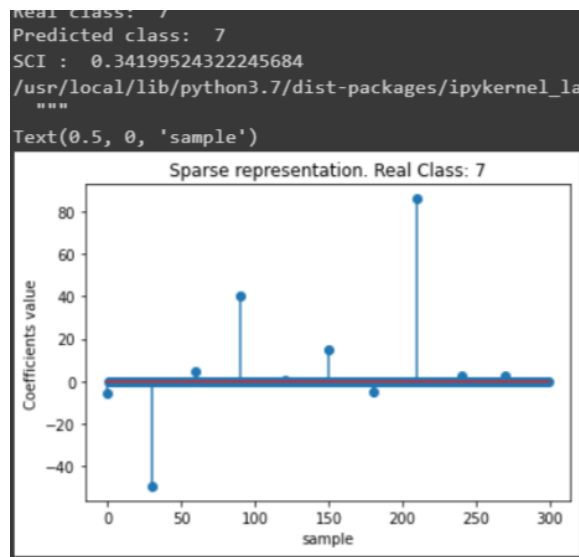
در مورد دادگان با برچسپ ۴ ، مشاهده می شود که طبقه بندی کننده مد نظر ما به درستی همه را ۴ دسته بندی کرده اما نکته این نمونه ها این است که کلاس اعداد با برچسپ ۷ نیز امتیاز بسیار بالایی به دست آورده اند.



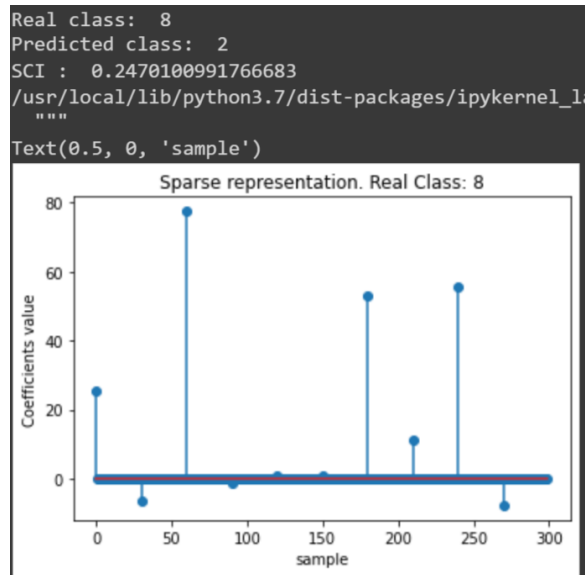
در مورد دادگان با برچسپ ۵ نیز همگی را درست دسته بندی کرده است ، بیشترین امتیاز بعد از ۵ مربوط به دادگان کلاس ۹ می باشد.



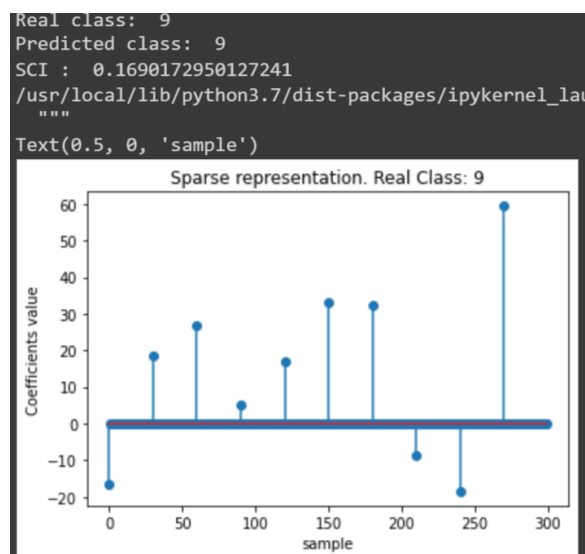
دادگان مربوط به کلاس ۶ به اشتباه کلاس ۰ دسته بندی شده اند ، مشاهده میشود کلاس ۱ و ۹ و ۶ دارای بیشترین امتیاز بعدی هستند.



دادگان مربوط به کلاس ۷ به درستی دسته بندی شده اند .



دادگان مربوط به کلاس ۸ نیز به اشتباه کلاس ۲ دسته بندی شده اند ، بیشترین امتیاز پس از کلاس ۲ مربوط به کلاس ۸ بوده است .



دادگان مربوط به کلاس ۹ نیز به درستی کلاس ۹ دسته بندی شده اند .

در انتها بر روی کل دادگان test اجرا میکنیم که مطابق انتظار از ۱۰ رقم ۴ رقم به اشتباه دسته بندی میشوند (رقم ۸ اشتباه برچسپ ۲ میخورد و رقم ۲ اشتباه برچسپ ۸ و رقم ۶ اشتباه برچسپ ۰ و رقم ۳ اشتباه برچسپ ۶)

```
total number of test sampels is : 70 and true predicted number is : 42
Test Score : 60.0 %
```

لازم به ذکر است افزایش دادگان آموزش و کاهش مقدار آلفا نتوانستند مشکل دقت ۶۰ درصد را حل کرده و آن را افزایش دهند

همچنین باید اضافه کرد که بنده در هنگام train، برچسپ دادگان ۸ و ۲ را برعکس تنظیم کردم، یعنی برای دادگان مربوط به رقم ۲ برچسپ ۸ تنظیم کردم و برای دادگان مربوط به رقم ۸ برچسپ ۲ تنظیم کردم تا نتیجه را مشاهده کنم:

```
if y_train[i] == 2:
    while(j3 <30):
        j3 += 1
        x.append(X_train[i].ravel())
        y.append(8)
for i in range(1000):
    if y_train[i] == 8:
        while(j9 <30):
            j9 += 1
            x.append(X_train[i].ravel())
            y.append(2)
```

در نهایت مشاهده شد که مقدار **accuracy** برابر شد با ۸۰ درصد:

```
total number of test sampels is : 70 and true predicted number is : 56
Test Score : 80.0 %
```

برای قسمت **b** نیز از تابع **OrthogonalMatchingPursuit** استفاده شد که البته در نهایت همان ۶۰ درصد **accuracy** حاصل شد و این روش نیز در تشخیصی همان ۴ رقم اشاره شده به مشکل خورده است.

```
[134] def FaceClass_noplot2(test_pic):

    y4 = xtest[:,test_pic]
    omp = OrthogonalMatchingPursuit(n_nonzero_coefs=10, normalize=False)
    omp.fit(xtrain, y4)
    coef = omp.coef_

    pred_class = np.argmin(residual(xtest[:,test_pic],xtrain,coef,ytrain))

    if ytest[test_pic] == pred_class:
        return "Good Prediction"
    else:
        return "Wrong Prediction"

total2 = 0

for i in range(len(xtest[0])):
    #res = FaceClass_noplot(1,alpha)
    res = FaceClass_noplot2(i)
    if res == 'Good Prediction':
        total2+=1
print('total number of test sampels is :', len(xtest[0]), ' and true predicted number is :', total2)
print('Test Score : %s' %(total2/len(xtest[0])*100) , ' %' )

total number of test sampels is : 70 and true predicted number is : 42
Test Score : 60.0 %
```

در کل به نظر می رسد الگوریتم دارد الگویی را یاد میگیرد ولی این الگو در مورد ۴ رقم ذکر شده درست نیست.