



University of Tehran

School of Electrical and Computer Engineering



---

# MACHINE LEARNING

---

## Assignment 2

Sasan Keshavarz  
810199253

Spring 2022

## فهرست

سوال ۱	..... ۳
سوال ۲	..... ۵
سوال ۳	..... ۷
سوال ۴	..... ۹
سوال ۵	..... ۱۰
سوال ۶	..... ۱۵

۱- الف)  $p(x, y)$   $f^*: X \rightarrow Y$

$$f^* = \arg \min_f E((f(x) - Y)^2)$$

$$f^*(x) = ? \quad E(Y|x) \rightarrow E((E(Y|x) - Y)^2) \leq E((f(x) - Y)^2)$$

برای همه  $f$  های برچیده

~~$$E((f(x) - Y)^2) = E((f(x) - E(Y|x) + E(Y|x) - Y)^2)$$~~

$$E((f(x) - E(Y|x) + E(Y|x) - Y)^2) = E((f(x) - E(Y|x))^2) + E((E(Y|x) - Y)^2) + \underbrace{E((f(x) - E(Y|x))(E(Y|x) - Y))}_{=0}$$

$$z \rightarrow E((f(x) - E(Y|x))(E(Y|x) - Y)) =$$

$$= E_x(E_{Y|x}((f(x) - E(Y|x))(E(Y|x) - Y) | x)) = 0$$

$$\rightarrow E((f(x) - Y)^2) = E((f(x) - E(Y|x))^2) + E((E(Y|x) - Y)^2)$$

$\geq 0$  همیشه مثبت یا صفر

$$\rightarrow E((f(x) - Y)^2) \geq E((E(Y|x) - Y)^2)$$

بنابراین  $f^*(x) = E(Y|x)$  است که این به صورت شهودی هم ملموس و

قابل درک است.

$$\sum_{i=1}^K P(Y=y_i | X) = 1 \quad (ب-1)$$

$$\rightarrow P(Y=y_K | X) = 1 - \sum_{i=1}^{K-1} P(Y=y_i | X)$$

$$\rightarrow P(Y=y_K | X) = \frac{1}{1 + \sum_{i=1}^{K-1} \exp(w_{K0} + \sum_{j=1}^d w_{Kj} X_j)}$$

$$P(Y=y_K | X) = \frac{\exp(w_{K0} + \sum_{j=1}^d w_{Kj} X_j)}{1 + \sum_{i=1}^{K-1} \exp(w_{K0} + \sum_{j=1}^d w_{Kj} X_j)}$$

1- ج) قانون طبقه بندی کلاس با بیشترین احتمال را باید انتخاب کند.

$$y = \hat{y}_K \quad K = \arg \max_K P(Y=y_K | X)$$



۲- الف) برای اینکه در حل ستقیم سئله رگرسیون  $Y = X\theta + \epsilon$   $ATA$

واردون پذیریت و نفقه روش نداشتن تاثر استفاده است از

روش رگرسیون استفاده می‌کنیم ~~در این روش~~

در ~~معادله~~ بهینه سازی نرم اول یارم  $\theta$  (وزرها) را اضافه

می‌کنیم. مزیت دیگر این روش این است که در فضای صحت و جری

نداشتن هم نتایج را محدود می‌کند و جواب بهتری در نهایت به دست

می‌دهد.

L-1 regularization:

$$(Lasso) \quad \hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n (y_i - x_i \theta)^2 + \lambda \|\theta\|_1$$

در روش L-1 و نرم اول  $\theta$  را در معادله های دهم.

L-2 regularization:

$$(Ridge) \quad \hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n (y_i - x_i \theta)^2 + \lambda \|\theta\|_2^2$$

در روش L-2 و نرم دوم  $\theta$  را اضافه می‌کنیم.

در روش L-1 راه حل sparse است و مجموع قدر مطلق وزرها جریبه می‌شود.

اما در روش L-2 مجموع مجذور وزرها جریبه می‌شود.

۳- به کمک تیلور مرتبه دوم  $J(w)$ :

$$J(w) = J(w_0) + (w - w_0)^T \nabla J(w_0) + \frac{1}{2} (w - w_0)^T \nabla^2_{J(w_0)} (w - w_0)$$

$$\frac{\partial J(w)}{\partial w} = \nabla J(w_0) + \nabla^2_{J(w_0)} (w - w_0) = 0$$

$$\rightarrow w^* = w_0 - H_{J(w_0)}^{-1} \cdot \nabla J(w_0)$$

$$y_i = \beta_0 + \varepsilon_i$$

(الف. ۳)

$$\hat{\beta}_0 = \text{Arg min}_{\beta_0} \sum_{i=1}^n (y_i - \beta_0)^2$$

$$\hat{\beta}_0 = \text{Arg min}_{\beta_0} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$\frac{\partial J}{\partial \beta_0} = 0 \rightarrow -2 \sum_{i=1}^n (y_i - \beta_0) \rightarrow \sum_{i=1}^n y_i = \beta_0 n$$

$$\rightarrow \beta_0 = \bar{y} = 24.1$$

$$y_i = \beta_1 x_i + \varepsilon_i$$

(ب. ۳)

$$\hat{\beta}_1 = \text{Arg min}_{\beta_1} \sum_{i=1}^n (y_i - \beta_1 x_i)^2 = \text{Arg min}_{\beta_1} \sum_{i=1}^n (y_i - \beta_1 x_i)^2$$

$$\frac{\partial J}{\partial \beta_1} = 0 \rightarrow -2 \sum_{i=1}^n (y_i - \beta_1 x_i) x_i = 0$$

۱۳۵۲  
کیفیت

$$\beta_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^{10} x_i^2} = \frac{11221}{4173} \approx 3 \quad (\text{ا. ۳-})$$

۳- ج) معادله  $\hat{y}$  به صورت یقینی است و جواب دل کالای خطی است و

احتمال اینکه مقادیر از خط بیرون باشند و برد ندارند.

امار برد  $\bar{y}$  در معادله  $\hat{y}$  باعث می شود که ماب این مدل احتمالاتی باشد

و بدین داده های اولیه در مدل دخیل شود

۳- د) ضربه چند و گرسیرد خطی ماب کالای دقیق به دست می دهد و

نزدیکترین خطی که به داده ها  $\hat{y}$  برازش شود را پیدا می کنند. پس همان است

تقراری خطی از  $y=22$  و برد داشت باشد.

صنعت اینکه جهت داده  $(x=6, y=?)$  خود خط  $\hat{y}$  را تغییر می دهند.

$$6^2 = \frac{\sum_{i=1}^n (\text{Error})^2}{n-1} = \frac{v}{15} = 0.466 \quad (\text{ه. ۳-})$$



#### سوال ۴

همانطور که خواسته شده ابتدا داده‌ها را تولید میکنیم و آن را به نویز گاوسی و پواسون آغشته میکنیم. برای فهم بهتر مسئله، نمودار داده‌های نویزی شده رسم شده است.

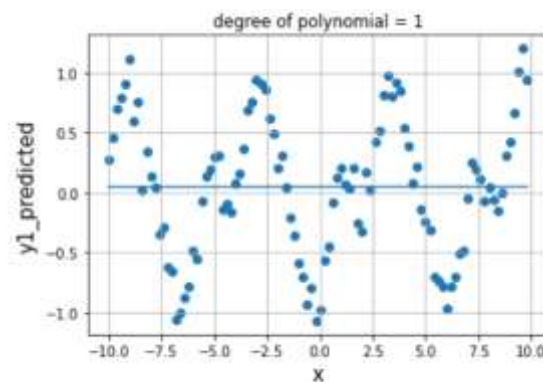
#### حالت اول : داده آغشته به نویز گاوسی

(الف)

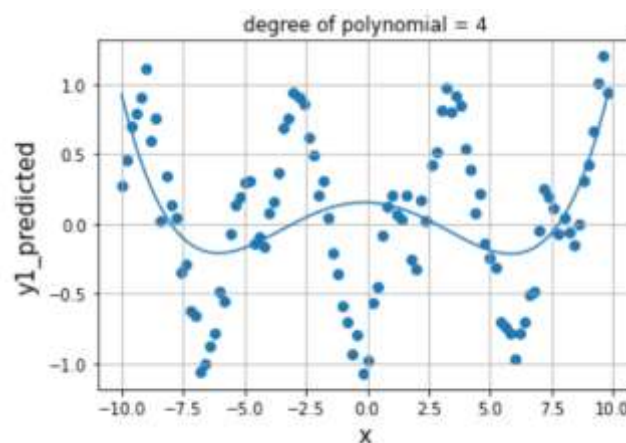
سپس با استفاده از یک حلقه برای تمام درجات ۱ تا ۱۵ برازش انجام شده و نمودار آن رسم گردید. مقادیر MSE هم برای هر درجه از برازش محاسبه شد. هم نمودار درجات و هم مقدار MSE حاکی از این هستند که بهترین درجه برازش برای این داده‌ها درجه ۱۵ است.

(ب)

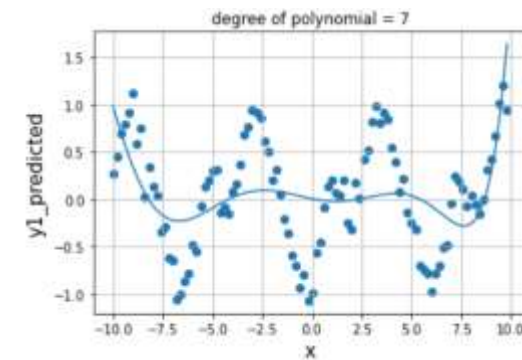
مقدار MSE درجه برازش ۱ : ۰,۲۸۸



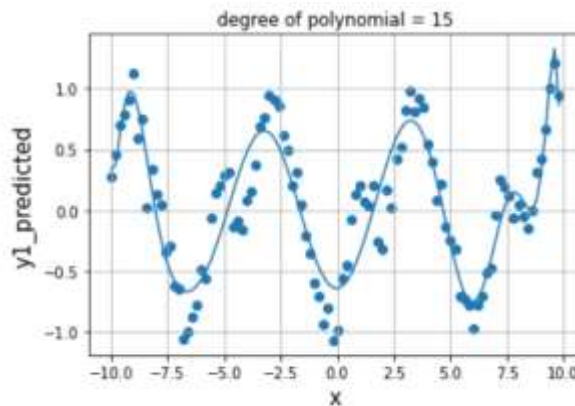
مقدار MSE درجه برازش ۴ : ۰,۲۳



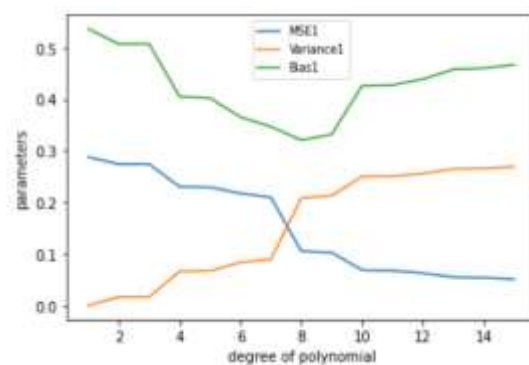
مقدار MSE درجه برازش ۷ : ۰,۲۰۹



مقدار MSE درجه برازش ۱۵ که بهترین درجه برازش است: ۰,۰۵



ج) در این مسئله هرچقدر درجه برازش افزایش یافته است، حداقل مربعات خطا کاهش یافته است زیرا منحنی برازش شده به داده ها نزدیکتر شده است و اختلاف داده های جدید با داده های اصلی کم شده است. واریانس با افزایش درجه برازش منحنی ها افزایش می یابد زیرا هرچه درجه آن زیاد میشود، بیشتر نویز داده ها مدل میشود و واریانس هم مستقیماً زیاد میشود. بایاس مدل ابتدا کم میشود و سپس زیاد میشود.



Variance:

1: 7.772504638878087e-08, 2: 0.016803172486435192, 3: 0.01681073655654209  
4: 0.06621247030260476, 5: 0.06743994177709274, 6: 0.0838462069938738,  
7: 0.08961936142479007, 8: 0.20858637879235706, 9: 0.21309078080784175,  
10: 0.2503936997499987, 11: 0.25070300570074805, 12: 0.25588780045436815,  
13: 0.2648893905181676, 14: 0.2660213101736884, 15: 0.2693578238798767

Bias:

1: 0.5368212924765313, 2: 0.5074011766755355, 3: 0.5074550102880495,  
4: 0.4052950581749932, 5: 0.4024505640598039, 6: 0.36564304309623263,  
7: 0.3469343165282122, 8: 0.32106858182217074, 9: 0.3317838060709222,  
10: 0.42617088455469554, 11: 0.42762651127363405, 12: 0.43925317306993017,  
13: 0.457997800958285, 14: 0.46033432515571077, 15: 0.4670052274579057}

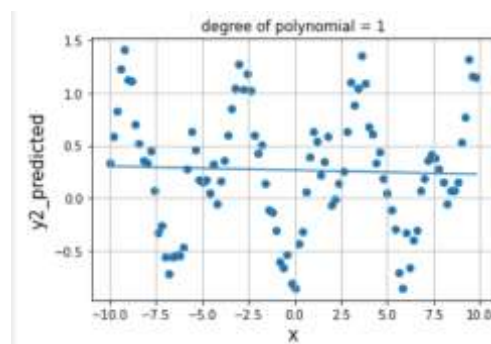
### حالت دوم : داده آغشته به نویز پواسون

الف) در این حالت هم مثل حالت قبل بهترین درجه برازش منحنی مطابق نمودار و MSE مربوط به درجه

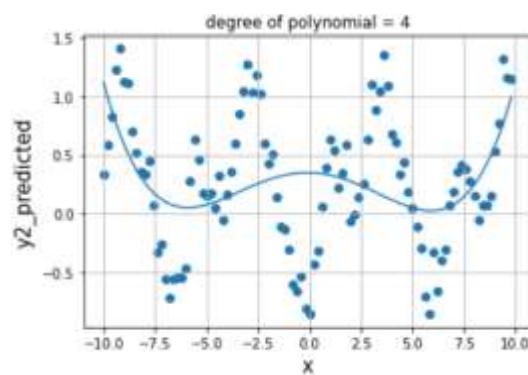
۱۵ است.

(ب)

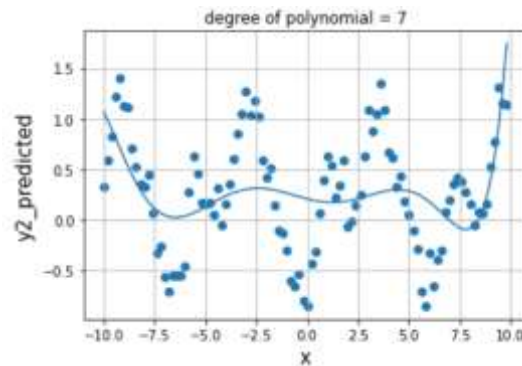
مقدار MSE درجه برازش ۱ : ۰,۳۳۹



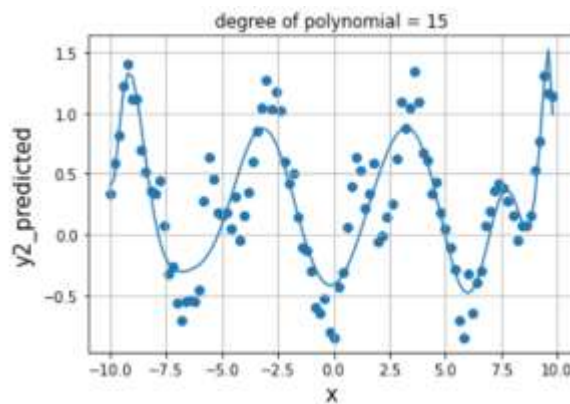
مقدار MSE درجه برازش ۴ : ۰,۲۸۱



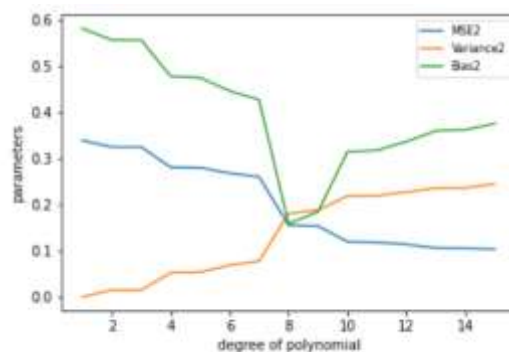
مقدار MSE درجه برازش ۷ : ۰,۲۶



مقدار MSE درجه برازش ۱۵ : ۰,۱۰۳



ج) تحلیل این قسمت هم مثل قسمت قبل است و تفاوت صرفا در مقادیر بایاس ، واریانس و MSE است اما روند کلی تغییرات مشابه است.



VARIANCE:

1: 0.0004831717949331432, 2: 0.015046185479767578, 3: 0.015136532300968855,  
4: 0.052639542331403735, 5: 0.0545460326621825, 6: 0.06898453410530515,  
7: 0.07772593662300592, 8: 0.18091223823627545, 9: 0.18816881578005382,  
10: 0.21885395128510543, 11: 0.21961945833112176, 12: 0.22751912438796523,  
13: 0.23606936100583295, 14: 0.2366232423237666, 15: 0.24505888523395342}

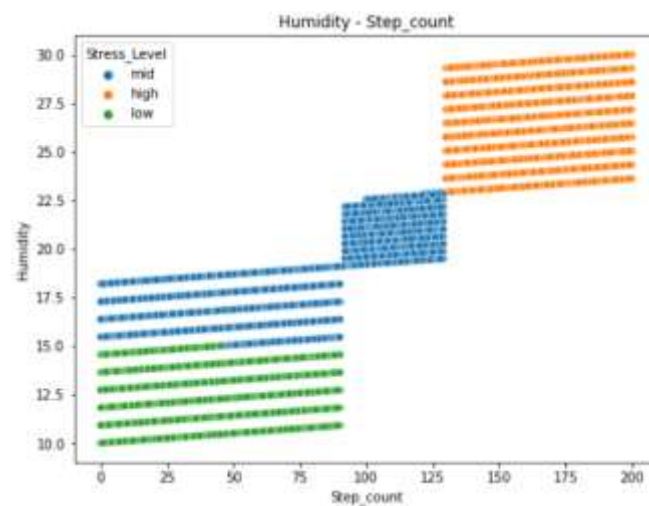
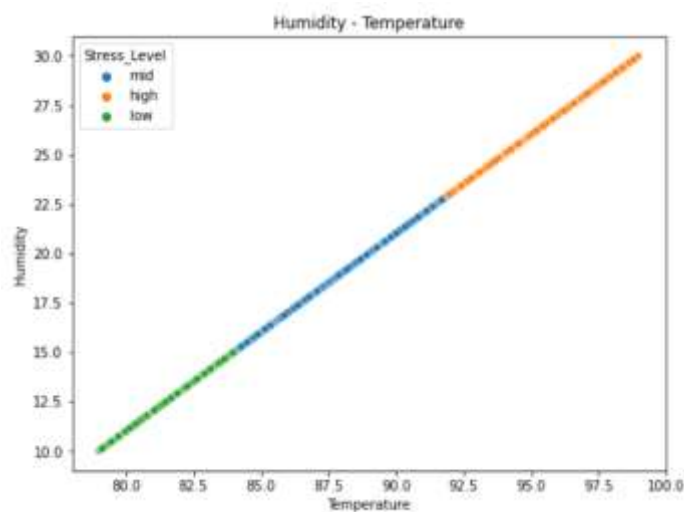


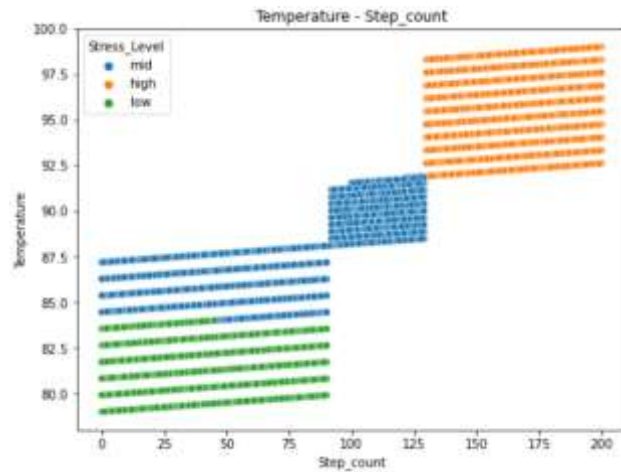
Bias:

1: 0.5819645891337089, 2: 0.5568506325642664, 3: 0.5566810683865461,  
4: 0.47817278109448835, 5: 0.47516917799341474, 6: 0.4463896373930628,  
7: 0.4276205249143854, 8: 0.1588777186698335, 9: 0.18471903086075858,  
10: 0.31465335220852847, 11: 0.31774503231290435, 12: 0.33675335797389533,  
13: 0.36026987285525, 14: 0.36234699732477715, 15: 0.37567504892918174

## سوال ۵

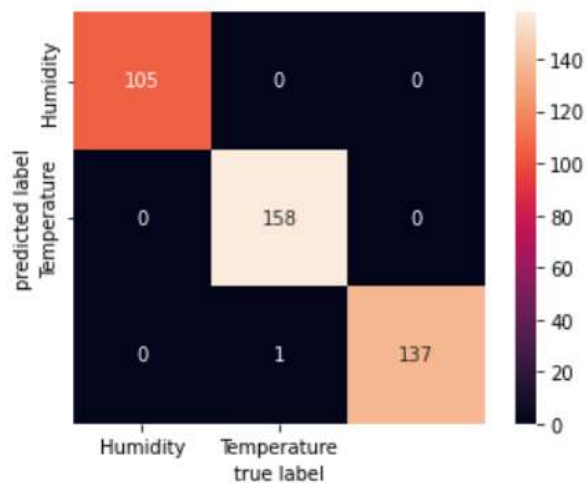
الف) ابتدا چند داده اول دیتاست را بررسی میکنیم برای اطمینان از صحت اطلاعات آن. سپس بر هر بار بر اساس دو ویژگی نمودار پراکندگی سطح استرس ( که با رنگ در نمودار ها مشخص است) را ترسیم میکنیم.





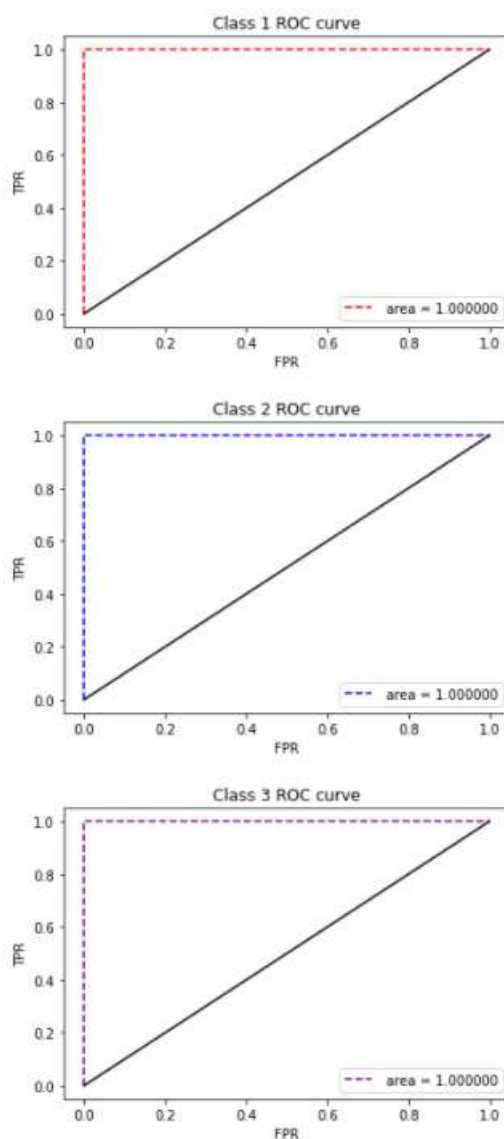
همانطور که مشخص است نموداری که بر اساس دما و رطوبت ترسیم شده است بهترین نمودار برای تفکیک سطح استرس داده ها میباشد.

ج) موارد خواسته شده سوال را با استفاده از کتابخانه انجام دادم. ابتدا داده ها را به نسبت ۸۰ درصد داده آموزش و بقیه تست تفکیک کردم. سپس تکنیک one against all را به کار بردم و برای هر نمونه، هر بار یک کلاس را از نظر احتمال وقوع با بقیه مقایسه کرده و در نهایت طبقه بندی کردم. ابتدا مدل را به داده آموزش ترین کرده و سپس داده های آزمون را به مدل داده و کیفیت طبقه بند مورد بررسی قرار گرفت. البته تمامی موارد به کمک توابع sklearn انجام شد.



The accuracy of model is 0.9975062344139651  
 f1\_score : 0.997736354077048  
 jaccard\_score : 0.9954881050041017

در نهایت نمودار roc را برای هر سه کلاس ترسیم کردم.



به دلیل اینکه داده ها بر اساس ویژگی های رطوبت و دما کاملاً از تفکیک شده بودند دقت طبقه بند بسیار بالا بود و هر کلاس roc ایده ال دارند. یعنی هم true positive rate همه کلاسها ۱ است و هم false positive rate همه صفر است. برای همین سطح زیر نمودار برای همه کلاسها چنانچه در شکل مشخص است ۱ شده است.

## سوال ۶

الف) برای تولید نقاط رندوم با توزیع یکسان در یک حلقه مشخص یک تابع به نام ring نوشتم. سپس از این تابع برای ساخت داده های دسته اول و دوم حالت اول و دسته دوم حالت دوم که چنین توزیعی در ضا داشتند استفاده کردم. برای تولید دسته اول از حالت دوم از یک تابع تولید رندوم ساده استفاده کردم. نتایج در ادامه آورده شده است.

