به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

ماشین لرنینگ

تمرین ۲

نام و نام خانوادگی	محمد مشرقی
شماره دانشجویی	810199492
تاریخ ارسال گزارش	

فهرست

۱–

الف

ML

۱۳۹۶
فروردین
۲۶
شنبه

HW2

محمد [illegible]
81019391

15 Apr. 2017 ▪ ۱۴۳۸ رجب ۱۷

$x = [5, 11, 18, 15, 21, 6, 17, 10, 24, 19]^T$

$y = [34, 47, 66, 52, 80, 35, 66, 48, 87, 81]^T$

الف) در اینجا چون $y = \beta + \varepsilon$ و β مقدار ثابتی است

پس $\underline{x = [1, \dots 1]^T_{1\times 10}}$ فرض می‌کنیم و شروع به حل مسئله می‌کنیم

$$\Theta = \underbrace{(\underset{1\times 10}{x^T}\ \underset{10\times 1}{x})^{-1}}_{1\times 1}\ \underset{1\times 10}{x}\ \underset{10\times 1}{y}$$

$$(x^T x) = \sum_{1}^{10} 1 = 10 \qquad x.y = \sum_{i=1}^{10} y_i = 596$$

$$\Theta = \frac{1}{10} \times 596 = 59.6$$

$$\boxed{y = 59.6 + \varepsilon}$$

ب

$y = \beta_1 x + \epsilon$

ب) در اینجا برای برازش داده ها از مقادیر داده شده استفاده میکنیم

بدون ضریب β_1 به آن شیب داده‌ام

$$(x^T x) = 5^2 + \dots 19^2 = 2498 \Rightarrow (x^T x)^{-1} = \frac{1}{2498}$$

$$x^T y = 5\times 34 + \dots + 19 \times 81 = 9774$$

$$\theta = \frac{9774}{2498} \Rightarrow \boxed{y = 3.9127x + \epsilon}$$

ج

$$\hat{Y} = b_0 + b_1 X \quad \text{و} \quad Y = \beta_0 + \beta_1 X + \epsilon$$

ج)

$\hat{Y}$ به مقادیر تقریبی هستش از Y به دست آمده است

اما ϵ یک متغیر تصادفی است که به ازای هر x یک مقدار مشخص دارد و باعث

می‌شود که $\hat{Y}$ که مقدار تقریبی است به مقدار واقعی آن یعنی Y برسد

د)

$$Var = \frac{1}{n}\sum_{i=1}^{n=10}(y_i - \hat{y}_i)^2$$

د

الف) $\Rightarrow \boxed{Var_0 = 331.8}$

ب) $\Rightarrow \boxed{Var_1 = 59.7}$

۵

۵) خیر چون ε یک متغیر تصادفی است و برای هر x مقدار نامشخصی دارد و به همین علت نمی‌توان دقیق پیش‌بینی کرد.

۲-

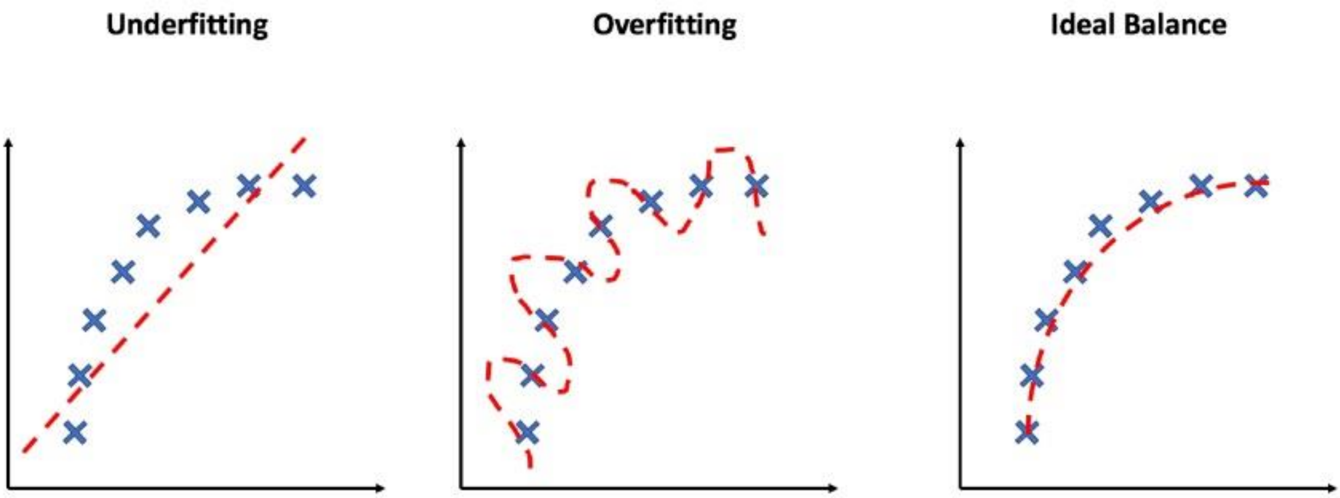
الف

ابتدا به تعریف می پردازیم:

در ماشین لرنینگ و دیپ لرنینگ برای اینکه از

۱. پیچیده تر شدن مدل(Overfitting).
۲. و نتایج بسیار نزدیک به داده های تمرینی نزدیک شود که در نهایت ممکن است باعث شود در داده های استفاده نشده (جدید) خطا زیاد شود.

جلوگیری شود از Regularization استفاده می شود



انواع آن :

- **L1 Regularization (Lasso Regression):** (Least Absolute Shrinkage and Selection Operator)

در اینجا با انتخاب مناسب lambda می توانیم برای هر ضریب به مقداری مناسب برسیم و آن را به کم کنیم(و نهایتا وزن به صفر متمایل کنیم)

این ویژگی وقتی ضریب های زیادی داشته باشیم و بخواهیم وزن ضریب های کم مهم صفر شوند بدرد می خورد.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

- **L2 Regularization(Ridge regression):**

در اینجا هم مانند L1 عمل می کند اما در انتهای تمرین نتایج فقط به صفر میل می کنن و صفر نمی شوند.

وقتی استفاده می کنیم که ویژگی های هم خطی و همبستگی داشته باشیم.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Comparison of L1 and L2 regularization	
L1 regularization	***L2 regularization***
Sum of absolute value of weights	Sum of square of weights
Sparse solution	Non-sparse solution
Multiple solutions	One solution
Built-in feature selection	No feature selection
Robust to outliers	Not robust to outliers (due to the square term)

ب

ب) 2

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n}(Y_i - X_i\beta)^2 + \lambda\|\beta\|_2^2$$

$$= \arg\min_{\beta} (Y - X\beta)^2 + \lambda\|\beta\|_2^2$$

$$loss = Y^TY - 2\hat{\beta}^T X^T Y + \beta^T X^T X \beta + \lambda \beta^T \beta$$

حال نسبت به β مشتق برابر صفر قرار میدهیم

$$\frac{\partial loss}{\partial \beta} = 0 = -2X^TY + 2X^TX\beta + 2\lambda\beta = 0$$

$$\Rightarrow (X^TX + \lambda I)\beta = X^TY$$

$$\Rightarrow \boxed{\hat{\beta} = (X^TX + \lambda I)^{-1} X^T Y}$$

الف

$$P(Y=y_k|X) \sim \exp(w_{k0} + \sum_{i=1}^{n} w_{ki} X_i)$$ from 1 to K-1

الف) چون چند کلاس داریم W_K بصورت بردار است

$$f(x;w) = \begin{bmatrix} P(y=1|x;w_1) \\ \vdots \\ P(y=y_k|x;w_k) \end{bmatrix}$$

~~K~~ و دیگر تمام آن‌ها باید مثبت باشند و برای محاسبه $P(Y=y_k|X)$

داریم:

$$P(y=k|X) = \frac{e^{w_k^T x + w_{k0}}}{\sum_{j=1}^{K} e^{w_j^T x + w_{j0}}}$$

در این صورت ما از softmax استفاده می‌کردیم

پس به جای logistic reg ما گوییم $\boxed{\text{softmax reg}}$

ب

ب) قانون طبقه‌بندی به این صورت است که ~~...~~

به ازای هر ورودی هر کلاسی که احتمال بیشتری دارد

متعلق به آن است ~~...~~

$$\boxed{K = \arg\max_{k_i} P(Y=k|x)}$$

۴

الف

28 Mar. 2017 ■ ۲۹ جمادی الثانیة ۱۴۳۸ ■

$$\hat{p}(x) = \frac{1}{n h_n}\sum_{i=1}^{n}\varphi\left(\frac{x-x_i}{h}\right)$$

$$\boxed{\bar{P}_n(x) = N(\mu, h_n^2+\sigma^2)}$$

اثبات

$$\bar{P}_n(x) = E[P_n(x)] = E\left[\frac{1}{n h_n}\sum_{i=1}^{n}\varphi\left(\frac{x-x_i}{h_n}\right)\right] = \frac{1}{n h_n}\sum_{i=1}^{n}E\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right]$$

$$= \frac{1}{h_n}\int_{-\infty}^{\infty}\varphi\left(\frac{x-u}{h_n}\right)P(u)\,du = \frac{1}{h_n}\int_{-\infty}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-u)^2}{h_n^2}}}_{\varphi\left(\frac{x-u}{h_n}\right)}\underbrace{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\mu-u}{\sigma}\right)^2}}_{P(u)}du$$

$$= \frac{1}{2\pi\sigma h_n}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(\frac{x-u}{h_n}\right)^2-\frac{1}{2}\left(\frac{\mu-u}{\sigma}\right)^2} = \frac{e^{-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}\right)}}{2\pi h_n\sigma}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}u^2\underbrace{\left(\frac{1}{h_n^2}+\frac{1}{\sigma^2}\right)}_{C_1}-u\underbrace{\left(\frac{x}{h_n^2}+\frac{\mu}{\sigma^2}\right)}_{C_2}}$$

$$= \frac{1}{2\pi h_n\sigma}e^{-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}\right)}\left(\frac{\sqrt{2\pi}}{\sqrt{C_1}}e^{\frac{C_2^2}{2C_1}}\right)$$

ساده سازی

$$\simeq \boxed{\bar{P}_n(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{h_n^2+\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}}}$$

ب

جلسه ۱ | ب ۱

فروردین ۷ دوشنبه ۱۳۹۶

27 Mar. 2017 ■ ۲۸ جمادی الثانیه ۱۴۳۸

$$P_h(x) - \bar{P}_h(x) \simeq \frac{1}{2}\left(\frac{h_n}{\sigma_f}\right)^2\left[1-\left(\frac{x-\mu}{\sigma}\right)^2\right]P(x)$$

$$\rightarrow \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x-\mu}{2\sigma^2}} - \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{h_n^2+\sigma^2}} e^{\left(-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)}$$

$$= \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right\}}\left[1 - \frac{\sigma}{\sqrt{h_n^2+\sigma^2}} e^{\left\{\frac{(x-\mu)^2}{2}\left(\frac{1}{\sigma^2}-\frac{1}{h_n^2+\sigma^2}\right)\right\}}\right]$$

گوسی کامل نیست

$$\frac{h_n^2}{\sigma^2(h_n^2+\sigma^2)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\left\{e^{\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}}\right\}\left(1 - \frac{\sigma}{\sqrt{h_n^2+\sigma^2}} e^{\frac{(x-\mu)^2}{2}\frac{h_n^2}{\sigma^2(h_n^2+\sigma^2)}}\right)$$

(3) (2) (1)

از تیلور

$$(1) \simeq 1 + \frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2} \qquad (2) \Rightarrow 1 - \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2$$

حل

$$(3) = P(x)$$

$$= P(x)\left(1 - \left(1-\frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\right)\left(1+\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)\right)$$

$$= P(x)\left[1 - \left(1+\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right) + \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left(1+\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)\right]$$

$$= P(x)\left[-\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2} + \frac{1}{2}\frac{h_n^2}{\sigma^2} + \frac{h_n^4}{4\sigma^4}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right]$$

قابل صرف نظر کردن: h_n خیلی کوچک است

۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴ ۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱ ۲۲ ۲۳ ۲۴ ۲۵ ۲۶ ۲۷ ۲۸ ۲۹ ۳۰ ۳۱

$$= P(x)\left(-\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}+\frac{1}{2}\frac{h_n^2}{\sigma^2}\right)$$

$$= P(x)\frac{h_n^2}{\sigma^2}\left(-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}+\frac{1}{2}\right)$$

$$\boxed{=\frac{h_n^2}{2\sigma^2}P(x)\left(1-\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)} \longrightarrow$$

$h_n \ll 1$ قابل صرفنظر

$$P(x)-\bar{P}_n(x) \simeq \frac{1}{2}\left(\frac{h_n^2}{\sigma^2}\right)\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]P(x)$$

۵

الف

$$D(a,b)=\sqrt{\sum_{k=1}^{d}(a_k-b_k)^2}$$

$A=D[a_1,\dots,a_n]$ از ق نامساوی کوشی شوارتز استفاده می‌کنیم

$$D(a,b)+D(b,c)\geq D(a,c) \qquad a-b=x \quad b-c=y$$

$$\|A(a-b)\|_2 \quad \|A(b-c)\|_2 \quad \|A(a-c)\|_2$$

$$\Rightarrow \|Ax\|_2+\|Ay\|_2 \geq \|A(x+y)\| \xrightarrow[Ay=y']{Ax=x'}$$

$$\|x'\|_2+\|y'\|_2 \geq \|x'+y'\| \xrightarrow[\text{هر دو طرف به توان 2}]{}$$

$$\|x'\|_2^2+\|y'\|_2^2+2\|x'\|_2\|y'\|_2 \geq \|x'\|_2+\|y'\|_2+2\sum_{i=1}^{d}x'y'=\|x'+y'\|_2$$

پس می‌توان نوشت

$$\|x'\|_2+\|y'\|_2 \geq \sum_{i=1}^{d}x'_iy'_i \Rightarrow \|x'\|_2+\|y'\| \geq \left|\sum_{i=1}^{d}x'_iy'_i\right|$$

حال اگر می‌دانیم که در فضای متریک بگوییم که هر نقطه‌ای به نام P با کمترین فاصله وجود دارد

$$\varepsilon=\min_x D(P,x)$$

که x در آن زیرمجموعه $S=\{x_1,\dots,x_n\}$ هستند حال اگر تعداد نمونه‌ها زیاد شود داریم

$$\text{if } n\to\infty \Rightarrow \boxed{D(P,x_n)=\varepsilon}$$

حال اگر از نامساوی کوشی شوارتز استفاده کنیم و x^* را نقطه همگرایی x_n در نظر بگیریم

داریم:

$$\underbrace{D(P, x_n) + D(x_n, x^*)}_{①} \geq \underbrace{D(P, x^*)}_{②}$$

$$\text{if } n \to \infty \Rightarrow ① = \varepsilon \Rightarrow \varepsilon \geq D(P, x^*) \quad ★$$

همچنین می‌دانیم که ε همواره کمترین مقدار هستش داریم

و x^* هم در زیرمجموعه S هستش:

$$D(P, x^*) \geq \varepsilon \quad ★★$$

★

★★

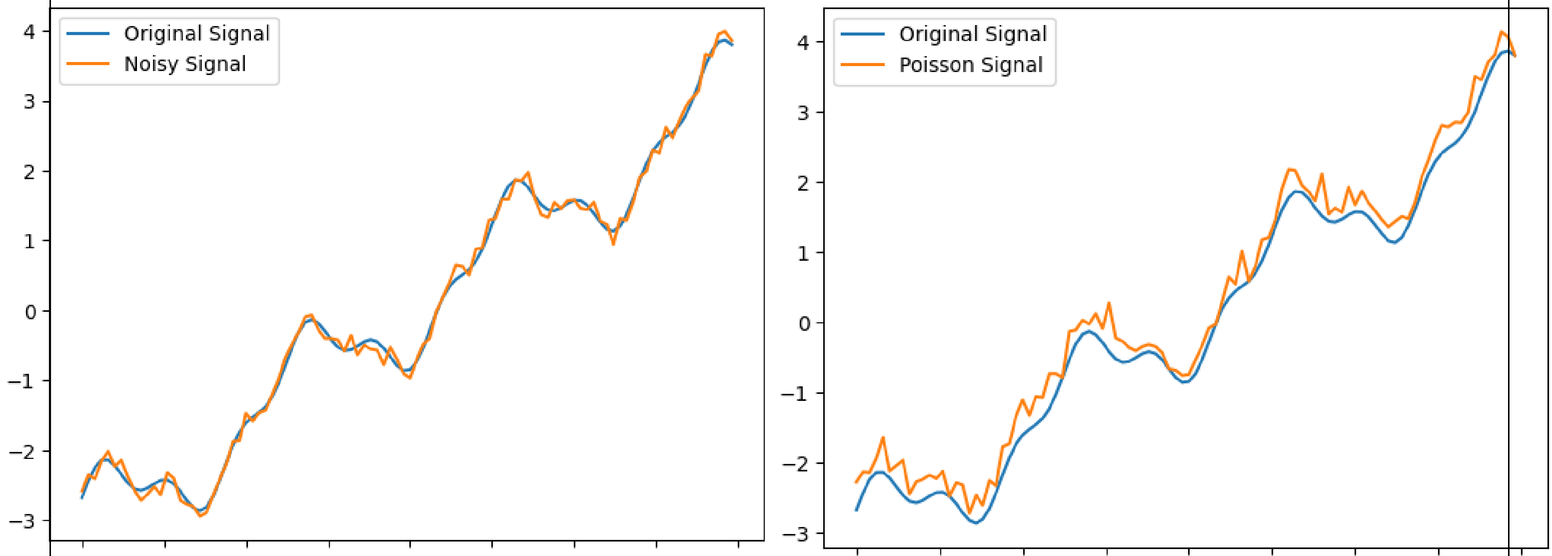
و

$D(x_i, x_i) = 0$

$$\Rightarrow P = x^* \Rightarrow \boxed{P\left(\frac{w_i}{P}\right) = P\left(\frac{w_i}{x^*}\right)}$$

ضریبی که در تمام مقادیر ضرب می‌شود، درسته است که ممکن است فاصله را زیاد یا حتی کم کند اما چون ما داریم را حسب K تا همسایه در نظر می‌گیریم برای همه به یک نسبت تغییر می‌کند و همسایه‌ها همان همسایه‌های قبلی باقی می‌مانند و تغییری ایجاد نمی‌کرده.

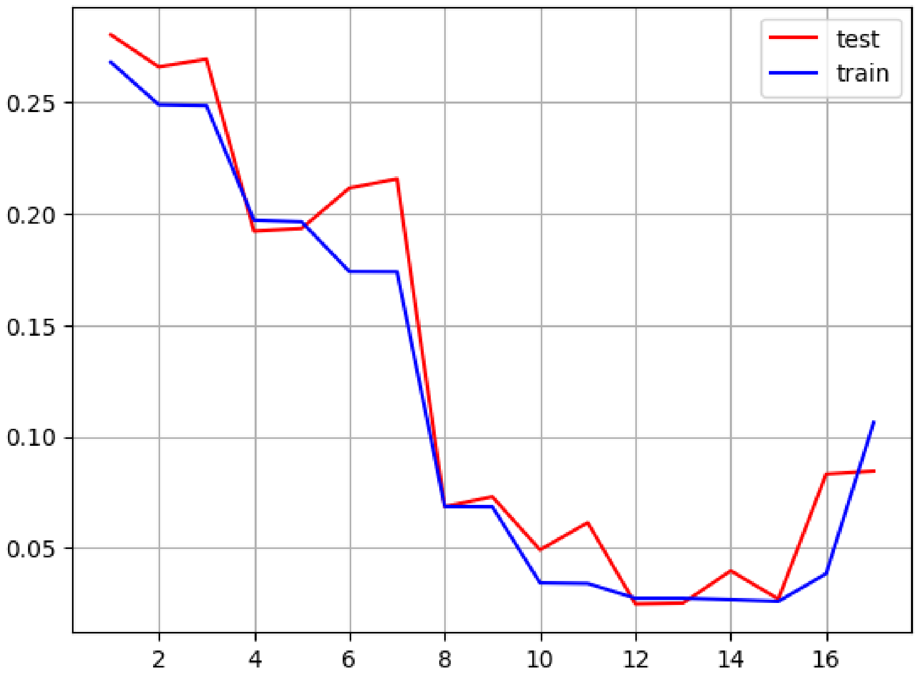
۶



الف

نویز سفید

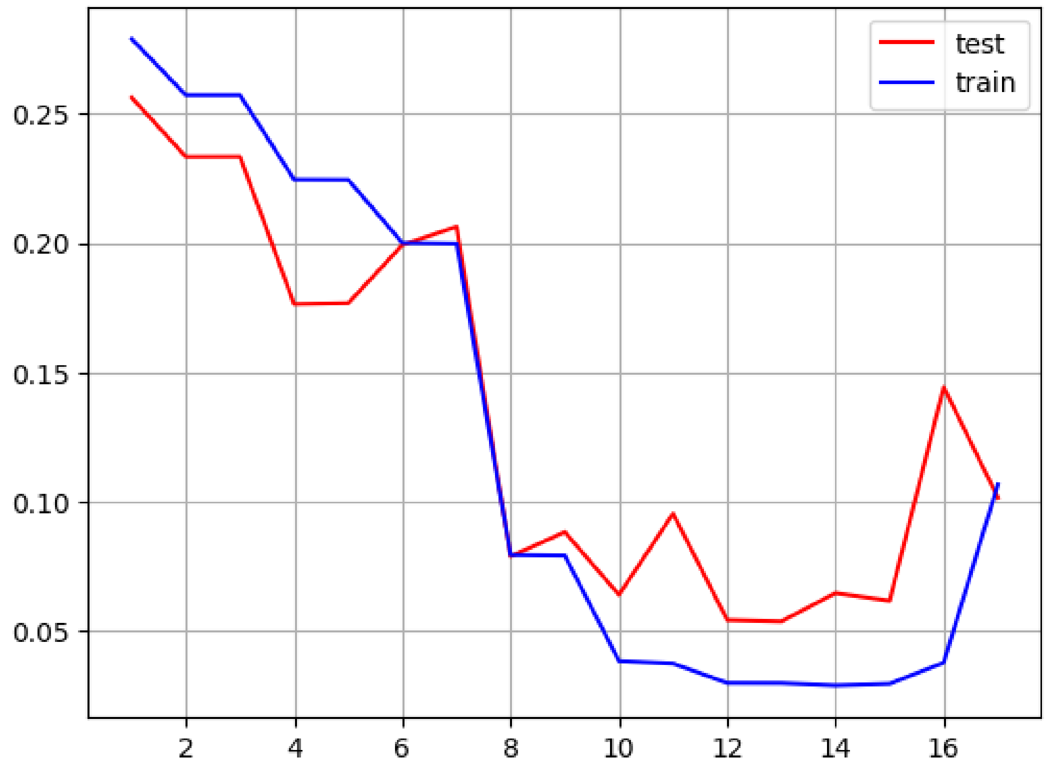
```
degree =  1 train mse =  0.26771108348933365 test mse =  0.28003451421474357
degree =  2 train mse =  0.2487550624892941 test mse =  0.2656941634748907
degree =  3 train mse =  0.2483644017310239 test mse =  0.26913772485455134
degree =  4 train mse =  0.196960668925193 test mse =  0.1920265128554382
degree =  5 train mse =  0.19619843565579803 test mse =  0.19326265967877293
degree =  6 train mse =  0.17394124599983093 test mse =  0.21134994978221727
degree =  7 train mse =  0.17386533041904073 test mse =  0.21548701725519565
degree =  8 train mse =  0.06862220305765028 test mse =  0.06854959188638675
degree =  9 train mse =  0.06848019794052991 test mse =  0.07296254515458053
degree =  10 train mse =  0.034349556166334984 test mse =  0.049174853025231266
degree =  11 train mse =  0.03408498065065172 test mse =  0.061359588850015835
degree =  12 train mse =  0.027480565064376482 test mse =  0.02491490438925687
degree =  13 train mse =  0.027472871099051943 test mse =  0.025313898035500094
degree =  14 train mse =  0.026815197900735683 test mse =  0.0397762309016583
degree =  15 train mse =  0.02606255491795073 test mse =  0.027113970890772117
```



بدترین درجه سه و بهترین درجه هشت

نویز پواسون

```
degree =  1 train mse =  0.27899075159837106 test mse =  0.2562824019970868
degree =  2 train mse =  0.2572387573237461 test mse =  0.23342995831507127
degree =  3 train mse =  0.2572374978964856 test mse =  0.2334719405449146
degree =  4 train mse =  0.22460239235841523 test mse =  0.1764418787938821
degree =  5 train mse =  0.22451549219816944 test mse =  0.17680117752547123
degree =  6 train mse =  0.20008539506632445 test mse =  0.19933412525812227
degree =  7 train mse =  0.19982793116955377 test mse =  0.2063716047648681
degree =  8 train mse =  0.07940775335109318 test mse =  0.07890323532953271
degree =  9 train mse =  0.07915632234536757 test mse =  0.08827466194308925
degree =  10 train mse =  0.038407013798615705 test mse =  0.0640046654436259
degree =  11 train mse =  0.037428579956825414 test mse =  0.09542133194367129
degree =  12 train mse =  0.029935434446202436 test mse =  0.054252790784845226
degree =  13 train mse =  0.02989593338534775 test mse =  0.0537285881892496
degree =  14 train mse =  0.029010239923529746 test mse =  0.06465889217803503
degree =  15 train mse =  0.029633805725978457 test mse =  0.06169862538875903
```



بهترین ۸ و بدترین ۴ (بین ۱ تا ۱۵)

برای بدست اوردن بهترین و بدترین :

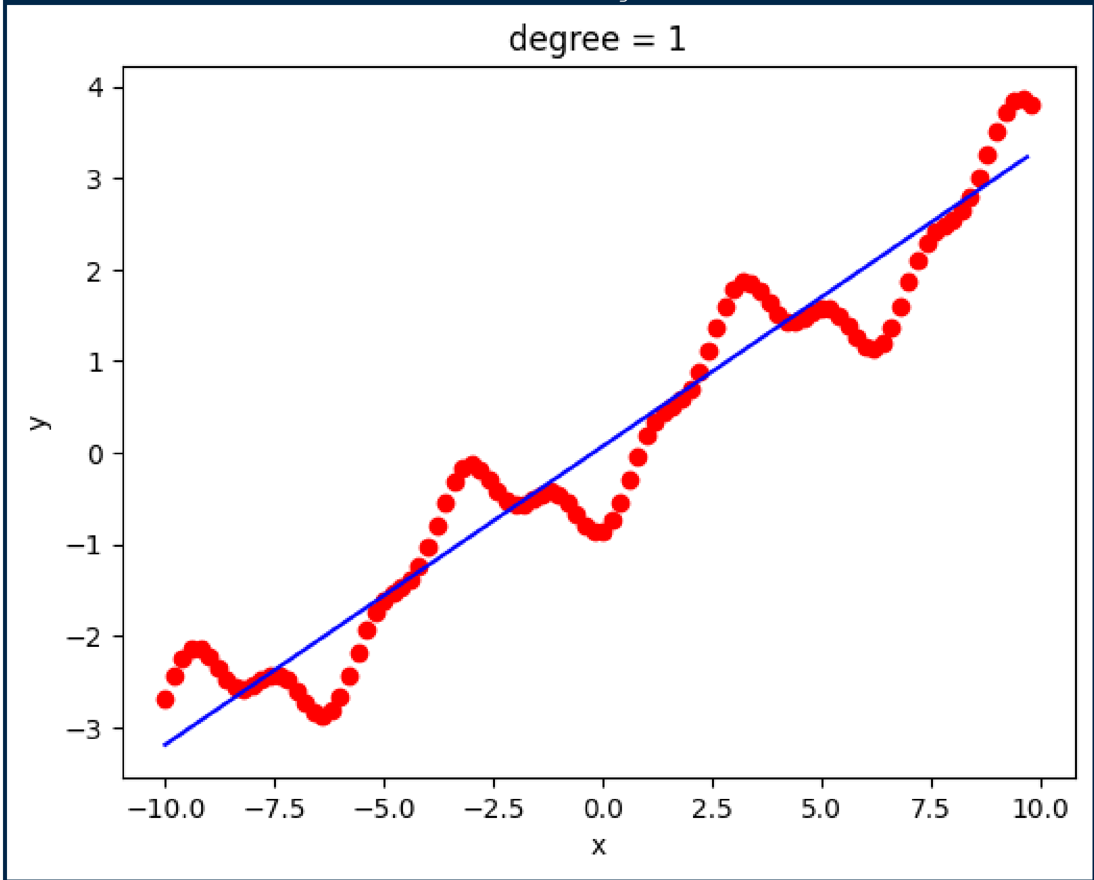
بهترین هرجا دو نمودار به هم نزدیک شد و در عین حال کمترین خطا رو داشت

بدترین هرجا از هم دورتر شد و در عین حال بیشترین خطا رو داشت

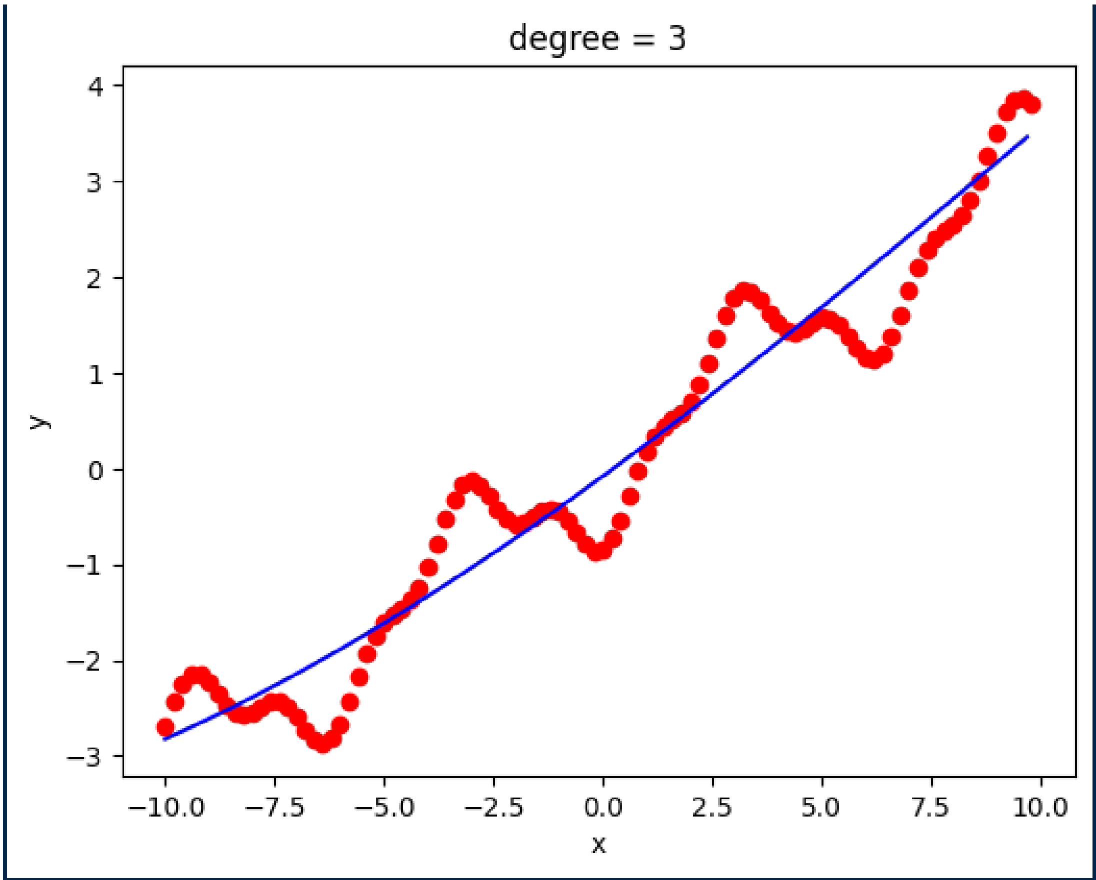
ب)

نویز سفید

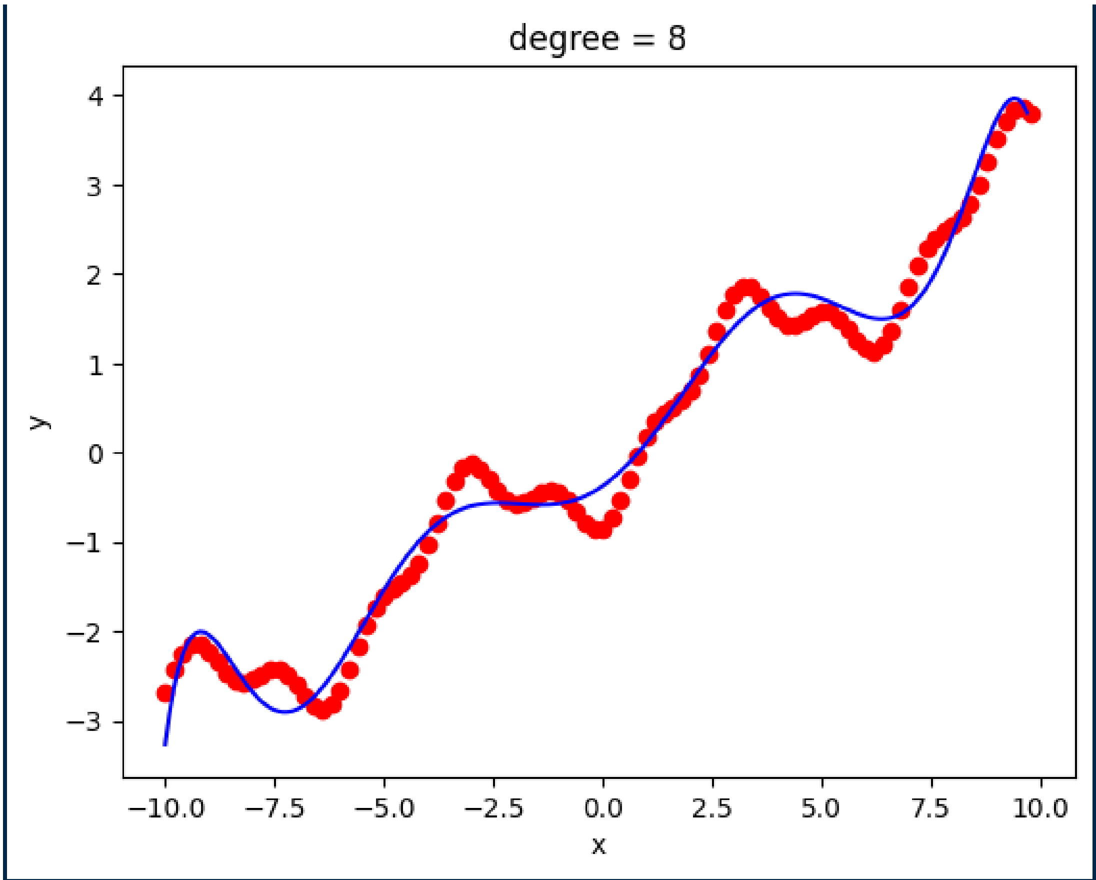
```
train error 0.26771108348933365 for degree =  1
test error 0.28003451421474357 for degree =  1
train bias 3.8059052729867036 for degree =  1
test bias 2.7489424186844857 for degree =  1
train variance 3.538194189497368 for degree =  1
test variance 3.402075099370019 for degree =  1
```



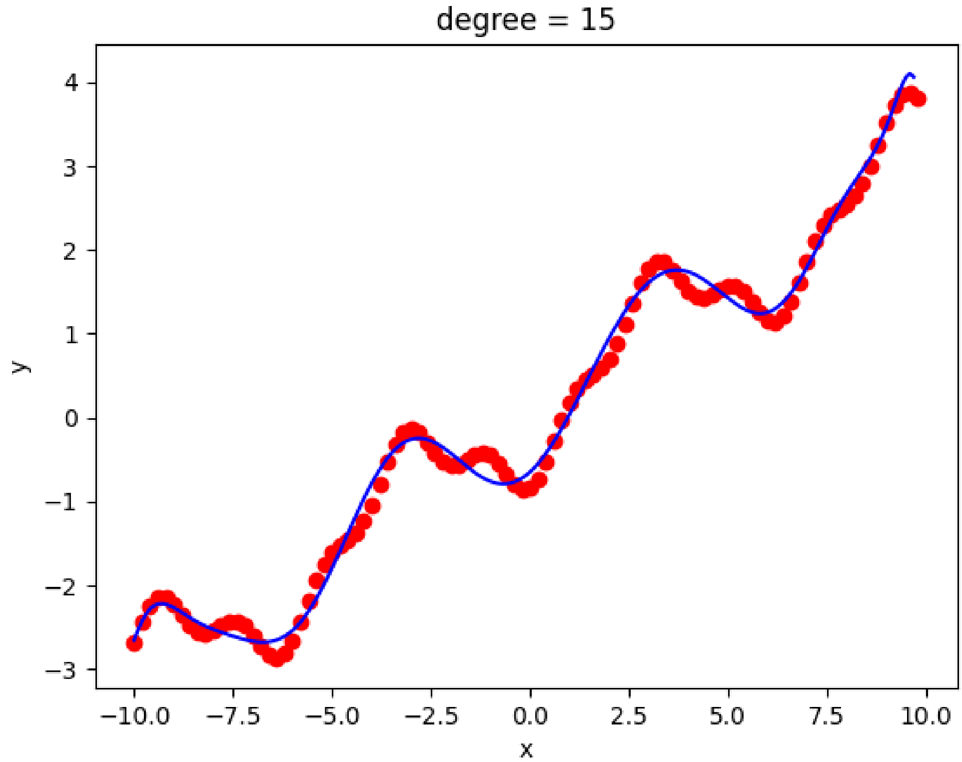

```
train error 0.24836440173102 39 for degree =  3
test error 0.26913772485455134 for degree =  3
train bias 3.8059052729867036 for degree =  3
test bias 2.750018463258655 for degree =  3
train variance 3.557540871255682 for degree =  3
test variance 3.2522898215326506 for degree =  3
```




```
train error 0.06862220305765028 for degree =  8
test error 0.06854959188638675 for degree =  8
train bias 3.805905272986705 for degree =  8
test bias 2.7356623615027362 for degree =  8
train variance 3.737283069715063 for degree =  8
test variance 3.1792038567492833 for degree =  8
```

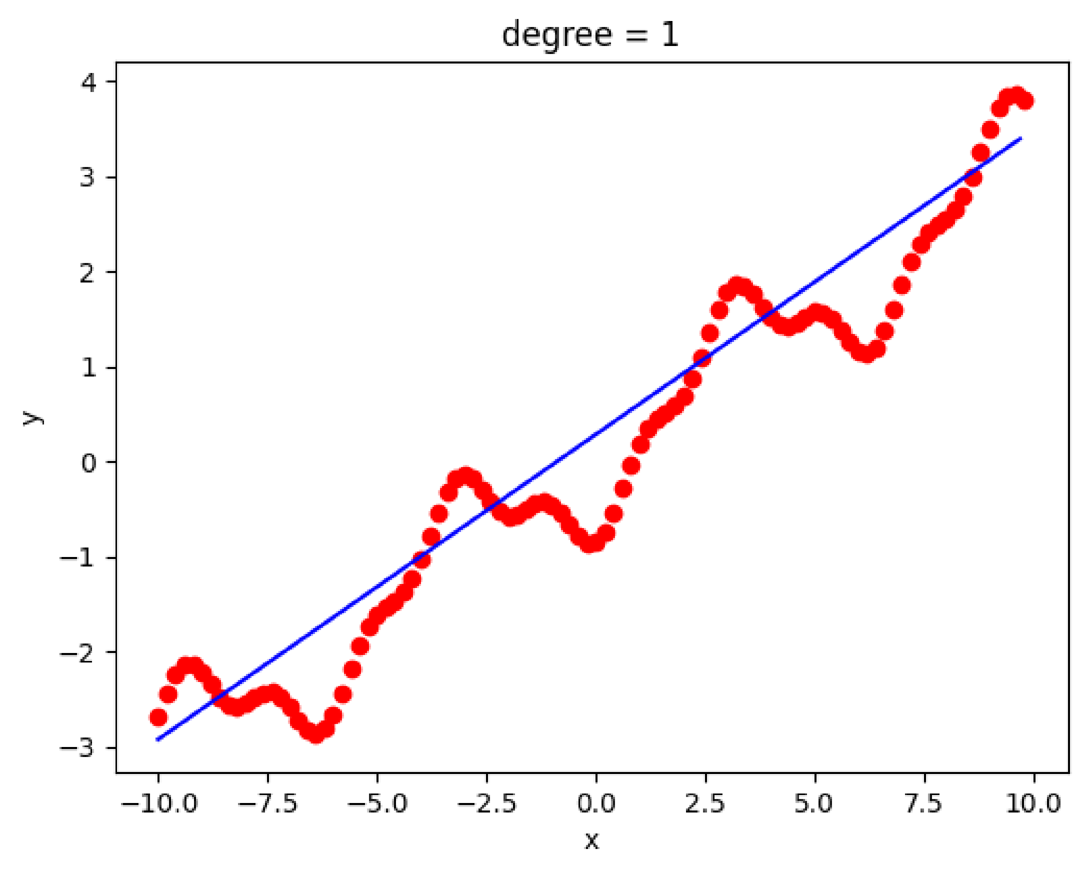



```
train error 0.02606255491795073 for degree =  15
test error 0.027113970890772117 for degree =  15
train bias 3.8059052729867036 for degree =  15
test bias 2.7347946410387145 for degree =  15
train variance 3.804939340687141 for degree =  15
test variance 2.8510911902094835 for degree =  15
```

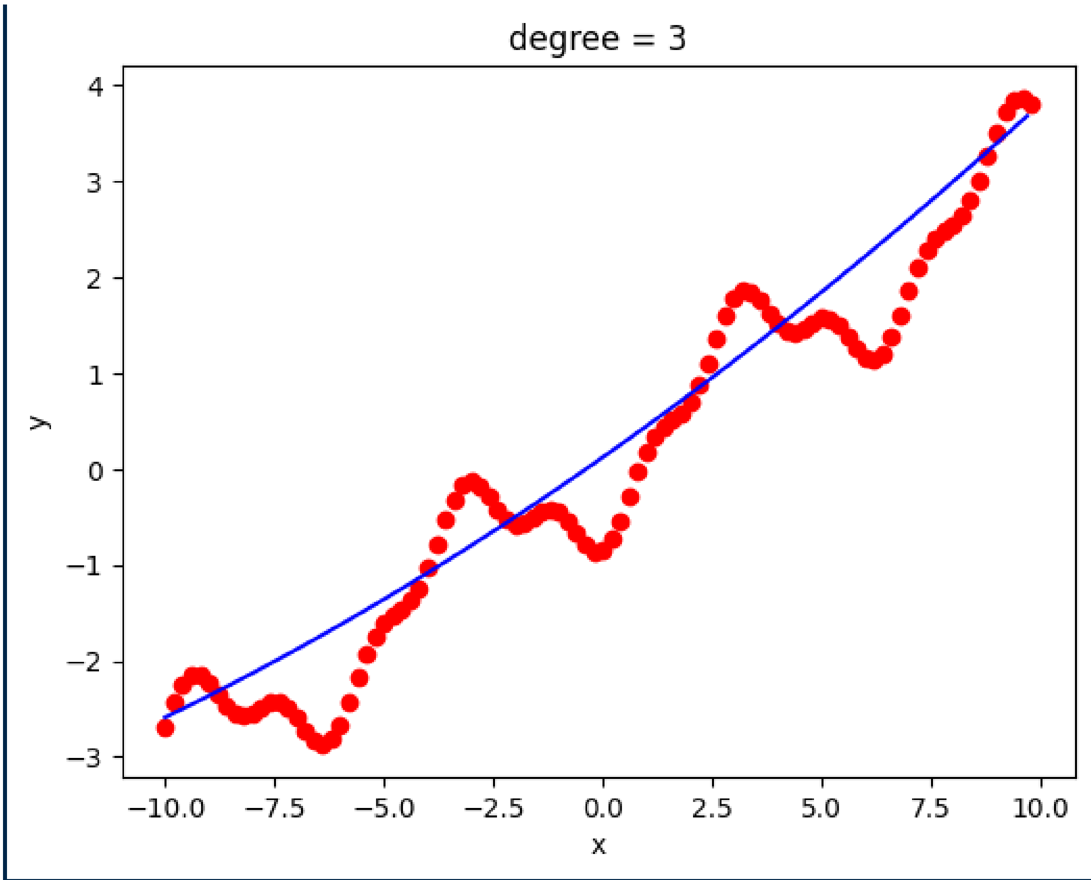


پوآسون

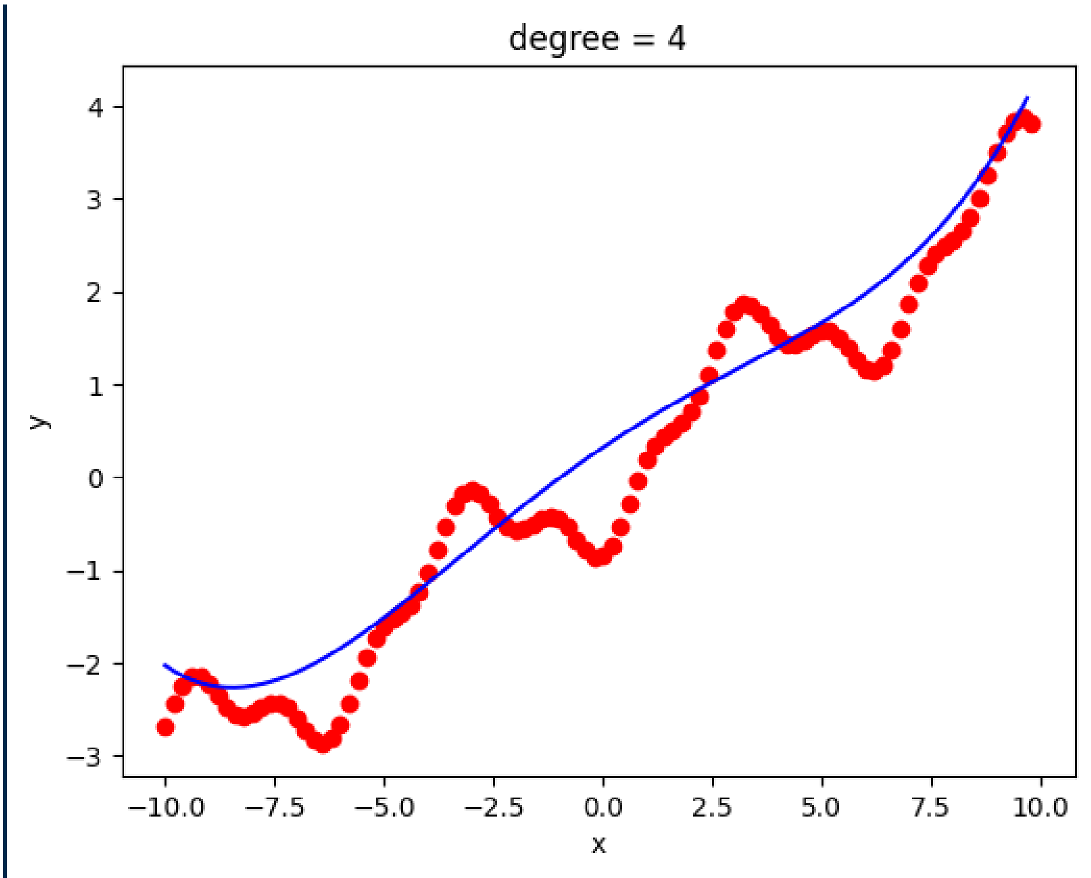
```
train error 0.27899075159837106 for degree =  1
test error 0.2562824019970868 for degree =  1
train bias 3.709691122170284 for degree =  1
test bias 2.6917654980254904 for degree =  1
train variance 3.4307003705719112 for degree =  1
test variance 3.298716712261696 for degree =  1
```



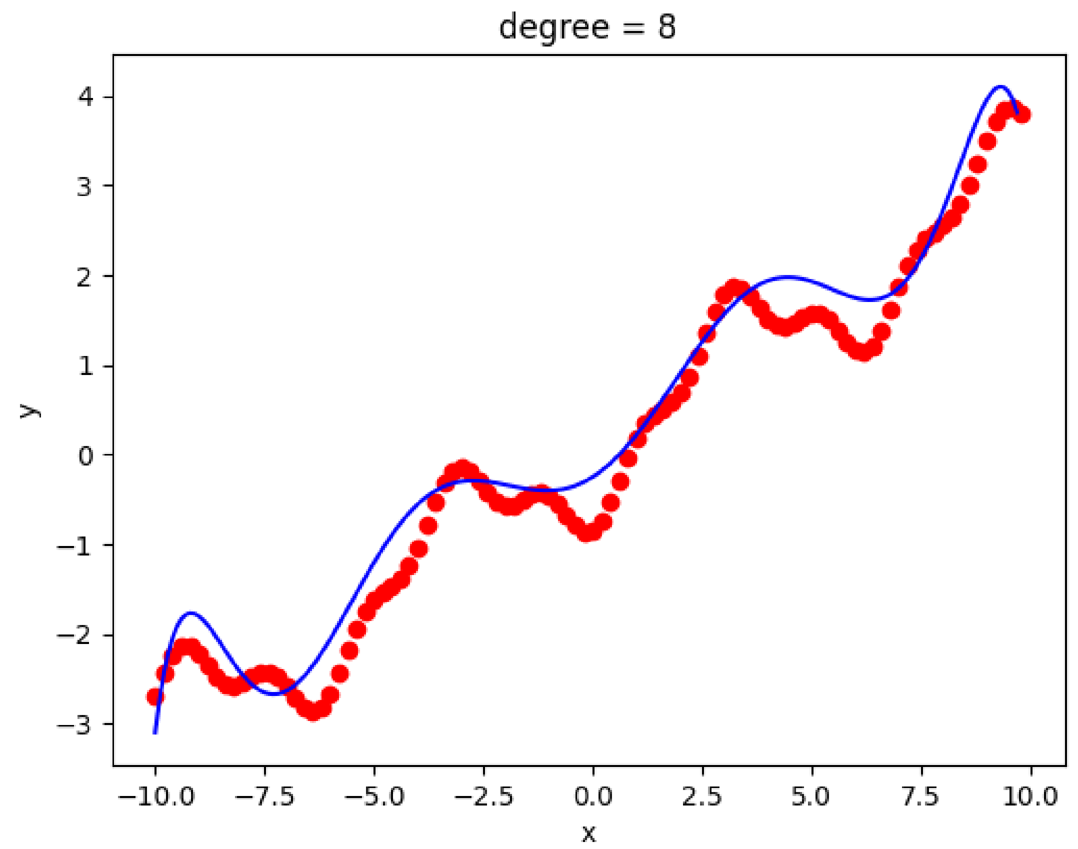

```
train error 0.2572374978964856 for degree =  3
test error 0.2334719405449146 for degree =  3
train bias 3.709691122170284 for degree =  3
test bias 2.691843051255867 for degree =  3
train variance 3.452453624273801 for degree =  3
test variance 3.1243181602293317 for degree =  3
```



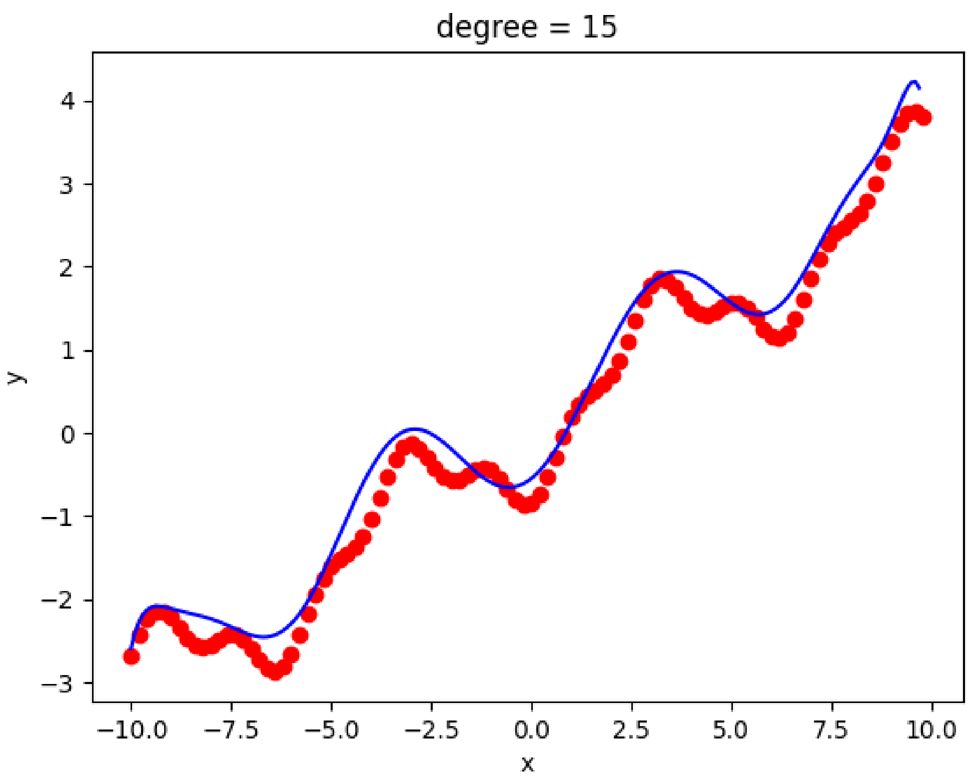

```
train error 0.22460239235841523 for degree =  4
test error 0.1764418787938821 for degree =  4
train bias 3.709691122170284 for degree =  4
test bias 2.689631528039265 for degree =  4
train variance 3.4850887298119004 for degree =  4
test variance 2.8276402208995917 for degree =  4
```




```
train error 0.07940775335109318 for degree =  8
test error 0.07890323532953271 for degree =  8
train bias 3.709691122170284 for degree =  8
test bias 2.689729284156079 for degree =  8
train variance 3.6302833682213773 for degree =  8
test variance 3.150727058311132 for degree =  8
```




```
train error 0.029633805725978457 for degree =  15
test error 0.06169862538875903 for degree =  15
train bias 3.709691122170284 for degree =  15
test bias 2.691260993173508 for degree =  15
train variance 3.7518429051758373 for degree =  15
test variance 2.829646644909956 for degree =  15
```



ج

می دانیم فرمول بایاس و واریانس به شرح زیر است:

Bias^2 = E[(y_true - y_pred)^2]

Variance = E[y_pred^2] - E[y_pred]^2

بایاس یک مدل، تفاوت بین مقدار مورد انتظار پیش‌بینی مدل و مقادیر واقعی متغیر هدف است. بایاس بالا به معنی زیر بودن مدل در مقابل داده‌ها است، به این معنی که نمونه های ترین را در داده‌ها به خوبی نمی‌گیرد.

واریانس یک مدل، میزان تغییر پیش‌بینی‌های آن برای نمونه‌های مختلف داده‌های آموزشی است. واریانس بالا به معنی بیش‌برازش شدن مدل به داده‌ها است، به این معنی که مدل نویز های داده را به جای نمونه های ترین ، یاد می‌گیرد.

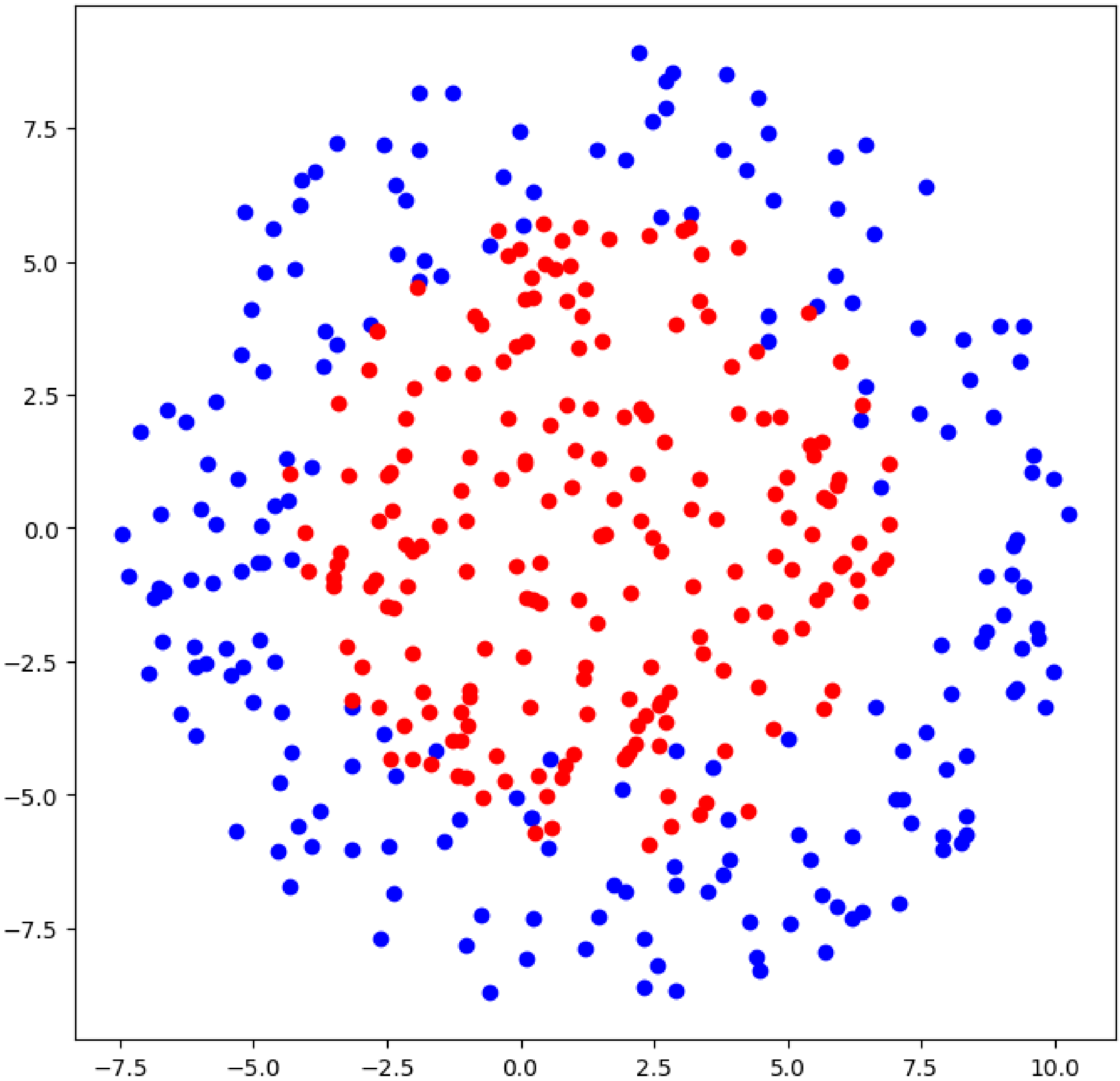
اگر یک مدل بایاس بالا و واریانس پایین داشته باشد، به این معنی است که مدل به اندازه کافی پیچیده نیست تا نمونه های ترین در داده‌ها را بگیرد، اما پیش‌بینی‌هایش برای نمونه‌های مختلف داده‌ها مطمئن هستند. از سوی دیگر، اگر یک مدل بایاس پایین و واریانس بالا داشته باشد، به این معنی است که مدل به اندازه کافی پیچیده است تا نمونه های ترین در داده‌ها را بگیرد، اما پیش‌بینی‌هایش برای نمونه‌های مختلف داده‌ها مطمئن نیستند.

بنابراین ما به دنبال مدلی هستیم که هم بایاس آن کم و واریانس آن کم باشد تا نه زیاد overfitشود یا برعکس

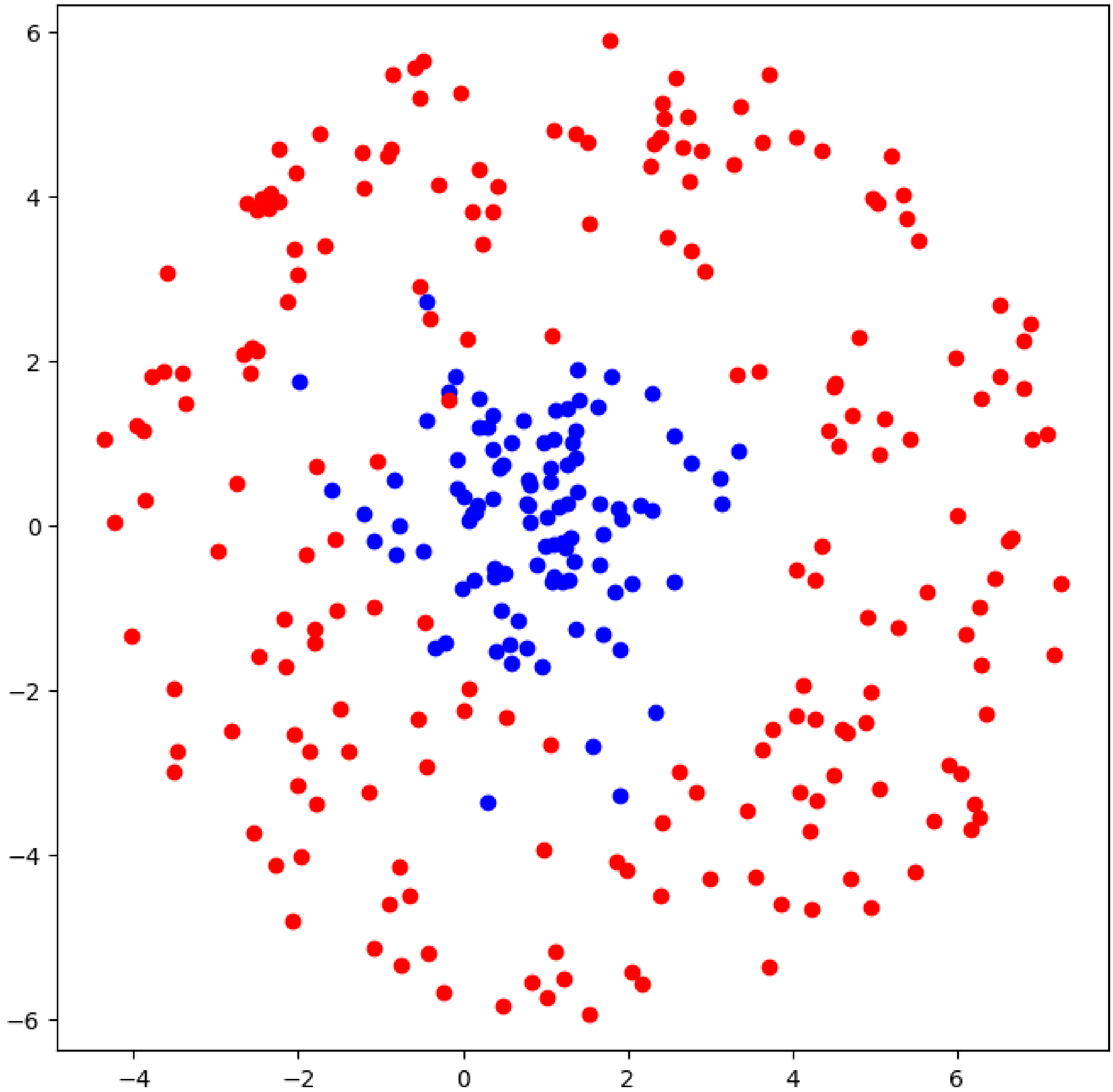
۷-

الف

حالت اول



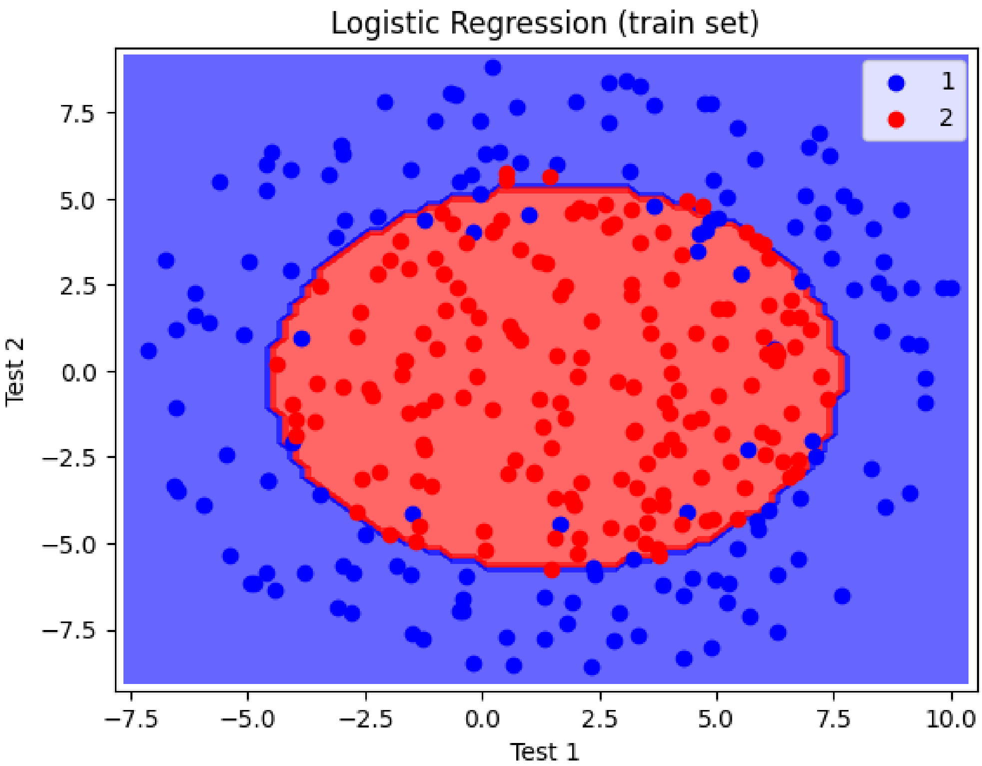
حالت دوم

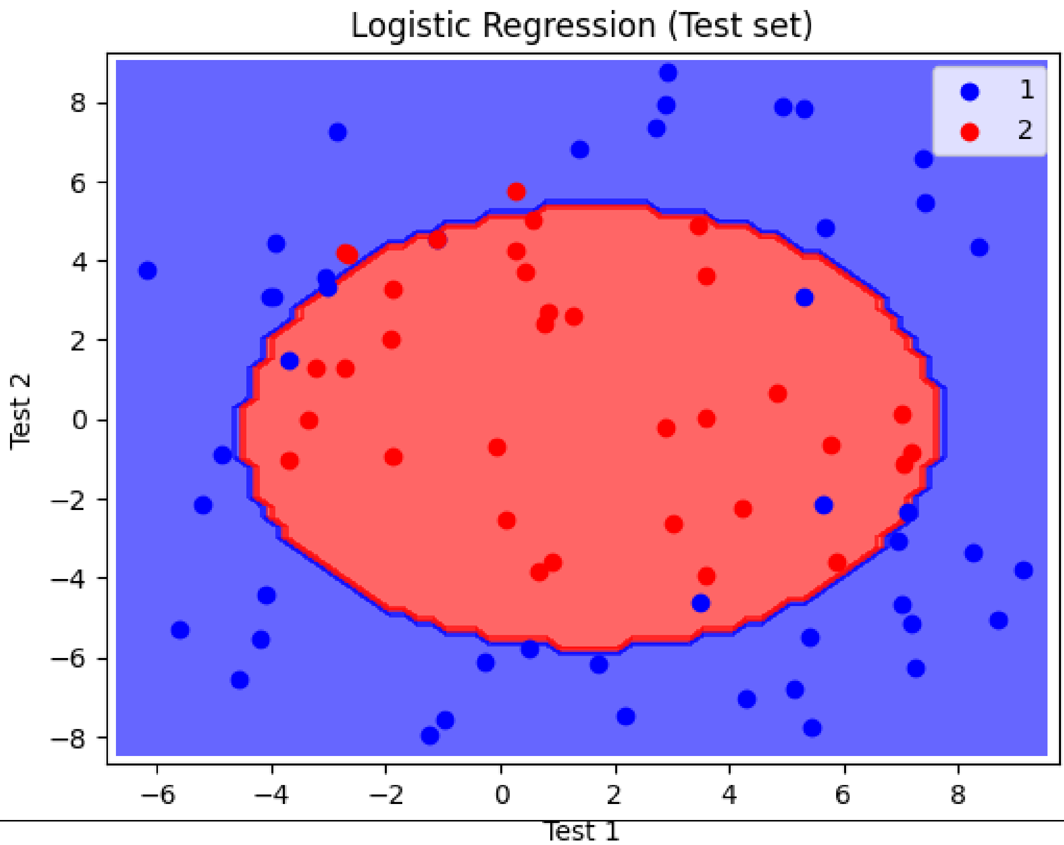


ب

حالت اول درجه دو:

```
confusion_matrix
= [[39  7]
 [ 3 31]]
accuracy_score = 0.875
```

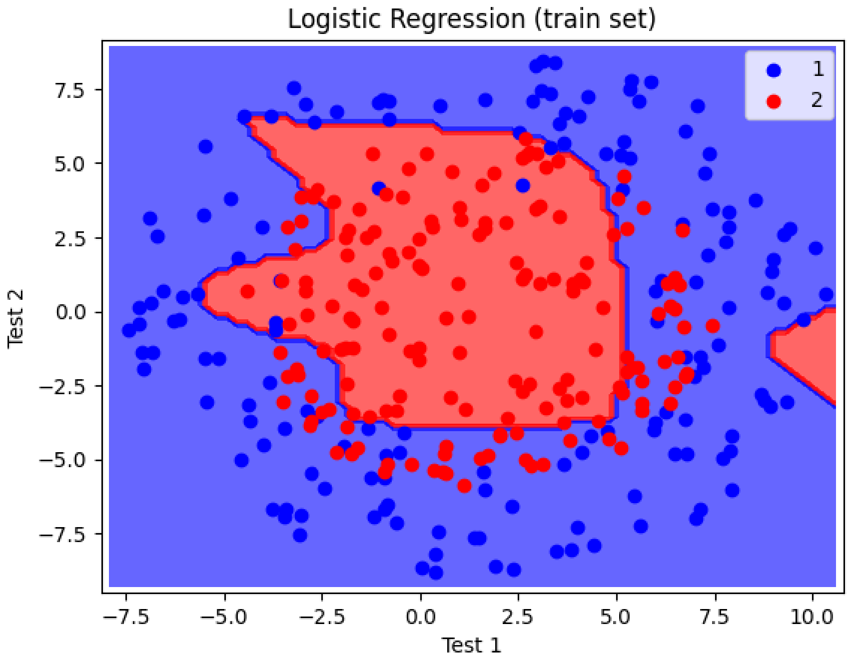


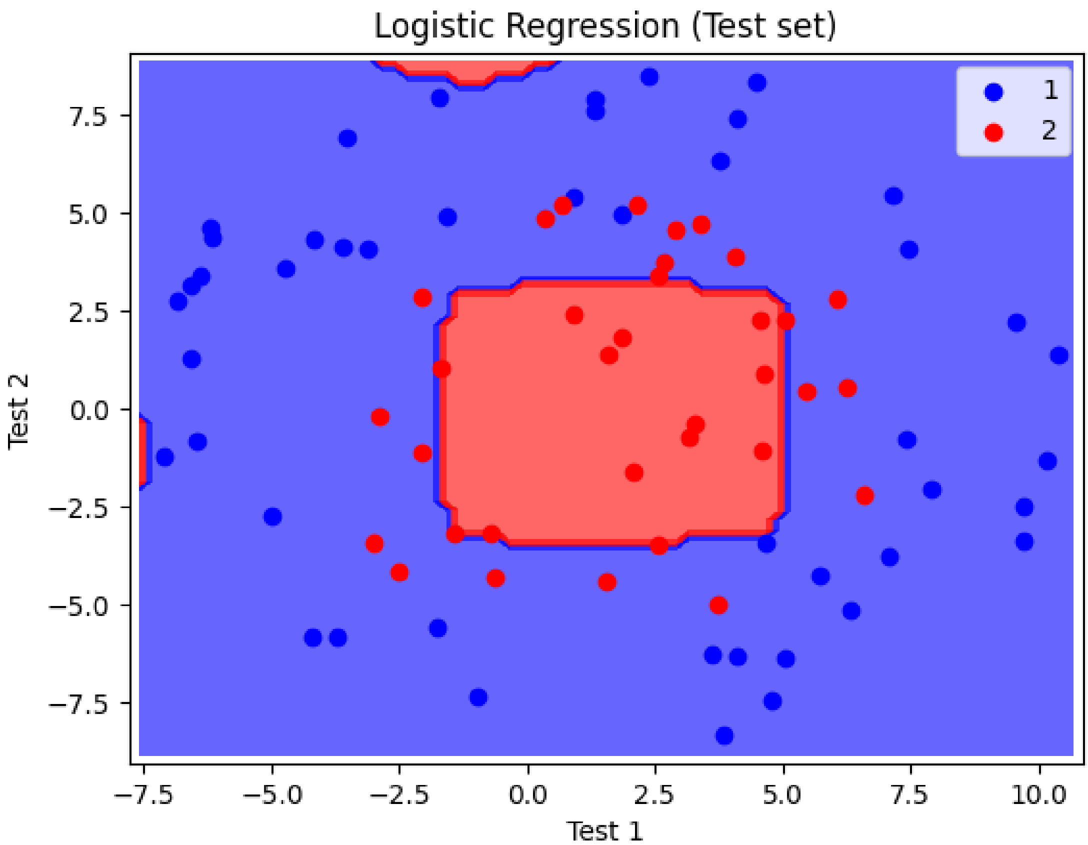


حالت اول درجه ۳۵ :

دقت و ماتریس آشفتگی

```
[[40  6]
 [12 22]]
0.775
```

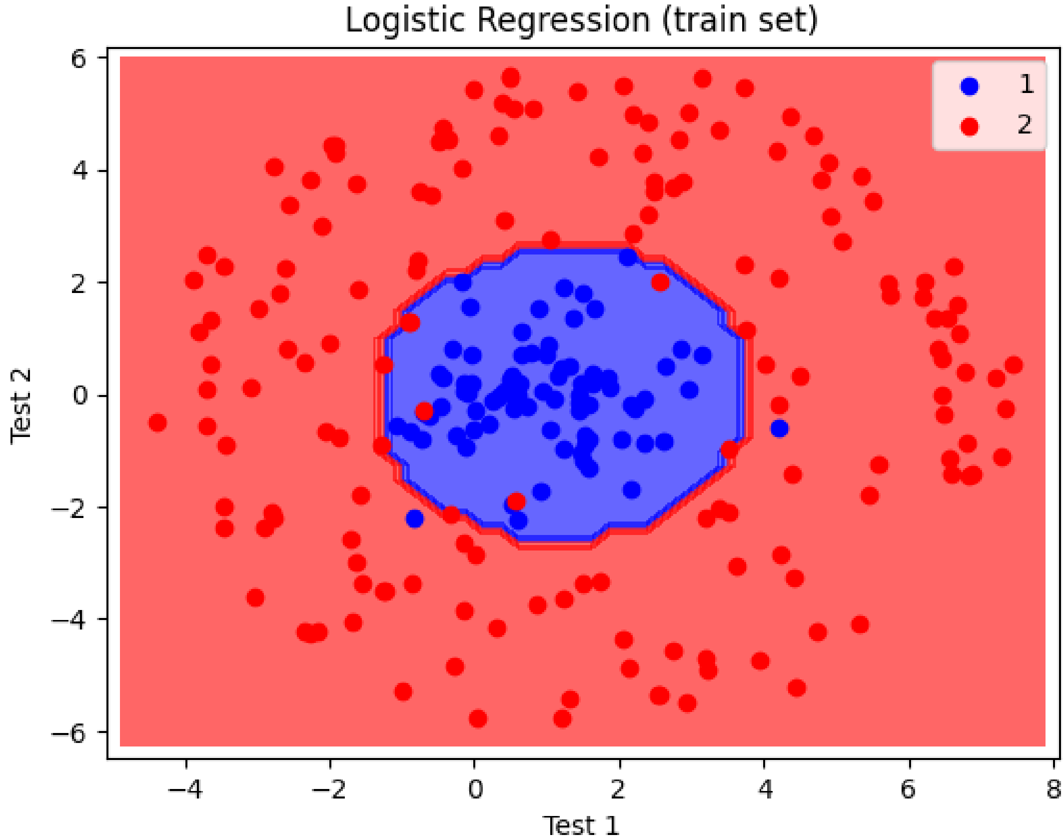


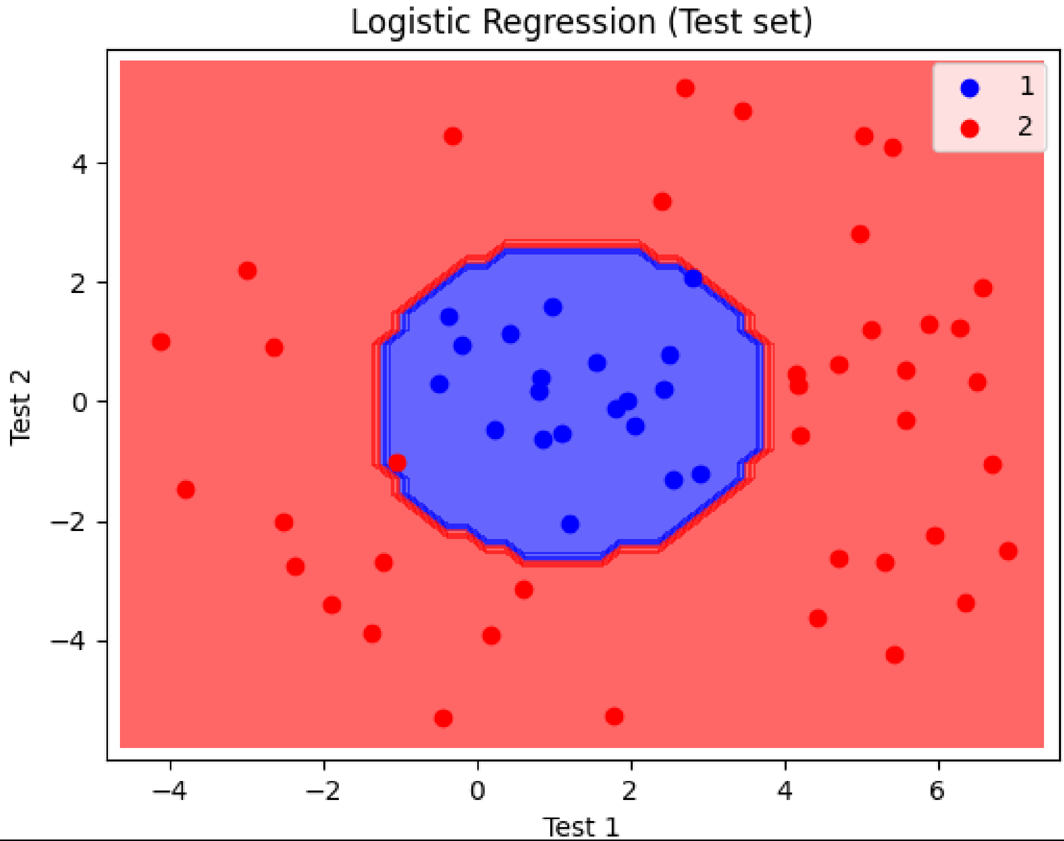


برای حالت دوم درجه 2:

ماتری آشفتگی و دقت:

```
[[20  0]
 [ 1 39]]
0.9833333333333333
```

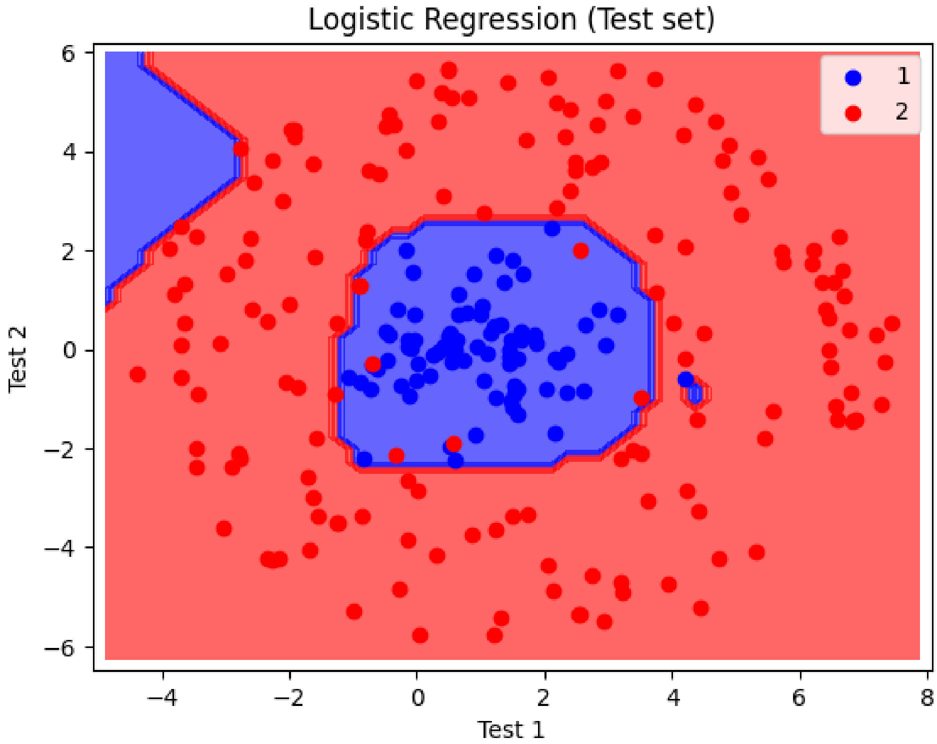


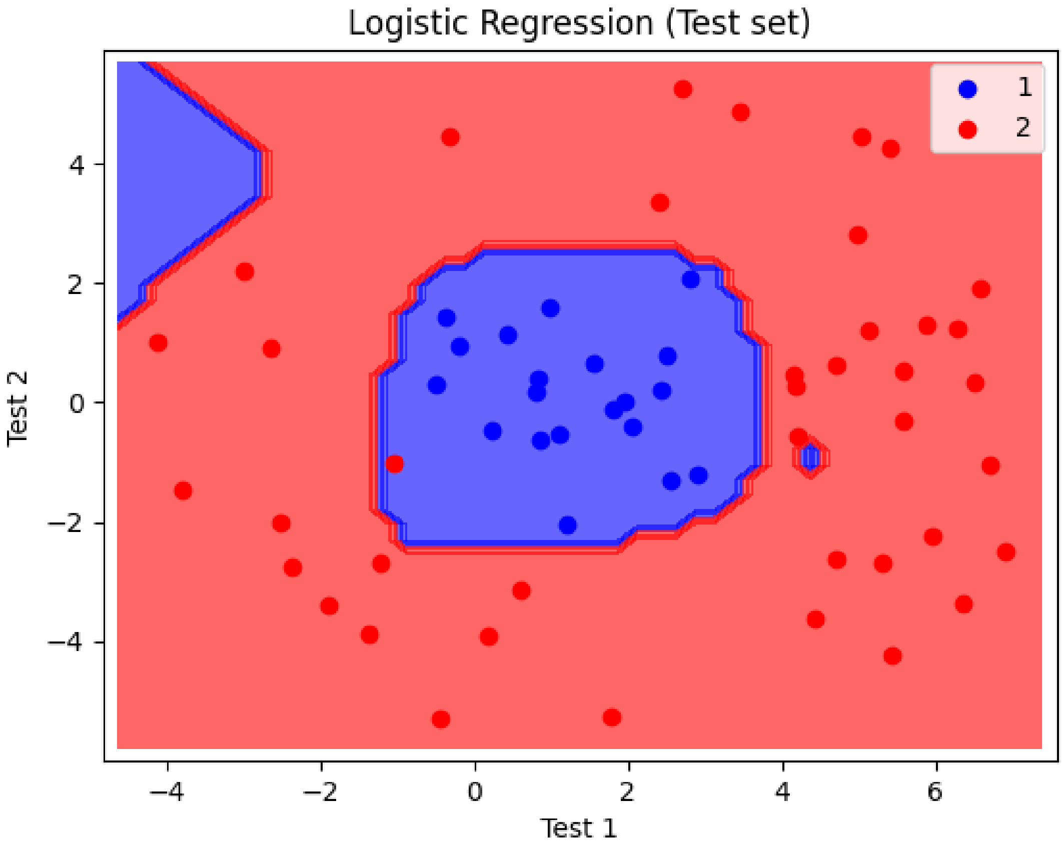


برای حالت دوم درجه 35:

ماتریس آشفتگی و دقت:

```
[[20  0]
 [ 1 39]]
0.9833333333333333
```



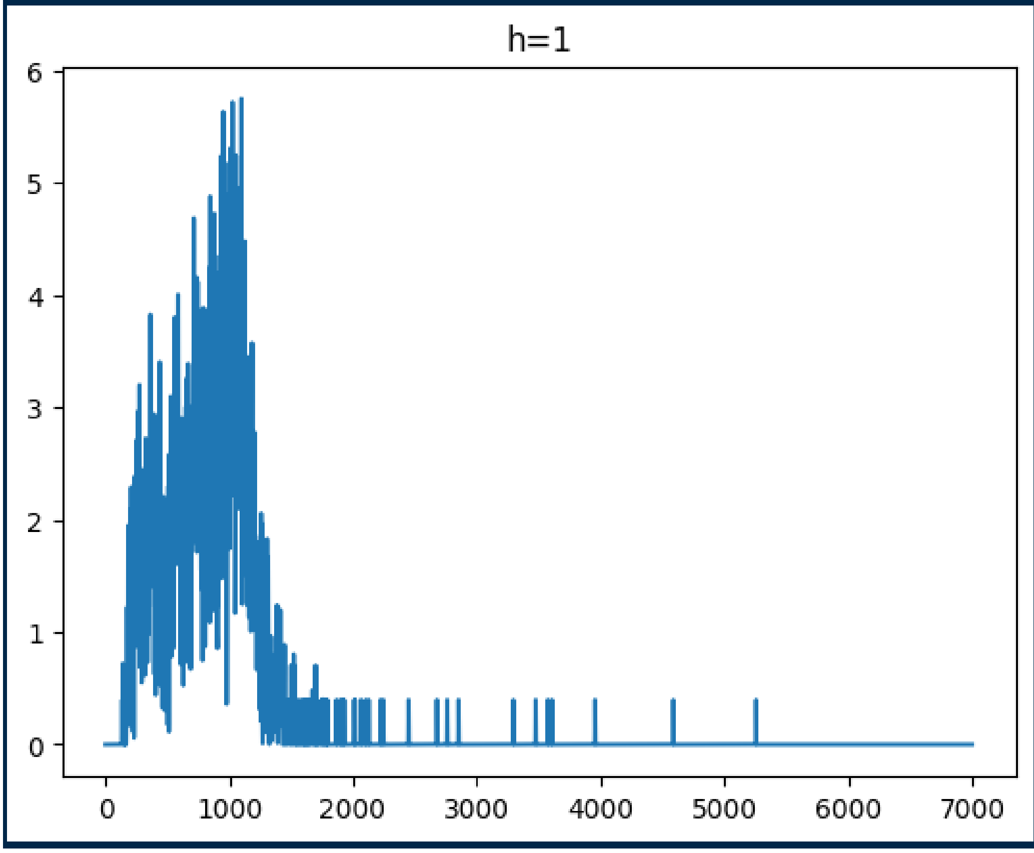


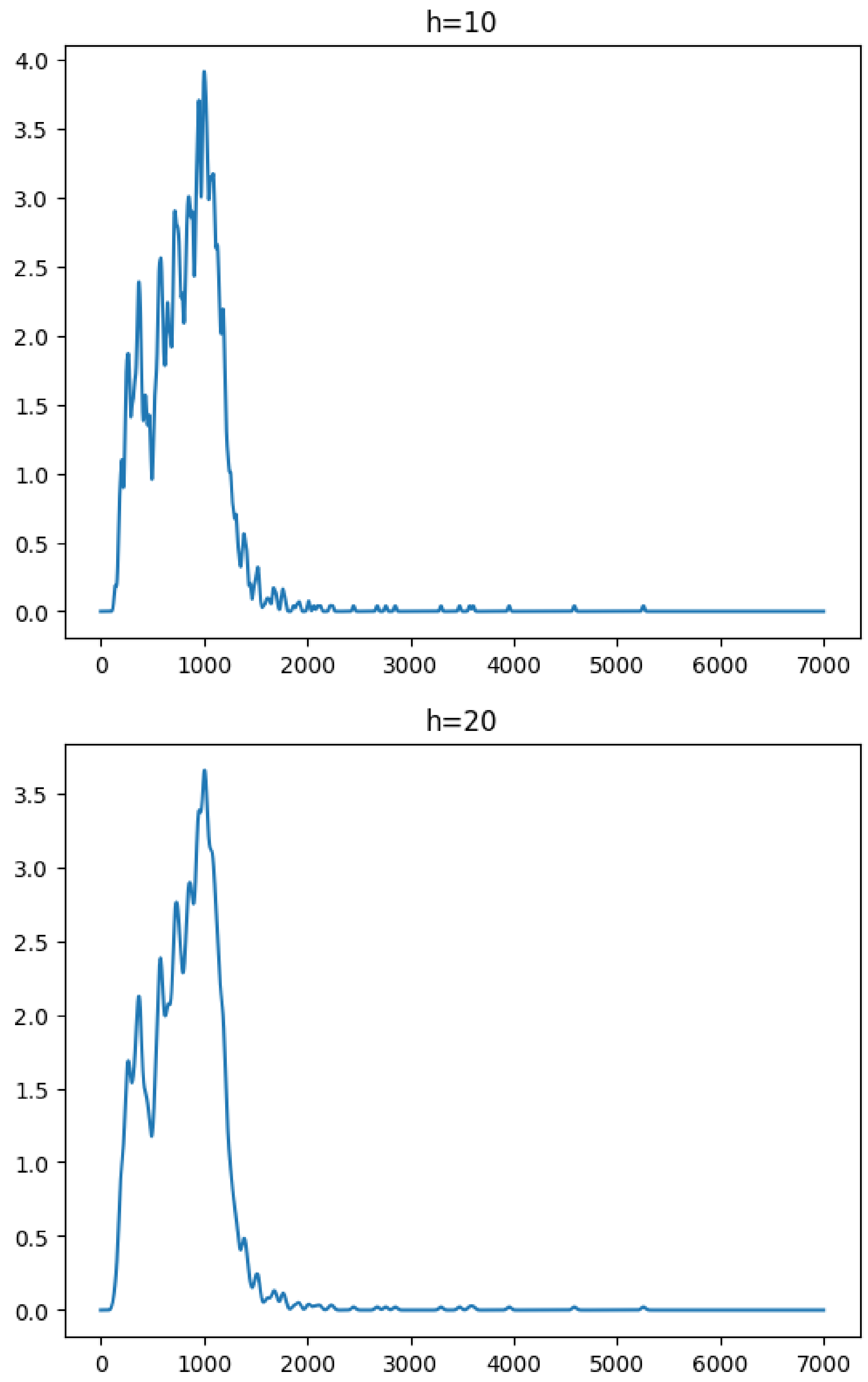
ج

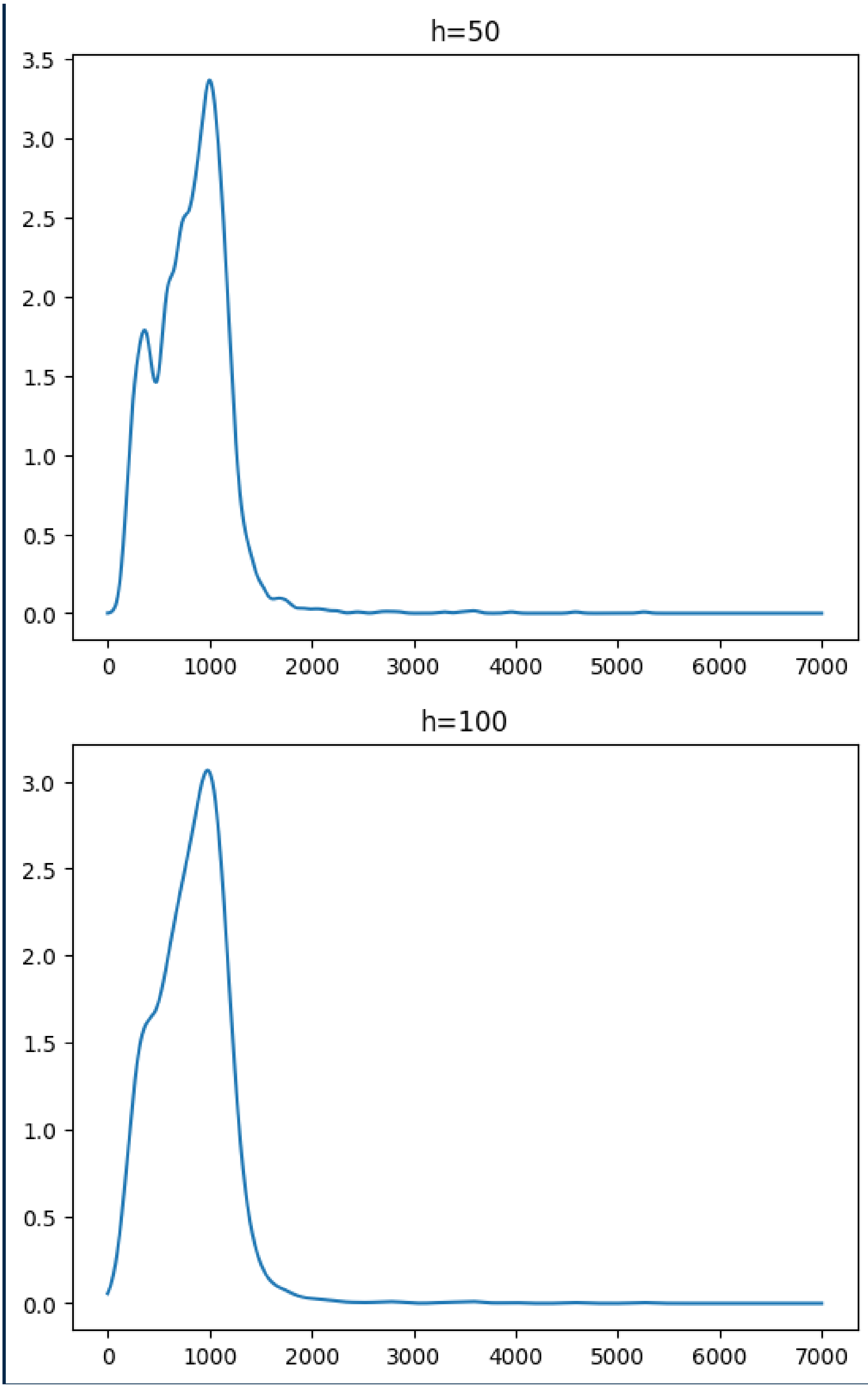
باتوجه به نتایج با افزایش فضای ابعاد مدل اول دقتش کاهش میابد و برای داده های تست خیلی بد عمل می کند

و برای حالت دوم برای دقت اتفاق خاصی نمی افتد اما مدل خیلی پیچیده می شود

۸-




h=10
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
6000
7000
h=20
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
6000
7000


h=50
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
6000
7000
h=100
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
6000
7000

با توجه به اینکه کرنل گوسی داریم ، داریم :

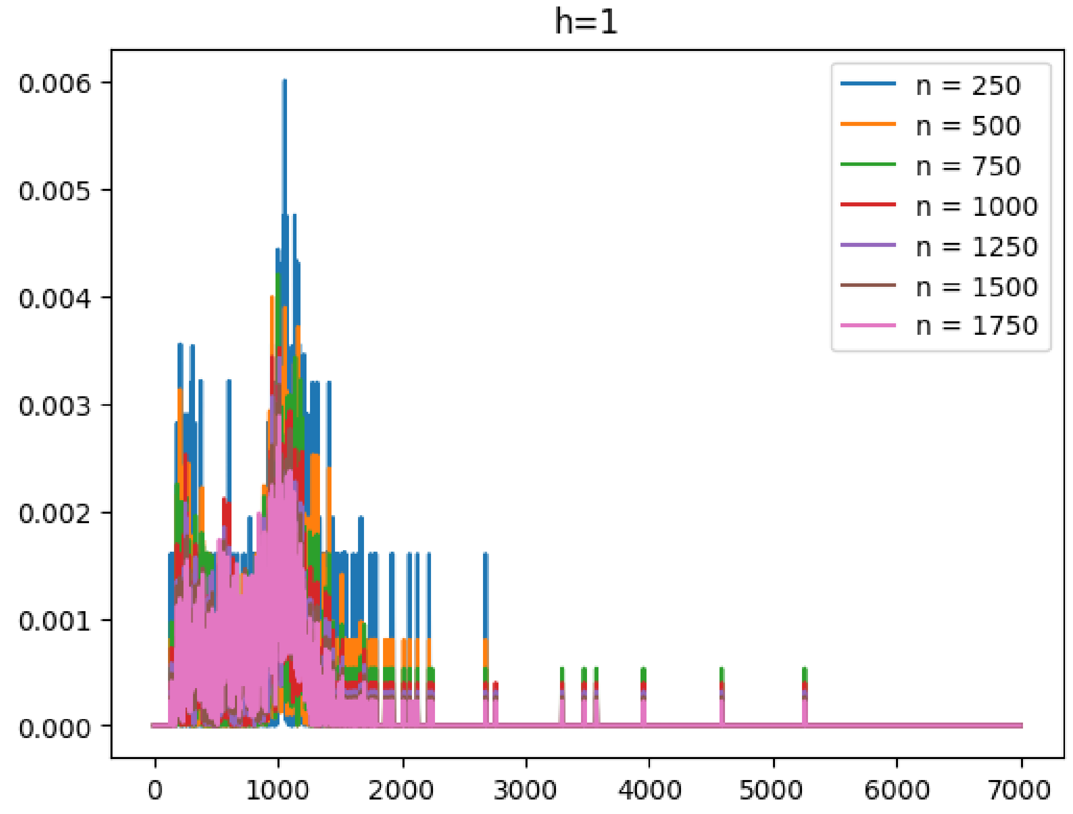
$$p_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^d}\phi\left[\frac{1}{(\sqrt{2\pi})^d h_n^d}\,exp\left[-\frac{1}{2}\left(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h_n}\right)^2\right]\right]$$

با توجه به فرمول واریانس برابرست با توان دو h که نتیجتا با افزایش h داده های بیشتری در پنجره قرار می گرد

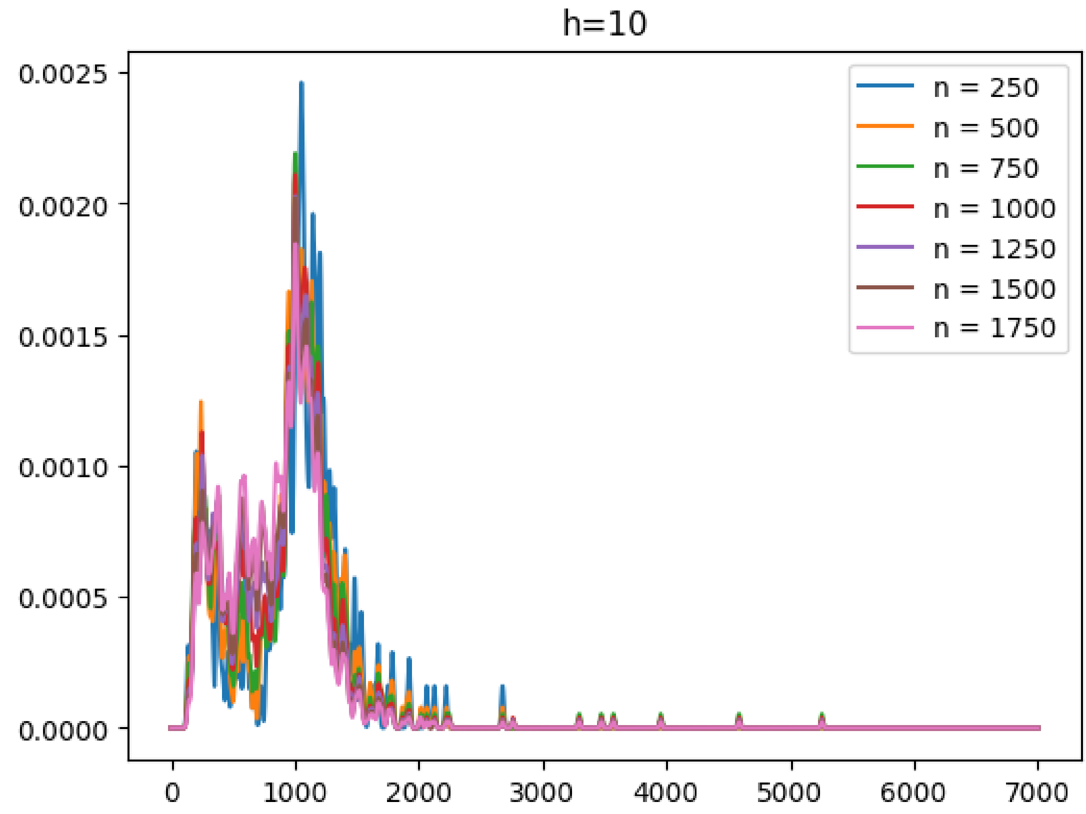
h، نرمی چگالی تخمینی را تعیین می کند. مقدار h بزرگتر منجر به تخمین هموارتر و تعمیم یافته تر می شود، در حالی که مقدار h کوچکتر منجر به تخمین چگالی دقیق تر و نوسانی تر می شود.

و حالت مطلوب برا ما بین دو مقدار است.

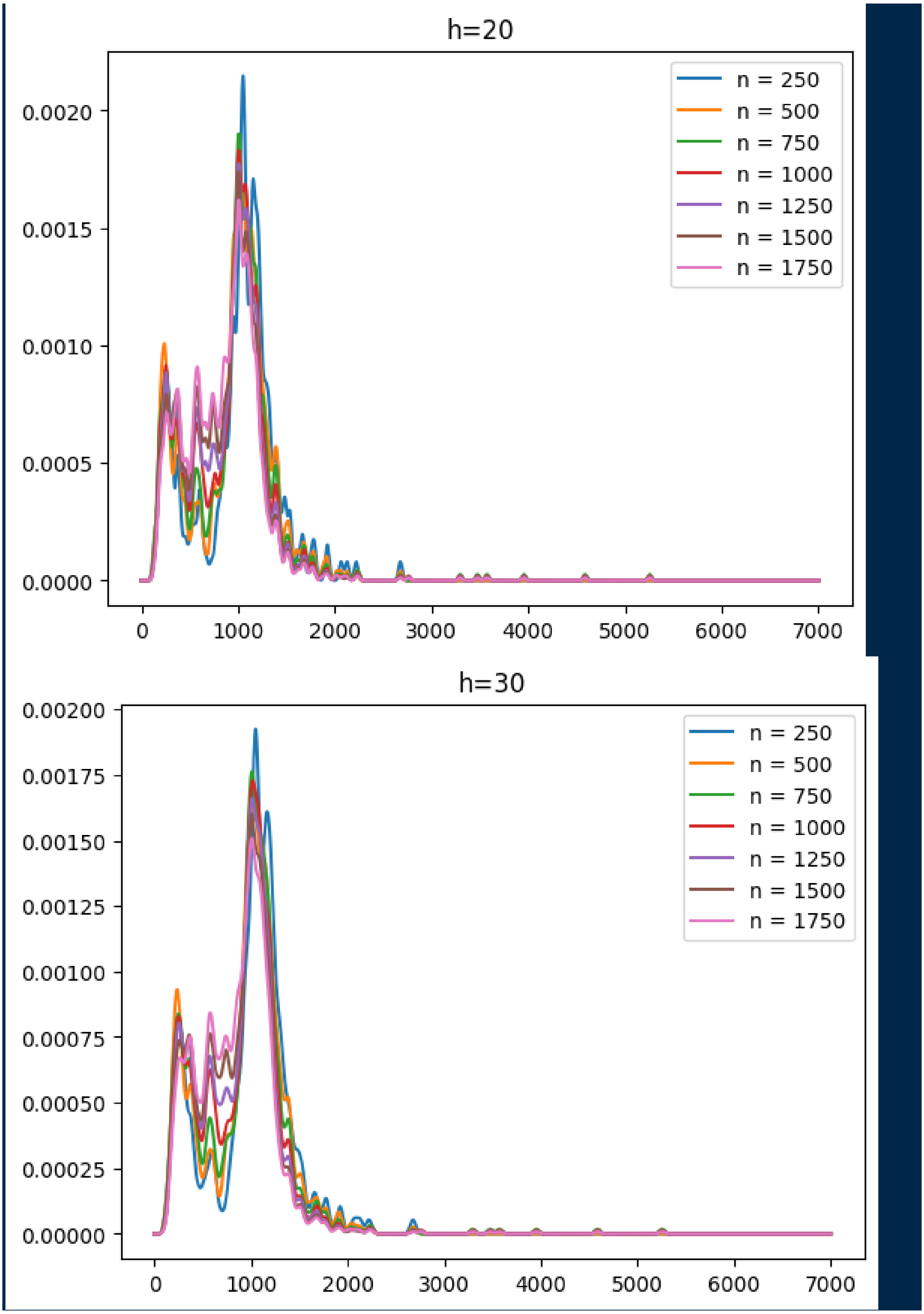
ج

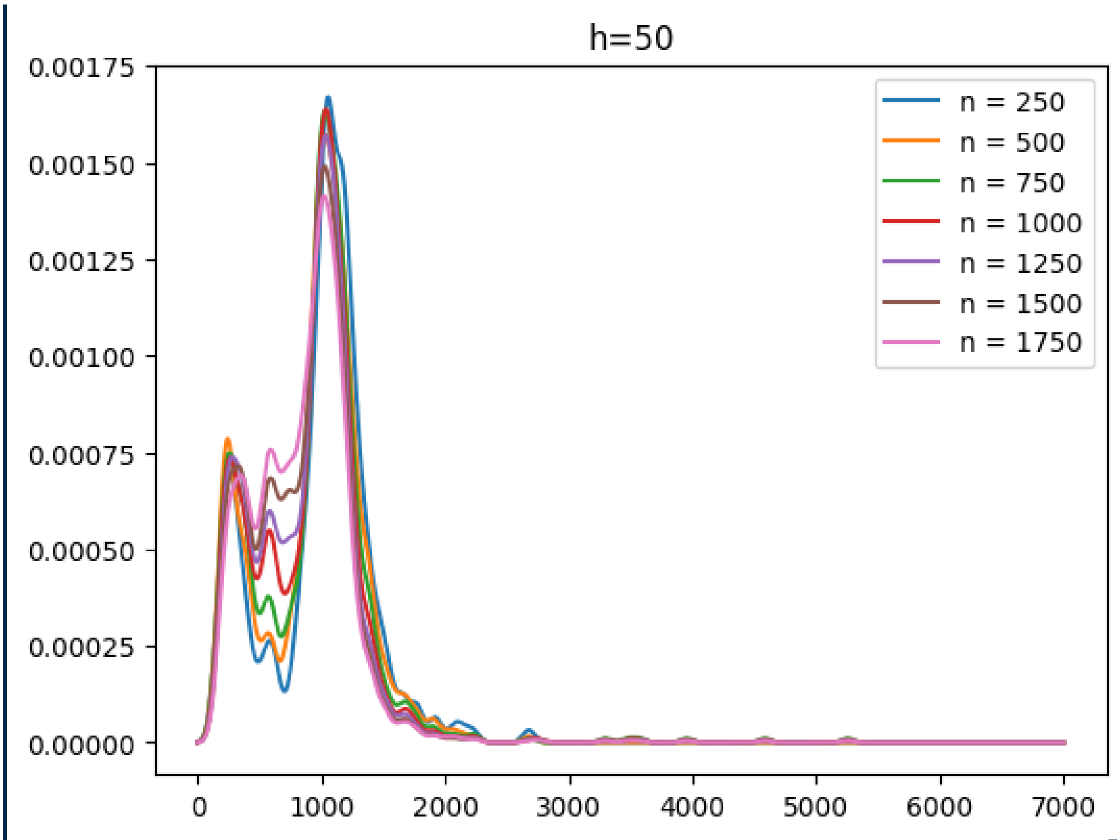


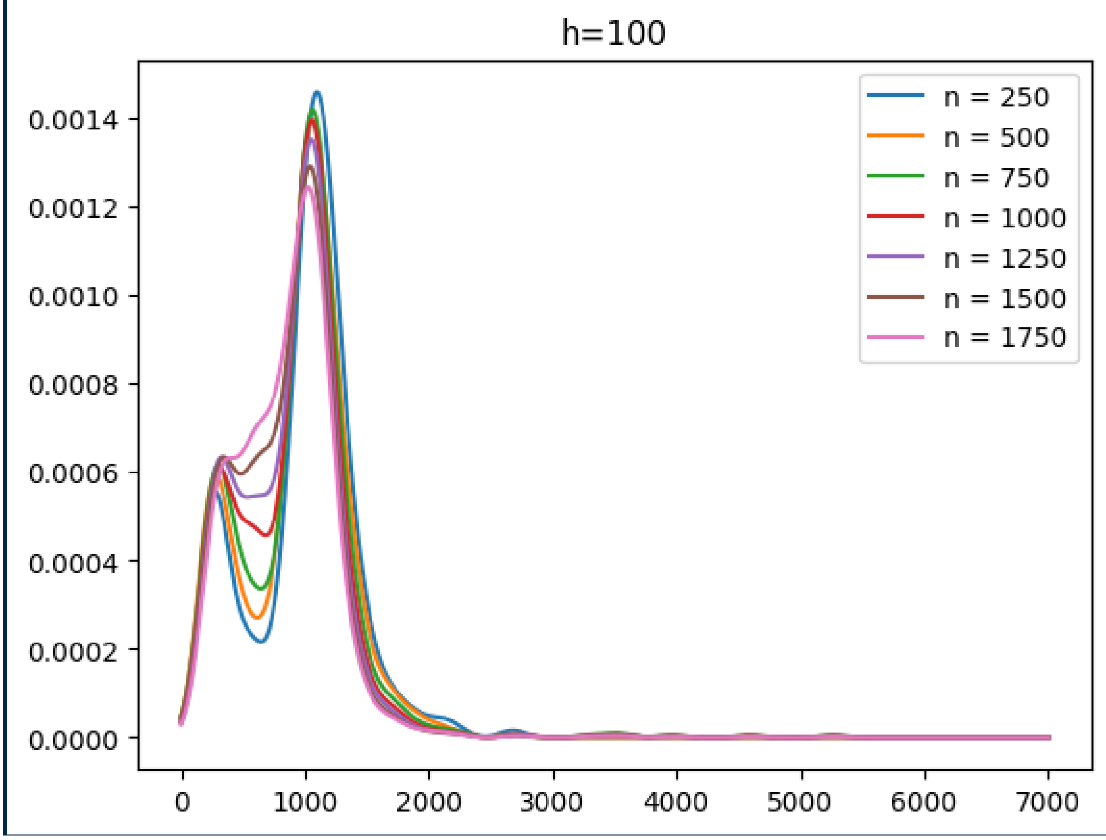
h=1
n = 250
n = 500
n = 750
n = 1000
n = 1250
n = 1500
n = 1750
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0 1000 2000 3000 4000 5000 6000 7000



h=10
n = 250
n = 500
n = 750
n = 1000
n = 1250
n = 1500
n = 1750
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0 1000 2000 3000 4000 5000 6000 7000


h=20
n = 250
n = 500
n = 750
n = 1000
n = 1250
n = 1500
n = 1750
0.0020
0.0015
0.0010
0.0005
0.0000
0 1000 2000 3000 4000 5000 6000 7000
h=30
n = 250
n = 500
n = 750
n = 1000
n = 1250
n = 1500
n = 1750
0.00200
0.00175
0.00150
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000
0 1000 2000 3000 4000 5000 6000 7000


h=50
0.00175
0.00150
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000
0
1000
2000
3000
4000
5000
6000
7000
n = 250
n = 500
n = 750
n = 1000
n = 1250
n = 1500
n = 1750


h=100
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
1000
2000
3000
4000
5000
6000
7000
n = 250
n = 500
n = 750
n = 1000
n = 1250
n = 1500
n = 1750

با توجه به داده ها هرچه قدر تعداد را بیشتر کنیم نمودار نرم تر می شود که در هر نمودار می توان آن را دید و مقایسه کرد

د

با توجه به نگاه کردن می فهمیم شکل نمودار ها یکسان هست و به خوبی تونستیم بدون استفاده از کتابخانه تخمین پارازن رو به خوبی پیاده کنیم.