



دانشگاه یزد

گروه مهندسی کامپیوتر

رشته تحصیلی: هوش مصنوعی و رباتیکز

نام درس: تشخیص الگو

گزارش تمرین عملی ۳

استاد مربوطه: دکتر مهدی یزدیان

تهیه کننده: حمیدرضا نادمی

پاییز ۱۳۹۷

Part1: Parametric Methods:

- b) Generate N samples with N=10 from this density;

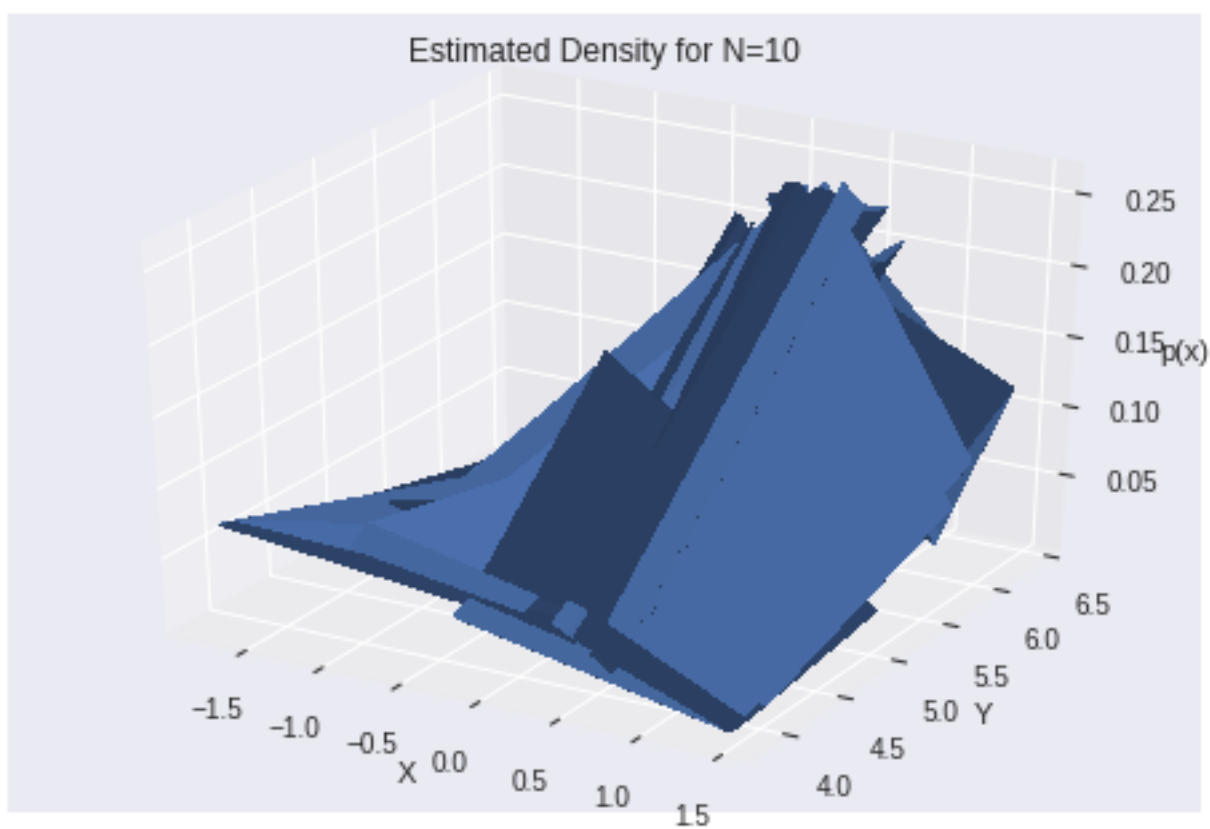
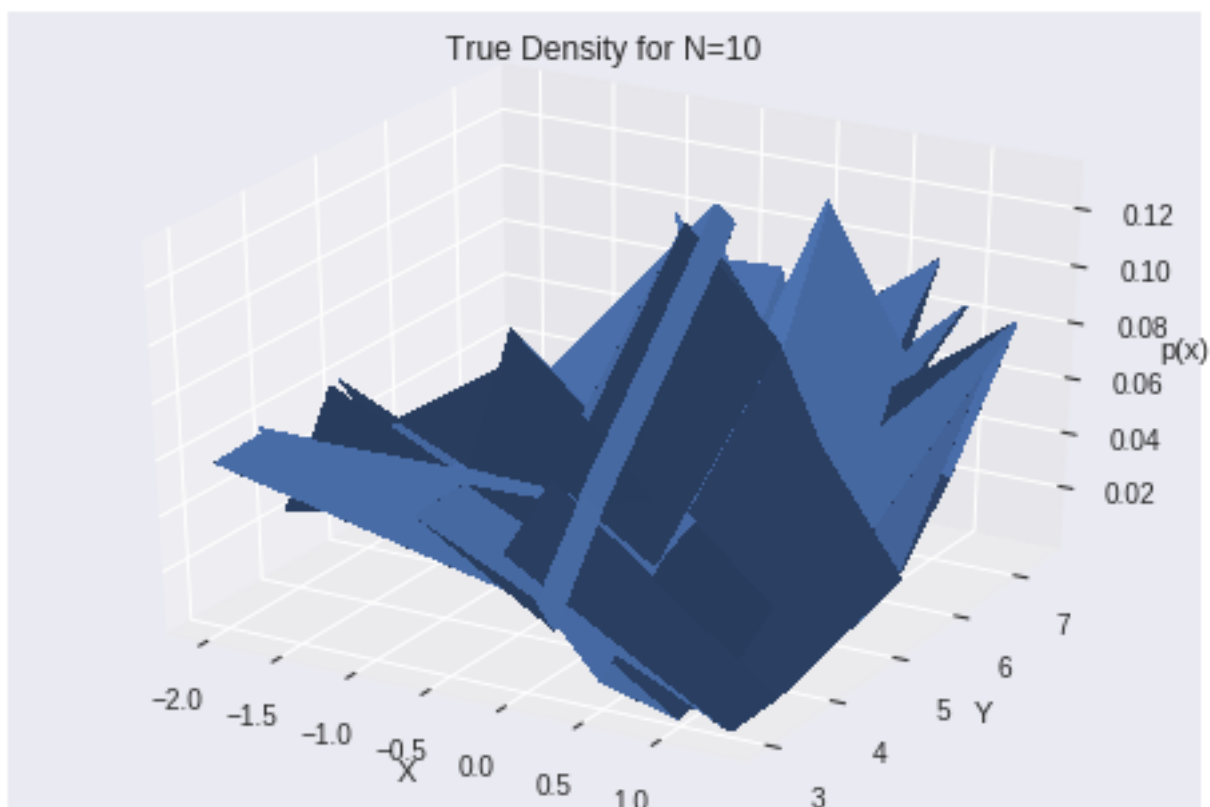
Now, we are going to estimate this density using an ML estimate. For this, we have to suppose a specific distribution for these samples. Based on the above simulation, we assume the density of the samples is Gaussian;

Estimate the parameter of the Gaussian using ML estimate; report the estimated parameters and compare them with the true values.

	Mean vector	Covariance matrix
True Value	$[0, 5]$	$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$
Estimated Value for N=10	$[-0.080, 4.746]$	$\begin{bmatrix} 0.862 & 0.981 \\ 0.981 & 2.203 \end{bmatrix}$

مقادیر تخمینی بدست آمده حدودا با مقادیر واقعی برابر می باشد.

- c) Plot the true density and the estimated density.



- d) Repeat Section a and b for 20 times; Compute and report the bias and variance; Compare and discuss about the results.

	Mean	Covariance matrix
Bias	[0.05, 0.12]	$[-0.24, -0.32]$
Variance	[0.08, 0.23]	[0.15, 0.57]

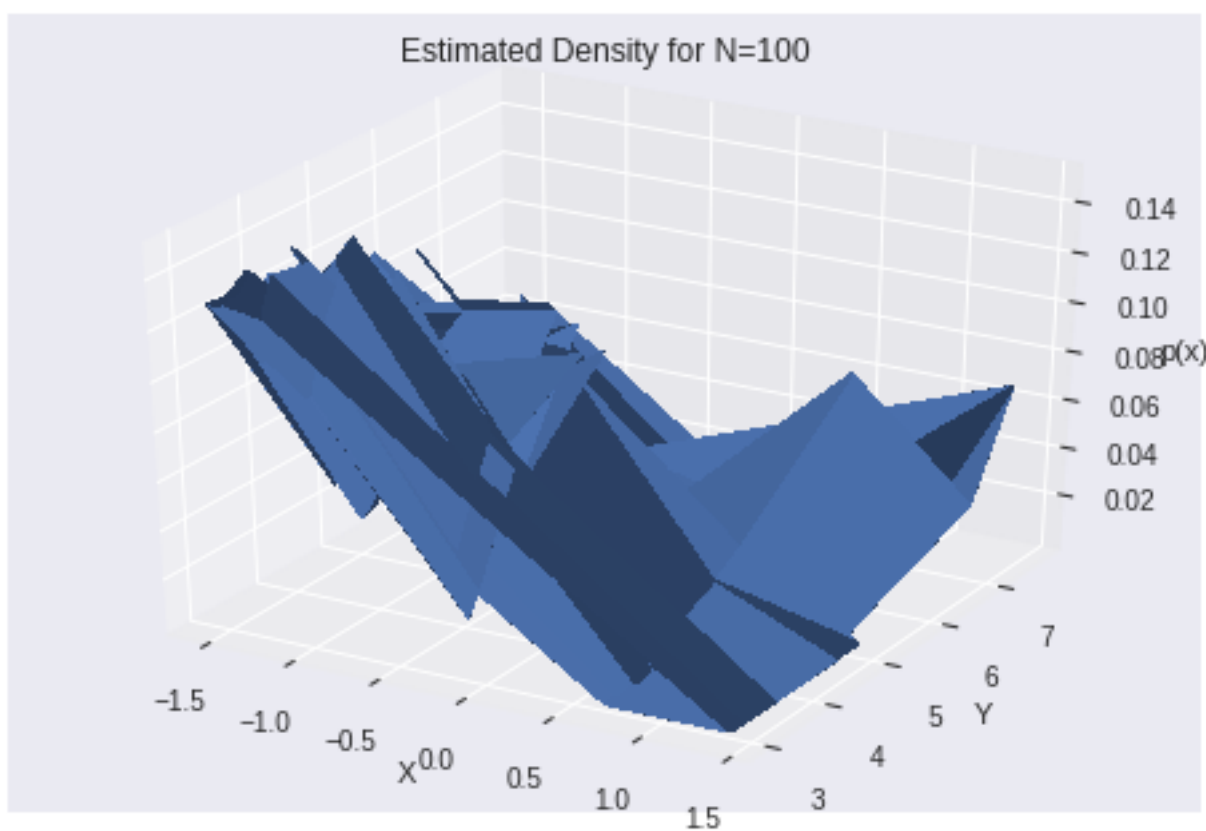
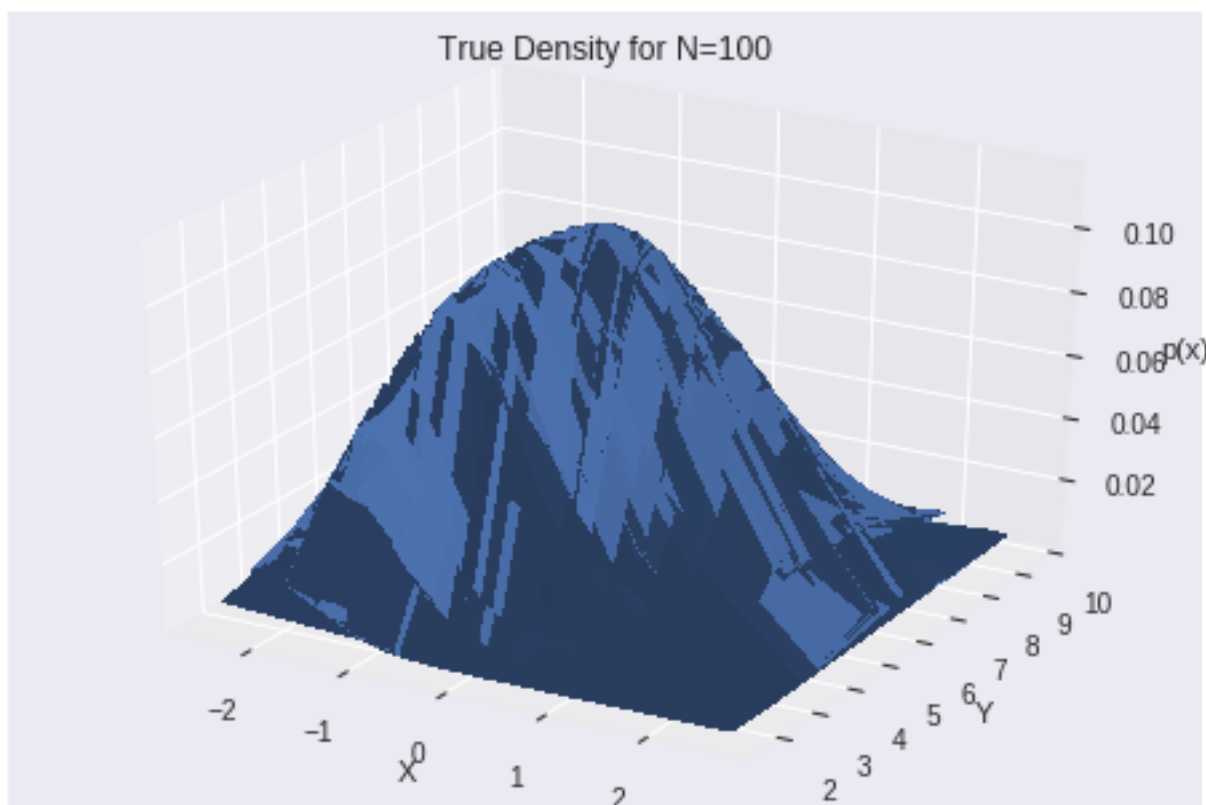
با رفتن به لینک زیر می توانید خروجی آنلاین کدها را در محیط Colab Research Google مشاهده کنید.

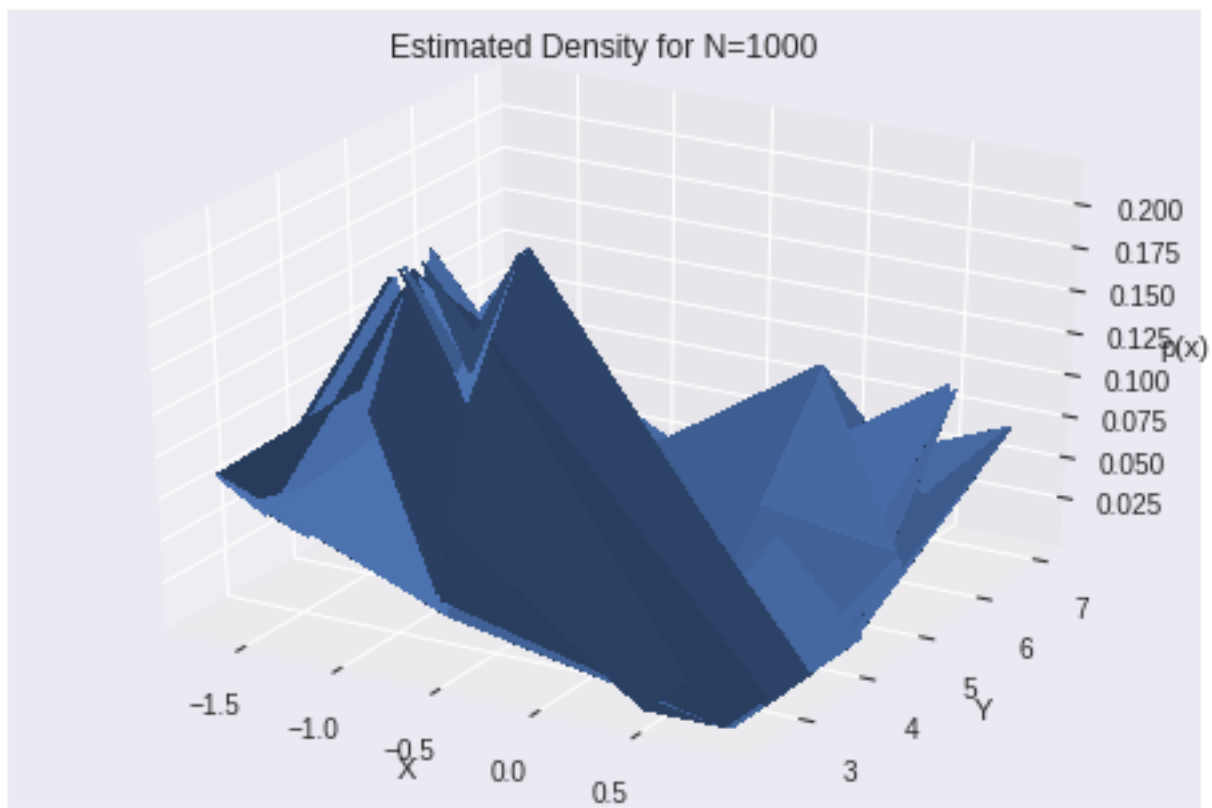
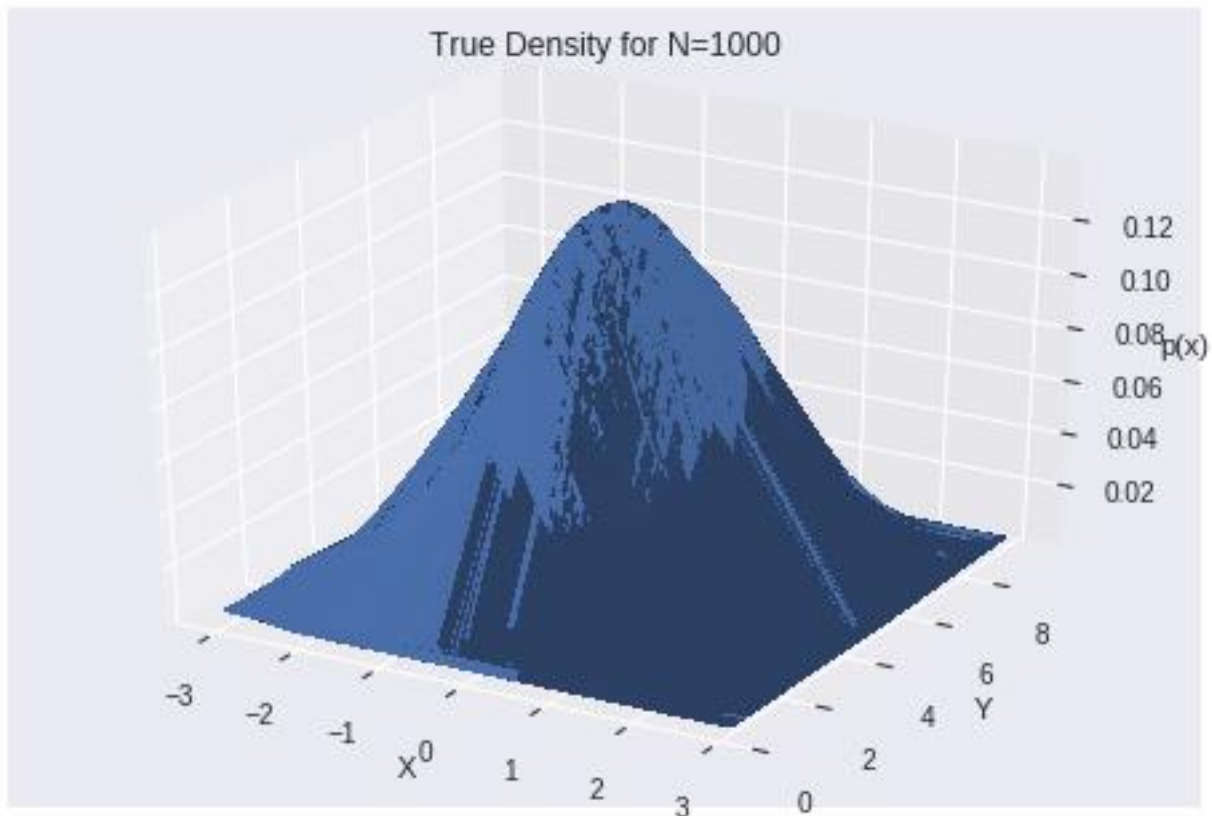
https://colab.research.google.com/drive/1_oBapPiQu7GMSZAv0w5YiOAotbxQWLyn

- e) Repeat Section a to d for $N=100$, $N=1000$. Discuss the effect of sample size in the performance of the estimation.

	Mean vector	Covariance matrix
True Value	[0, 5]	$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$
Estimated Value for $N=100$	[0.028, 5.022]	$\begin{bmatrix} 1.159 & 0.999 \\ 0.999 & 1.981 \end{bmatrix}$
Estimated Value for $N=1000$	[0.035, 5.042]	$\begin{bmatrix} 1.023 & 1.077 \\ 1.077 & 2.150 \end{bmatrix}$

	Mean		Covariance matrix	
	Bias	Variance	Bias	Variance
$N=100$	[0.023, 0.049]	[0.005, 0.015]	[0.020, 0.117]	[0.016, 0.089]
$N=1000$	$[-9.945, -0.006]$	[0.001, 0.001]	$[-0.001, 0.013]$	[0.001, 0.009]





هرچه تعداد sample تولید شده بیشتر باشد، بردار میانه و ماتریس کوواریانس تخمین زده شده به مقدار واقعی نزدیکتر است و نمودار تولید شده به شکل گوسی نزدیکتر می شود.

با رفتن به لینک زیر می توانید خروجی آنلاین کدها را در محیط Colab Research Google مشاهده کنید.

<https://colab.research.google.com/drive/1ljsDZiSzK-qFvLP2x09wzH47rnxLBtc>

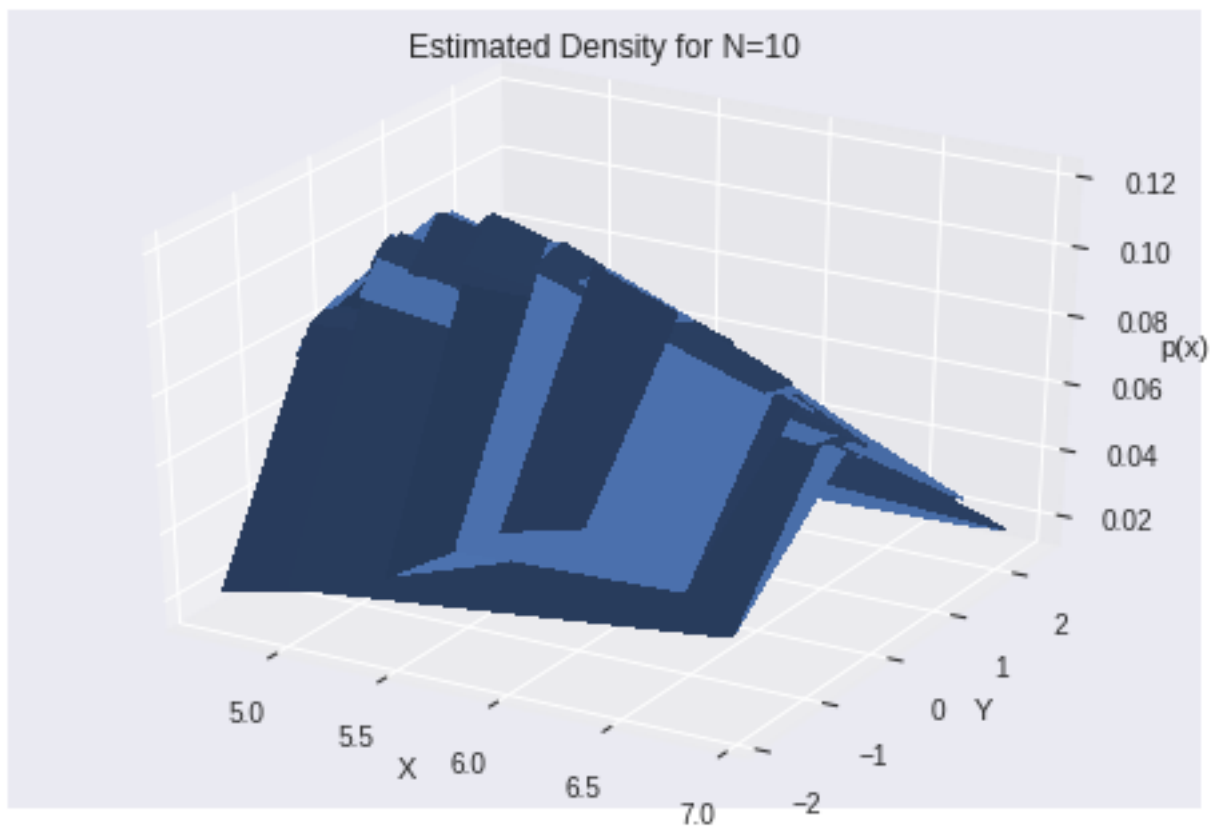
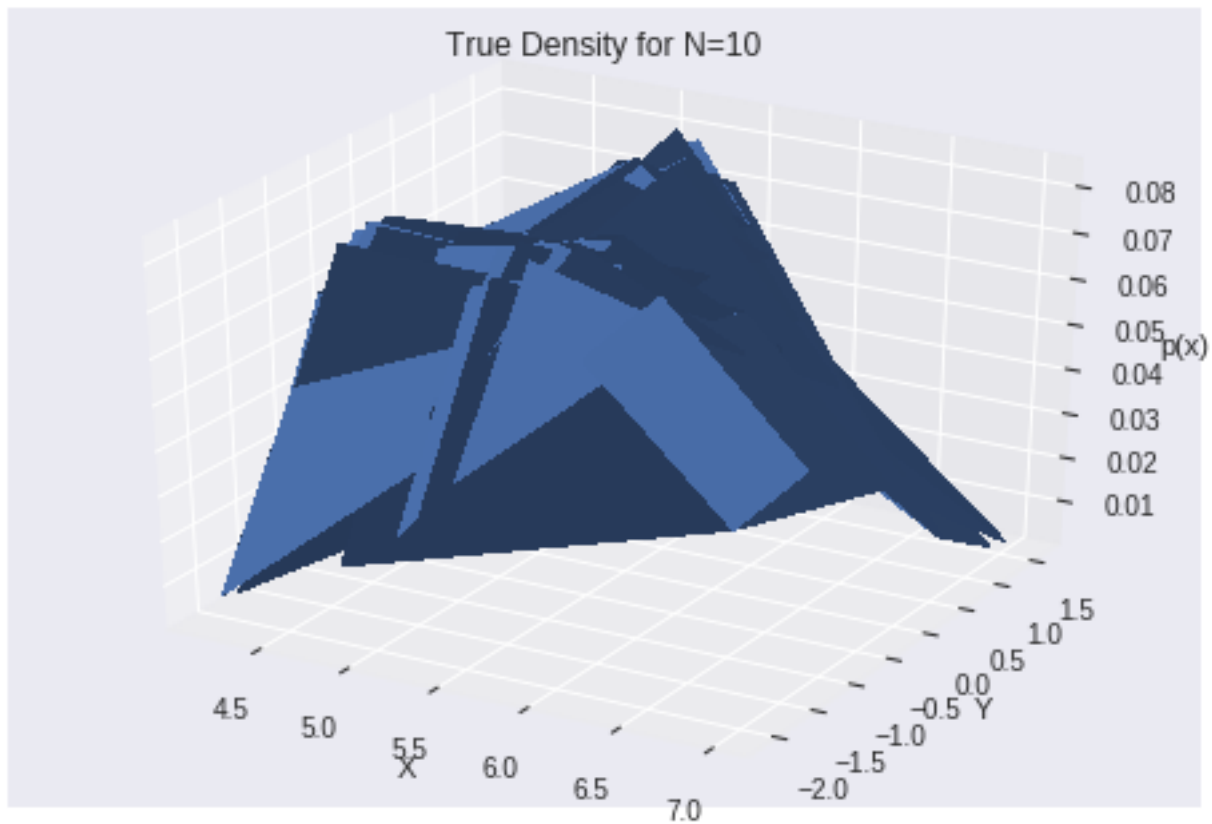
- f) Repeat all the previous sections with the following mean and covariance; Compare the results with the previous ones.

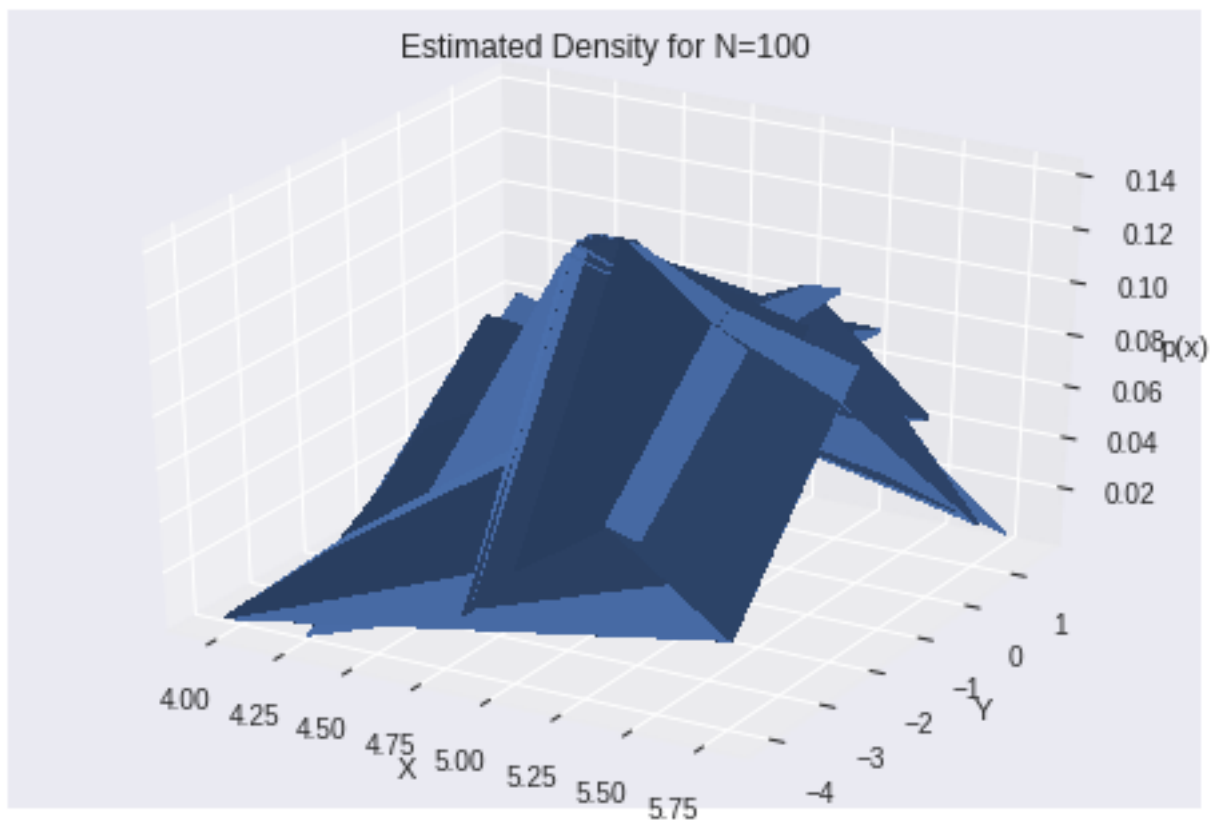
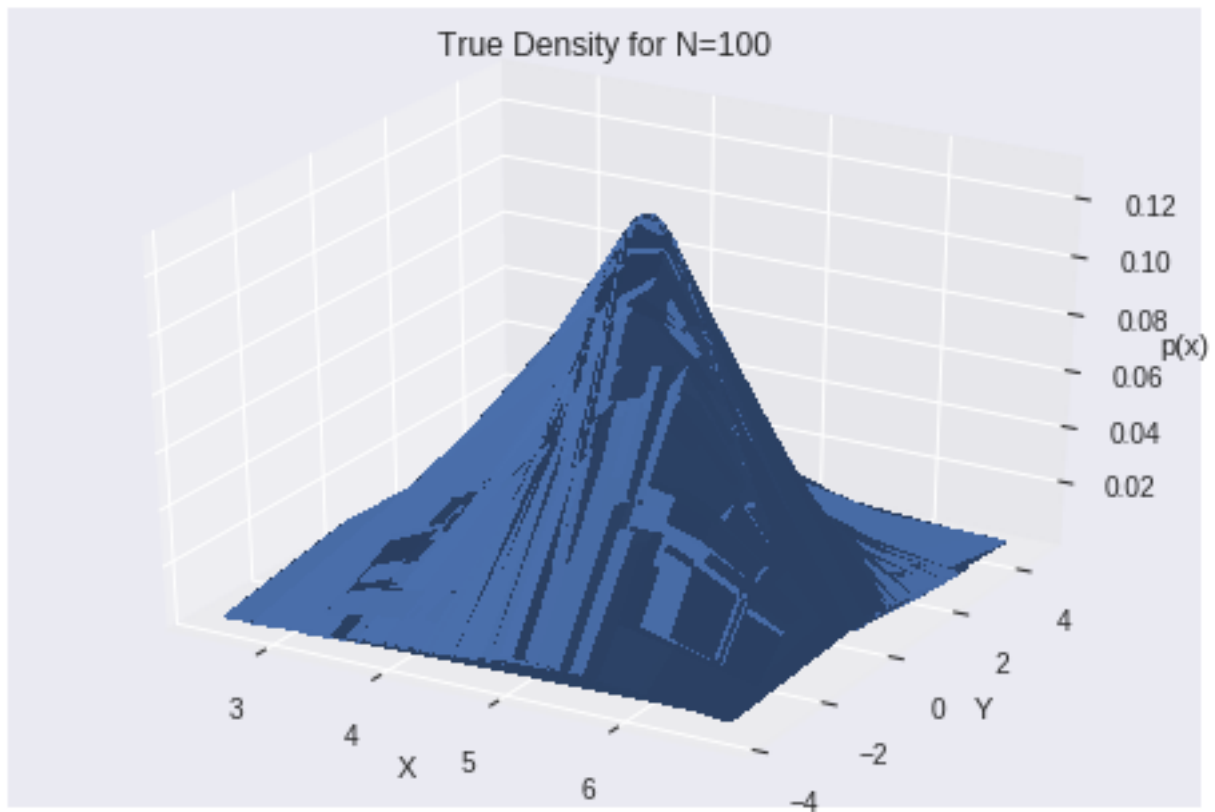
$$\mu_2 = [5 \ 0]^T$$

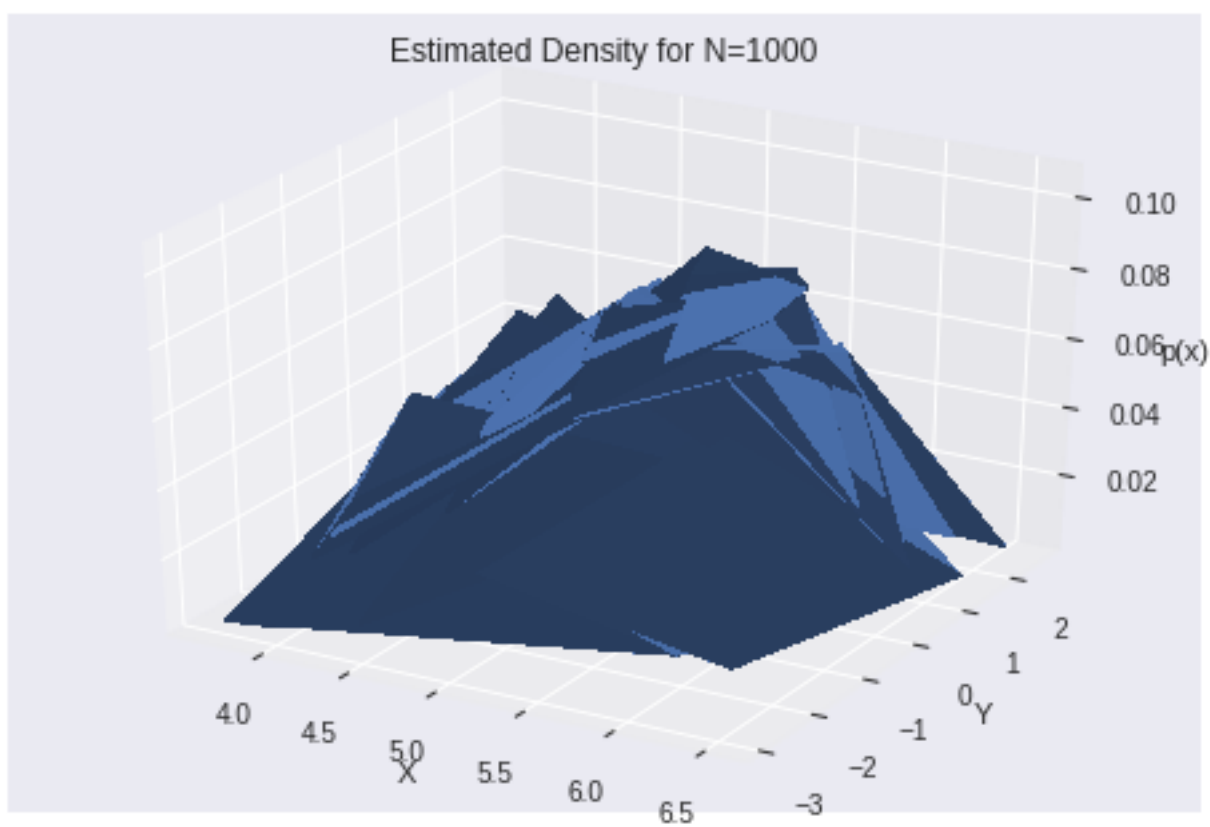
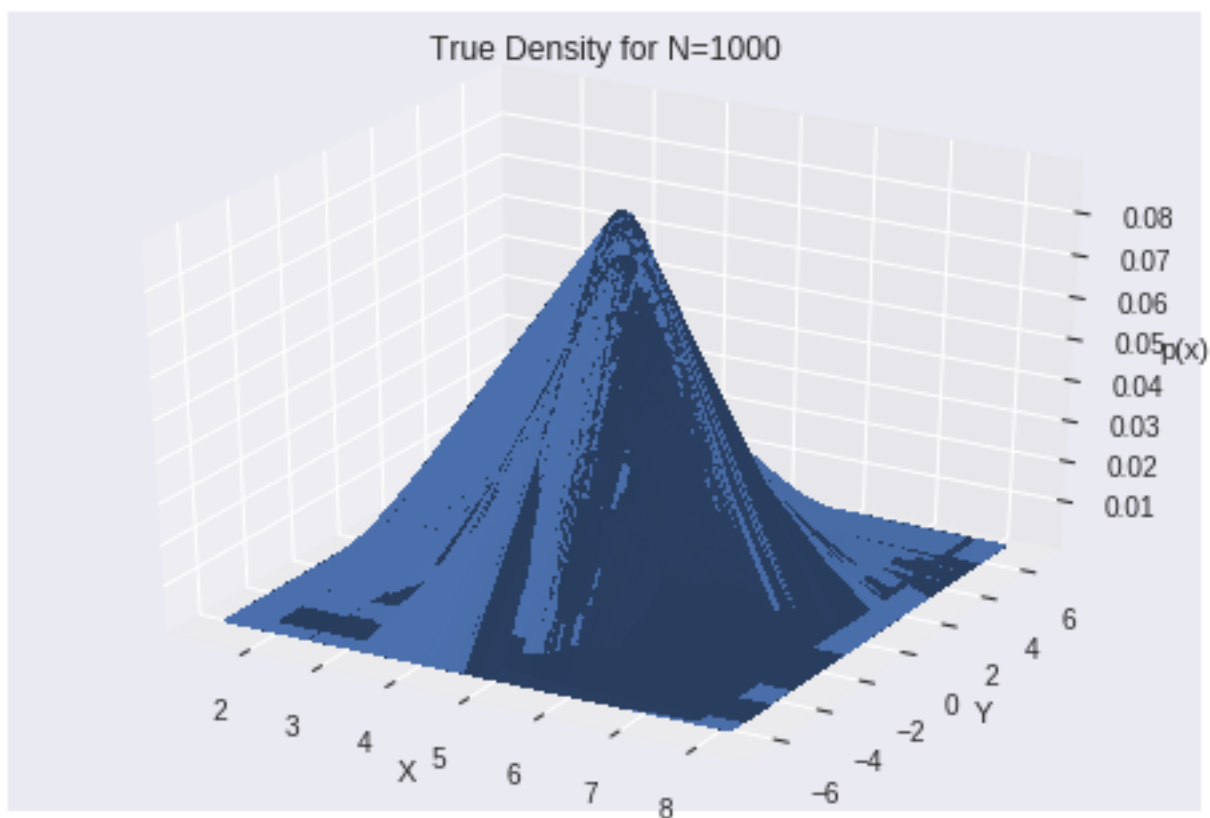
$$\Sigma_2 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

	Mean vector	Covariance matrix
True Value	[5,0]	$\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$
Estimated Value for N=10	[5.492, -0.315]	$\begin{bmatrix} 0.758 & -0.421 \\ -0.421 & 1.373 \end{bmatrix}$
Estimated Value for N=100	[4.849, 0.131]	$\begin{bmatrix} 0.754 & -0.489 \\ -0.489 & 2.507 \end{bmatrix}$
Estimated Value for N=1000	[4.964, 0.066]	$\begin{bmatrix} 0.947 & -1.025 \\ -1.025 & 3.838 \end{bmatrix}$

	Mean		Covariance matrix	
	Bias	Variance	Bias	Variance
N=10	[5.062, -4.881]	[0.098, 0.222]	[-0.316, 1.728]	[0.055, 2.627]
N=100	[5.000, -5.047]	[0.007, 0.034]	[-0.017, 1.953]	[0.017, 0.452]
N=1000	[4.996, -5.027]	[0.000, 0.002]	[-0.005, 1.968]	[0.001, 0.036]







با توجه به ماتریس کوواریانس متوجه میشویم که میزان پخش شدن داده‌ها در محور Y ۴ برابر محور X است و اگر توزیع گوسی شبیه به کلاه برعکس در نظر بگیریم، نوک کلاه مایل به محور Y است (برای پلات های $N=100$ و $N=1000$ قابل مشاهده است) و این مایل بودن نسبت به نمودارهای قسمت e دو برابر است و در جهت برعکس است زیرا ضریب ماتریس کوواریانس منفی است. با رفتن به لینک زیر می توانید خروجی آنلاین کدها را در محیط Colab Research Google مشاهده کنید.

https://colab.research.google.com/drive/1E7FepJ9EO_FsfB4myanxPp2bN3RFQ7hE#scrollTo=PR9sH9PK4y5S

B) [optional] MAP Parameter Estimate

Consider a Normal Distribution $N(\mu_1, \Sigma_1)$ as defined in Section A. Assume Σ_1 is known and μ_1 is unknown; Estimate the mean parameter of this density using MAP estimate. Suppose the prior of the mean is as follows:

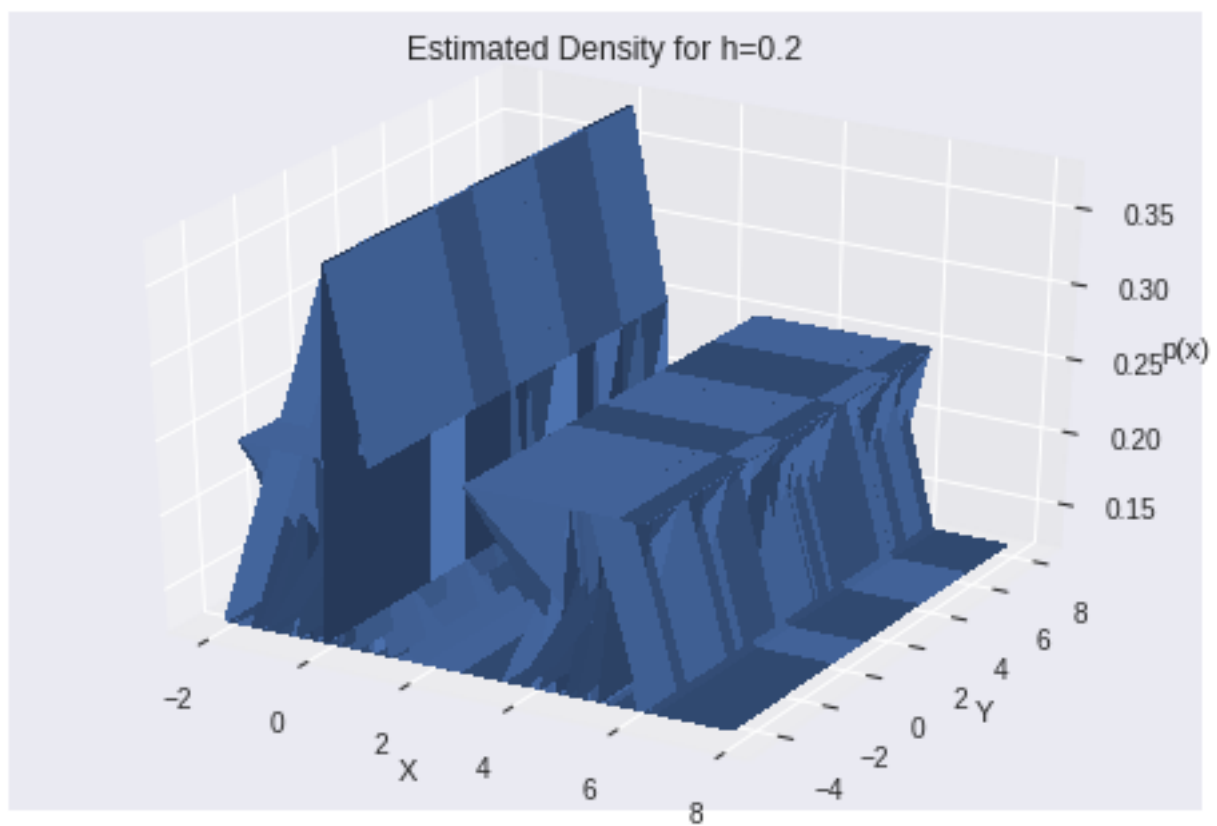
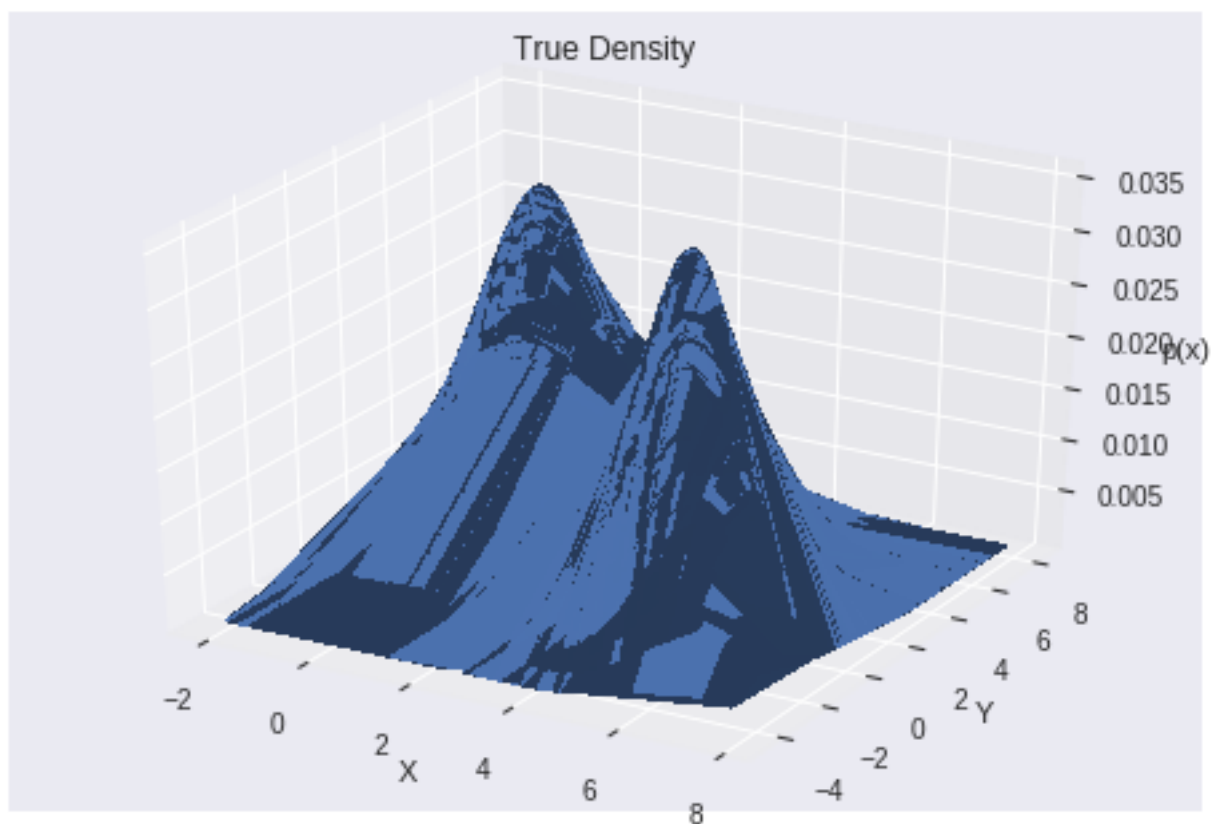
$$P(\mu) = N(\mu_\mu, \Sigma_\mu) = \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix} \right)$$

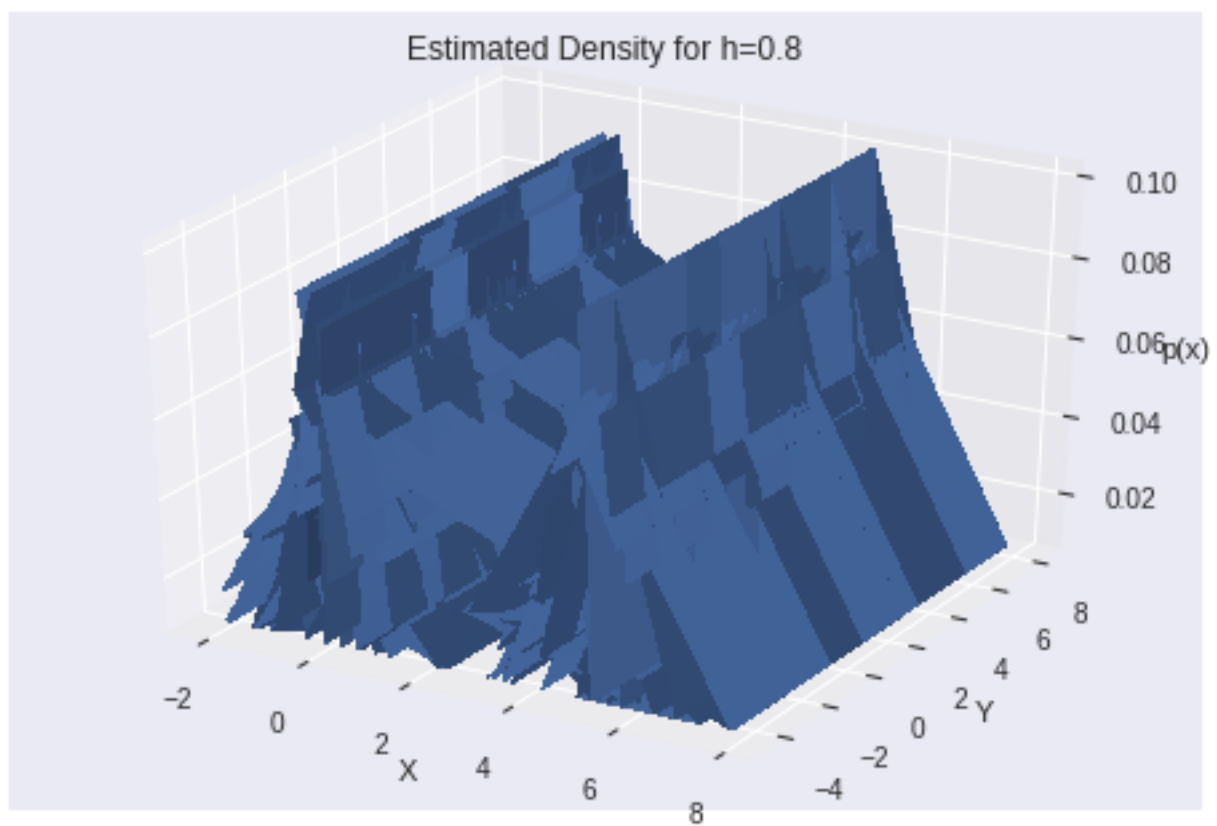
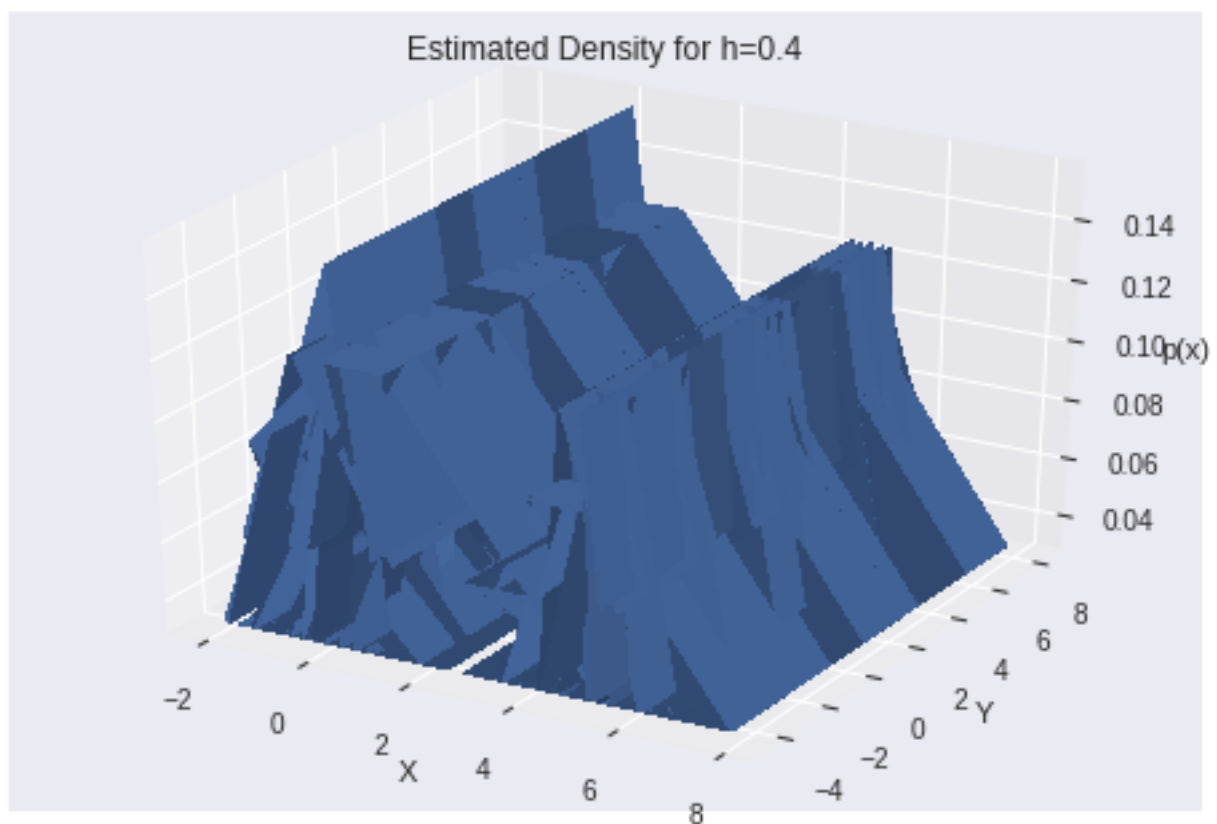
Note: repeat this section for $N=10, 100, 1000$

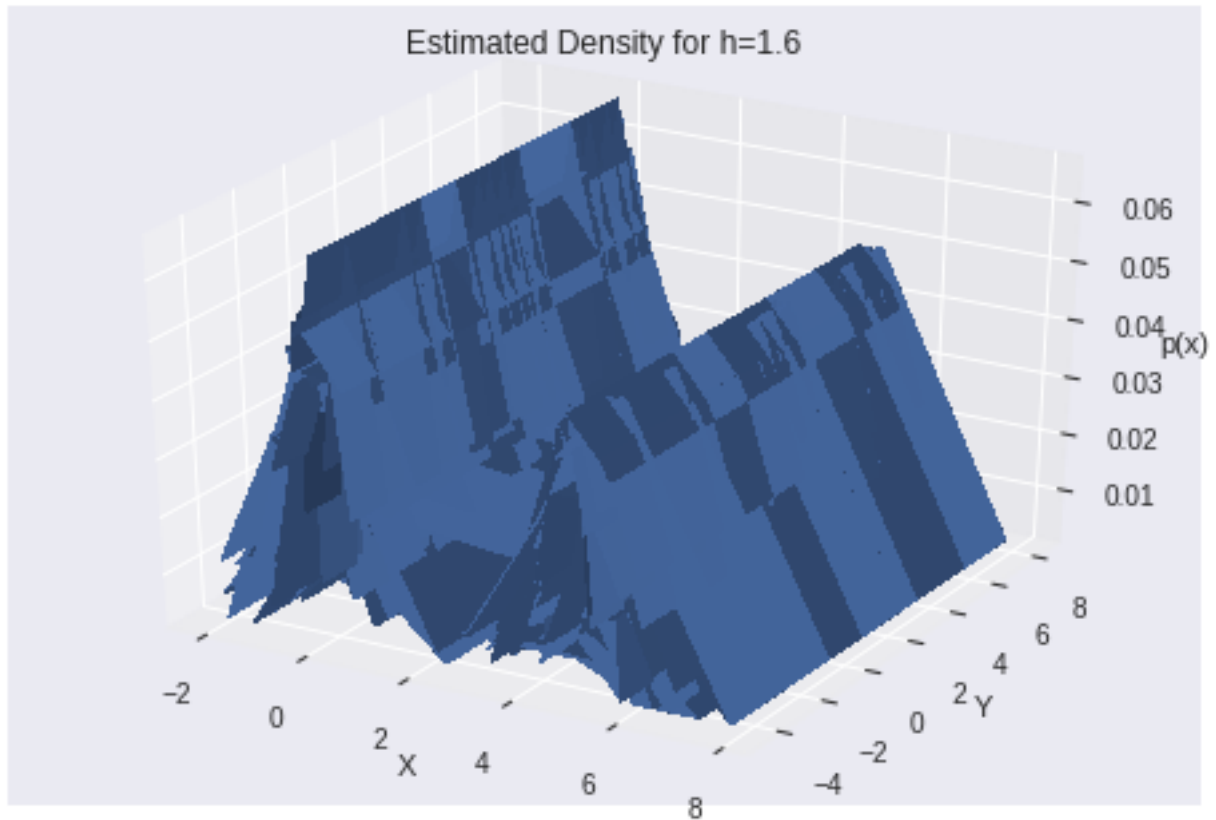
	Mean
N=10	$[-0.0930, 4.8959]$
N=100	$[0.0033, 4.9409]$
N=1000	$[0.0569, 5.0654]$

[Part2: Non-Parametric Methods]

- b) Using the generated samples in Section (a), estimate their density via the Parzen Window method for $h=0.2, 0.4, 0.8, 1.6$; plot the true density and the estimated densities for different h values.



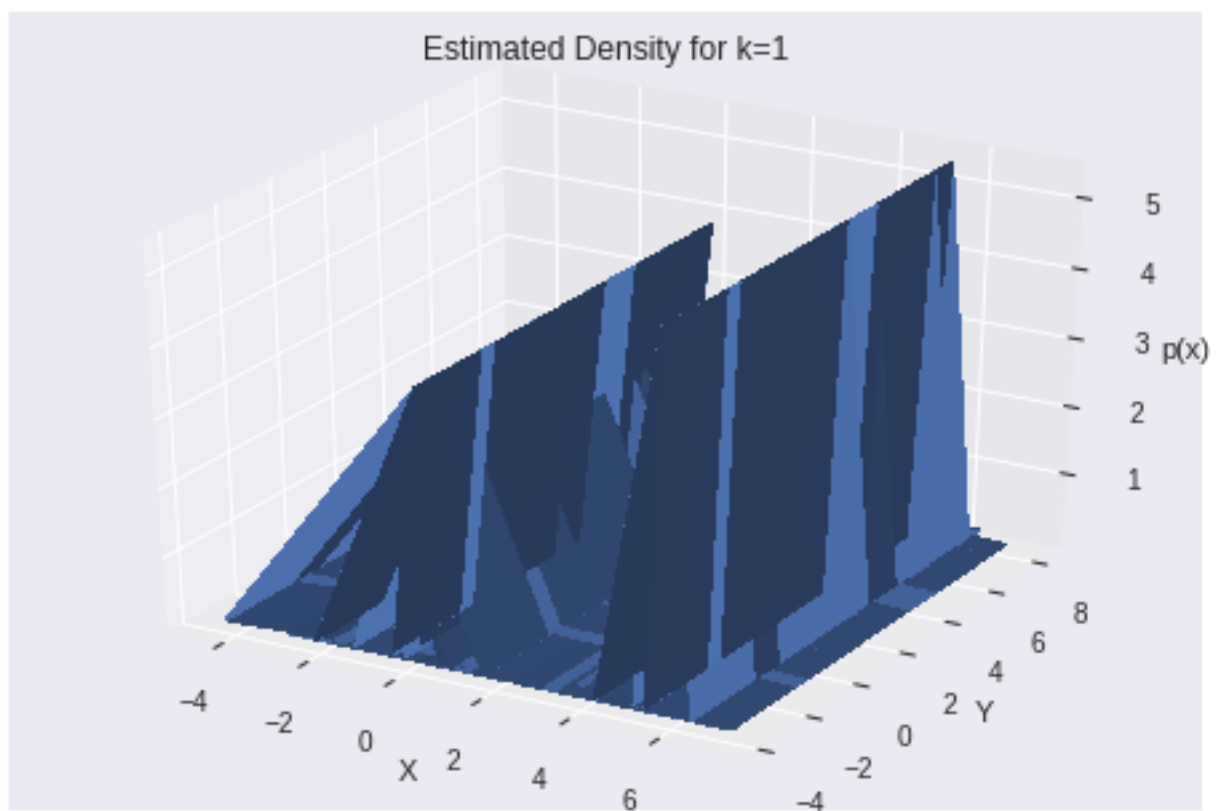
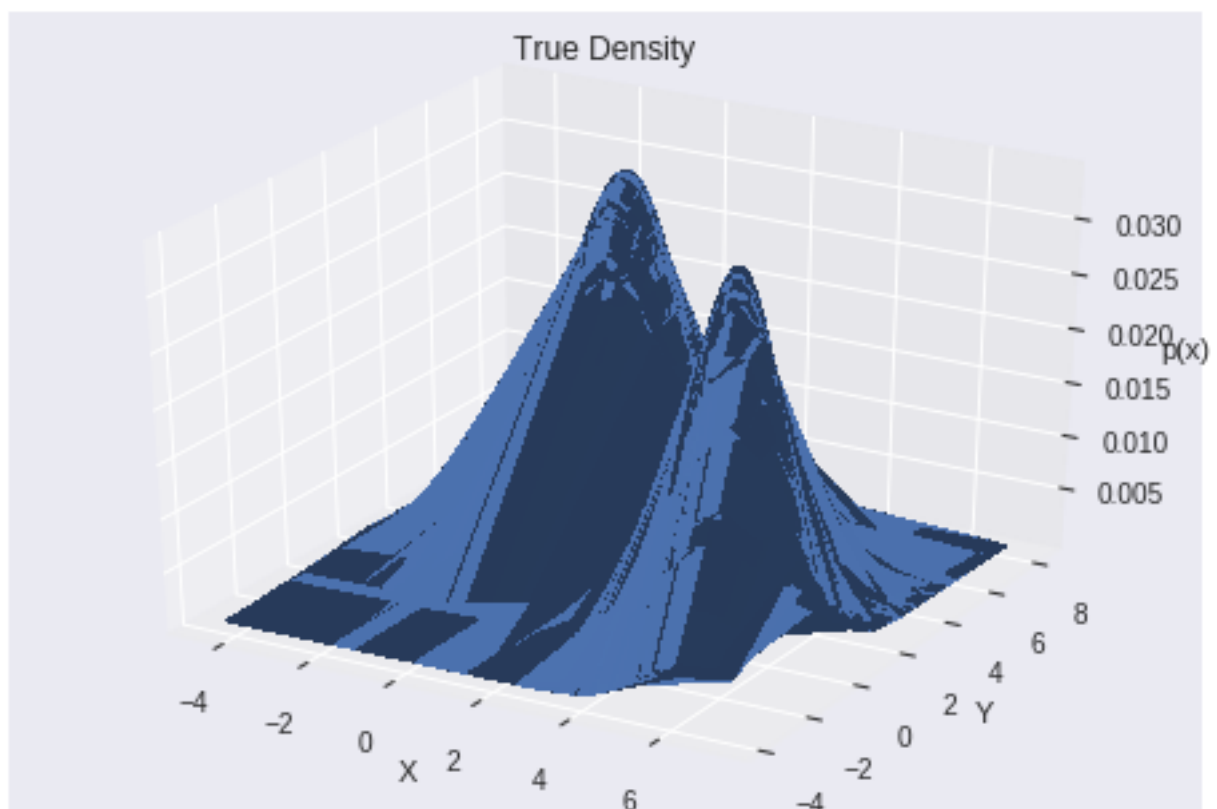


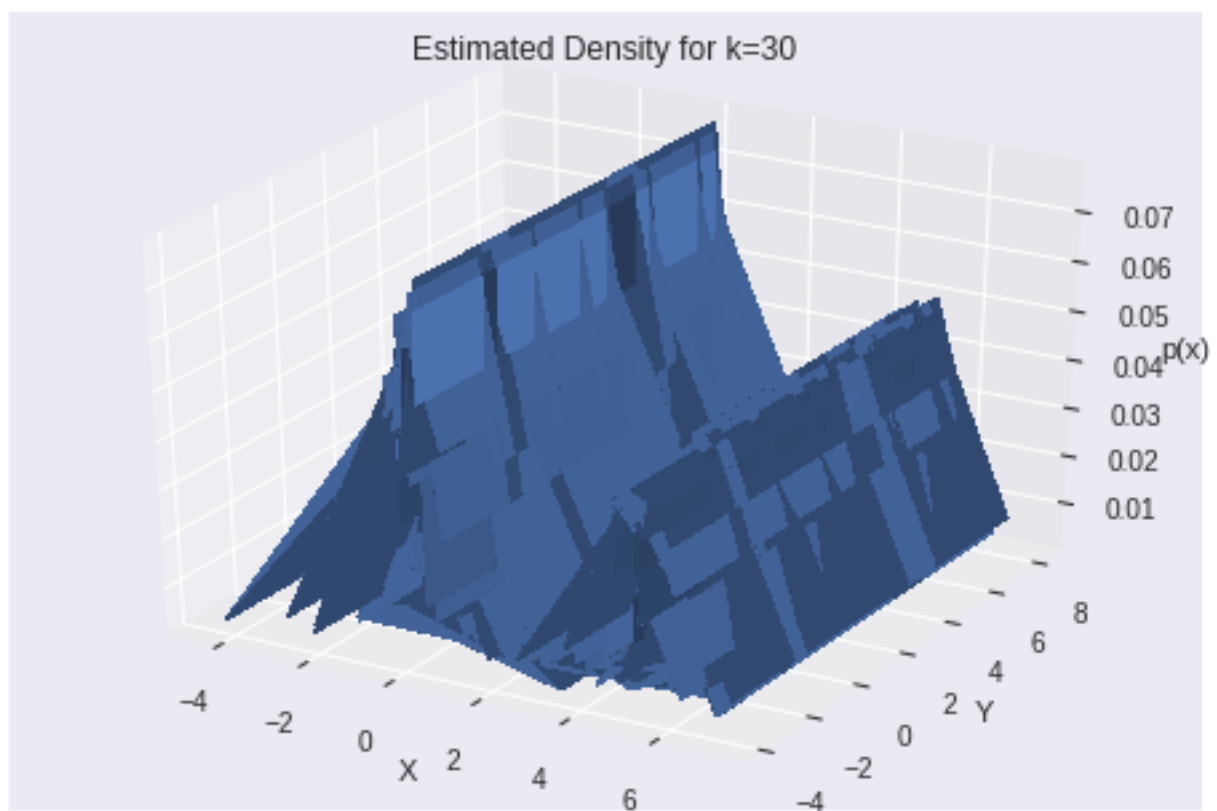
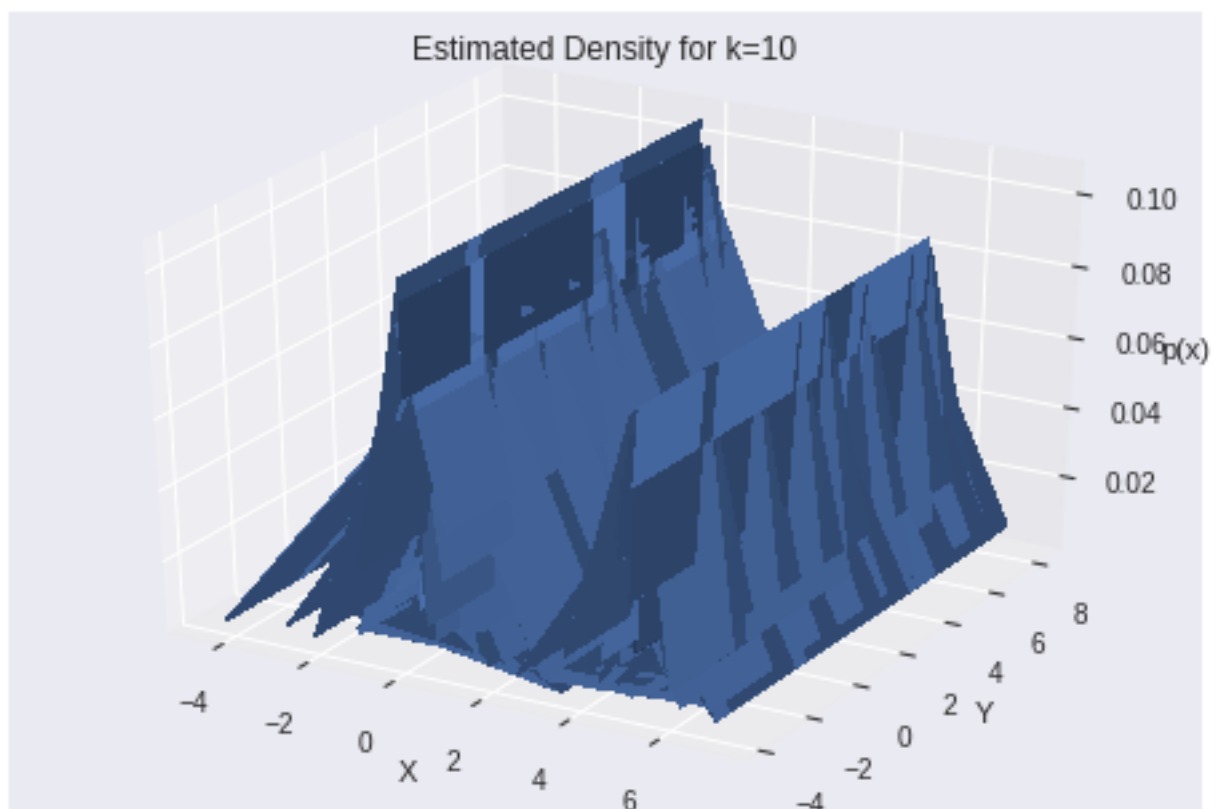


با رفتن به لینک زیر می توانید خروجی آنلاین کدها را در محیط Colab Research Google مشاهده کنید.

https://colab.research.google.com/drive/1jd15oyFDLb6TQlaX0WKfW_jQXoNYPYFH

- c) Estimate their density via the k-nearest neighbors (KNN) method for $k=1, 10, 30$; plot the true density and the estimated densities for different k values.





با رفتن به لینک زیر می توانید خروجی آنلاین کدها را در محیط Colab Research Google مشاهده کنید.

<https://colab.research.google.com/drive/1-XMNvWa0A6j8gNjn2X9Qjxmv6dphu4V6>

d) Compare and discuss about the results.

در خروجی قسمت b هرچه مقدار h کوچکتر باشد شکل نمودار سیخ سیخی تر یا (spiky تر) است و شکل اصلی داده‌ها را از دست می‌دهیم، مقدار $h = 1.6$ نزدیکترین پلات به پلات واقعی (*True Density*) می باشد و بهترین مقدار h همین مقدار می باشد (از بین h های داده شده)

در خروجی سوال c هرچه مقدار k کوچکتر باشد شکل نمودار سیخ سیخی تر یا (spiky تر) است و شکل اصلی داده‌ها را از دست می‌دهیم و پلات ایجاد شده شباهت کمی به پلات واقعی دارد با افزایش مقدار k پلات به شکل گوسی داده‌ها نزدیکتر می شود و تحلیل و تفسیر آن راحت تر می شود، مقدار $k = 10, k = 30$ مقادیر مناسبی جهت محاسبه *Density Estimation* می باشد.