

Tareas de Cómputo Científico para Probabilidad y Estadística

Alumnos del curso, Ago-Dic 2017

07 AUG 2024

Contents

1	Descomposición LU y Cholesky	2
2	Descomposición QR y mínimos cuadrados	2
3	Estabilidad	3
4	Cálculo de eigenvalores	4
5	Simulación Estocástica, introducción	4
6	MCMC: Metropolis-Hastings	5
7	MCMC: Metropolis-Hastings II	6
8	MCMC: MH con Kerneles Híbridos y Gibbs Sampler	7
9	MCMC: Tarea Final	10
10	Deep neural networks	12

1 Descomposición LU y Cholesky

1. Implementar los algoritmos de *Backward* y *Forward substitution*.
2. Implementar el algoritmo de eliminación gaussiana con pivoteo parcial LUP, 21.1 del Trefethen (p. 160).
3. Dar la descomposición LUP para una matriz aleatoria de entradas $U(0, 1)$ de tamaño 5×5 , y para la matriz

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 1 \\ -1 & -1 & 1 & 0 & 1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 \end{pmatrix} \quad (1)$$

4. Usando la descomposición LUP anterior, resolver el sistema de la forma

$$Dx = b \quad (2)$$

donde D son las matrices del problema 3, para 5 diferentes b aleatorios con entradas $U(0, 1)$. Verificando si es o no posible resolver el sistema.

5. Implementar el algoritmo de descomposición de Cholesky 23.1 del Trefethen (p. 175).
6. Comparar la complejidad de su implementación de los algoritmos de factorización de Cholesky y LUP mediante la medición de los tiempos que tardan con respecto a la descomposición de una matriz aleatoria hermitiana definida positiva. Graficar la comparación.

2 Descomposición QR y mínimos cuadrados

1. Implementar el algoritmo de Gram-Schmidt modificado 8.1 del Trefethen (p. 58) para generar la descomposición QR .
2. Implementar el algoritmo que calcula el estimador de mínimos cuadrados en una regresión usando la descomposición QR .

3. Generar \mathbf{Y} compuesto de $y_i = \text{sen}(x_i) + \epsilon_i$ donde $\epsilon_i \sim N(0, \sigma)$ con $\sigma = 0.11$ para $x_i = \frac{4\pi i}{n}$ para $i = 1, \dots, n$.

Hacer un ajuste de mínimos cuadrados a \mathbf{Y} , con descomposición QR , ajustando un polinomio de grado $p - 1$.

- Considerar los 12 casos: $p = 2, 4, 6, 100$ y $n = 100, 1000, 10000$.
 - Graficar el ajuste en cada caso.
 - Medir tiempo de ejecución de su algoritmo, comparar con descomposición QR de scipy y graficar los resultados.
4. Hacer $p = 0.1n$, o sea, diez veces más observaciones que coeficientes en la regresión, ¿Cual es la n máxima que puede manejar su computadora?

3 Estabilidad

1. Sea Q una matriz unitaria aleatoria de 20×20 (eg. con A una matriz de tamaño 20×20 aleatoria calculen su descomposición QR), . Sean $\lambda_1 > \lambda_2 > \dots \geq \lambda_{20} = 1 > 0$ y

$$B = Q^* \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{20}) Q \text{ y } B_\epsilon = Q^* \text{diag}(\lambda_1 + \epsilon_1, \lambda_2 + \epsilon_2, \dots, \lambda_{20} + \epsilon_{20}) Q,$$

con $\epsilon_i \sim N(0, \sigma)$, con $\sigma = 0.025\lambda_{20} = 0.01$.

- (a) Comparar la descomposición de Cholesky de B y de B_ϵ usando el algoritmo de la tarea 1. Considerar los casos cuando B tiene un *buen* número de condición y un *mal* número de condición.
 - (b) Con el caso mal condicionado, comparar el resultado de su algoritmo con el del algoritmo de Cholesky de scipy.
 - (c) Medir el tiempo de ejecución de su algoritmo de Cholesky con el de scipy.
2. Resolver el problema de mínimos cuadrados,

$$y = X\beta + \epsilon, \quad \epsilon_i \sim N(0, \sigma)$$

usando su implementación de la descomposición QR ; β es de tamaño $n \times 1$ y X de tamaño $n \times d$.

Sean $d = 5, n = 20, \beta = (5, 4, 3, 2, 1)'$ y $\sigma = 0.12$.

- (a) Hacer X con entradas aleatorias $U(0, 1)$ y simular y . Encontrar $\hat{\beta}$ y compararlo con el obtenido $\hat{\beta}_p$ haciendo $X + \Delta X$, donde las entradas de ΔX son $N(0, \sigma = 0.01)$. Comparar a su vez con $\hat{\beta}_c = ((X + \Delta X)'(X + \Delta X))^{-1}(X + \Delta X)'y$ usando el algoritmo genérico para invertir matrices `scipy.linalg.inv`.
- (b) Lo mismo que el anterior pero con X mal condicionada (ie. con casi colinealidad).

4 Cálculo de eigenvalores

1. Dado el siguiente

Teorema (Gershgorin):

Dada una matriz $A = a_{ij}$ de $m \times m$, cada eigenvalor de A está en al menos uno de los discos en el plano complejo con centro en a_{ii} y radio $\sum_{j \neq i} |a_{ij}|$. Además, si n de estos discos forman un dominio conexo, disjunto de los otros $m - n$ discos, entonces hay exactamente n eigenvalores en ese dominio.

Deduce estimaciones de los eigenvalores de

$$A = \begin{pmatrix} 8 & 1 & 0 \\ 1 & 4 & \epsilon \\ 0 & \epsilon & 1 \end{pmatrix}$$

con $|\epsilon| \leq 1$.

2. Implementa la iteración QR. Aplícala a la matriz A del Ejercicio 1 con $\epsilon = 10^{-N}$ para $N = 0, 1, 2, 4, 5$. Compara con lo obtenido por el algoritmo de `scipy`.

5 Simulación Estocástica, introducción

1. Definir la cdf inversa generalizada F_X^- y demostrar que en el caso de variables aleatorias continuas esta coincide con la inversa usual. Demostrar además que en general para simular de X podemos simular $u \sim U(0, 1)$ y $F_X^-(u)$ se distribuye como X . [1 punto]

- Implementar el siguiente algoritmo para simular variables aleatorias uniformes:

$$x_i = 107374182x_{i-1} + 104420x_{i-5} \pmod{2^{31} - 1}$$

regresa x_i y recorrer el estado, esto es $x_{j-1} = x_j; j = 1, 2, 3, 4, 5$; ¿parecen $U(0, 1)$? [1 punto]

- ¿Cuál es el algoritmo que usa `scipy.stats.uniform` para generar números aleatorios? ¿Cómo se pone la semilla? ¿y en R? [1 punto]
- ¿En `scipy` que funciones hay para simular una variable aleatoria genérica discreta? ¿tienen preproceso? [1 punto]
- Implementar el algoritmo Adaptive Rejection Sampling y simular de una $Gama(2, 1)$ 10,000 muestras. ¿cuando es conveniente dejar de adaptar la envolvente? (vea alg. A.7, p. 54 Robert y Casella, 2da ed.) [6 puntos]

6 MCMC: Metropolis-Hastings

- Simular $n = 5$ y $n = 50$ v.a's Bernoulli $Be(1/3)$; sea r el número de éxitos en cada caso.
- Implementar el algoritmo Metropolis-Hastings para simular de la posterior

$$f(p|\bar{x}) \propto p^r(1-p)^{n-r} \cos(\pi p) I_{[0, \frac{1}{2}]}(p),$$

con los dos casos de n y r de arriba. Para ello poner la propuesta $(p'|p) = p' \sim Beta(r+1, n-r+1)$ y la distribución inicial de la cadena $\mu \sim U(0, \frac{1}{2})$.

- Argumentar porque la cadena es f -irreducible y porque es ergódica. Implementar el algoritmo con los datos descritos y discutir los resultados.
- Implementar el algoritmo Metropolis-Hastings con la posterior de arriba tomando una propuesta diferente.

7 MCMC: Metropolis-Hastings II

Con el algoritmo Metropolis-Hastings (MH), simular lo siguiente:

1. Sean $x_i \sim Ga(\alpha, \beta); i = 1, 2, \dots, n$. Simular datos x_i con $\alpha = 3$ y $\beta = 100$ considerando los casos $n = 5$ y 40 .

Con $\alpha \sim U(1,4)$, $\beta \sim \exp(1)$ distribuciones a priori, se tiene la posterior

$$f(\alpha, \beta | \bar{x}) \propto \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} r_1^{\alpha-1} e^{-\beta(r_2+1)} 1(1 \leq \alpha \leq 4) 1(\beta > 1),$$

$$\text{con } r_2 = \sum_{i=1}^n x_i \text{ y } r_1 = \prod_{i=1}^n x_i.$$

En ambos casos, grafica los contornos para visualizar dónde está concentrada la posterior.

Utilizar la propuesta

$$q\left(\begin{pmatrix} \alpha_p \\ \beta_p \end{pmatrix} \mid \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix},$$

donde

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right).$$

2. Simular de la distribución $\text{Gamma}(\alpha, 1)$ con la propuesta $\text{Gamma}([\alpha], 1)$, donde $[\alpha]$ denota la parte entera de α .

Además, realizar el siguiente experimento: poner como punto inicial $x_0 = 950$ y graficar la evolución de la cadena, es decir, $f(X_t)$ vs t .

3. Implementar Random Walk Metropolis Hasting (RWMH) donde la distribución objetivo es $\mathcal{N}_2(\mu, \Sigma)$, con

$$\mu = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

Utilizar como propuesta $\varepsilon_t \sim \mathcal{N}_2(\mathbf{0}, \sigma I)$. ¿Cómo elegir σ para que la cadena sea eficiente? ¿Qué consecuencias tiene la elección de σ ?

Como experimento, elige como punto inicial $x_o = \begin{pmatrix} 1000 \\ 1 \end{pmatrix}$ y comenta los resultados.

Para todos los incisos del ejercicio anterior:

- Establece cual es tu distribución inicial.
- Grafica la evolución de la cadena.
- Indica cuál es el Burn-in.
- Comenta qué tan eficiente es la cadena.
- Implementa el algoritmo MH considerando una propuesta diferente.

8 MCMC: MH con Kerneles Híbridos y Gibbs Sampler

1. Aplique el algoritmo de Metropolis-Hastings considerando como función objetivo la distribución normal bivariada:

$$f_{X_1, X_2}(\bar{x}) = \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu) \right\}$$

donde,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Así, se tienen las siguientes distribuciones condicionales:

$$X_1|X_2 = x_2 \sim N \left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2), \sigma_1^2 (1 - \rho^2) \right)$$

$$X_2|X_1 = x_1 \sim N \left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1), \sigma_2^2 (1 - \rho^2) \right)$$

Considere las siguientes propuestas:

$$q_1((x'_1, x'_2)|(x_1, x_2)) = f_{X_1|X_2}(x'_1|x_2)1(x'_2 = x_2)$$

$$q_2((x'_1, x'_2)|(x_1, x_2)) = f_{X_2|X_1}(x'_2|x_1)1(x'_1 = x_1)$$

A partir del algoritmo MH usando Kerneles híbridos simule valores de la distribución normal bivariada, fijando $\sigma_1 = \sigma_2 = 1$, considere los casos $\rho = 0.85$ y $\rho = 0.99$ ¹

2. Consideré los tiempos de falla t_1, \dots, t_n con distribución *Weibull*(α, λ):

$$f(t_i|\alpha, \lambda) = \alpha \lambda t_i^{\alpha-1} e^{-t_i^\alpha \lambda}$$

Se asumen como a priori $\alpha \sim \text{exp}(c)$ y $\lambda|\alpha \sim \text{Gama}(\alpha, b)$, por lo tanto, $f(\alpha, \lambda) = f(\lambda|\alpha)f(\alpha)$ ². Así, para la disitribución posterior se tiene:

$$f(\alpha, \lambda|\bar{t}) \propto f(\bar{t}|\alpha, \lambda)f(\alpha, \lambda)$$

A partir del algoritmo MH usando Kerneles híbridos simule valores de la distribución posterior $f(\alpha, \lambda|\bar{t})$, considerando las siguientes propuestas:

Propuesta 1:

$$\lambda_p|\alpha, \bar{t} \sim \text{Gama} \left(\alpha + n, b + \sum_{i=1}^n t_i^\alpha \right) \text{ y dejando } \alpha \text{ fijo.}$$

Propuesta 2:

¹Ver la tesis de Cricelio Montesinos para una explicación más extensa del Gibbs, Montesinos, C (2016) “Distribución de Direcciones en el Gibbs Sampler Generalizado”, MSc Dissertation, CIMAT. https://www.cimat.mx/es/Tesis_digitales/. También vean la Enciclopedia de Estadística de Wiley, la entrada de *Gibbs Sampler*: https://www.cimat.mx/~jac/2016WileyStatsRef_GibbsSampling.pdf.

²Este ejemplo aparece en Kundu, D. (2008), “Bayesian Inference and Life Testing Plan for the Weibull Distribution in Presence of Progressive Censoring”, *Technometrics*, **50**(2), 144–154

$\alpha_p | \lambda, \bar{t} \sim \text{Gama}(n + 1, -\log(b) - \log(r_1) + c)$, con $r_1 = \prod_{i=1}^n t_i$ y dejando λ fijo.

Propuesta 3:

$\alpha_p \sim \text{exp}(c)$ y $\lambda_p | \alpha_p \sim \text{Gama}(\alpha_p, b)$.

Propuesta 4 (RWMH):

$\alpha_p = \alpha + \epsilon$, con $\epsilon \sim N(0, \sigma)$ y dejando λ fijo.

Simular datos usando $\alpha = 1$ y $\lambda = 1$ con $n = 30$. Para la a priori usar $c = 1$ y $b = 1$.

3. Considere el ejemplo referente al número de fallas de bombas de agua en una central nuclear³, donde p_i representa el número de fallas en el tiempo de operación t_i , con $i = 1, \dots, n$.

Se considera el modelo $p_i \sim \text{Poisson}(\lambda_i t_i)$, (las λ_i son independientes entre si), con distribuciones a priori $\lambda_i | \beta \sim \text{Gama}(\alpha, \beta)$ y $\beta \sim \text{Gama}(\gamma, \delta)$, por lo tanto:

$$f(\lambda_1, \dots, \lambda_n, \beta) = f(\lambda_1 | \beta) f(\lambda_2 | \beta) \dots f(\lambda_n | \beta) f(\beta)$$

Para la distribución posterior se tiene:

$$f(\lambda_1, \dots, \lambda_n, \beta | \bar{p}) \propto L(\bar{p}, \bar{\lambda}, \beta) f(\lambda_1, \dots, \lambda_n, \beta)$$

Simule valores de la distribución posterior $f(\lambda_1, \dots, \lambda_n, \beta | \bar{p})$, usando un kernel híbrido, considerando las propuestas:

$$\lambda_i | \bar{\lambda}_{-i}, \beta, \bar{t} \sim \text{Gama}(p_i + \alpha, \beta + t_i)$$

³Este ejemplo fué usado en el artículo original del Gibbs sampler del Gelfand y Smith (1990). Vea también Norton, R.A., Christen, J.A. y Fox, C. (2017), “Sampling hyperparameters in hierarchical models: improving on Gibbs for high-dimensional latent fields and large data sets” *Communications in Statistics - Simulation and Computation*, <http://dx.doi.org/10.1080/03610918.2017.1353618>.

Bomba (i)	1	2	3	4	5	6	7	8	9	10
T. de uso (t_i)	94.32	15.72	62.88	125.76	5.24	31.44	1.05	1.05	2.1	10.48
# de fallas (p_i)	5	1	5	14	3	20	1	1	4	22

Table 1: Datos de bombas de agua en centrales nucleares (Robert y Casella, p. 385) para el ejemplo 8.3.

$$\beta|\bar{\lambda}, \bar{t} \sim \text{Gama} \left(n\alpha + \gamma, \delta + \sum_{i=1}^n \lambda_i \right).$$

Verifique que estas son propuestas Gibbs.

Use los datos del Cuadro 1 con los parámetros a priori $\alpha = 1.8$, $\gamma = 0.01$ y $\delta = 1$.

9 MCMC: Tarea Final

En ambos problemas hay que diseñar e implementar el MCMC, investigar sobre su convergencia y tener algún grado de certeza sobre si sí se está simulando de la posterior correspondiente. Más aun, recuerde que se trata de un problema de inferencia: Hay que hablar del problema en si, comentar sobre las posteriores simuladas y posibles estimadores (a partir de la muestra de posterior) que se pueden proporcionar de cada parámetro.

1. **(Problema en ecología)** Sean X_1, \dots, X_m variables aleatorias donde X_i denota el número de individuos de una especie en cierta región. Suponga que $X_i|N, p \sim \text{Binomial}(N, p)$, entonces

$$f(\bar{x}|N, p) = \prod_{i=1}^m \frac{N!}{x_i!(N - x_i)!} p^{x_i} (1 - p)^{N - x_i}.$$

Asumiendo la distribución a priori $p \sim \text{Beta}(\alpha, \beta)$ y $N \sim h(\cdot)$, donde h es una dist. discreta en $\{0, 1, 2, \dots, N_{max}\}$, se tiene definida la distribución posterior $f(N, P|\bar{x})$.

A partir del algoritmo MH, simule valores de la distribución posterior usando un kernel híbrido. Para ello considere **como sugerencia** la

siguiente distribución inicial para el MCMC

$$p \sim U(0, 1) \quad \text{y} \quad N \sim U_d \left\{ \max_{i \in \{1, \dots, m\}} (x_i), \max_{i \in \{1, \dots, m\}} (x_i) + 1, \dots, N_{max} \right\}$$

y las propuestas

- Propuesta 1: De la condicional total de p (kernel Gibbs).
- Propuesta 2: De la a priori.
- Propuesta 3: Propuesta hipergeométrica (j ?).
- Propuesta 4: Poisson: $N_p \sim \max_{i \in \{1, \dots, m\}} (x_i) + \text{Poisson}(?)$.
- Propuesta 5: Caminata aleatoria

$$N_p = N + \epsilon, \quad \mathbb{P}(\epsilon = 1) = \frac{1}{2} = \mathbb{P}(\epsilon = -1).$$

Los datos son estos: 7, 7, 8, 8, 9, 4, 7, 5, 5, 6, 9, 8, 11, 7, 5, 5, 7, 3, 10, 3. Los datos son estos: 7, 8, 6, 5, 2, 8, 6, 6, 7, 4, 8, 8, 6, 4, 8, 8, 10, 5, 4, 2.

A priori, esperamos que sea difícil observar a los individuos entonces $\alpha = 1, \beta = 20$. La especie no es muy abundante y entonces $N_{max} = 1000$ y $h(N) = 1/(N_{max} + 1); N \in \{0, 1, 2, \dots, N_{max}\}$.

Las propuestas y distribución inicial para el MCMC de arriba son **solamente sugerencia**, propongan otras propuestas, experimenten y comenten.

2. (**Estudio de mercado**) Se tiene un producto y se realiza una encuesta con el fin de estudiar cuánto se consume dependiendo de la edad. Sea Y_i el monto de compra y X_i la covariable la cual representa la edad. Suponga que $Y_i \sim Po(\lambda_i)$ (distribución Poisson con intensidad λ_i)

$$\lambda_i = cg_b(x_i - a)$$

para g_b la siguiente función de liga

$$g_b(x) = \exp\left(-\frac{x^2}{2b^2}\right).$$

O sea, se trata de regresión Poisson con una función liga no usual. Si $\lambda_i = 0$ entonces $P(Y_i = 0) = 1$. a = años medio del segmento (años), c = gasto promedio (pesos), b = “amplitud” del segmento (años).

Considere las distribuciones a priori

$$a \sim N(35, 5), \quad c \sim Gama(3, 3/950), \quad b \sim Gama(2, 2/5).$$

El segundo parámetro de la normal es desviación estandar y el segundo parámetro de las gammas es tasa (*rate*).

Usando MH simule de la distribución posterior de a, c y b .

Los datos son estos, $n = 100$:

```
X = array([ 27, 19, 21, 51, 16, 59, 16, 54, 52, 16, 31, 31, 54, 26, 19, 13, 59, 48, 54, 23, 50, 59, 55, 37, 61, 53, 56,
31, 34, 15, 41, 14, 13, 13, 32, 46, 17, 52, 54, 25, 61, 15, 53, 39, 33, 52, 65, 35, 65, 26, 54, 16, 47, 14, 42, 47, 48,
25, 15, 46, 31, 50, 42, 23, 17, 47, 32, 65, 45, 28, 12, 22, 30, 36, 33, 16, 39, 50, 13, 23, 50, 34, 19, 46, 43, 56, 52,
42, 48, 55, 37, 21, 45, 64, 53, 16, 62, 16, 25, 62])
```

```
Y = array([1275, 325, 517, 0, 86, 0, 101, 0, 0, 89, 78, 83, 0, 1074, 508, 5, 0, 0, 0, 1447, 0, 0, 0, 0, 0, 0, 87, 7,
37, 0, 15, 5, 6, 35, 0, 158, 0, 0, 1349, 0, 35, 0, 0, 12, 0, 0, 2, 0, 1117, 0, 79, 0, 13, 0, 0, 0, 1334, 56, 0, 81, 0, 0,
1480, 177, 0, 29, 0, 0, 551, 0, 1338, 196, 0, 9, 104, 0, 0, 3, 1430, 0, 2, 492, 0, 0, 0, 0, 0, 0, 0, 1057, 0, 0, 0, 68,
0, 87, 1362, 0])
```

- Investiga y describe muy brevemente los softwares OpenBugs, Nimble, JAGS, DRAM, Rtwalk, Mcee Hammer, PyMCMC.

10 Deep neural networks

Escoja UNO de los tres ejercicios siguientes:

- Usando la base de datos MNIST realice lo siguiente y explique cada decisión tomada:
 - Diseñe una red neuronal de una sola capa oculta para la clasificación de las imágenes. Use una función de pérdida predefinida.
 - Entrene la red neuronal
 - Presente la matriz de confusión (Confusion matrix),
 - Describa que es la precisión y la exhaustividad (precision and recall) y calculelos a partir de su matriz de confusión

2. Usando la base de datos Fashion MNIST realice lo siguiente y explique cada decisión tomada:
 - Diseñe una red neuronal de dos capas ocultas para la clasificación de las imágenes, cada una con diferente función de activación y use una función de pérdida en la que el error de clasificar mal un calzado sea el doble que el resto de prendas.
 - Entrene la red neuronal
 - Presente la matriz de confusión (Confusion matrix)

3. Con la base de datos CIFAR-10 realice lo siguiente:
 - Diseñe una red neuronal de una capa oculta completamente conectada de al menos 10,000 neuronas para la clasificación de las imágenes.
 - Entrene la red neuronal
 - Presente la matriz de confusión (Confusion matrix)
 - Entrene una segunda red neuronal usando la estructura de LeNet 5.
 - Presente la matriz de confusión (Confusion matrix)
 - Compare ambas redes. Para esta comparación será necesario que al menos presente las curvas de error de entrenamiento y predicción para ambas redes, los tiempos de entrenamiento y que tome en cuenta las matrices de confusión.
 - Usando la base de datos MNIST y la estructura LeNet 5 realice lo siguiente:
 - Entrene la red neuronal
 - Implemente el algoritmo *Deepfool* y construya un ejemplo adversario para cada una de las categorías, presente la imagen original, el ruido que añadió y la nueva imagen.