| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: | DNI:1803657848 | |  |
| Apellidos: MAYORGA PEREZ DIEGO FERNANDO |

| ESTUDIO | ASIGNATURA | CONVOCATORIA |
| --- | --- | --- |
| MÁSTER UNIVERSITARIO EN INGENIERÍA MATEMÁTICA Y COMPUTACIÓN (PLAN 2016) | 4391020006.- TÉCNICAS MULTIVARIANTES | Ordinaria Número periodo 1823 |

| FECHA | MODELO | CIUDAD DEL EXAMEN |
| --- | --- | --- |
| 14-16/01/2022 | Modelo - C |  |

| Etiqueta identificativa |
| --- |
|  |

**INSTRUCCIONES GENERALES**

1. Ten disponible tu documentación oficial para identificarte, en el caso de que se te solicite.
2. Si tu examen consta de una parte tipo test, indica las respuestas en la plantilla según las características de este.
3. Debes contestar en el documento adjunto, respetando en todo momento el espaciado indicado para cada pregunta. Si este es en formato digital, los márgenes, el interlineado, fuente y tamaño de letra vienen dados por defecto y no deben modificarse. En cualquier caso, asegúrate de que la presentación es suficientemente clara y legible. Entrega toda la documentación relativa al examen, revisando con detenimiento que los archivos o documentos son los correctos. El envío de archivos erróneos o un envío incompleto supondrá una calificación de “no presentado”.
4. Durante el examen y en la corrección por parte del docente, se aplicará el Reglamento de Evaluación Académica de UNIR que regula las consecuencias derivadas de las posibles irregularidades y prácticas académicas incorrectas con relación al plagio y uso inadecuado de materiales y recursos.
5. No está permitido el uso de Internet ni ningún tipo de comunicación con otra persona.Durante todo el examen tu teléfono móvil debe estar en modo avión.
6. La parte principal de cada pregunta consiste en interpretar y comentar los resultados obtenidos. Si te limitas a hacer los cálculos no vas a poder superar el examen.
7. Es fundamental que las respuestas estén debidamente redactadas, de forma clara y precisa y sin faltas de ortografía.
8. Para hacer el examen puedes utilizar los apuntes del curso y los scripts que hayas preparado y Python para hacer los cálculos.

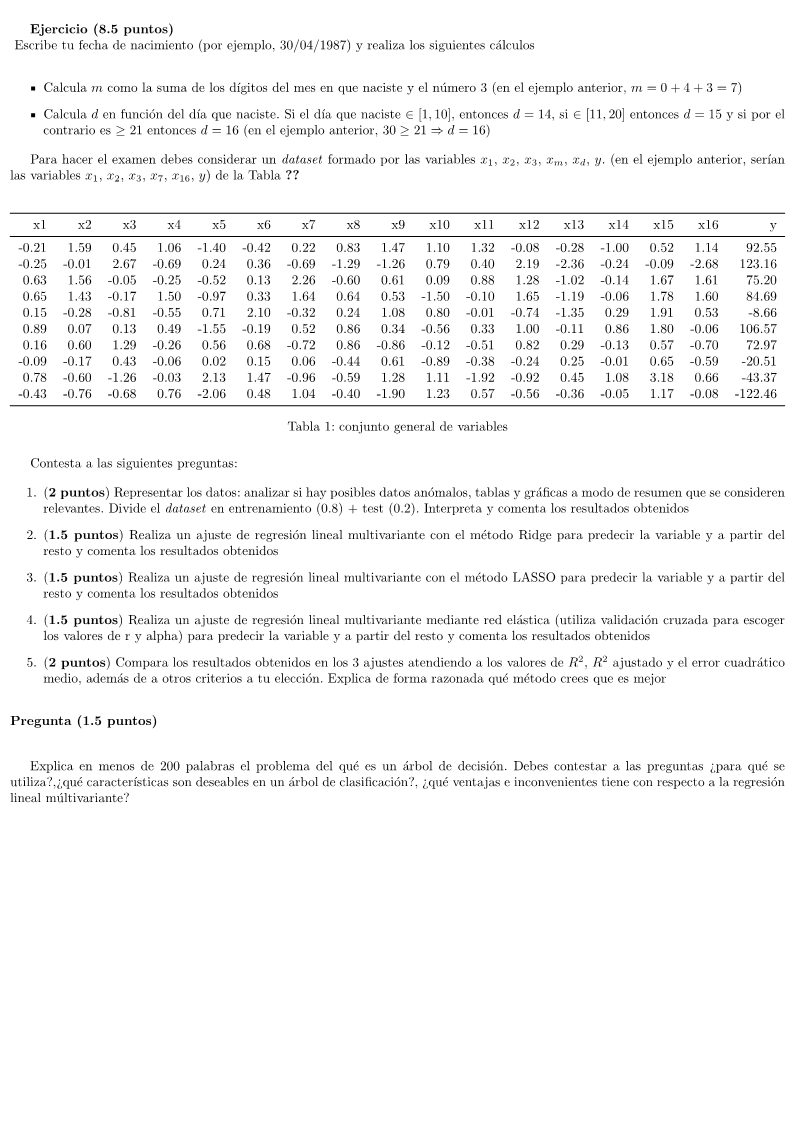
**Puntuación**

**Preguntas**

* Puntuación máxima 10.00 puntos

El examen constará de un ejercicio práctico (8,5 puntos) y una pregunta teórica (1,5 puntos). Los enunciados están en la página 14 y el espacio para responder el examen está entre las práginas 4 y 13.  
  
**1.** Pregunta

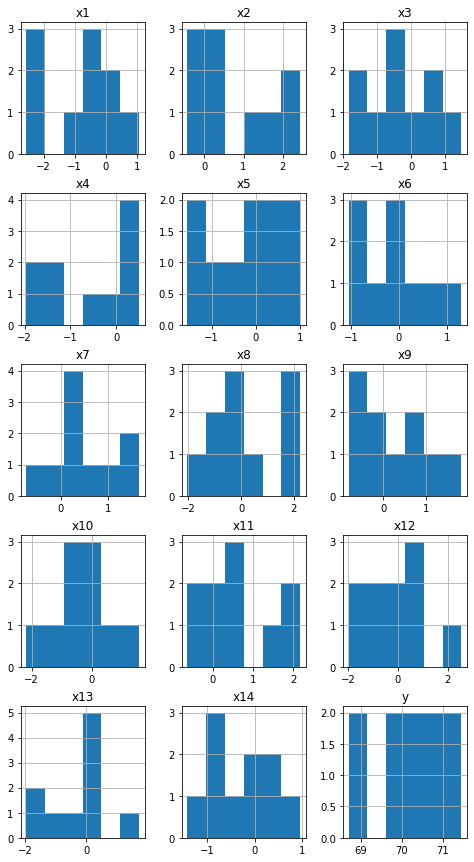
 (Responder en 10 caras)

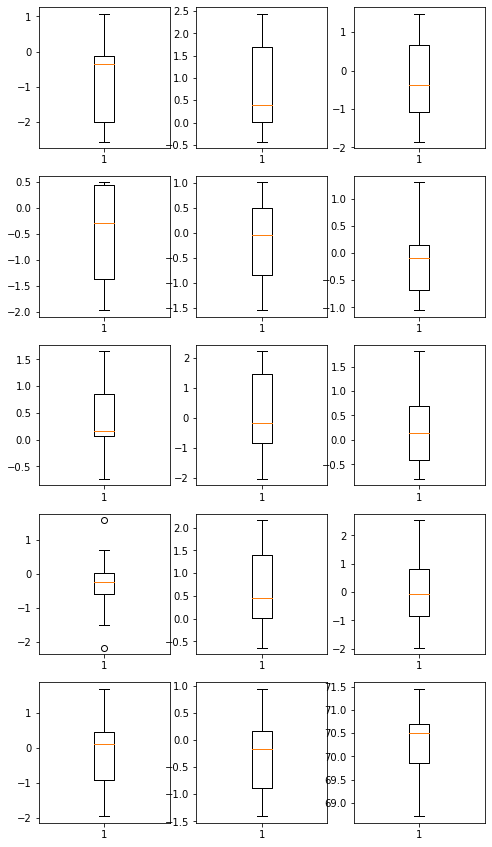


Pregunta 1

Los valores de las variables son m=6 y d=14

Para detectar los datos anómalos e identificar la distribución de los valores de las variables se realizan los histogramas y boxplot de cada variable, de igual manera de la variable de respuesta, se obtienen las siguientes gráficas.

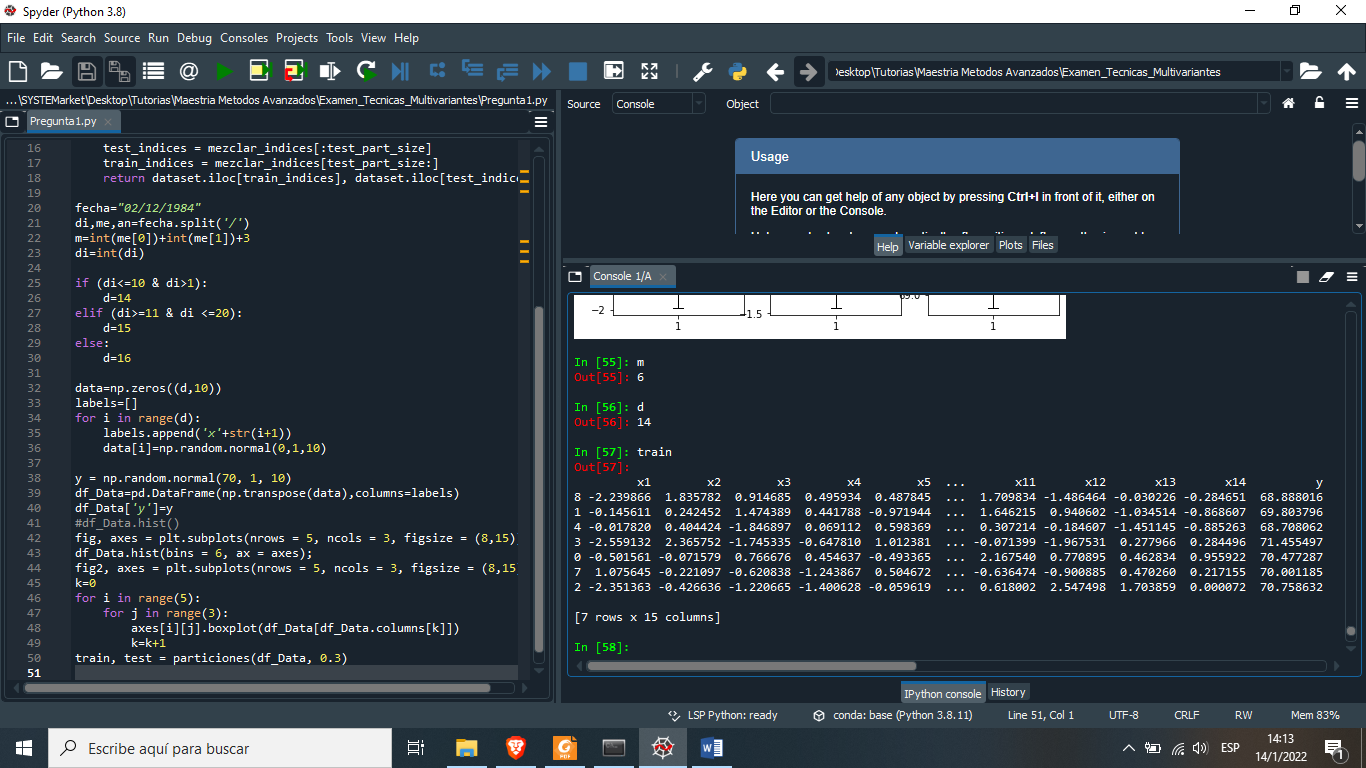




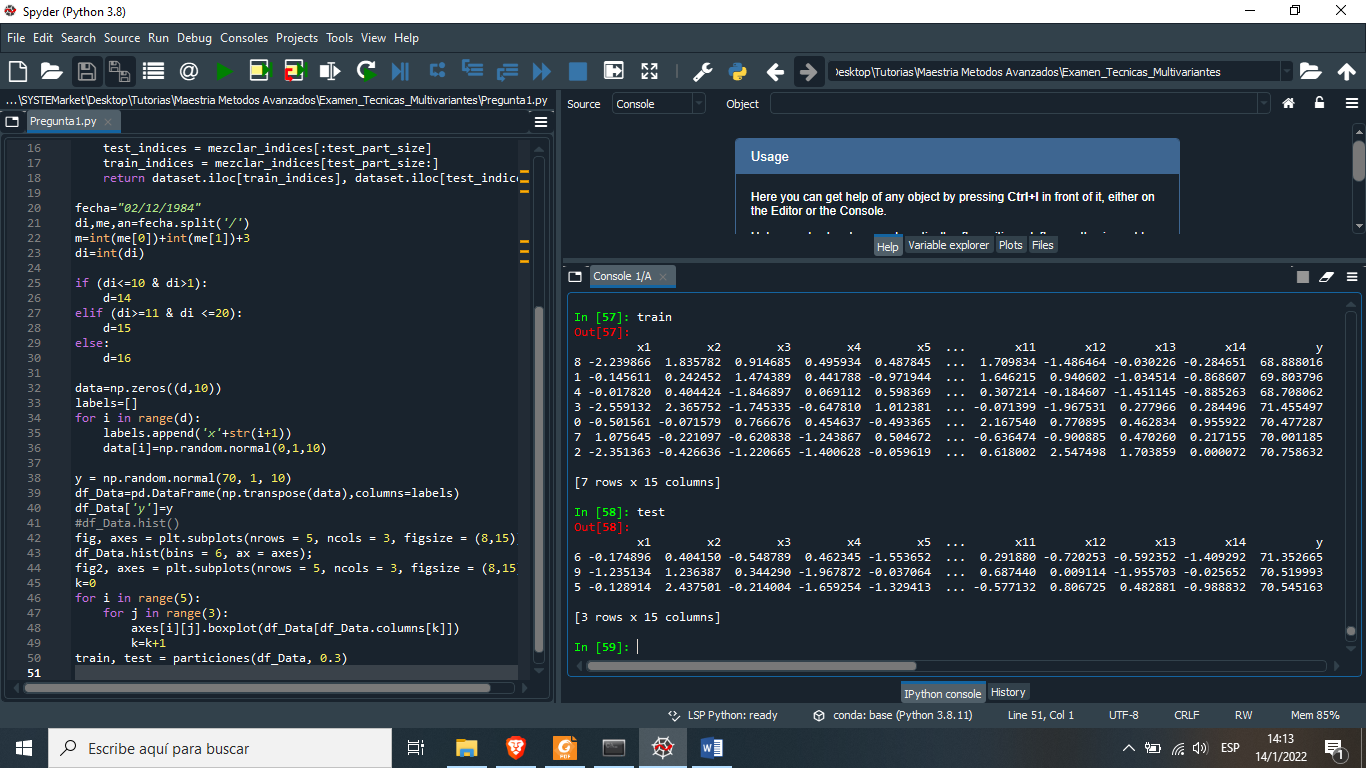
Con los boxplots podemos observar que solo la variable x10 presenta dos datos anómalos mientras que el resto si presenta una distribución normal según los histogramas.

Se realizo la división de los datos a 0.8 obtiendo las siguientes tablas

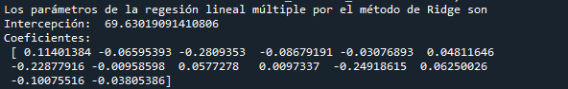
El de entrenamiento es



El de testeo es

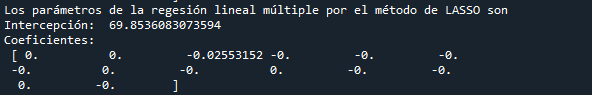


**PREGUNTA 2**



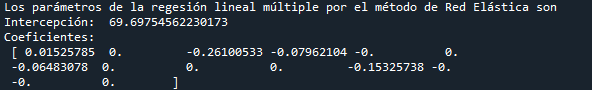
Como estamos en un modelo de regesión lineal Ridge, la cual es un tipo de regresión penalizada de tipo cuadratico a los coeficientes, pero el método no elimina variables ya que los coeficientes se esogen de tal manera que se ajuste de mejor maenra el modelo. Por esta razón podemos observar que todos los coeficientes con tomados enceutna y tienen valores parecidos, mientras que el intercepto es el que tiene un valor cercano a los valores de la variable de respuesta.

**PREGUNTA 3**



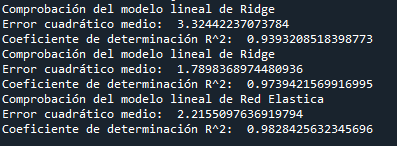
En el caso de la regresión de LASSO existe una mayor restricción para que pasen los coeficientes al modelo y esto se basa una norma de y tiene la forma de |Bi|<s, donde Bi son los coeficientes. En este caso vemos que tan solo una variable supera la resttricción del metodo de lasso y el intercepto tiene un valor parecido al obtenido con el metodo de Ridge.

**PREGUNTA 4**



Al igual que en el mpetodo de LASSO la red elastica tiene trestricciones para considerar las variables, es una regresión lineal múltiple penalizada y dicha restriccón utiliza al mismo tiempo las restricción de LASSO y la de Ridge pero no sobre los coeficientes sino que actúa sobre el hiperparámetro r. Vemos que existen mas variables consideradas en este modelo comparado con el de LASSO, las varialbes eliminadas se agrupan en otras varialbes que tienen colinealidad por esta razón tenemos menos coeficientes que el modelo de Ridge.

**PREGUNTA 5**



LASSO

Analizando los coeficientes de determinacion R^2 podemos observar que el modelo de Red Elastica tiene un valor mas cercano a 1 por tanto indica que se ajusta mejor al modelo planteado.

**PREGUNTA TEÓRICA: ARBOL DE DECISIONES**

Los arboles de decision son una serie de condiciones que se aplican a varias variables predictoras y las variables de respuesta, estas condiciones pueden generar distintos caminos similares a las ramas de un arbol que indicaràn las divisiones del espacio para diferentes regiones, por lo que una variable puede tener diferentes probabilidades de ocurrencia. Algunas de las caracteristicas que poseen son su analisis grafico que presentan los sucesos, la toma de decisiones de forma probabilistica. Una vez que el árbol a sido diseñado, se implementa las herramientas estadisticas en la region que pertenece la muestra. Los métodos con los que se trabaja los arboles son faciles de implementar, pero de manera externa, es necesario realizar ajustes como la regresión penalizada cuando las predicciónes son peores a las esperadas, ademas de los métodos de ensamble que permite fusionar otros arboles. Dentro de las ventajas sería la facilidad de implementación, entendimiento e interpretación, ademas que no solo se utiliza para la regresión sino ademas para la clasificación, sin embargo, las desventajas como se explico anterioremente, son su sensiblidad a la rotaciones de los datos de la muestra de entramiento para el arbol.

**ANEXO CODIGO**

**import numpy as np**

**import matplotlib.pyplot as plt**

**import pandas as pd**

**from sklearn.metrics import mean\_squared\_error, r2\_score**

**from sklearn.linear\_model import Ridge**

**def particiones(dataset, test\_part):**

**test\_part\_size = int(len(dataset) \* test\_part)**

**mezclar\_indices = np.random.permutation(len(dataset))**

**test\_indices = mezclar\_indices[:test\_part\_size]**

**train\_indices = mezclar\_indices[test\_part\_size:]**

**return dataset.iloc[train\_indices], dataset.iloc[test\_indices]**

**np.random.seed(20)**

**fecha="02/12/1984"**

**di,me,an=fecha.split('/')**

**m=int(me[0])+int(me[1])+3**

**di=int(di)**

**if (di<=10 & di>1):**

**d=14**

**elif (di>=11 & di <=20):**

**d=15**

**else:**

**d=16**

**data=np.zeros((d,10))**

**labels=[]**

**for i in range(d):**

**labels.append('x'+str(i+1))**

**data[i]=np.random.normal(0,1,10)**

**y = np.random.normal(70, 1, 10)**

**df\_Data=pd.DataFrame(np.transpose(data),columns=labels)**

**df\_Data['y']=y**

**df\_Data.hist()**

**fig, axes = plt.subplots(nrows = 5, ncols = 3, figsize = (8,15))**

**df\_Data.hist(bins = 6, ax = axes);#fig2, axes = plt.subplots(nrows = 5, ncols = 3, figsize = (8,15))**

**k=0**

**for i in range(5):**

**for j in range(3):**

**axes[i][j].boxplot(df\_Data[df\_Data.columns[k]])**

**k=k+1**

**train, test = particiones(df\_Data, 0.3)**

**# separamos la variable respuesta**

**respuesta=train['y']**

**data\_prep=train.drop('y',axis=1)**

**# CREAMOS EL MODELO DE REGRESION POR RIDGE**

**ridge\_reg = Ridge(alpha = 1, solver = "auto")**

**ridge\_reg.fit(data\_prep, respuesta)**

**# obtener coeficientes del modelo--------------------------------**

**# intercepto**

**print('Los parámetros de la regesión lineal múltiple por el método de Ridge son')**

**print('Intercepción: ',ridge\_reg.intercept\_)**

**# coeficientes de regresion**

**print('Coeficientes: \n',ridge\_reg.coef\_)**

**### CREAMOS EL MODELO DE REGRESIÓN POR LASSO**

**from sklearn.linear\_model import Lasso**

**# ajustar el modelo----------------------------------------------**

**lasso\_reg = Lasso()**

**lasso\_reg.fit(data\_prep, respuesta)**

**print('Los parámetros de la regesión lineal múltiple por el método de LASSO son')**

**#for elem in list(lasso\_reg.coef\_):**

**# print(str(elem))**

**print('Intercepción: ',lasso\_reg.intercept\_)**

**print('Coeficientes: \n',lasso\_reg.coef\_)**

**### CREAMOS EL MODELO DE LA RED ELASTICA**

**# importar clase-------------------------------------------------**

**from sklearn.linear\_model import ElasticNetCV**

**# ajustar el modelo----------------------------------------------**

**e\_net = ElasticNetCV()**

**e\_net.fit(data\_prep, respuesta)**

**# obtener coeficientes del modelo--------------------------------**

**print('Los parámetros de la regesión lineal múltiple por el método de Red Elástica son')**

**print('Intercepción: ',e\_net.intercept\_)**

**print('Coeficientes: \n',e\_net.coef\_)**

**# Relizamos el cálculos para comprobar los modelos**

**from sklearn.metrics import mean\_squared\_error, r2\_score**

**import statsmodels.api as sm**

**from scipy import stats**

**### CON EL MODELO DE RIDGE**

**### REALIZAMOS LAS PREDICCIONES CON EL TEST DATA**

**y\_test=test['y']**

**test=test.drop('y',axis=1)**

**ridgePred=ridge\_reg.predict(test)**

**print("Comprobación del modelo lineal de Ridge")**

**print("Error cuadrático medio: " , mean\_squared\_error(y\_test, ridgePred))**

**print("Coeficiente de determinación R^2: ", r2\_score(y\_test, ridgePred))**

**### CON EL MODELO DE LASSO**

**lassoPred=lasso\_reg.predict(test)**

**print("Comprobación del modelo lineal de Ridge")**

**print("Error cuadrático medio: " , mean\_squared\_error(y\_test, lassoPred))**

**print("Coeficiente de determinación R^2: ", r2\_score(y\_test, lassoPred))**

**### CON EL MODELO DE Red Elastica**

**NetPred=e\_net.predict(test)**

**print("Comprobación del modelo lineal de Red Elastica")**

**print("Error cuadrático medio: " , mean\_squared\_error(y\_test, NetPred))**

**print("Coeficiente de determinación R^2: ", r2\_score(y\_test, NetPred))**