

Examen 1

Desarrollo de Software Estadístico Descriptivo.

Mauricio de Garay Hernández
Taller de Desarrollo de Aplicaciones - Grupo A
César Arturo Ángeles Ruiz
Universidad Iberoamericana
Ciudad de México, México

mauriciodg00@gmail.com

I. INTRODUCCIÓN

El futuro de la tecnología de algoritmos apunta a que cada vez van a haber más y más modelos predictivos. Estos modelos son conocidos como algoritmos de inteligencia artificial, y actualmente se utilizan para identificar patrones, crear predicciones, optimizar procesos, entre otros.

El gran problema de los algoritmos de inteligencia artificial es que requieren una enorme cantidad de datos para poder hacer sus predicciones, después de todo, ¿cómo podría saber google lo que te gusta si no tiene muchísimos datos recopilados?. Claramente, esta gran cantidad de información requiere demasiado poder de procesamiento, ya que solo son datos crudos.

Para poder facilitar este proceso, todos estos datos crudos son procesados por técnicas de procesamiento de datos masivos (Big Data) por medio de un proceso llamado ETL. Esto permite que, en vez de tener muchísima “raw data”, se tengan pocos estadísticos que sean muy descriptivos para el fenómeno que se quiere entender y predecir con estos algoritmos.

El objetivo de este examen es poder generar los estadísticos dados un archivo de datos crudos que se tiene. Aquí, obtendremos lo siguiente:

1) Medidas de valor central:

- Media. $\mu = \frac{1}{N} \sum_{n=0}^{N-1} X_n$
- Mediana. $med(sort(X_n))$
- Moda.

2) Medidas de dispersion:

- Desviación estandar $\sigma = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (X_n - \mu)^2}$
- Rango. $\min(X), \max(x)$
- Distancia intercuartil. $Sort(X[95\%]) - sort(X[25\%])$

3) Histograma de los datos

->Sort(X) counts(per range)

Este es uno de los primeros pasos en cualquier algoritmo de inteligencia artificial, y es importantísimo para poder describir el fenómeno de la manera más completa, pero que a su vez no tenga datos inútiles. Se espera que en el examen se logren obtener estos datos estadísticos para así comprender mejor los datos.

II. Metodología

El programa va a estar dividido en cuatro módulos principales: TDA_Statics_Controlador, TDA_Statistics_Vista, TDA_Statistics_Controlador y TDA_Statistics_Pipes. El módulo de vista va a consistir en desplegar un menú al usuario en donde se le va a poder pedir los siguientes datos:

NombreArchivo: Es el nombre del archivo, que tiene que existir dentro del directorio y ser csv.

Por otro lado, el módulo controlador se encargará de recibir los datos que recibe del módulo de vista para poder mandar a llamar al módulo del modelo, si es que existe el archivo.

III. Pseudocódigo

Módulo Vista:

La función de vista se encargará de desplegar un menú en el cual el usuario podrá ingresar el nombre de su archivo o salir del programa. Si el usuario mete un csv que no existe, se le avisará y lo retornará al menú.

@params

@returns

Módulo Modelo:

La función de modelo se encargará de recibir los datos del módulo controlador para poder obtener un resultado a partir de nuestra ecuación de predicción, además de escribir el CSV.

@params

tiempo discreto, masa de la persona, constante del resorte, tiempo máximo

@returns

valor futuro (predicción)

Módulo Controlador:

El controlador se encargará de recibir los datos de la vista, validarlos, y, si son válidos, mandarlos al modelo.

@params

NombreArchivo

@returns

Diagrama de módulos:

Diseño 1.-


```
"clear" no se reconoce como un comando interno o externo,  
programa o archivo por lotes ejecutable.
```

```
Menu:
```

- 1.-Ingresar Nombre del Archivo
- 2.-Salir

```
1
```

```
Dime el nombre de tu archivo sin extension:
```

```
dsadas
```

```
ERROR: No existe el archivo llamado dsadas.csv.
```

```
"clear" no se reconoce como un comando interno o externo,  
programa o archivo por lotes ejecutable.
```

```
Menu:
```

- 1.-Ingresar Nombre del Archivo
- 2.-Salir

```
2
```

Igual clear no jala en la terminal de windows, es "cls", pero deje todo en términos de bash de linux/mac.

Resultados de archivo histograma:

```
TDA-Exams > Examen1 > data_statistic.csv
```

1	4.000000	4
2	5.000000	45
3	6.000000	1
4		

Windows igualmente redondea los floats con .5 al convertir float->int para arriba/abajo en lugar de solo truncar la parte decimal, por eso hay un 6 correspondiente a un 5.63.

Resultados de Estadísticos.txt:

```
TDA-Exams > Examen1 > Estadisticos.txt
```

1	Media: 4.955432
2	Mediana: 5.031065
3	Moda: 4.162632
4	Desviacion Estandar: 0.280655
5	Minimo: 4.162632
6	Maximo: 5.636707
7	Distancia Intercuartil: 0.554859