

Computación Blanda

Soft Computing

Autor: Daniel Giraldo Muñoz

IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia

Correo-e: daniel.giraldo2@utp.edu.co

Resumen— Este documento presenta un resumen sobre un programa de machine learning básico, basados en la demanda de solicitudes web por hora que tiene una empresa en su página web, esto con el fin de predecir si se necesitara mejorar el límite de solicitudes permitidas para un óptimo desempeño de la infraestructura actual, pero teniendo cuidado de no asignar demasiados recursos que después no serán utilizados.

Palabras clave— algoritmo, arreglo, machine learning, solicitud.

Abstract— This document presents a summary of a basic machine learning program, based on the demand for web requests per hour that a company has on its web page, this in order to predict if it will be necessary to improve the limit of requests allowed for optimal performance. of the current infrastructure, but taking care not to allocate too many resources that will not be used later.

Key Word— algorithm, array, machine learning, request

I. INTRODUCCIÓN

Debido al exitoso crecimiento de la empresa dedicada a proporcionar el servicio de aprendizaje automático a través de http aumento la demanda de una mejor infraestructura para atender todas las solicitudes web entrantes.

Se desea asignar los recursos necesarios ya que si se asignan demasiados recursos de los necesarios sería demasiado costoso, en cambio si no se logra una buena infraestructura se perdería dinero ya que no se podría atender todas las solicitudes entrantes.

Queremos conocer cuando se alcanzará el límite de la infraestructura actual, se estima en 100.000 solicitudes por hora.

Se desea predecir en qué momento se tendrá que solicitar servidores adicionales en la nube para atender todas las solicitudes con éxito.

I.1 METODOLOGIA

Se comienza con un paquete de datos a ser procesados, tenemos una columna que nos indica el número de horas, otra columna que nos indica el número de tareas ejecutadas.

Haciendo uso de la librería numpy asignamos los valores a una variable llamada “data”, con las herramientas que nos ofrece numpy dividimos el arreglo en dos vectores, columnas x,y.

Con las siguientes instrucciones que nos permiten utilizar todas las filas con la columna cero y viceversa “x= data[:,0], y=[:,1]”

A continuación, verificamos la dimensión de los vectores con la función “n.dim” la cual nos proporciona el número de columnas, seguido de función “shape” la cual nos muestra el número de filas.

Sabemos que en un conjunto de datos se pueden presentar datos erróneos o fuera de unos parámetros establecidos, por lo tanto, debemos conocer cuántos valores erróneos existen en nuestros vectores para que no afecten los siguientes procedimientos.

Con las siguientes funciones nos permite conocer el número exacto de datos erróneos “(np.sum(np.isnan()))”

Np.sum nos devuelve una sumatoria de los valores encontrados por la función np.isnan que nos dice verdadero si un dato esta fuera de sus parámetros o falso si se encuentra dentro de sus parámetros

Ahora deseamos eliminar los datos erróneos de los vectores, negando la función “np.isnan()”

X= x[~np.isnan(y)]

```
y= y[~np.isnan(y)]
```

lo que me produce es que ya los valores erróneos son negativos, ¿por qué solo se hace esto en el vector y?, porque los valores erróneos solo se encuentran en el vector y.

importamos la librería matplotlib.pyplot para realizar las gráficas, dibujamos los puntos x, y con un radio de 10 “plt.scatter(x,y, s=10)” ponemos los respectivos nombres a sus ejes y procedemos a realizar el cálculo de las horas por semanas.

```
Plt.xticks([w*7*24 for w in range(10)], ['semana %i' % w for w in range(10)])
```

Esta función nos permite conocer las horas por semana mientras a su vez nos da la etiqueta por semana para realizar la respectiva gráfica.



I.2 CONCLUSIONES

Con este análisis y tratamiento que le dimos al conjunto de datos principales tenemos una base para partir y deducir que observamos un incremento exponencial que se dispara entre la semana 3 y la semana 4 casi afirmando que llega a la semana 5. Con estos datos podríamos proceder a calcular unas rectas polinomiales las cuales nos describan la trayectoria y el comportamiento de estos datos y de esta manera tomar una decisión acertada en cuanto a la adquisición de servidores en la nube.



Giraldo Daniel. Nació en Pereira, Colombia en 2000. Bachiller del colegio La Anunciación (2016), estudiante becado de la Universidad Tecnológica de Pereira (2016) cursando actualmente pregrado en Ingeniería en Sistemas y Computación.