

## Uso de técnicas de Machine Learning: Kmeans + SVM

AUTOR: Robledo, Nallely.<sup>a,\*</sup>

<sup>a</sup> Universidad Autonoma de Nuevo León, Facultad de Fisicomatematicas  
Coordinación de maestria MCD, Pedro de Alba S/N, Niños Héroes, Ciudad Universitaria, San Nicolás de los Garza, N.L.

**To cite this article:** Robledo, Nallely. 2023. Calorie burn prediction during exercise. Universidad Autonoma de Nuevo León, Facultad de Ciencias Fisicomatematicas 00, 1-5. <https://doi.org/10.4995/riai.2020.7133>

### Resumen

Si la gente respondiera honestamente a la pregunta '¿Cuáles son las razones por las que haces ejercicio?', una respuesta frecuente sería quemar calorías. De hecho, según el Departamento de Salud y Servicios Humanos de EE. UU. (1992), el 26 por ciento de los adultos estadounidenses entre 20 y 74 años tienen sobrepeso, lo que demuestra claramente el impacto de esta preocupación nacional.

Se sabe que la reducción de la grasa corporal puede revertir varios procesos de enfermedades (p. ej., diabetes tipo II, enfermedades cardíacas, etc.), el ejercicio aumenta el gasto calórico total y también maximiza la pérdida de grasa corporal y el mantenimiento o aumento de masa muscular, la participación en el ejercicio es una estrategia muy consecuente y gratificante para perder grasa corporal y mejorar su salud.

El ejercicio como medio para quemar calorías ha sido reconocido por la industria del fitness. Hay muchos tipos de modalidades de ejercicio que se comercializan con el reclamo de "quemar más calorías", y el consumidor se pregunta qué es lo que determina la cantidad de calorías quemadas durante el ejercicio. Esta situación es la razón fundamental para escribir este artículo.

### Calorie burn prediction during exercise -

#### Abstract

If people were to honestly answer the question 'What are the reasons you exercise?' a frequent answer would be to burn calories. In fact, according to the US Department of Health and Human Services (1992), 26 percent of American adults ages 20-74 are overweight, clearly demonstrating the impact of this national concern.

It is known that reducing body fat can reverse various disease processes (eg, type II diabetes, heart disease, etc.), exercise increases total caloric expenditure and also maximizes body fat loss and maintenance or gaining muscle mass, engaging in exercise is a very consistent and rewarding strategy for losing body fat and improving your health.

Exercise as a means of burning calories has been recognized by the fitness industry. There are many types of exercise modalities that are marketed as "burning more calories." and the consumer wonders what determines the number of calories burned during exercise. This situation is the fundamental reason for writing this article.

### 1. Introducción

La principal directriz para la creación de este artículo es representar la solución a la interrogante: *¿Cuál es el factor de mayor influencia en la quema de calorías durante el ejercicio?*

Se evalúan diferentes variables que involucran las características físicas de la persona que realiza el ejercicio, contra otras variables, que describen la manera de hacer el ejercicio. Las

variables de estudio serán: Edad, Sexo, Peso, Estatura, Pulso cardíaco, Temperatura y Duración.

En reposo, el cuerpo gasta energía para mantener las funciones de las células que son esenciales para la vida. El bombeo continuo de sangre por parte del corazón exige energía, al igual que la ventilación continua (movimiento de aire hacia adentro y hacia afuera) de los pulmones. Además, mantener un entorno de soporte vital dentro y alrededor de las células

requiere una descomposición constante de ciertas moléculas liberadoras de energía. Esta energía también se utiliza para formar las moléculas necesarias para reparar las células, almacenar energía (glucógeno y triglicéridos), combatir infecciones y procesar los nutrientes obtenidos de la digestión. Estas funciones exigentes de energía se combinan para formar la tasa metabólica basal del cuerpo, que puede variar de aproximadamente 800 a 1500 Kcal dependiendo del tamaño del cuerpo y la ingesta calórica total (cantidad ingerida de alimentos)..

### 1.1. Descomposición química

El trifosfato de adenosina (ATP) es la molécula principal que el cuerpo utiliza como medio para utilizar la energía química para realizar el trabajo celular. El ejercicio aumenta el gasto calórico del cuerpo, ya que la contracción muscular implica la necesidad de formar y descomponer ATP repetidamente. La energía liberada por la descomposición del ATP alimenta la contracción del músculo esquelético, lo que aumenta las demandas de energía del cuerpo y aumenta el gasto calórico. Las investigaciones han demostrado que durante el ejercicio el aumento del gasto calórico se debe casi en su totalidad a la contracción del músculo esquelético; el equilibrio se debe a un aumento en las demandas de energía del corazón y los músculos utilizados durante la ventilación.

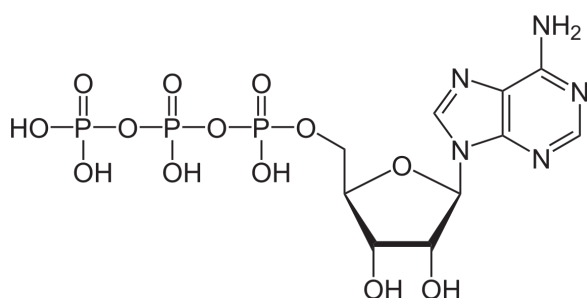


Figura 1: Estructura química del trifosfato.

## 2. Técnicas de Machine Learning

Este documento hace una recopilación de conjuntos apropiados para enseñar a nuestros modelos de aprendizaje automático para que logre saber cuál es la cantidad de calorías que el individuo gasta para quemar. De manera simultánea se divide el conjunto de datos en conjunto de entrenamiento y prueba. Aquí usamos SelectKBest, Kmeans y SVM como modelos de aprendizaje automático para comparar y luego evaluar estos modelos. La herramienta utilizada es Google Colaboratory o Google Colab es una herramienta basada en web y un servicio basado en la nube.

### 2.1. Selección de características

La primera fase del análisis de características, se presentará a través de un ANOVA, esta herramienta es de gran utilidad, pues es una fórmula estadística que sirve para comparar las varianzas entre las medias (o el promedio) de diferentes grupos. También se utiliza para determinar si existe alguna diferencia entre las medias de los diferentes grupos.

A continuación se presentan los resultados del valor de F, ordenando de las variables que tienen un valor más alto, al menos representativo.

Tabla 1: Resultados de ANOVA, F-Value por variable

Resultados de ANOVA	
Variable	F Value
Duration	157053.43
Heart Rate	62387.94
Body Temp	31855.44
Age	18.904356
Weight	366.25
Gender	7.5
Height	4.61

En base a esta evaluación se interpreta que un valor F alto, indica alta relación lineal; valores menores, lo contrario. Por lo tanto asumimos que existen 3 variables con mayor relación con la variable de respuesta *Burned Calories* y estas son Duration, Heart Rate y Body Temp.

Se busca una segunda opción para la selección de características, esto con la intención de revisar si existen resultados coincidentes, el método seleccionado es el de información mutua.

La información mutua mide la cantidad de información transferida cuando  $x^i$  = (variable de interés) es transmitido y  $y^i$  = (variable de respuesta) es recibido.

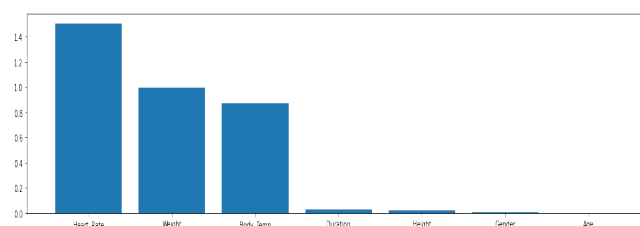


Figura 2: Título de la Figura 1. Recuerde que los títulos de figura que ocupan una sola línea van centrados. La figura es un fichero eps y gracias al paquete `epstopdf` se convierte automáticamente a pdf.

Con el gráfico anterior podemos observar que las variables Heart Rate, Weight y Body Temp comparten información mutua relevante, en comparación con el resto de las variables.

En base a lo anterior se cree que la eliminación de las variables Age y Gender pudiera ser conveniente para el agrupamiento de características.

### 2.2. Análisis de grupos

En esta sección se realizará el análisis de grupos o también conocido como clustering, es la tarea de agrupar objetos por similitud, en conjuntos de manera que los miembros del mismo grupo tengan características similares.

Es la tarea principal de la minería de datos exploratoria y es una técnica común en el análisis de datos estadísticos. Se puede realizar a través del aprendizaje automático No Supervisado y el Supervisado.

En primer instancia es importante mencionar que para esta sección del análisis, las variables ya fueron filtradas por medio

de la selección de características precedente, por lo que solo se tomarán en cuenta aquellas que aportan mayor información al modelo.

### Parte 1. Agrupamiento mediante K-Means

A continuación se mostrarán los resultados del análisis por el método KMeans. Este algoritmo es bastante popular en su uso para la clasificación no supervisada (clusterización), inicialmente agrupa objetos en  $k$  grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster.

#### 1) Definición de $K$ y número de centroides

Para definir la cantidad de clústers a aplicar, se efectúa la curva de codo. De manera lógica los primeros conglomerados agregarán mucha información, ya que los datos en realidad consisten en gran cantidad de grupos, pero una vez que el número de conglomerados exceda el número real de grupos en el data, la información agregada caerá bruscamente, porque solo está subdividiendo los grupos reales. Suponiendo que esto suceda, habrá un codo pronunciado en el gráfico de la variación explicada frente a los conglomerados, es el dato que se tomará para definir la cantidad de clústers.

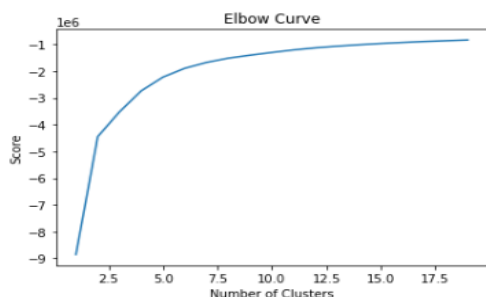


Figura 3: Gráfico de Curva de Codo.

El codo de la curva decrece en 4, por lo que este será el número de clústers a utilizar.

#### 2) Calcular distancia y asignar centroides

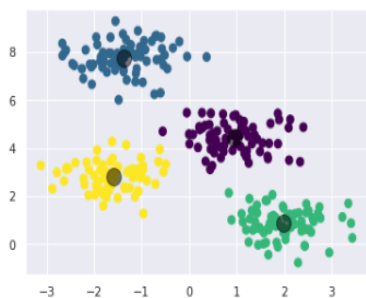


Figura 4: Agrupaciones antes del algoritmo..

#### 3) Actualizar la posición de los centroides

Weight	Duration	Heart_Rate
56.483912	15.701416	95.509974
98.374893	23.668100	105.202064
75.241119	7.905537	86.788847
89.614104	16.988573	97.105452
68.866768	24.652075	106.136277

### Parte 2. Agrupamiento mediante SVM

Support Vector Machine funciona correlacionando datos a un espacio de características de grandes dimensiones de forma que los puntos de datos se puedan categorizar, incluso si los datos no se puedan separar linealmente de otro modo.

#### 1) Kernel

Es definido como el núcleo o corazón del sistema operativo. El Kernel lineal es el que mejor se ajusta al sistema operativo el porcentaje que ofrece es de 7.4.

### 2.3. Creación del modelo matemático

PENDIENTE

## 3. Conclusiones

El objetivo de este artículo fue explicar la relación entre el ejercicio y el gasto calórico durante mediante la demostración de un estudio de caso. Si el compromiso con la actividad física relacionada con la salud es el objetivo para usted, estas son las recomendaciones:

PENDIENTE

Esperamos haberle proporcionado la información suficiente para que pueda tomar decisiones óptimas en la realización de ejercicio físico. Cabe mencionar que los beneficios fisiológicos a largo plazo del ejercicio regular de la parte superior e inferior del cuerpo no se han dilucidado completamente en los resultados de la investigación.

## Agradecimientos

Este trabajo ha sido realizado con el apoyo de los maestros de la Facultad de Ciencias Fisicomatéticas de la UANL.

## 4. Apendice

PENDIENTE

## 5. Referencias

PENDIENTE