

Detección de géneros musicales y modulación

Jhon Jairo Castaño Murcia 816015

Pablo Chaves Tabares 815013

Universidad Nacional de Colombia

Sede Manizales

5 agosto 2019

I. Resumen

En el presente documento se hará una descripción detallada de cómo se llevó a cabo el proyecto de detecciones de géneros musicales, gracias a las similitudes que presentan algunos géneros entre sus canciones es posible hacer un modelo estadístico y clasificar una canción dependiendo de su espectro en frecuencia.

Cada artista, instrumento y los distintos tonos que caracterizan a una canción, tienen un espectro determinado, por lo tanto estas similitudes al compararlos con otros espectros, nos permiten clasificarlos mediante puntos en su frecuencia.

II. Objetivos

- Obtener una base de datos a partir de archivos en formato mp3.
- Usar el análisis de componentes principales (PCA) para la reducción de dimensionalidades al momento de análisis de grandes cantidades de datos.
- Realizar el proceso de modulación y demodulación para una señal audible.
- Usar un modelo estadístico para realizar la clasificación.

III. Procedimiento

Se tienen en cuenta los siguientes ítems:

- A. Creación de base de datos.
- B. Transmisión por medio de modulación.
- C. Extracción de características con análisis de Fourier.
- D. Análisis de componentes principales.
- E. Clasificador para toma de decisión.

A. Creación de base de datos:

Debido al manejo de grandes cantidades de información es necesario tener un archivo el cual contenga la información, la cual se encuentra indexada.

Para la selección inicial de las canciones se eligieron 4 géneros musicales, los cuales fueron reggaeton, electronica, rock y salsa.

Para cada uno de los géneros se seleccionaron 5 canciones en formato mp3.

A cada una de las canciones se les extrajo 5 segmentos en diferentes partes de la canción, donde cada segmento fue de 5 segundos.

Debido a que las canciones poseen 2 canales, izquierdo y derecho respectivamente, se optó por realizar un promedio de los dos canales.

Géneros	Matriz de base datos
Género 1 (Reggaeton)	Segmento 1
	...
	Segmento 30
Género 2 (Electrónica)	Segmento 31
	...
	Segmento 60
Género 3 (Rock)	Segmento 61
	...
	Segmento 90
Género 4 (Salsa)	Segmento 91
	...
	Segmento 120

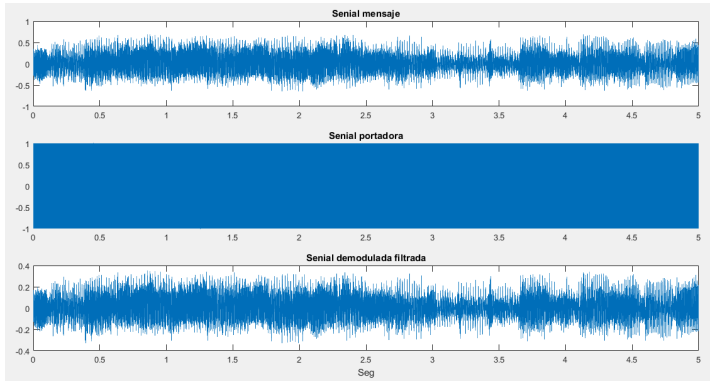
Ya que la frecuencia de muestreo fue elegida de 44100, según Nyquist donde la frecuencia de muestreo para este caso es de 44100 porque el espectro audible va hasta los 20K Hz, por tanto el tamaño de cada segmento para 5 segundos es de $44100 * 5 = 220500$

B. Transmisión por medio de modulación:

El proceso de modulación se realizó por medio de una simulación. Donde la señal mensaje es la canción nueva la cual se desea transmitir, por medio de la multiplicación de una señal portadora de alta frecuencia. La alta frecuencia es

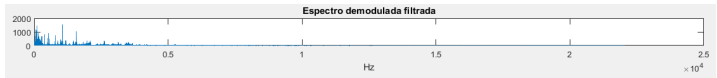
necesaria en cuanto a transmisión por medio de antenas, debido a que a mayor frecuencia, más pequeña es la antena, lo cual reduce costos.

A continuación se observan las etapas más importantes de la transmisión en tiempo.



Es más claro observar por medio del espectro de fourier la modulación, con objeto de ver cómo se desplaza el mensaje en frecuencia, justo en la frecuencia donde se encuentra la señal portadora la cual es una señal senoidal.

Al obtener la señal demodulada por medio de demodulación síncrona, se puede observar que dicha, se encuentra con un espectro remanente, el cual es el mismo mensaje desplazado en dos veces la frecuencia de la señal portadora, por lo cual se realiza un filtro con el fin de obtener únicamente el mensaje.

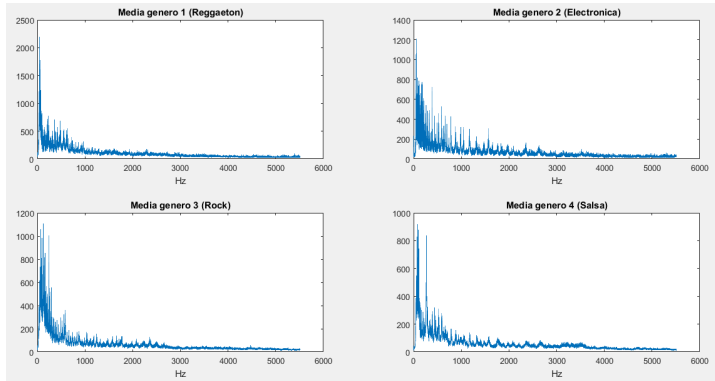


C. Extracción de características con análisis de Fourier:

Para el proceso de generar características para la clasificación de los datos es necesario utilizar Fourier, el cual nos indica las características claves del sonido en el espectro frecuencial, que comparado con la representación temporal, nos permite distinguir muchas de las características de los géneros musicales. Para cada género, las componentes espectrales en frecuencia son distintas unas de otras, debido a que los instrumentos, la velocidad de bits por minuto y la intensidad son distintos entre sí, por tanto la base de datos va tener la misma estructura pero en el espectro de fourier. A continuación se observan algunos datos clave en la representación espectral de algunos de los instrumentos:

Instrumento	Frecuencia fundamental	Armónicos
Clarinete	165-1568	2-10 KHz
Trompeta	165-988	1-7.5 KHz
Trombón	73-587	1-4 KHz
Tambor	100-200	1-20 KHz
Platillos	300-587	1-15 KHz
Violín	196-3136	4-15 KHz
Cello	65-698	1-6.5 KHz
Bajo eléctrico	41-300	1-7 KHz
Guitarra acústica	82-988	1-15 KHz
Guitarra eléctrica	82-1319	1-3.5 KHz
Piano	28-4196	5-8 KHz
Saxo Soprano	247-1175	2-12 KHz
Saxo alto	175-698	2-12 KHz
Cantante	87-392	1-12 KHz

En el siguiente gráfico se obtuvo un promedio del género en Fourier representado como un vector para los cuatro géneros:



Se observa que las componentes frecuenciales de cada género varían notablemente. En el caso de la música electrónica se encuentran picos de energía en los 1600 Hz, a diferencia de los demas generos. Debido a que dichas características se encuentran repetidas, es decir, para cada género existen componentes frecuenciales parecidas, se encuentra redundancia en la información, por tanto es necesario realizar un proceso para descartar información repetida, este proceso es llamado PCA.

D. Análisis de componentes principales

El análisis de componentes principales consiste en un método de reducción dimensional, usado para reducir la dimensión de los datos, transformando grandes cantidades

de datos en pequeñas cantidades, aun conteniendo la mayor parte de la información de los datos.

- **Estandarización:**

En el proceso es crítico realizar un proceso de estandarización del rango de las variables, para que todas contribuyen igualmente al análisis. Por ejemplo, si hay grandes diferencias entre los rangos de las variables iniciales, esas variables con grandes rangos, dominarán sobre los rangos más pequeños.

Matemáticamente esto se hace sustrayendo la media y dividiendo entre la desviación estándar de cada valor de cada variable.

$$Xstd = \frac{X-u}{o}$$

Donde:

X es la matriz base de datos

u es la media calculada por columnas (un vector)

o es la desviación calculada por columnas (un vector)

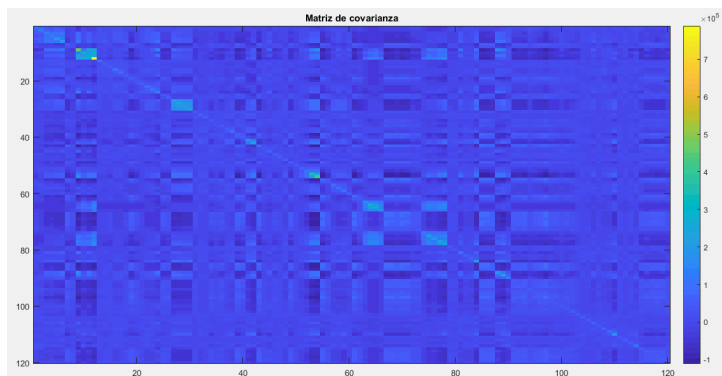
- **Computar la matriz de covarianza:**

Para este paso se debe entender como las variables de los datos de entrada van variando con respecto a la media con respecto a las otras o, en otras palabras, ver si hay alguna relación entre dichas variables. Ya que a veces las variables están muy correlacionadas, se obtiene información redundante.

Para identificar estas relaciones, se computa la matriz de covarianza.

$$Covariance\ Matrix = Xstd * Xstd^T$$

Se grafica la matriz de covarianza:



Esta matriz nos determina la relación entre cada uno de los 120 segmentos, donde la diagonal de este matriz relaciona los segmentos entre sí mismos.

- **Computar eigenvectores y eigenvalores de la matriz de covarianza:**

El objetivo de los componentes principales, es generar unas nuevas variables construidas como combinaciones lineales de las variables iniciales. Estas combinaciones están hechas para que las nuevas variables (componentes principales) entre si no están correlacionadas y la mayor parte de la información se encuentra comprimida en las primeras componentes.

Organizando la información en componentes principales de esta forma, podremos reducir dimensionalmente sin perder mucha información, descartando los componentes que posean baja información y considerando dichas componentes como sus nuevas variables.

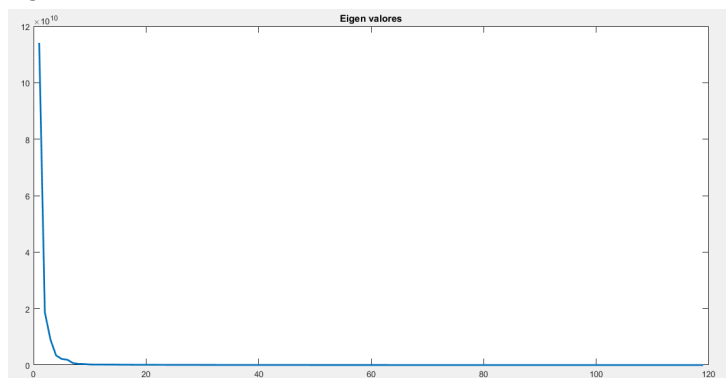
$$[Vec\ pca, Val\ pca] = eig(Covariance\ Matrix)$$

Donde:

Vec pca son los vectores propios

Val pca son los valores propios

A continuación se observa la gráfica obtenida de los eigenvalores:



El PCA intenta poner el máximo posible de información de la primera componente.

- **Construcción de PCA'S:**

Las componentes principales están construidas de tal forma que los mayores valores posibles de la varianza de los datos se encuentren en los primeros componentes principales.

Volviendo a los eigenvectores y eigenvalores, dichos vienen en pares, es decir que, para cada eigenvector, hay un eigenvalor.

- **Feature Vector**

Lo que hacemos en este paso es escoger para contener todos los componentes o descartar los menos significativos y formar una matriz de vectores la cual llamamos Feature vector.

Dicho Feature vector es simplemente una matriz la cual tiene las columnas de eigenvectores de los componentes los cuales hemos decidido mantener.

Para este caso hemos decidido mantener 100 eigen.

- **Refundir los datos a lo largo de los ejes de los componentes principales**

En los pasos previos, aparte de la estandarización, no se hacen ningunos cambios a los datos, solo se seleccionaron los componentes principales y se formó el feature vector, pero en los datos de entrada se mantiene siempre los ejes originales.

$$V = Xstd^T * Vec Pr * Val Pr^{-0.5}$$

En este último paso, se usa el feature vector formado, usando los eigenvectores de la matriz de covarianza para reorientar los datos desde los ejes originales. Esto se puede realizar multiplicando los datos originales por el feature vector, obteniendo la proyección Z.

$$Z = Xstd * V$$

- **Proyección para los nuevos datos**

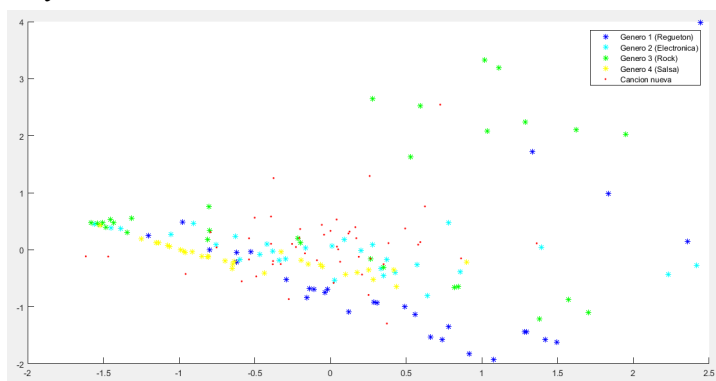
Se estandariza el vector de la nueva canción Xnew:

$$Xstd\ new = \frac{Xnew - u}{o}$$

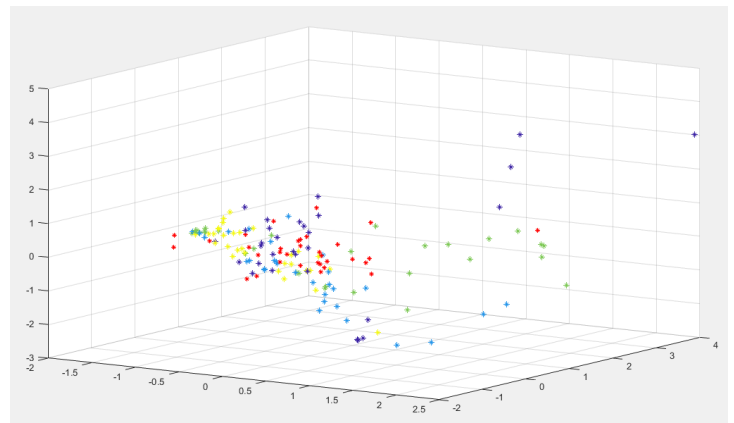
Se proyecta en Z:

$$Znew = Xstd\ new * V$$

Proyección en 2D:



Proyección en 3D:

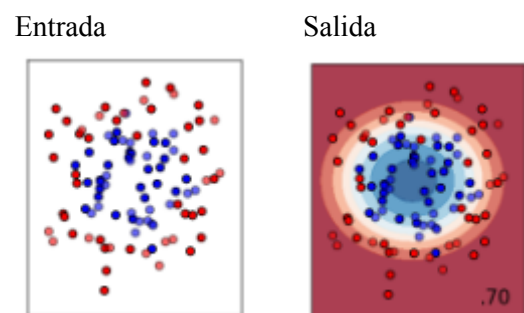
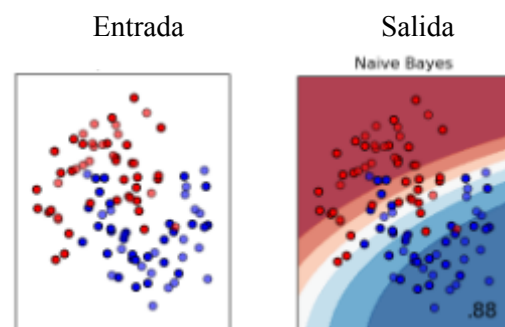


E. Clasificadores usados:

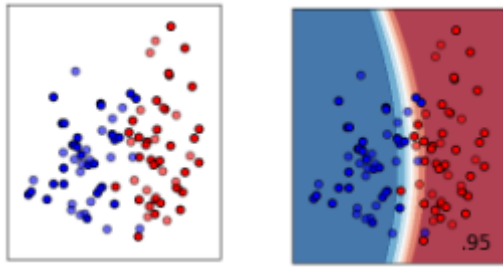
- **Bayes lineal**

Particularmente para espacios de alta dimensión, los datos se pueden separar más fácilmente de forma lineal y la simplicidad del clasificador ingenuo Bayes lineal podría conducir a una mejor generalización que la que obtienen otros clasificadores.

A continuación se observa la respuesta natural del clasificador de Naive Bayes tomando como ejemplo distintos datos de entrada:



Entrada Salida



Para realizar la detección se usa el teorema de Bayes: Evaluando cuatro probabilidades evaluando Z_{new} con respecto a cada género:

$$P(Z|Género) = \frac{1}{\sqrt{2\pi}Og} e^{-\frac{(Z_{new}-Ug)^2}{2Og^2}}$$

Donde:

Ug es la media de cada género

Og es la desviación de cada género

Dicha fórmula, es la cual matlab usa internamente.

Primero se realiza el entrenamiento del clasificador ingresando los datos obtenidos del proceso de PCA, es decir la matriz Z .

Posteriormente se obtiene la matriz de confusión con el fin de determinar la capacidad del sistema de detectar coherentemente.

Confusión	Reggaeton	Electrónica	Rock	Salsa
Reggaeton	28	0	1	1
Electrónica	0	25	4	1
Rock	0	0	30	0
Salsa	1	0	2	27

Después se ingresan los nuevos datos teniendo en cuenta el entrenamiento realizado, para al final tomar la máxima probabilidad obteniendo la decisión final.

Un clasificador es un algoritmo utilizado en el aprendizaje de máquina(machine Learning), su función es la de determinar el grupo al que pertenece o mas se asemeja determinada entrada, este algoritmo utiliza una base de datos con el que puede hacer comparaciones de determinados elementos, como lo pueden ser números , colores, formas entre otros.En nuestro caso es espectro en frecuencia lo que usamos para hacer la comparación.

Existen bastantes algoritmos de clasificación pero los que poseen características más eficientes para este tipo de elementos son:

- **K- vecinos más cercanos (algoritmo de neighbors)**

Este algoritmo busca las aproximaciones más cercanas de lo que se está tratando de categorizar y lo clasifica a partir de un punto de interés de los datos que lo rodean.

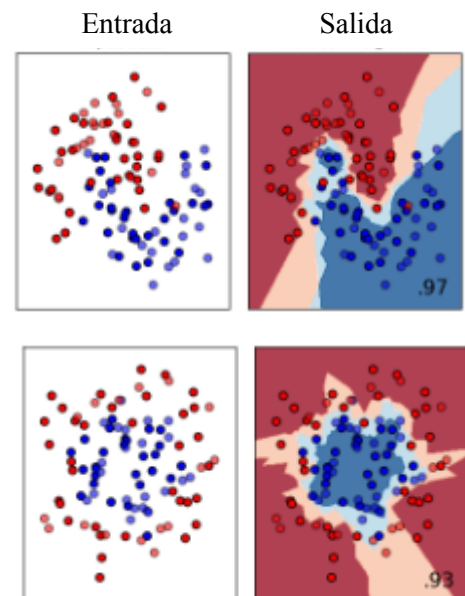
Es un algoritmo sencillo de implementar aunque tienes sus contras, uno de estos es la gran capacidad de almacenamiento y procesamiento que se necesita para hacer la clasificación , por lo que es altamente eficiente en problemas que no involucren un gran tamaño de datos.

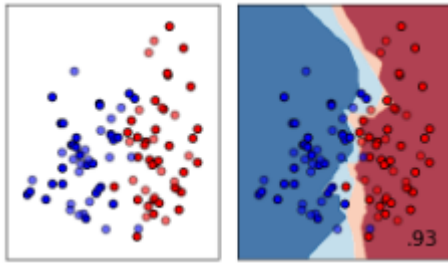
-Funcionamiento:

1. Se calcula la distancia del item seleccionado a los grupos de comparación
2. Selecciona los los elementos más cercanos
3. Determina la probabilidad de que el ítem tenga la similitud con alguno de estos.

Una de las formas más populares para medir cercanía y la que resulta más útil para el proyecto, es la medición por distancia euclidiana, hay que tener en cuenta que el concepto de lo que tenemos de “distancia” es más abstracto en este algoritmo ya que involucra más dimensiones de las que son reales, ya que con estas puede hacer las mejores clasificaciones entre las características relaciones.

A continuación se observa la respuesta natural del clasificador neighbors, para distintas entradas

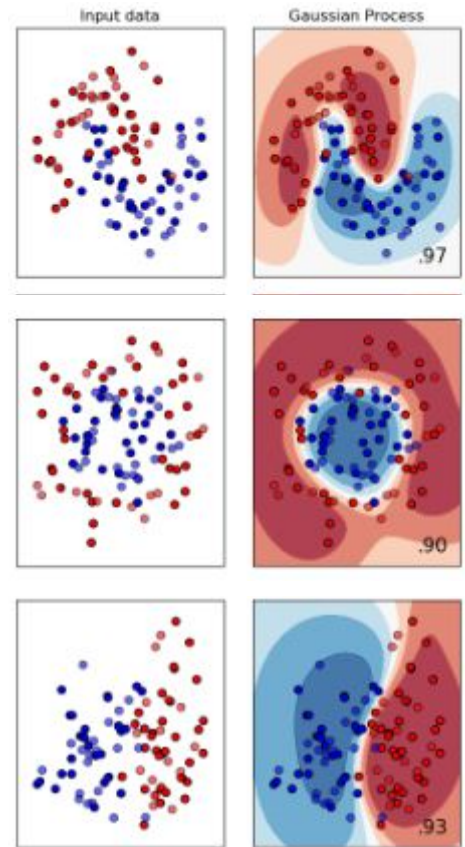
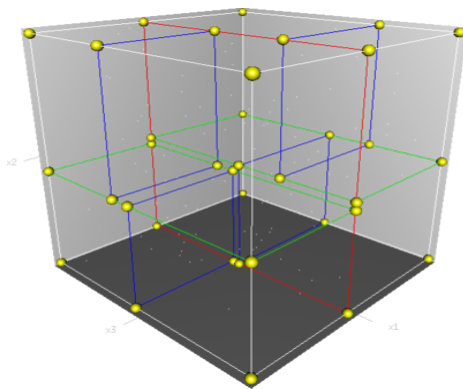




- **Kd -tree**

Los objetos modelados del algoritmo Kd-tree se encarga de almacenar los datos de la búsqueda del elemento más cercano, por medio del método de algoritmo kd. El resultado de este algoritmo arrojan resultados de cercanía, datos de entrenamiento y número de datos por nodos.

Este algoritmo Kd es más eficiente que el algoritmo de búsqueda exhaustivo, cuando K es pequeño (menor a 10), las bases de entrenamiento tienden a tener bastantes observaciones.



2. Redes Neuronales

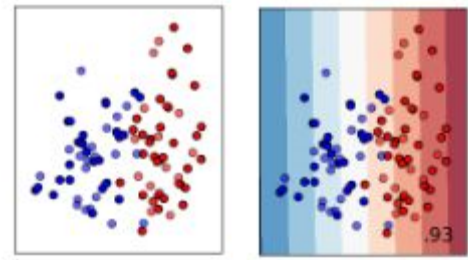
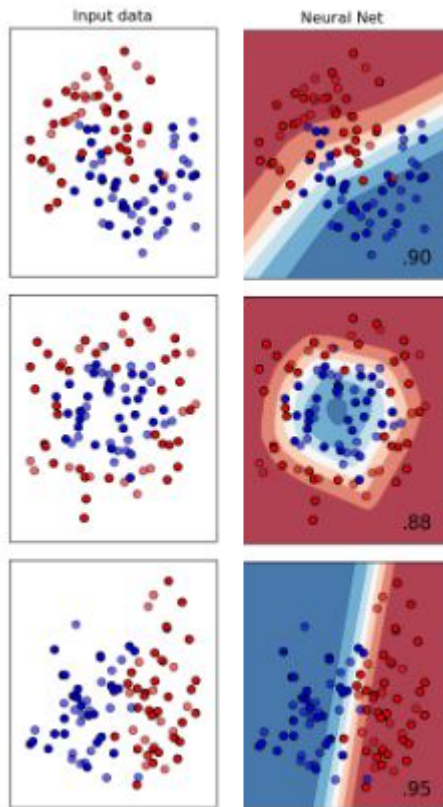
Son un conjunto de nodos estructurados en capas, cuya salida depende de la entrada y el peso que tenga, estos pesos se ajustan durante el entrenamiento, este tipo de clasificadores son buenos cuando los ítems tienen bastantes características

IV. Otros Clasificadores No usados

1. Procesos Gaussianos.

es un método genérico de aprendizaje supervisado, diseñado para resolver problemas de regresión y clasificación, su clasificación viene dada en las predicciones se dan de forma probabilística por lo que se puede formar intervalos de confianza y en base de estos hacer ajustes de línea para una región de interés.

Sus desventajas vienen cuando pierden eficiencia al usarse varias dimensiones más o menos cuando se superan la docena.



Existen muchos otros tipo de clasificadores que dependiendo de la aplicación pueden ser más o menos eficientes , principalmente estos son los clasificadores que mejor desempeñan la tarea de comparar espectros en frecuencia de las canciones, debido a una alta tasa de datos y dimensiones que se pueden llegar a alcanzar.

-Estado del arte:

como ya se ha dicho existen una gran variedad de clasificadores, cuya eficiencia dependen de la aplicación unos clasificadores son más para un conjunto de datos pequeños y bastante puntuales, en cambio cuando se posee una gran cantidad de datos y con muchas dimensiones en un espacio euclidiano, por lo tanto el clasificador que mejor se adapta es el de K neighbors o vecinos más cercanos que a partir de un grupo de puntos éste determina su cercanía y con un modelo probabilístico la categoriza.

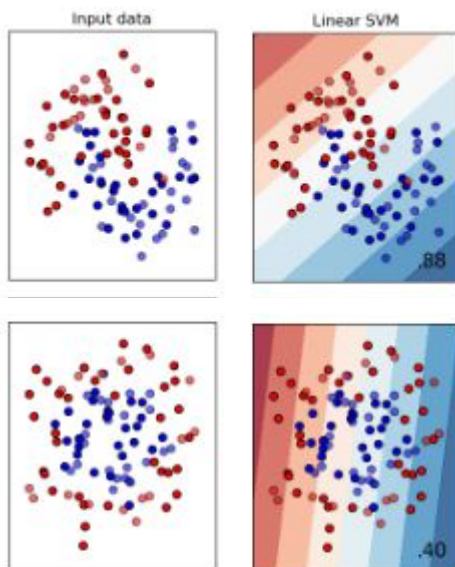
V. Bitacora montaje analogico

Para el proceso de transmisión y recepción de la señal vía aérea por modulación AM, se propusieron varios montajes y circuitos con el fin de transmitir con mayor potencia. En sí la idea es transmitir una señal de audio que oscila entre los 500 mV de amplitud, y aumentarle la potencia con una señal portadora con la frecuencia suficientemente alta para que pueda ser recibida en otro circuito de modulación , seguido por un circuito de acondicionamiento con arduino para la entrada analógica al computador, por medio de un programa de enlace de arduino podemos hacer una comunicación con el ambiente de programación que estamos usando y el arduino.

primero se usó un circuito integrado AD633 para hacer las veces de un circuito modulador AM lo más sencillo posible

3. Máquina de vectores de soporte

Es un clasificador que suma fronteras lineales en el espacio de los valores , por lo que dependiendo de un grupo de datos, se clasifica dependiendo de la frontera en la que se encuentren, estas fronteras se modelan por ecuaciones diferenciales.



recibir archivos que tengan tanto peso, solo están hechos para recibir instrucciones, el punto es que esto limitaría muchísimo la cantidad de información que necesitamos que entre para hacer la respectiva comparación.

VI. Conclusiones

- Fourier juega un factor clave en la extracción de características del proceso ya que permite distinguir en distintas frecuencias, las magnitudes.
- El clasificador de Bayes lineal, es llamado Bayes ingenuo debido a que se resume en la hipótesis de independencia entre las variables predictoras, y no es muy eficiente.
- Los componentes propios de los géneros, son valores sin dimensión física, con los que se buscan hacer resaltar las principales características de algún género, y por medio de los vectores propios marcar una tendencia.
- Los moduladores de radiofrecuencia necesitan una etapa de potencia para que se pueda realizar efectivamente la transmisión de señales en el aire.

VII. Bibliografía:

- <https://la.mathworks.com/help/stats/nearest-neighbors-1.html>
- <https://towardsdatascience.com/a-step-by-step-explanation-of-principal-component-analysis-b836fb9c97e2>