

# INTRODUCTION TO MACHINE LEARNING

## ISABELLE GUYON

### 2008-02-31

Notas tomadas por: María Eugenia Rojas

#### Qué es Machine Learning?

El proceso de aprendizaje de maquina consiste en tener una gran base de datos que se coloca como entrada a un algoritmo que se ejecuta dentro en una maquina entrenada para obtener una salida, en el proceso de utilización el usuario realiza preguntas o consultas a la maquina y obtiene respuestas.

#### Para que el aprendizaje de maquina?

El aprendizaje de maquina es útil en diferentes tipos de aplicaciones como

1. Clasificación de problemas: para predecir una salida que es binaria o categórica, esto quiere decir que tiene un número finito de valores.
2. Predicción de series de tiempo se utilizan valores continuos para predecir la salida.
3. Regresión: Se resuelven mas o menos los mismos problemas que se resuelven en la predicción
4. Clustering : No se sabe cual será el valor final, solo se trata de encontrar grupos de datos.

#### Algunos aspectos del aprendizaje de maquina

El aprendizaje de maquina incluye los siguientes métodos:

1. Modelos Lineales
2. Métodos de Kernel
3. Redes neuronales
4. Arboles de decisión

Que son los modelos más usados hasta hoy .

#### Aplicaciones

Las aplicaciones pueden ser desarrolladas de la siguiente forma:

Sobre el eje X colocamos las variables de entrada y en el eje Y se colocan los ejemplos de entrenamiento.

En este espacio de dos dimensiones se puede dibujar la complejidad de los problemas que pueden existir. Se puede tener una gran cantidad de variables con pocos ejemplos de entrenamiento lo cual dificulta la ejecución de los algoritmos de aprendizaje de maquina como en el caso de las aplicaciones de bioinformática, o se pueden tener una

cantidad mínima de variables pero muchos ejemplos como en los problemas de diagnóstico.

En el caso de las aplicaciones de análisis de mercados y ecología se tiene muchos datos de análisis y suficientes variables para realizar los procesos, este es el mejor caso de análisis.

### **Aplicaciones en Banca, comunicaciones y Ventas**

Otras de las aplicaciones están en las áreas de Banca, Comunicaciones y Ventas y sirven para identificar hechos como:

Perspectivas sobre los clientes  
Identificar clientes insatisfechos  
Identificar los buenos clientes  
Identificar los clientes que no pagan

Se desea obtener

Mensajes más efectivos sobre los clientes  
Menor riesgo al otorgar los créditos  
Disminuir los fraudes  
Disminuir la tasa de cambio de los clientes entre compañías que indica la tasa de permanencia de un cliente dentro de una compañía

### **Aplicaciones en Biomédicina y Biometría**

Otras aplicaciones en medicina ayudan a examinar a las personas para determinar el riesgo de sufrir alguna enfermedad  
Diagnosticar y pronosticar algunas enfermedades  
Ayudar en el descubrimiento de nuevas drogas

Seguridad

Resolver problemas de seguridad como:  
Reconocimiento de las facciones de una persona  
Reconocimiento de la firma, procesamiento de imágenes de las huellas digitales y procesos de verificación de una persona a través del iris  
Identificar las huellas de ADN

### **Aplicaciones en Computación e Internet**

En las Interfaces desarrolladas para los computadores se tienen

- Guías para el uso de aplicaciones
- Identificación de patrones de escritura y de voz
- Estudio de ondas cerebrales para controlar los computadores directamente sin usar los demás sentidos sino solamente el poder del cerebro.

Para las aplicaciones por Internet

- Buscadores que presentan las páginas de acuerdo al uso que de ellas realizan los usuarios empezando desde las más usadas a las menos usadas
- Filtros para el correo spam, que permiten seleccionar el correo no deseado

- Categorización de texto
- Traducciones de texto
- Recomendaciones sobre consultas de paginas

## **Cambios**

Recientemente se han realizado cambios en el espectro de la distribución de los datos y para efectos del ejemplo que se va a trabajar los datos se distribuyen de una forma uniforme que facilita la realización de experimentos

## **Diez tareas de clasificación**

Se realiza un experimento con diez personas a las cuales se les dan unos problemas para ser clasificados y se mide el desempeño de cada uno de ellos en la realización de la tarea.

Se muestran histogramas con el rendimiento de cada uno de los participantes en el experimento de clasificar estos problemas y se puede ver que Arcene realiza la tarea bastante bien, pero Dorothea no tuvo el mismo rendimiento, los datos indican que algunas personas lo hacen bastante bien y otras lo hacen bastante mal.

En la realidad lo que se quiere es que los problemas de clasificación sean realizados de la forma mas certera posible, ya que la clasificación se hace teniendo en cuenta el punto de vista de la persona que realiza el proceso lo cual la puede hacer que salga muy bien o muy mal dependiendo del criterio de cada persona.

## **Cambiando a métodos ganadores**

Se muestran las formas como una maquina puede aprender a resolver el problema de la clasificación. Existen cuatro categorias para el aprendizaje de maquina

- Métodos lineales y de kernel
- Redes neuronales
- Arboles de decisión /RF
- Naive Bayes

Se nota que cuando se usa cualquiera de los métodos los resultados en el proceso de clasificación mejoran ostensiblemente. Ninguno de los métodos es mejor que el otro

El objetivo de el aprendizaje de maquina es poder realizarlo de la mejor posible y obtener los mejores resultados,

Arboles de decisión es un método que hemos usado cada uno de nosotros en algún momento para tomar alguna decisión

El método de Naive Bayes es independiente de las entradas

El objetivo es tener la mejor decisión en el menor tiempo posible usando el método más eficiente

## **Convenciones**

Se tienen las siguientes convenciones en el eje X se representan las variables a tener en cuenta y en el eje Y cada uno de los pacientes que son ejemplo, es decir las líneas representan los valores de las variables para cada uno de los pacientes y las columnas

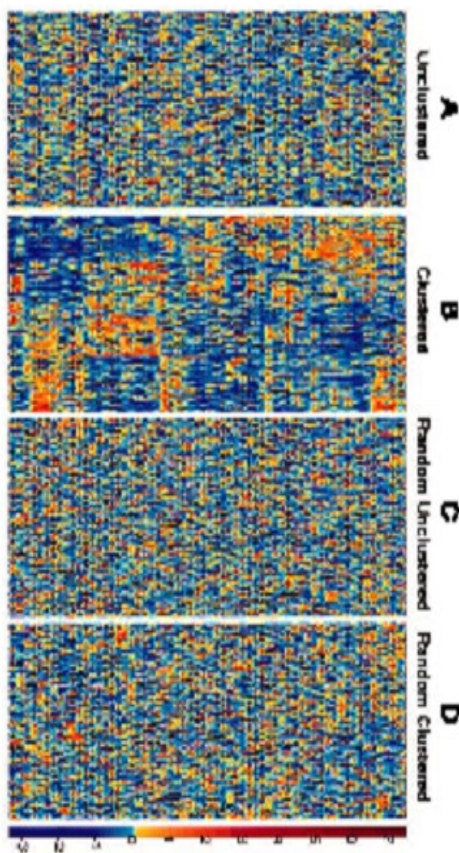
representan las variables, en el eje X están los valores de cada variable para cada uno de los pacientes que son relevantes para el estudio.

W determina la función de decisión que se usa para realizar las predicciones, este indica el peso de cada una de las variables dentro del modelo.

El vector  $\alpha$  indica el peso de cada uno de los pacientes dentro del estudio, dependiendo de las características propias de cada paciente.

El vector Y define la función objetivo que indica de alguna manera que si se ingresa un nuevo paciente es decir una nueva fila o un nuevo ejemplo el resultado que se obtendrá indicara si el paciente podrá ser saludable o no

### Problemas de aprendizaje



Dada la matriz X se quiere supervisar como aprender de los datos que aquí aparecen y ver si conservan alguna estructura o muestran algún patrón. Para poder visualizar mejor los datos se le da color naranja a los valores positivos y azul a los valores negativos. Usando operaciones de agrupamiento se trata de ver la relación entre los datos, se aplican tantas operaciones de agrupamiento como sean necesarias para poder ver características comunes a los datos y luego verificar la estructura que tienen

### Modelos Lineales

Las ecuaciones de los modelos lieneales están dadas por las siguientes ecuaciones:

- $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = \sum_{j=1:n} w_j x_j + b$

Linearity in the parameters, NOT in the input components.

- $f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b = \sum_j w_j \phi_j(\mathbf{x}) + b$  (Perceptron)

- $f(\mathbf{x}) = \sum_{i=1:m} \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b$  (Kernel method)

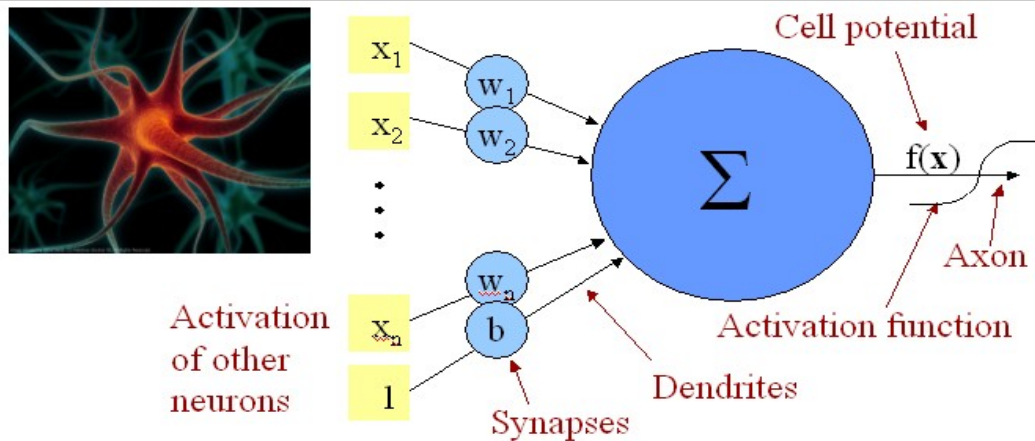
Con el uso de los modelos lineales quiere construir una función que permita realizar predicciones cuando se agregan nuevos ejemplos, en el caso de los pacientes quiere decir que ingresa un nuevo paciente que requiere ser clasificado a partir de los datos que ya existen y de sus propios datos.

Para definir esta función de predicción una forma de hacerlo es usando el modelo lineal que se obtiene a partir de la realización del producto punto entre los vectores  $\mathbf{W}$  y  $\mathbf{X}$  al cual se le suma en vector constante  $\mathbf{B}$  que depende de la forma como se define el modelo, es necesario aclarar que la linealidad a la que se refiere el modelo esta en los parámetros mas no en los valores de los coeficientes.

En el caso en que los coeficientes y los parámetros son lineales el modelo que mas se ajusta es el Perceptron, pero la definición es similar a la del modelo lineal.

En el método de Kernel se transforman las entradas teniendo en cuenta el peso de cada una de las filas y sus propios datos con referencia a los ejemplos ya existentes.

# Artificial Neurons



McCulloch and Pitts, 1943

$$f(x) = w \cdot x + b$$

## Neuronas Artificiales

Las personas tienden a imitar la forma como el cerebro realiza sus funciones para implementar este proceso a las maquinas, con el fin de que las maquinas puedan realizar predicciones. La mínima unidad dentro del cerebro es la neurona y se trata de determinar como realiza cada una de ellas el proceso de aprendizaje.

Lo que se tiene son las variable de entrada, el vector  $W$  realizaría el mismo papel que se realiza en un proceso de sinopsis y la función de predicción seria la misma función de activación dentro de la neurona para poder tener al final el posible resultado, la función de predicción se obtiene usando el método lineal, lo cual permitiría clasificar cada uno de los ejemplos en una de dos clases, para el ejemplo de los pacientes nos permite decir si el paciente es saludable o no.

## Las fronteras en los procesos de decisión lineales

Se requiere separar los datos existentes en dos clases separadas por una línea, existen datos que claramente están de un lado o de otro pero se presentan datos para los cuales la pertenencia a un conjunto o no es bastante difusa ya que los datos en ciertas vecindades tienden a ser similares, estos son los problemas típicos de clasificación, el problema consiste en encontrar una función simple que pueda definir para cada dato a que grupo pertenece.

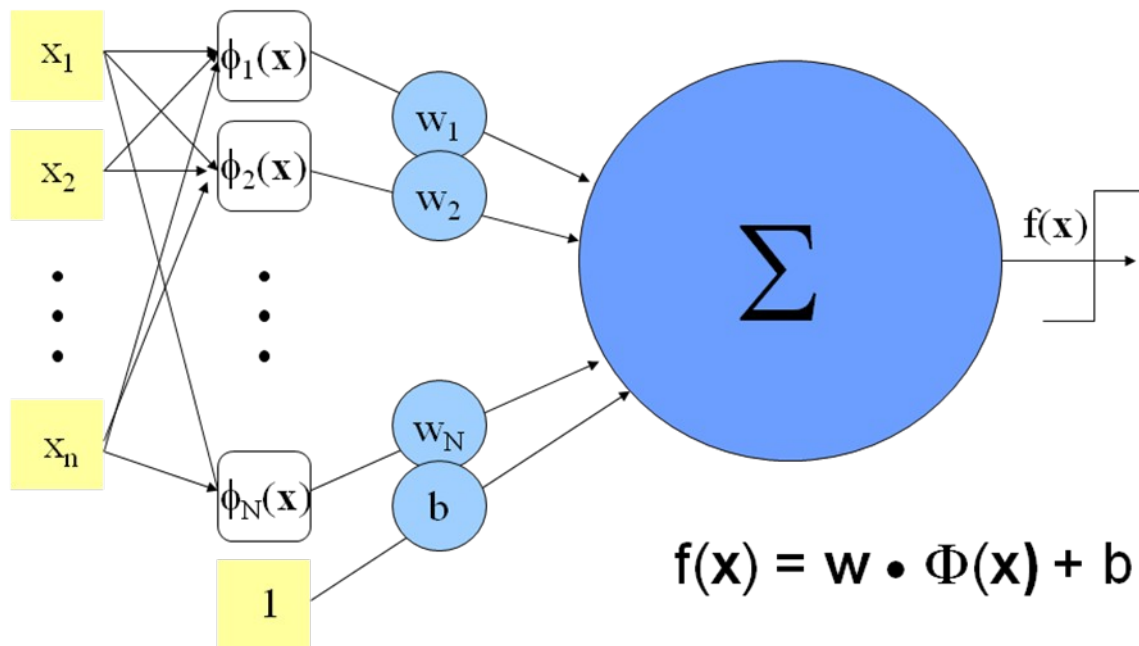
El mismo problema se presenta cuando tenemos un espacio tridimensional, solo que en este caso las variables no están cerca de una vecindad que tiene forma de línea sino que están en un plano y a medida que se agregan variables se genera el mismo problema pero presente en hiperplanos cuya dimensión depende de la cantidad de variables que se estén analizando.

En el ejemplo de los pacientes cada punto representa un paciente que tiene ciertas características dadas por los valores que toman cada una de las variables y se quiere predecir si un paciente esta o no en el grupo de personas saludables.

Sin embargo se presenta el caso de un paciente que este saludable pero por los coeficientes en las variables la función determine que está en un grupo equivocado, se pretende que este tipo de cosas ocurra con la menor frecuencia posible

### El modelo Perceptron

Gráficamente puede verse de la siguiente forma:



Este modelo funciona exactamente como el modelo lineal la diferencia esta en que las entradas corresponderán a una transformación de las entradas originales, de esto se obtiene un nuevo vector  $\phi$  que se opera con el vector  $W$  aplicando el modelo lineal.

El numero original de entradas se puede disminuir para mejorar le modelo y aplicar los procesos de regresión y obtener en la salida un valor que predice mucho mejor el grupo en el cual debe quedar el evento

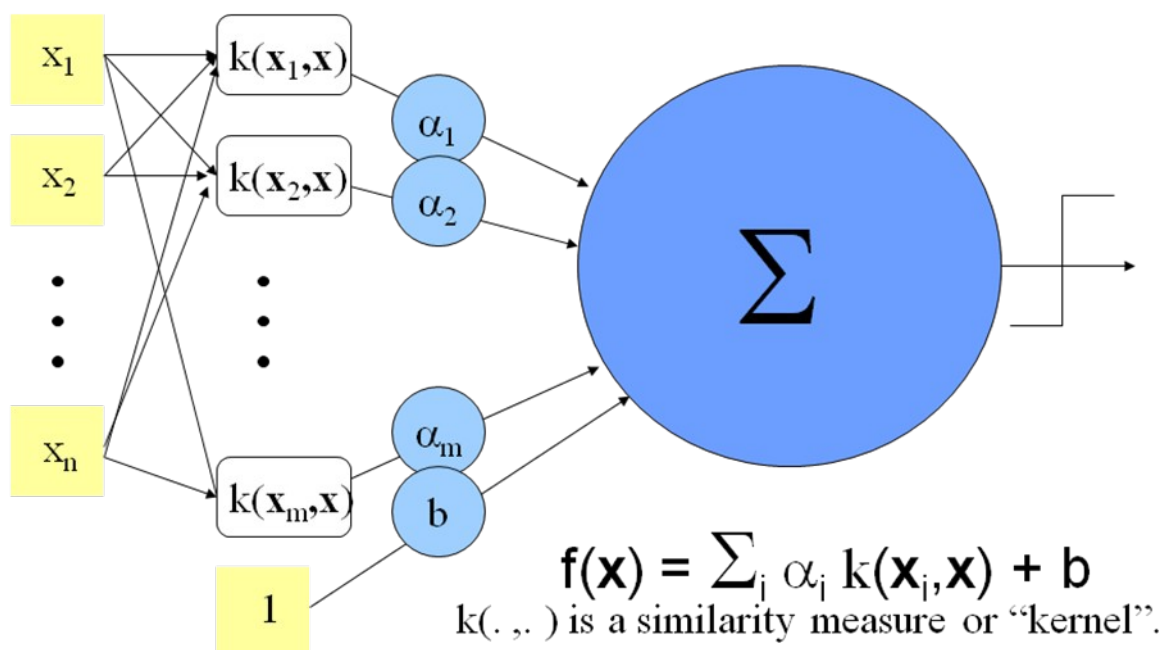
### La vecindad de decisión no lineal NL

Los datos pueden esta clasificados perfectamente en dos o mas dimensiones pero normalmente existen datos que están en ciertas vecindades y que no se pueden clasificar fácilmente.

En dos dimensiones se obtiene no una línea sino una curva y en mas dimensiones un plano que esta ondulado en diferentes partes, de acuerdo con la concentración de los datos.

### El método de Kernel

Gáficamente puede verse de la siguiente forma:



El método de Kernel es un caso especial de los modelos lineales, y es muy similar al perceptron, la diferencia esta en que las entradas son el resultado de comparar la nueva fila con los datos de entrenamiento, usando una k-función (es el resultado de aplicar el producto punto sobre cada uno de los vectores ya existentes con el nuevo vector) donde cada  $X_i$  representa los datos ya existentes y  $X$  representa el dato para el nuevo paciente.

Se realizan los cálculos aplicando el producto punto entre las componentes del vector  $\alpha$  y los datos del nuevo paciente, comparado con cada uno de los datos ya existentes, lo que se trata de determinar es, si el nuevo paciente esta o no en la vecindad de los pacientes saludables o no.

Para esto es necesario primero determinar cuantos pacientes pertenecen a cada grupo y determinar si el nuevo paciente esta cerca o no, el problema esta ahora en comparar si el nuevo registro se parece a los que están en una determinada grupo o no, dicha comparación se puede hacer de dos formas, una empírica basada en el conocimiento del tema y la otra definiendo un método adicional de clasificación que permita determinar de forma objetiva si el nuevo paciente pertenece o no a la vecindad que se analiza