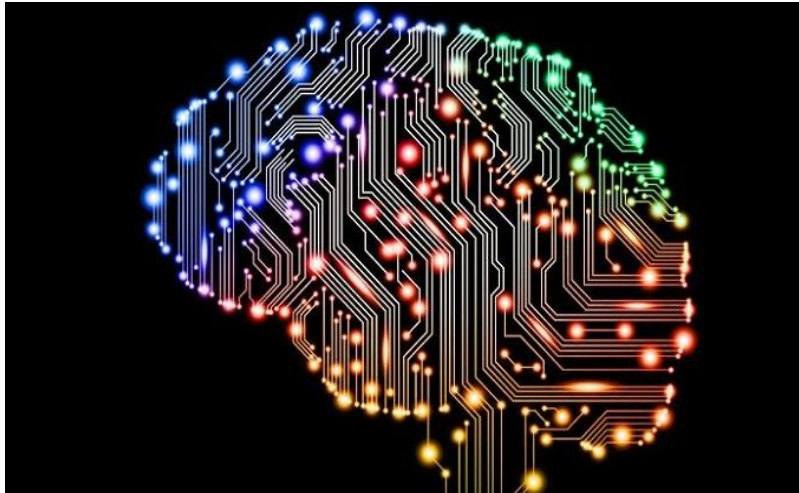


## Lección 9: fases que componen el aprendizaje



Fuente: muycomputerpro.com.

Se ha mencionado a lo largo del desarrollo del tema del machine learning qué hace éste, la importancia de los datos, que son los algoritmos y los usos potenciales de esta avanzada tecnología, pero al hablar de las fases que componen en aprendizaje automático se hace necesario profundizar en cada una de ellas con el fin de entender cómo y en qué orden deben ser llevadas a cabo.

Dependiendo de la fuente de donde se extraiga la información podemos encontrar que el proceso se divide en tres, cuatro o más fases, pero lo esencial es que se incluyan cada una de las actividades necesarias. A continuación, se presentan 7 fases del proceso de aprendizaje automático:

### **Fase 1: recolección de datos.**

Una vez identificada la problemática o tema de estudio se debe llevar a cabo la recolección de datos, esto se logra al realizar una investigación donde se determine qué datos serán necesarios y de qué manera se van a obtener, ya sea que estos se encuentren en una base de datos o que esta se deba crear. Es necesario tener en cuenta la cantidad y calidad de los datos puesto que de estos de dependerá a calidad del funcionamiento del modelo.

## **Fase 2: preparar los datos.**

Esta fase es de vital importancia ya que de la calidad de estos dependerá la credibilidad del modelo. Una vez obtenidos los datos se deben extraer las características que estos presentan y convertirlas en una secuencia de números dado que los ordenadores solo procesan este tipo de información. Por ejemplo, al trabajar con texto para hacer un análisis de sentimientos, al ordenador no se le pueden dar palabras como “bueno” o “bonito”, estas características se deben transformar en un número, lo que será en un vector de aprendizaje servirá para entrenar el sistema.

También es importante revisar si existen correlaciones entre las características de los datos, se debe hacer un balance de la cantidad de datos que se tiene para cada resultado con el fin de que esta sea representativa y no se cree algún tipo de tendencia hacia una respuesta.

Como última tarea de la fase 2, los datos deberán ser separados en dos grupos: el primer grupo de datos será utilizado para el entrenamiento, el segundo se utilizará para evaluar el modelo; esta proporción puede variar según la cantidad de datos con los que se cuente, pero generalmente será una relación de 80/20.

## **Fase 3: elección del modelo**

En esta fase los datos que fueron recopilados y preparados son recibidos y procesados por un algoritmo que sea capaz de construir un modelo que realice predicciones o lo que el usuario espera que este haga. Actualmente existen diferentes modelos que el usuario elige de acuerdo con su objetivo.

#### **Fase 4: ejecutar el aprendizaje automático**

Con el conjunto de datos de datos que fueron destinados con anterioridad a “entrenamiento”, se procede a ejecutar el modelo ingresando los datos, observando el resultado que este arroja el cual debe ir presentando mejoras incrementales en la predicción.

#### **Fase 5: evaluación.**

Esta acción se lleva a cabo con el set de datos que anteriormente fueron destinados a “evaluación”, este conjunto está formado por entradas que el modelo no conoce lo que permitirá hacer una verificación del modelo después de ser entrenado. Si al realizar esta evaluación el modelo arroja un 90% o más de exactitud se puede confiar en los resultados que el modelo otorga, en el caso contrario el modelo deberá ser ajustado.

Si la exactitud es menor o igual al 50% ese modelo no será útil ya que sería como lanzar una moneda al aire para tomar decisiones. Si alcanzamos un 90% o más podremos tener una buena confianza en los resultados que nos otorga el modelo.

El arquitecto Fabián Contreras en el documento “introducción al machine learning” ofrece las siguientes recomendaciones para mejorar el rendimiento de los Modelos de ML:

- Obtener más ejemplos de entrenamiento.
- Tratar con un conjunto de características más pequeño.
- Tratar de obtener características adicionales.
- Añadir características polinomiales ( $x_1^2$ ,  $x_2^2$ ,  $x_1x_2$ ).

- Ajustar lambda.

Para aumentar la velocidad de convergencia del Algoritmo, se debe tener en cuenta lo siguiente:

- Podemos utilizar dos técnicas para escalar los valores de las características (que todas las características tengan la misma escala de valores):
- Los rangos más utilizados son  $-1 \leq x \leq 1$  ó  $-0.5 \leq x \leq 0.5$
- Escalar las características.  $x_s = (x(i)_j) / (x_{\max} - x_{\min})$ .
- Normalización media.  $x_s = (x(i)_j - u_i) / (x_{\max} - x_{\min})$

### **Fase 6: configuración de parámetros**

En caso de no obtener las predicciones esperadas y no contar con la precisión mínima que se desea podrían estarse presentando problemas de overfitting u underfitting, lo que hará necesario regresar al paso numero 4 realizando una nueva configuración de los parámetros al modelo. Se puede aumentar la cantidad de iteraciones con los datos de entrenamiento. Existe un parámetro de gran utilidad conocido como “learning rate” o “taza de aprendizaje” que por lo regular es un valor que multiplica al gradiente para aproximarlo al mínimo global. Otra opción es poner un máximo error permitido al modelo.

Este proceso puede durar horas como puede durar días, lo importante es lograr obtener un modelo de alta calidad capaz de llevar a cabo el autoaprendizaje. Cada algoritmo presenta sus propios parámetros que deben ser ajustados que además al ser combinados abrirán paso a más opciones.

## Fase 7: predicción.

En esta fase el modelo de aprendizaje automático se encuentra listo para utilizarse, se podrá ingresar nueva información y realizar predicciones si es el caso, pues ya se evidencio anteriormente que los fines pueden variar de acuerdo con el tema de estudio.



## Referencias:

- Contreras, F. (mayo de 2016). Recuperado el 03 de octubre de 2018, de [https://www.zemsania.com/recursos-zemsania/whitepapers/DTS/Machine\\_learning.pdf](https://www.zemsania.com/recursos-zemsania/whitepapers/DTS/Machine_learning.pdf)
- Ignacio, J. (11 de septiembre de 2017). Aaprendemachinelearning.com. Recuperado el 03 de octubre de 2018, de <http://www.aprendemachinelearning.com/7-pasos-machine-learning-construir-maquina/>
- Roas, A. d. (09 de diciembre de 2015). Recuperado el 03 de octubre de 2018, de <https://www.clavei.es/blog/aprendizaje-automatico-futuro-alcance/>