

In [3]:

```
%load_ext watermark  
%watermark
```

The watermark extension is already loaded. To reload it, use:

```
%reload_ext watermark  
2019-05-30T21:29:36+02:00
```

CPython 3.6.5
IPython 6.4.0

```
compiler   : GCC 7.2.0  
system     : Linux  
release    : 5.1.5-arch1-2-ARCH  
machine    : x86_64  
processor  :  
CPU cores  : 4  
interpreter: 64bit
```

Introducción a Machine Learning

Definición.

El Aprendizaje Automático consiste en una disciplina de las ciencias informáticas, relacionada con el desarrollo de la Inteligencia Artificial, y que sirve, como ya se ha dicho, para crear sistemas que pueden aprender por sí solos.

Es una tecnología que permite hacer automáticas una serie de operaciones con el fin de reducir la necesidad de que intervengan los seres humanos. Esto puede suponer una gran ventaja a la hora de controlar una ingente cantidad de información de un modo mucho más efectivo.

Lo que se denomina aprendizaje consiste en la capacidad del sistema para identificar una gran serie de patrones complejos determinados por una gran cantidad de parámetros.

Es decir, la máquina no aprende por sí misma, sino un algoritmo de su programación, que se modifica con la constante entrada de datos en la interfaz, y que puede, de ese modo, predecir escenarios futuros o tomar acciones de manera automática según ciertas condiciones. Como estas acciones se realizan de manera autónoma por el sistema, se dice que el aprendizaje es automático, sin intervención humana.

Tipos de Machine Learning

Podemos definir a estos tipos de machine learning como *Tipos de Algoritmos Machine Learning*. Se puede dividir en 3 grandes familias:

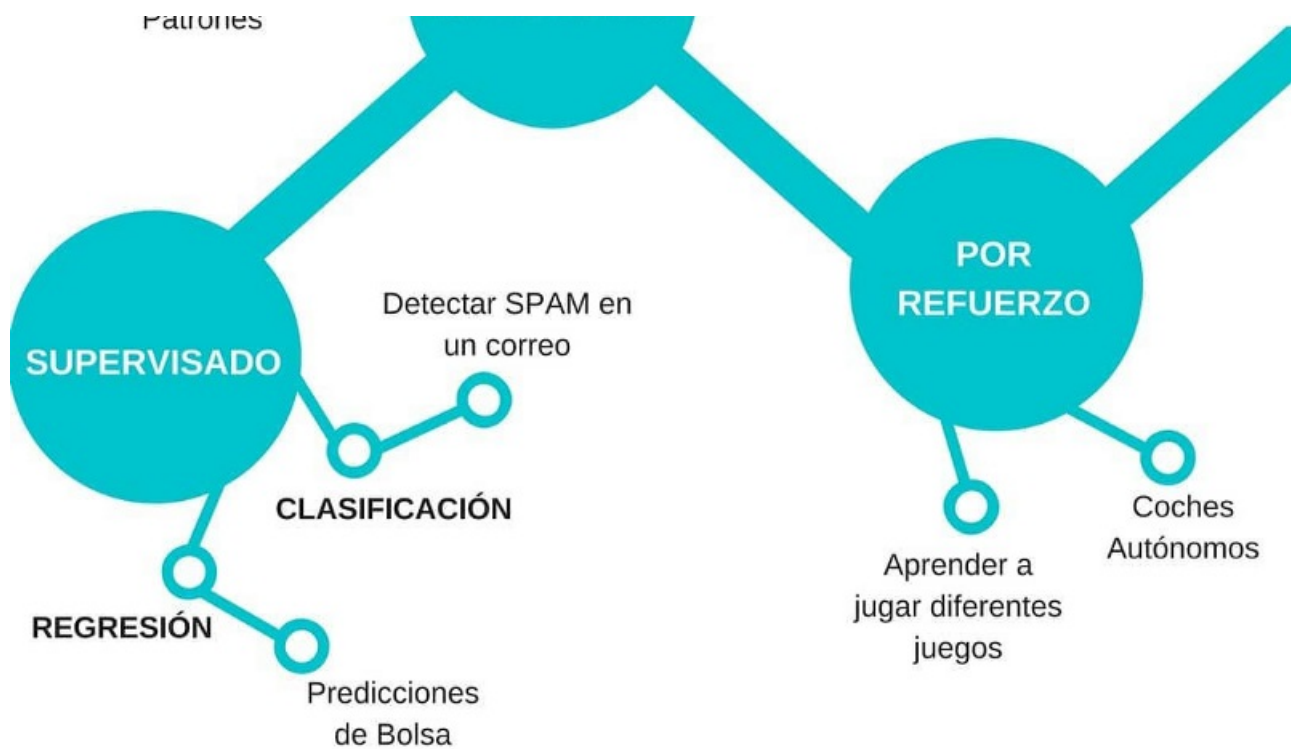
In [4]:

```
from IPython.display import Image  
Image("../..../RESOURCES/tipos_machine_learning.png")
```

Out[4]:

TIPOS DE MACHINE LEARNING





Aprendizaje Supervisado

Utilizamos la información contenida en el **dataset** de entrenamiento con el objetivo de declarar unas **labels** o etiquetas que permitan clasificar los datos (por ejemplo identificar y distinguir un coche de una moto). Este tipo de problemas, conocidos como **Problemas de clasificación**, podría ser el realizado en el conjunto de datos de la planta Iris data de los años '30, que es empleado con frecuencia como ejemplo por diferentes librerías que trabajan con datos o gráficos como pandas o el propio scikit-learn. De cada planta de la especie Iris (setosa, versicolor y virginica) se han tomado medidas de longitud y ancho de sépalo y pétalo.

Otro tipo de problema sería el que permitiese utilizar el número de habitaciones, coordenadas geográficas y tamaño total de una vivienda para estimar su precio en el mercado (**Problema de regresión**).

Después del proceso de entrenamiento, dado un elemento sin **label**, el modelo podrá determinar si se trata de un elemento u otro. A este tipo de aprendizaje se le denomina supervisado debido a que es necesario "entrenar" el modelo con muchos ejemplos para poder crear un "molde" o etiqueta para nuevos casos observados.

Aprendizaje No Supervisado.

En este caso y a diferencia del tipo de algoritmo para aprendizaje supervisado, no existen "labels" o etiquetas extraídas del ajuste del modelo, sino que más bien abstrae patrones de la información directamente para tomar las decisiones (**Problema de clustering**).

Aprendizaje por Refuerzo.

Aquí el modelo se "entrena" a partir de la experiencia. Aquí la máquina toma una serie de decisiones, de tal forma que cuando toma una mala decisión se le castiga y cuando por el contrario toma una buena decisión se le premia. Es básicamente una prueba de ensayo-error que utiliza una función de "premio" para que se vaya optimizando con el tiempo. En este tipo de algoritmos se está apostando mucho ya que no requiere de grandes cantidades de datos.

En este proyecto se realizarán pruebas con ScikitLearn en el ámbito de los problemas de regresión y de clasificación.