



# **Título**

Trabajo Fin de Máster

Escuela Técnica Superior de Ingeniería Informática

**Ingeniería de Sistemas de Decisión**

Curso 2015–2016

*José Ignacio Escribano Pablos*

Tutores:  
*Tutores*



# Índice general



# Índice de figuras



# Índice de tablas





# Introducción



# Introducción al Machine Learning

El Machine Learning o Aprendizaje Automático es la rama de las Ciencias de la Computación, y en particular, de la Inteligencia de la Artificial que se encarga del reconocimiento de patrones y de la teoría computacional del aprendizaje. El machine learning se basa en la construcción de modelos que permitan aprender y hacer predicciones sobre unos datos, de forma automática, es decir, sin intervención humana.

El machine learning permite resolver infinidad de problemas que se pueden clasificar de la siguiente manera:

- **Clasificación:** las entradas son divididas entre dos o más clases y el modelo debe aprender a asignar a cada entrada una de las clases.

Un ejemplo clásico de problema de clasificación es el filtro de spam: el modelo debe ser capaz de identificar un correo electrónico como “spam” o “no spam”, es decir, éstas serán las clases en las que se deberán dividir las entradas (por ejemplo, número de apariciones o frecuencia de distintas palabras en el correo) para identificar el spam.

- **Regresión:** las salidas son continuas. En contraposición con la clasificación la regresión tiene una salida continua, es decir, puede tomar todos los valores reales o un intervalo de ellos.

Por ejemplo, el valor de las acciones de una determinada empresa a lo largo del tiempo puede ser un ejemplo de regresión, ya que el valor de las acciones pueden tomar cualquier valor en el intervalo  $[0, \infty)$ .

- **Clustering:** un conjunto de entrada se divide en grupos (los grupos no están fijados de antemano).

Un ejemplo clásico es la segmentación del mercado, es decir, encontrar grupos con similares características de una población para ofrecer ofertas personalizadas.

- Reducción de dimensionalidad: simplifica las entradas por medio de una función a un espacio de dimensión inferior.

En el machine learning puede clasificarse por tipo de aprendizaje:

- Aprendizaje supervisado: consiste en construir un modelo a partir de un conjunto de entrenamiento que contiene los datos de entrada y la salida (o etiqueta) esperada. El algoritmo produce un modelo para inferir la salida de nuevos ejemplos.

En este tipo de aprendizaje se incluye la tarea de clasificación y la regresión.

- Aprendizaje no supervisado: en este caso no existen etiquetas predefinidas de antemano, y se trata de encontrar una función que describa la estructura oculta de los datos.

En este tipo de aprendizaje se incluye el clustering.

- Aprendizaje por refuerzo: un ordenador interactúa con un entorno dinámico para conseguir una determinada, sin que nadie le diga como de lejos está de conseguirla.

## 2.1 Aprendizaje supervisado

En el aprendizaje supervisado existe un conjunto de entrenamiento que consiste en un conjunto de datos de entrada junto con su correspondiente salida, que es la respuesta que un algoritmo de machine learning debería producir para esa entrada. Normalmente se representa como  $(\mathbf{x}, \mathbf{y})$ , donde  $x_i$  son las entradas, y  $y_i$  son las salidas.

Una característica importante de los algoritmos de machine learning es la capacidad de generalización: el algoritmo debería de producir salidas sensatas para entradas que no se introdujeron en el entrenamiento. También es importante que el algoritmo pueda tratar con el ruido, es decir, con las imprecisiones que se obtienen al medir cualquier variable del mundo real.

Dentro de este tipo de aprendizaje tenemos varios tipos de problemas, entre los que se encuentran la clasificación y la regresión.

### Clasificación

El problema de clasificación consiste en tomar las entradas y decidir en cuál de las  $n$  clases pertenece cada entrada, basado en el entrenamiento con ejemplares de cada clase.

Un punto clave en el problema de clasificación es que es discreto, es decir, cada ejemplo pertenece a una sola de las clases y el conjunto de clases cubre por completo el espacio de salida.

**Ejemplo 1.** Consideremos la clasificación de un correo electrónico como “spam” o “no spam”. En este caso el conjunto de clases vendría dado por  $C = \{\text{spam}, \text{no spam}\}$ . Las entradas podrían venir dadas, por ejemplo, por la frecuencia de aparición de distintas palabras claves del correo electrónico.

## Regresión

El problema de regresión consiste en obtener un valor de salida a partir de las entradas. En contraposición con el problema de clasificación, las salidas toman valores sobre un intervalo continuo.

**Ejemplo 2.** Supongamos que queremos estimar el valor de las acciones de una determinada empresa a partir de una serie de variables como el número de empleados, los ingresos, etc. En este caso estamos ante un problema de regresión ya que la variable salida (valor de las acciones) toma un valor continuo en el intervalo  $[0, \infty)$ .

## Regresión lineal

La regresión lineal viene dada por

$$y_i = \sum_{i=1}^n \alpha_i x_i + \alpha_0 \quad (2.1)$$

donde  $y_i$  es la variable salida y  $x_i$  es la variable de entrada.

El método de mínimos cuadrados nos garantiza que los parámetros  $\alpha_i$  que minimizan el error cuadrático vienen dados por

$$\alpha = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.2)$$

En el caso bidimensional el modelo viene dado por

$$y = \alpha_1 x + \alpha_0 \quad (2.3)$$

Se tiene que  $\alpha_0$  y  $\alpha_1$  vienen dados por este sistema de ecuaciones lineales:

$$\begin{cases} (\sum_{i=1}^n x_i^2) \alpha_1 + (\sum_{i=1}^n x_i) \alpha_0 = \sum_{i=1}^n x_i y_i \\ (\sum_{i=1}^n x_i) \alpha_1 + n \alpha_0 = \sum_{i=1}^n y_i \end{cases} \quad (2.4)$$

**Ejemplo 3.** Supongamos que disponemos de los datos de la Tabla ?? y queremos ajustar un modelo lineal de la forma  $y = \alpha_1 x + \alpha_0$ .

De acuerdo a lo anterior,

$$\begin{cases} 92\alpha_1 + 20\alpha_0 = 25 \\ 20\alpha_1 + 8\alpha_0 = 37 \end{cases}$$

Resolviendo el sistema lineal, se tiene que,

$$\alpha_1 \approx -1.607$$

$$\alpha_0 \approx 8.642$$

Por tanto,  $y = -1.607x + 8.642$ .

Si quisiéramos predecir el valor para  $x = 7$ , tendríamos que  $y = -1.607 \cdot 7 + 8.642 = -2.607$ .

Tabla 2.1: Datos para regresión lineal

$x$	-1	0	1	2	3	4	5	6
$y$	10	9	7	5	4	3	0	-1

Este método permite ajustar modelos que, en principio, no son lineales como, por ejemplo,

$$y = ax^m$$

Este modelo no puede ajustarse como regresión lineal, pero se pueden linealizar las variables  $x, y$  para convertirlo en un modelo lineal.

Tomando logaritmos a ambos lados de la igualdad,

$$\log y = \log(ax^m) = \log a + m \log x$$

Haciendo  $Y = \log y$ ,  $X = \log x$ ,  $\alpha_1 = \log a$  y  $m = \alpha_0$ , tenemos un modelo lineal.

En la Tabla ?? se pueden encontrar cómo linealizar distintos modelos.

### 2.1.1 El proceso de Machine Learning

El proceso general para resolver un problema usando aprendizaje supervisado consiste en los siguientes pasos:

1. Obtención de datos y preparación: consiste en obtener y preparar los datos que se usarán para obtener un modelo de machine learning adecuado para los datos. Consiste en obtener unos datos que sean relevantes, tarea que es difícil cuando se dispone de una cantidad de datos muy grande y que contiene outliers y datos faltantes.

Tabla 2.2: Linealización de distintos modelos

$y = f(x)$	Forma linealizada $y = \alpha_1 x + \alpha_0$	Cambio de variables y constantes
$y = \frac{\alpha_1}{x} + \alpha_0$	$y = \alpha_1 \frac{1}{x} + \alpha_0$	$X = \frac{1}{x}; Y = y$
$y = \frac{1}{\alpha_1 x + \alpha_0}$	$\frac{1}{y} = \alpha_1 x + \alpha_0$	$Y = \frac{1}{y}; X = x$
$y = \alpha_1 \log x + \alpha_0$	$y = \alpha_1 \log x + \alpha_0$	$Y = y; X = \log x$
$y = \alpha_1 e^{\alpha_0 x}$	$\log y = \log \alpha_1 + \alpha_2 \log x$	$Y = \log y; X = \log x; \alpha_1 = \log \alpha_1$
$y = (\alpha_0 + \alpha_1 x)^2$	$\sqrt{y} = \alpha_0 + \alpha_1 x$	$Y = \sqrt{y}; X = x$

2. Selección de características: consiste en la identificación de características que sean más útiles para el problema en cuestión.
3. Elección del algoritmo: dado el conjunto de datos, consiste en elegir uno o varios algoritmos adecuados que permita resolver el problema de manera satisfactoria.
4. Selección del modelo y sus parámetros: la mayoría de los algoritmos de Machine Learning tienen parámetros que deben ser fijados manualmente, o que requieren experimentación para obtener valores adecuados.
5. Entrenamiento: dado el conjunto de datos, el algoritmo y los parámetros, el entrenamiento deberá construir un modelo a partir de los datos para predecir las salidas de nuevos datos.
6. Evaluación: antes de que el sistema sea desplegado, necesita ser probado y evaluado con datos con los que no ha sido entrenado. A veces, incluye una comparación con expertos humanos en el campo y la selección de métricas apropiadas para esa comparación.

De los 6 puntos anteriores, nos centraremos en la elección del algoritmo. De entre todos los algoritmos entraremos en detalle de las redes neuronales, los Support Vector Machine (SVM) y los árboles de decisión.

### 2.1.2 Redes neuronales

Las redes neuronales fueron propuestas en 1943 por McCulloch y Pitts.

**2.1.3 Support Vector Machine****2.1.4 Árboles de decisión****2.2 Aprendizaje no supervisado****2.3 Aprendizaje por refuerzo**



## Redes parenclíticas



## **Diseño de la aplicación**

### **4.1 Tecnología utilizada**



**5**

## **Aplicaciones**



## Conclusiones

### 6.1 Mejoras y futuro trabajo





# Teoría de grafos

La teoría de grafos es la rama de las matemáticas que estudia los grafos, objetos matemáticos que constan de dos elementos: los nodos o vértices y las aristas.

A continuación, introduciremos el concepto de grafo, algunas operaciones que se pueden realizar con ellos y algunos de los tipos de grafos que utilizaremos a lo largo de la memoria.

**Definición 1.** Un grafo (o grafo no dirigido) es un par  $G = (V, E)$  de conjuntos que satisfacen que  $E \subseteq V^2$  y  $V \cap E = \emptyset$ . Los elementos de  $V$  se denominan vértices (o nodos) del grafo  $G$  y los elementos de  $E$  se denominan arcos (o aristas). Una arista entre los vértices  $x, y \in V$  se denota como  $xy$  o  $yx \in E$ .

La forma usual de representar un grafo es dibujar un punto (o círculo) por cada vértice y unir dos de estos dos puntos (o círculos) con una línea para formar un arco. Cómo estén dibujados los vértices y los arcos es irrelevante. sólo importa qué pares de nodos forman una arista y cuáles no.

**Ejemplo 4.** La Figura ?? muestra la representación gráfica de un grafo. Matemáticamente, el grafo es el par  $(V, E)$  donde

$$V = \{A, B, C, D\}$$

$$E = \{\{A, B\}, \{A, C\}, \{B, C\}, \{C, D\}\}$$

**Definición 2.** Se llama orden de un grafo  $G$  al número de vértices de dicho grafo. Se denota como  $|G|$ .

Un grafo  $G$  se dice que es finito si  $|G| < \infty$ . Si  $|G| = \infty$  se dice que el grafo  $G$  es infinito.

**Ejemplo 5.** El grafo de la Figura ?? es un grafo finito, puesto que el número de vértices del grafo es 4.

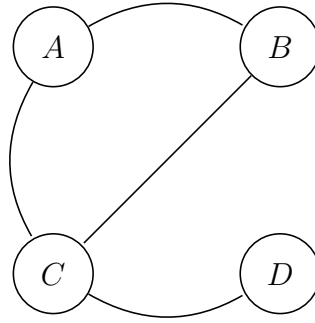


Figura A.1: Ejemplo de grafo

**Definición 3.** Dos vértices  $x, y \in V$  del grafo  $G = (V, E)$  se dicen adyacentes si existe una arista entre  $x$  e  $y$  (o  $xy \in E$ ).

**Definición 4.** Un grafo se dice completo si todos sus vértices son adyacentes.

**Ejemplo 6.** El grafo de la Figura ?? es completo ya que todos sus vértices son adyacentes. En efecto, el vértice  $A$  tiene una arista que lo une con los nodos  $B, C$  y  $D$ . De la misma forma, se comprueba para los vértices  $B, C$  y  $D$ .

Figura A.2: Ejemplo de grafo completo

**Definición 5.** Sean  $G = (V, E)$  y  $G' = (V', E')$  dos grafos. Decimos que  $G$  y  $G'$  son isomorfos, y escribimos  $G \simeq G'$ , si existe una biyección  $\phi : V \rightarrow V'$  tal que  $xy \in E \iff \phi(x)\phi(y) \in E' \forall x, y \in V$ . La aplicación  $\phi$  recibe el nombre de isomorfismo. Si  $G = G'$ ,  $\phi$  se dice que es un automorfismo.

Podemos definir operaciones sobre grafos, como la unión o la intersección.

**Definición 6.** Sean  $G = (V, E)$  y  $G' = (V', E')$  dos grafos, se definen la unión y la intersección de grafos como

$$\begin{aligned} G \cup G' &:= (V \cup V', E \cup E') \\ G \cap G' &:= (V \cap V', E \cap E') \end{aligned}$$

Si  $G \cap G' = \emptyset$ , entonces  $G$  y  $G'$  son disjuntos.

**Ejemplo 7.** La Figura ?? muestra la unión e intersección de grafos. Si  $G$  y  $G'$  son respectivamente

$$\begin{aligned} G &= (\{A, B, C, D, E\}, \{\{A, B\}, \{B, C\}, \{B, D\}, \{C, E\}, \{D, E\}\}) \\ G' &= (\{C, D, E, F\}, \{\{C, D\}, \{C, E\}, \{D, F\}, \{E, F\}\}) \end{aligned}$$

Por definición,

$$G \cup G' = (\{A, B, C, D, E, F\}, \{\{A, B\}, \{B, C\}, \{B, D\}, \{C, E\}, \{D, E\}, \{C, D\}, \{D, F\}, \{E, F\}\})$$

$$G \cap G' = (\{C, E\}, \{\{C, E\}\})$$

Figura A.3: Ejemplo de unión de unión e intersección de grafos

**Definición 7.** Sean  $G = (V, E)$  y  $G' = (V', E')$  dos grafos. Si  $V' \subseteq V$  y  $E' \subseteq E$ , se dice que  $G'$  es un subgrafo de  $G$  (y  $G$  es un supergrafo de  $G'$ ).

**Ejemplo 8.** La figura ?? muestra algunos de los subgrafos de  $G = (V, E)$  donde

$$V = \{A, B, C, D, E\}$$

$$E = \{\{A, B\}, \{A, C\}, \{A, E\}, \{B, D\}, \{B, E\}, \{C, D\}, \{D, E\}, \{C, E\}\}$$

Del mismo modo,  $G' = (V', E')$  y  $G'' = (V'', E'')$  donde

$$V' = \{A, B, C, D\}$$

$$E' = \{\{A, B\}, \{A, C\}, \{B, D\}, \{C, D\}\}$$

$$V'' = \{A, B, C, D, E\}$$

$$E'' = \{\{A, B\}, \{A, C\}, \{B, D\}, \{C, D\}, \{B, E\}\}$$

Se ve claramente que  $V' \subseteq V$  y  $E' \subseteq E$ , por lo que  $G'$  es un subgrafo de  $G$  (o  $G$  es un supergrafo de  $G'$ ). Análogo para  $G''$ .

Figura A.4: Ejemplo de subgrafos de un grafo  $G$

**Definición 8.** Sea  $G = (V, E)$  un grafo (no vacío). El grado de un vértice  $v \in V$ , denotado por  $d_G(v) = d(v)$ , se define como el número de vértices adyacentes a  $v$ .

Si todos los vértices de  $G$  tienen el mismo grado  $k$ , el grafo  $G$  es regular.

**Definición 9.** Se define el grado medio de un grafo  $G = (V, E)$  como el número

$$d(G) = \frac{1}{|V|} \sum_{v \in V} d(v) \quad (\text{A.1})$$

**Definición 10.** Un camino es un grafo no vacío  $P = (V, E)$  de la forma

$$V = \{x_0, x_1, \dots, x_k\} \quad E = \{x_0x_1, x_1x_2, \dots, x_{k-1}x_k\}$$

donde  $x_i \neq x_j \forall i \neq j$ .

Los vértices  $x_0$  y  $x_k$  se denominan final del camino  $P$ . Los vértices  $x_1, \dots, x_k$  se denominan vértices interiores del camino  $P$ .

El número de aristas del camino se denomina longitud del camino.

**Definición 11.** Un grafo no vacío  $G$  se dice conexo si cualquier par de vértices están unidos por un camino de  $G$ .

**Definición 12.** Sea  $G = (V, E)$  un grafo. Un subgrafo conexo maximal de  $G$  se llama componente conexa de  $G$ .

**Definición 13.** Un clique es un conjunto de nodos mutuamente conectados entre sí.

**Ejemplo 9.** Un triángulo es un clique formado por tres nodos.

**Definición 14.** Un grafo dirigido (o digrafo) es un par  $(V, E)$  de conjuntos disjuntos (de vértices y de aristas) junto con dos funciones  $\text{init} : E \rightarrow V$  y  $\text{ter} : E \rightarrow V$  que asigna a cada arista  $e$  un vértice inicial  $\text{init}(e)$  y un vértice terminal  $\text{ter}(e)$ .

La arista  $e$  se dice dirigida desde  $\text{init}(e)$  hasta  $\text{ter}(e)$ .

Si  $\text{init}(e) = \text{ter}(e)$ , la arista  $e$  se dice que es un bucle.

**Ejemplo 10.** La Figura ?? muestra la representación gráfica un grafo dirigido. Matemáticamente, es el par  $(V, E)$  donde

$$V = \{A, B, C, D\}$$

$$E = \{\{A, B\}, \{A, C\}, \{C, C\}, \{B, C\}, \{C, D\}\}$$

junto con las funciones  $\text{init} : E \rightarrow V$  y  $\text{ter} : E \rightarrow V$  definidas de la siguiente manera:

$\text{init} :$	$E \rightarrow V$ $\{A, B\} \mapsto A$ $\{A, C\} \mapsto A$ $\{C, C\} \mapsto C$ $\{B, C\} \mapsto B$ $\{C, D\} \mapsto C$	$\text{ter} :$	$E \rightarrow V$ $\{A, B\} \mapsto B$ $\{A, C\} \mapsto C$ $\{C, C\} \mapsto C$ $\{B, C\} \mapsto C$ $\{C, D\} \mapsto D$
-----------------	---	----------------	---

Además la arista  $\{C, C\}$  es un bucle porque  $\text{init}(\{C, C\}) = \text{ter}(\{C, C\}) = C$ .

Figura A.5: Ejemplo de grafo dirigido con bucle

**Definición 15.** Un grafo dirigido  $D = (V', E')$  es una orientación de un grafo (no dirigido)  $G = (V, E)$  si  $V = V'$  y  $E = E'$  y  $\{\text{init}(e), \text{ter}(e)\} = \{x, y\} \forall e = xy \in E$ .

**Definición 16.** Un grafo ponderado es un grafo con una función  $w : E \rightarrow \mathbb{R}$ , es decir, que  $w$  asocia un número real a cada arista. Esta función recibe el nombre de función peso.

**Ejemplo 11.** El grafo  $G = (V, E)$  con

$$V = \{A, B, C\} \quad (\text{A.2})$$

$$E = \{\{A, B\}, \{A, C\}, \{B, C\}\} \quad (\text{A.3})$$

es un grafo. Si le añadimos la función  $w : E \rightarrow \mathbb{R}$  definida de la siguiente forma,  $G$  es un grafo ponderado (ver Figura ??).

$$\begin{array}{lll} w : & E & \rightarrow \mathbb{R} \\ & \{A, B\} & \mapsto e \\ & \{A, C\} & \mapsto 5 \\ & \{B, C\} & \mapsto \pi \end{array}$$

Figura A.6: Ejemplo de grafo ponderado

**Definición 17.** Dado un grafo ponderado  $G = (V, E)$  y  $w : E \rightarrow \mathbb{R}$ , decimos que una arista  $e \in E$  incide en el vértice  $v \in V$ , si  $\text{ter}(e) = v$ .

**Definición 18.** La fuerza de un vértice en un grafo ponderado se define como la suma de todos los pesos de sus aristas incidentes.

**Definición 19.** Un grafo de dominancia  $G = (V, E)$  es un grafo dirigido tal que para todo  $x, y \in V$  se cumple una de las dos condiciones siguientes, pero no ambas simultáneamente:

- $\text{init}(xy) = x$  y  $\text{ter}(xy) = y$
- $\text{init}(xy) = y$  y  $\text{ter}(xy) = x$

**Ejemplo 12.** Si consideramos el grafo de la Figura ??, vemos que para todo vértice  $x, y \in \{A, B, C, D\}$  se cumple alguna de las dos condiciones anteriores, pero no ambas simultáneamente. Por ejemplo, para los vértices  $A, D$  se cumple que  $\text{init}(\{A, D\}) = A$  y  $\text{ter}(\{A, D\}) = D$ , pero no se cumple la otra condición. De la misma se comprueban los vértices restantes.

Figura A.7: Ejemplo de grafo de dominancia

**Definición 20.** Llamamos grafo complementario de  $G = (V, E)$ , y lo denotamos como  $\overline{G}$  al grafo que tiene como conjunto de vértices  $V$  y como conjunto de aristas, las aristas que no están unidas de  $G$ .

**Ejemplo 13.** La Figura ?? muestra un grafo  $G$  y su correspondiente grafo complementario  $\overline{G}$ . Las aristas que no están unidas en  $G$ , sí lo están en  $\overline{G}$ .

Figura A.8: Ejemplo de grafo complementario