

Minería de Datos

IIC2433

Tópicos avanzados y cierre

Vicente Domínguez

¿Qué vimos las clases pasadas?

- Text Analysis

¿Qué veremos esta clase?

- Tópicos avanzados
- Cierre de los contenidos del curso

Tópicos Avanzados

- Las áreas de Aprendizaje de Máquinas y Minería de datos son de las que más se han desarrollado en el último tiempo.
- El aumento de la tecnología (hardware) ha permitido la implementación de modelos cada vez más complejos.
- Los resultados que han demostrado tener estos algoritmos han hecho que los científicos y la industria pongan sus ojos y esfuerzos en desarrollar más estas áreas.
- De aquí en adelante queda un gran camino por delante, pero quiero al menos contarles sobre 3 áreas que me han parecido relevantes en mi formación.
- Estas son:
 - Recommender Systems
 - Reinforcement Learning
 - Deep Learning

Sistemas Recomendadores

- Área que ha tenido un gran crecimiento en los últimos años debido al impacto de la personalización.
- Varias compañías lo tienen integrado como parte de su modelo de negocios.



Sistemas Recomendadores

- De forma simple, la idea es priorizar un conjunto de productos para un usuario, dentro de un mar de productos.
- A diferencia de todos los métodos que hemos visto hasta ahora, en los algoritmos de recomendación hay entidades que definen al problema: los usuarios y los productos.

	userId	movieId	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931
5	1	70	3.0	964982400
6	1	101	5.0	964980868

Sistemas Recomendadores

- Más formalmente, una definición matemática del problema sería la siguiente (Adomavicius et al. 2007)

$$\forall c \in C, s'_c = \operatorname{argmax}_{s \in S} u(c, s)$$

$u : C \times S \rightarrow R$, *funcion de utilidad*

R : *conjunto recomendado de items*

C : *conjunto de usuarios*

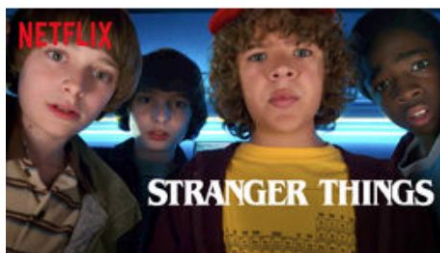
S : *conjunto de items*

Sistemas Recomendadores

- Dentro de los problemas típicos de esta área se encuentran: rating prediction y ranking prediction.
- Hay un curso completo que pueden tomar si están interesados en este tema. Sistemas Recomendadores - IIC3633
- También hay librerías con los algoritmos clásicos de predicción de rating y ranking. <https://github.com/gasevi/pyreclab>

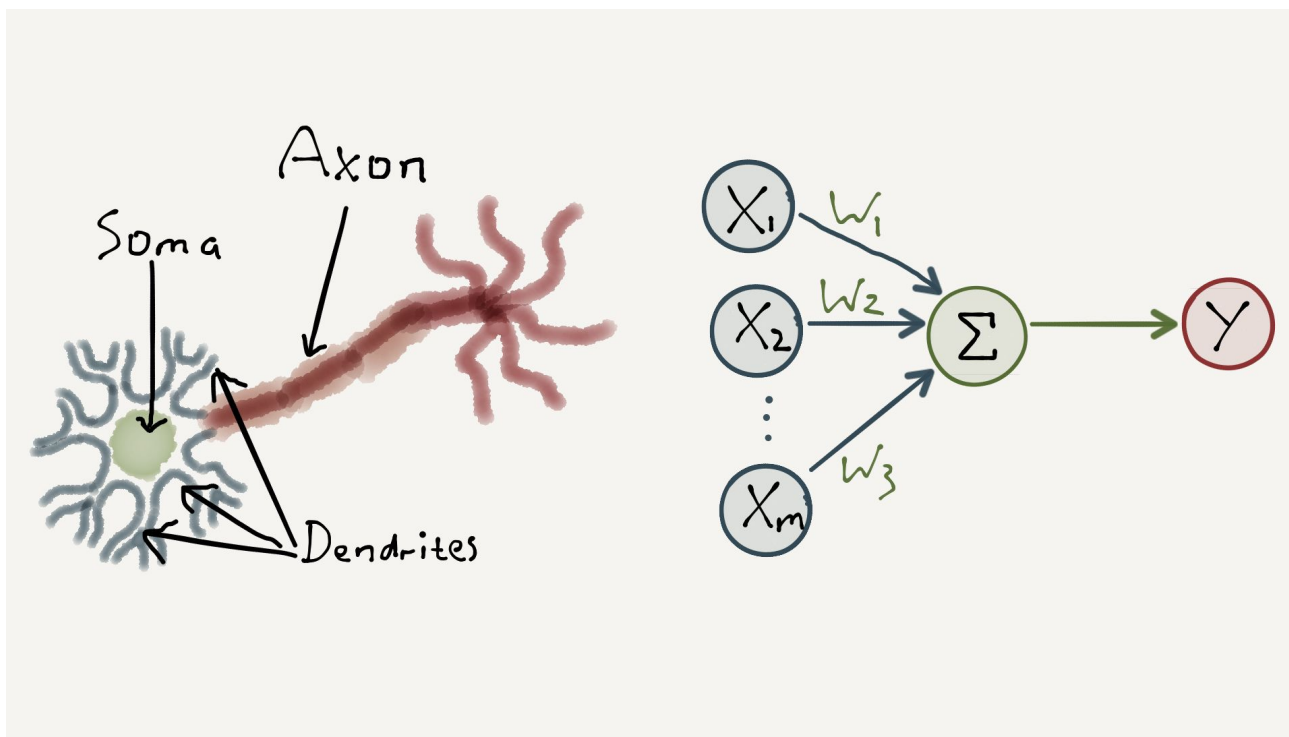
Sistemas Recomendadores (Aplicación)

- A finales del 2017 Netflix publicó un artículo donde explicaban que estaban llegando a un nivel de personalización tan granular que hasta decidían qué imagen mostrar en cada programa o película para cada usuario.
- [Post del artículo](#)



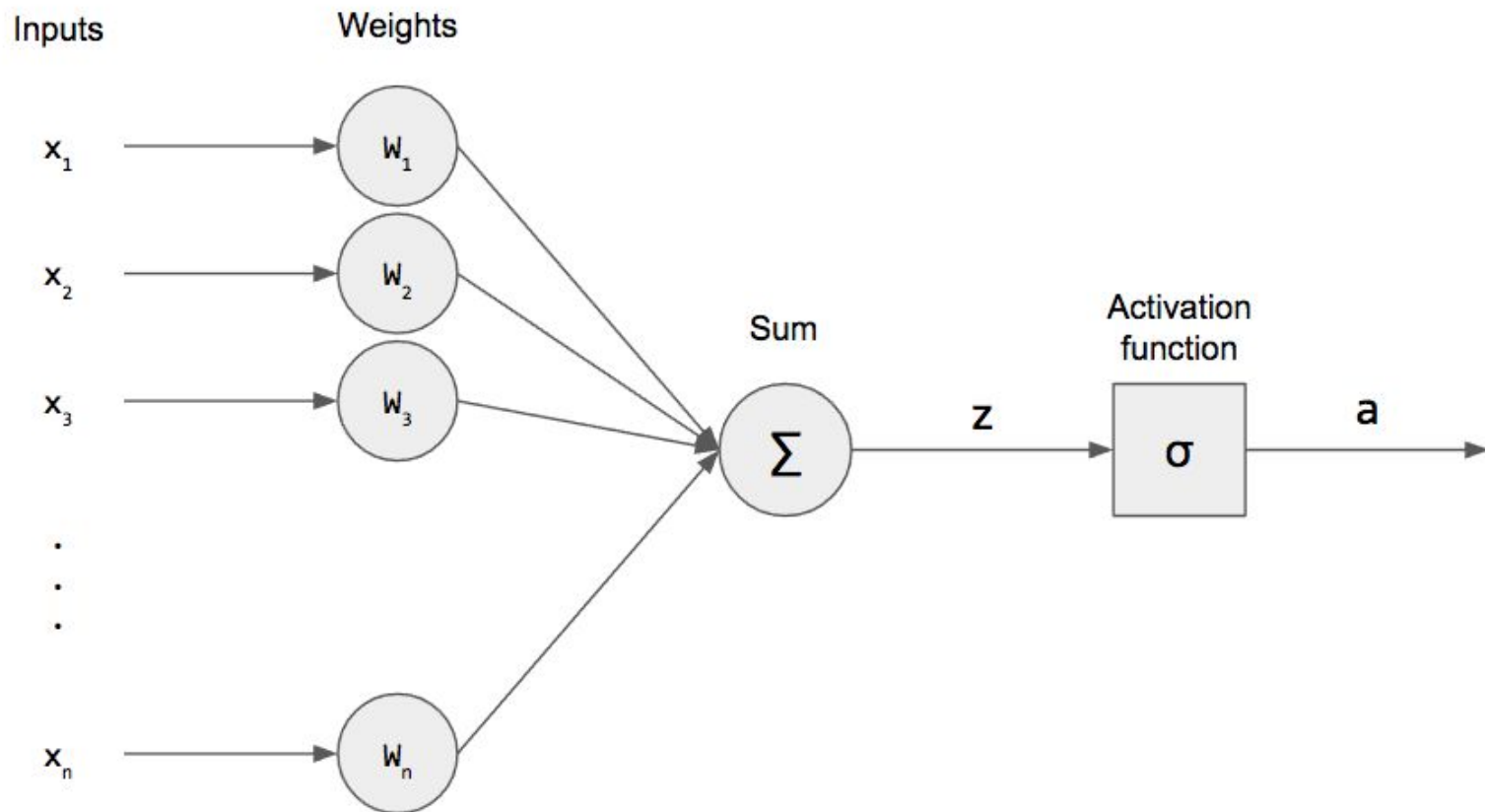
Deep Learning

- Área de Machine Learning enfocada en el desarrollo de modelos de redes neuronales profundas.
- Estos modelos tratan de modelar y simular el comportamiento de las neuronas del cerebro humano, en el proceso de aprendizaje.



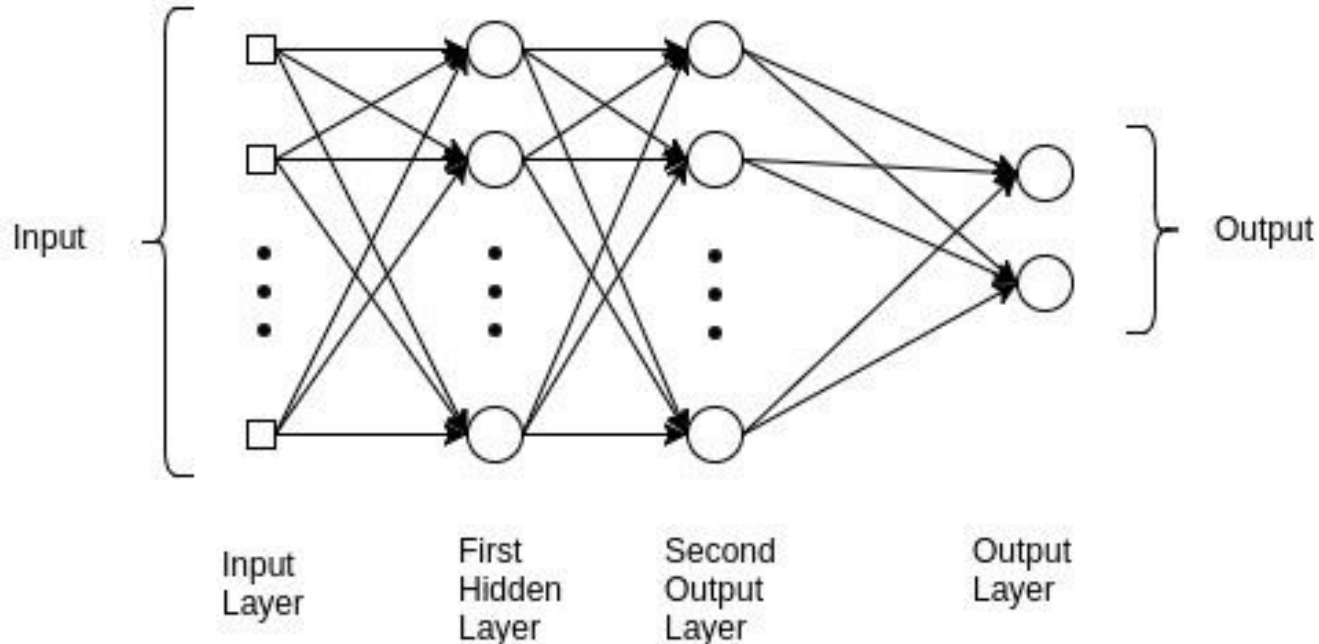
Deep Learning

- La unidad básica es el perceptrón, la cual es muy similar a la regresión logística con unas cuantas variaciones.



Deep Learning

- Los modelos finales consisten en una red profunda de capas de perceptrones



Deep Learning

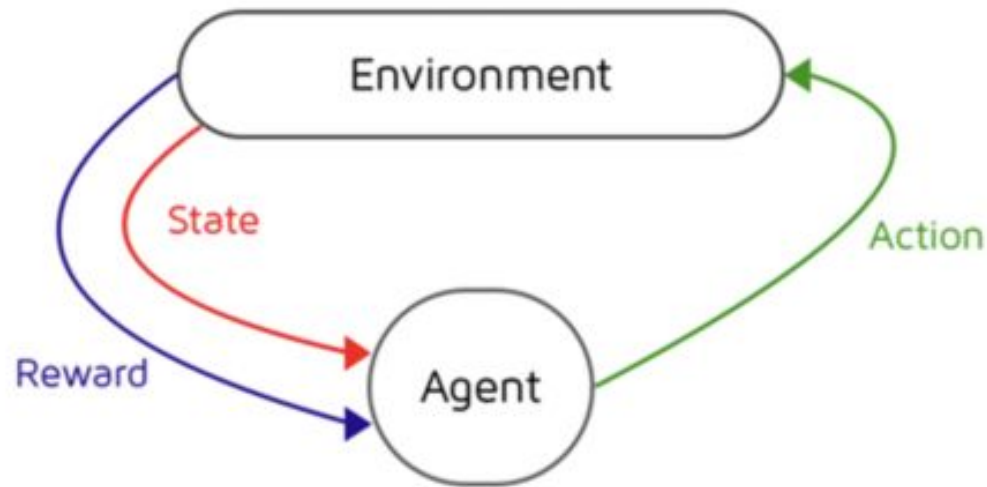
- Las ideas de las redes neuronales surgieron en los 90's, pero no pudieron ser implementadas debido a la diferencia entre la teoría y el hardware disponible.
- Gracias al desarrollo de poder de cómputo de las tarjetas gráficas, se pueden ejecutar estos modelos en paralelo en tiempos razonables.
- Actualmente nuevas tarjetas diseñadas para este tipo de tareas, son conocidas como TPUs (Tensor Processing Units).
- Si quieren saber más sobre estos temas, existe el curso de Aprendizaje Profundo - IIC3697.

Deep Learning (Aplicación)

- Una de las principales tareas que han mostrado grandes avances gracias al Deep Learning ha sido el reconocimiento facial.
- No solo eso, ya crear modelos que generen caras que nunca han existido de forma automática.
- Cada vez que ingresan a este [sitio](#), encontrarán una persona diferente que no existe.
- El diario New York Times realizó un artículo sobre este modelo, explicando cuales son las intuiciones detrás de él. Pueden ver el artículo [aquí](#).

Reinforcement Learning

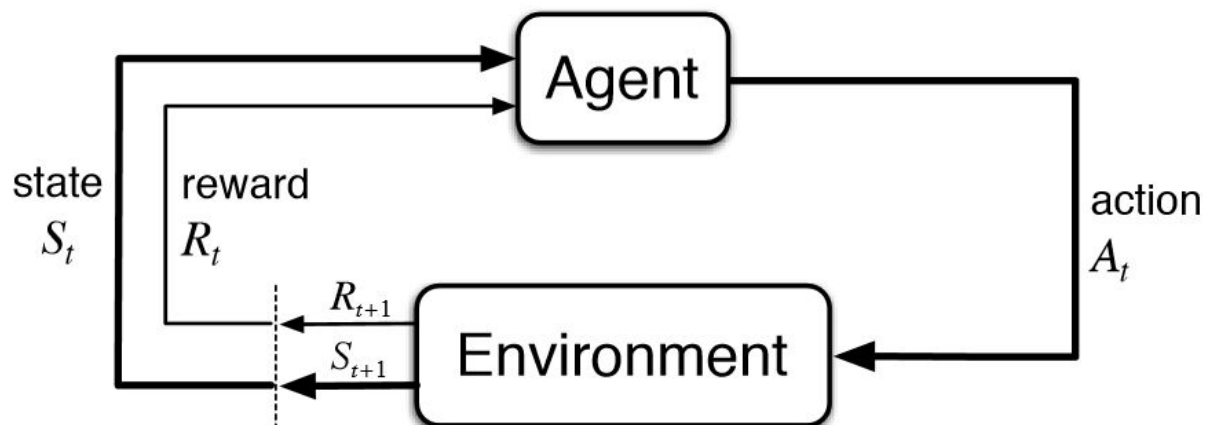
- Tipo de aprendizaje que surge de la psicología y la matemática.
- Se sustenta en el hecho de que los seres interactúan con un ambiente. Y de esas interacciones pueden surgir aprendizajes.
- Es un tipo de aprendizaje distinto a los vistos en el curso.



Reinforcement Learning

- Un conjunto de datos para este tipo de aprendizaje sería el siguiente

Timestep #	State	Action	Reward
1	S1	A1	-1
2	S2	A2	0
3	S3	A3	1
4	S4	A4	0



Reinforcement Learning

- Considere los siguientes datos



Veces jugado	70	10	30	50
% Éxito	55%	55%	25%	50%

**¿Qué máquina deberíamos jugar ahora
dada la evidencia?**

Reinforcement Learning

- Ahora, considere los siguientes datos

Ítem recomendado	1	2	3	4
Veces recomendado	70	10	30	50
% Éxito	55%	55%	25%	50%

**¿Qué máquina deberíamos jugar
ahora dada la evidencia?**

Reinforcement Learning

- Dado que tenemos acciones, y no siempre sabemos cual será la recompensa que nos dará el ambiente, surge de aquí el problema de **Explotación vs Exploración**

Exploración

Dada la evidencia, podemos decidir explorar y probar una máquina al azar.

Permite probar nuevas opciones y encontrar una jugada mejor que la evidencia.

Explotación

Podemos seguir la evidencia y jugar la máquina que mejor probabilidad de ganancia tiene hasta el momento.

Permite maximizar la ganancia del sistema.

Reinforcement Learning

- No hay cursos disponibles en la UC para aprender más de este tema.
- Pero hay libros y cursos en línea, les dejo un link a un libro muy bueno [acá](#).

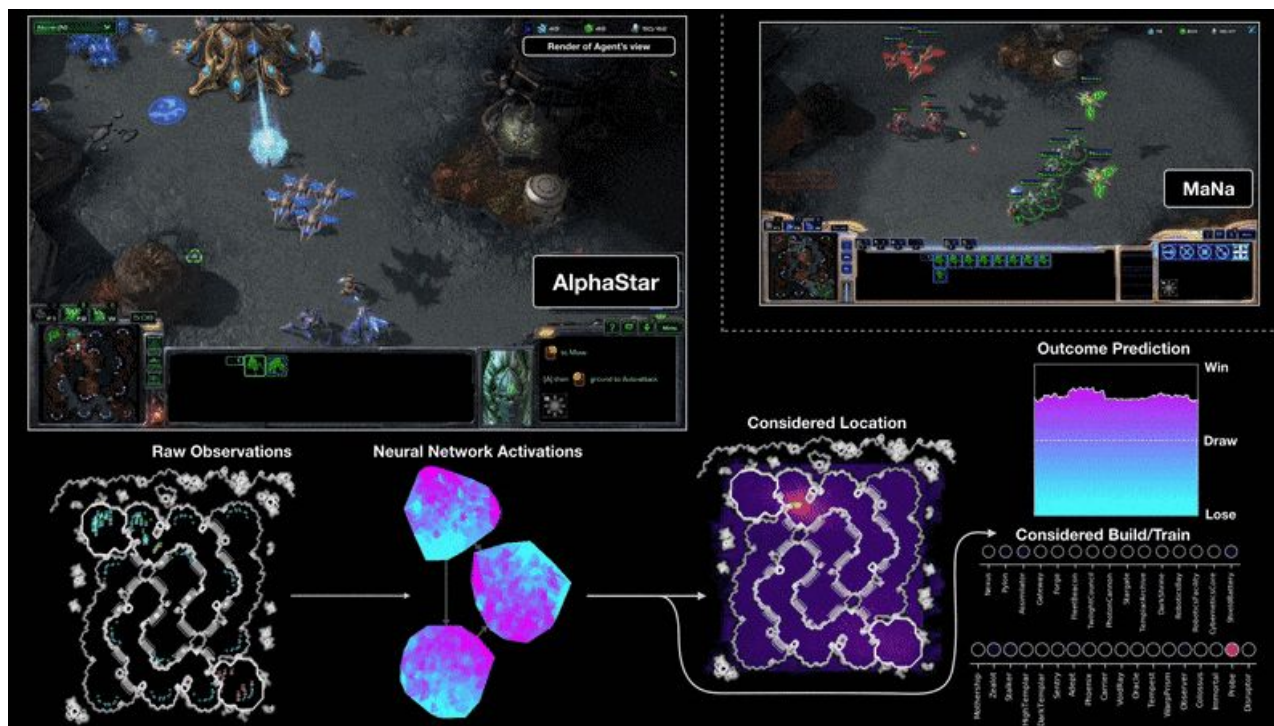
Reinforcement Learning (Aplicación)

- Una de sus aplicaciones más conocidas es en el ámbito de juegos.
- Debido a que pueden ir aprendiendo de los errores, interactuando con el ambiente, no es necesario conocer todo el espacio de acciones. Funciona bien en juegos de estrategia.



Reinforcement Learning (Aplicación)

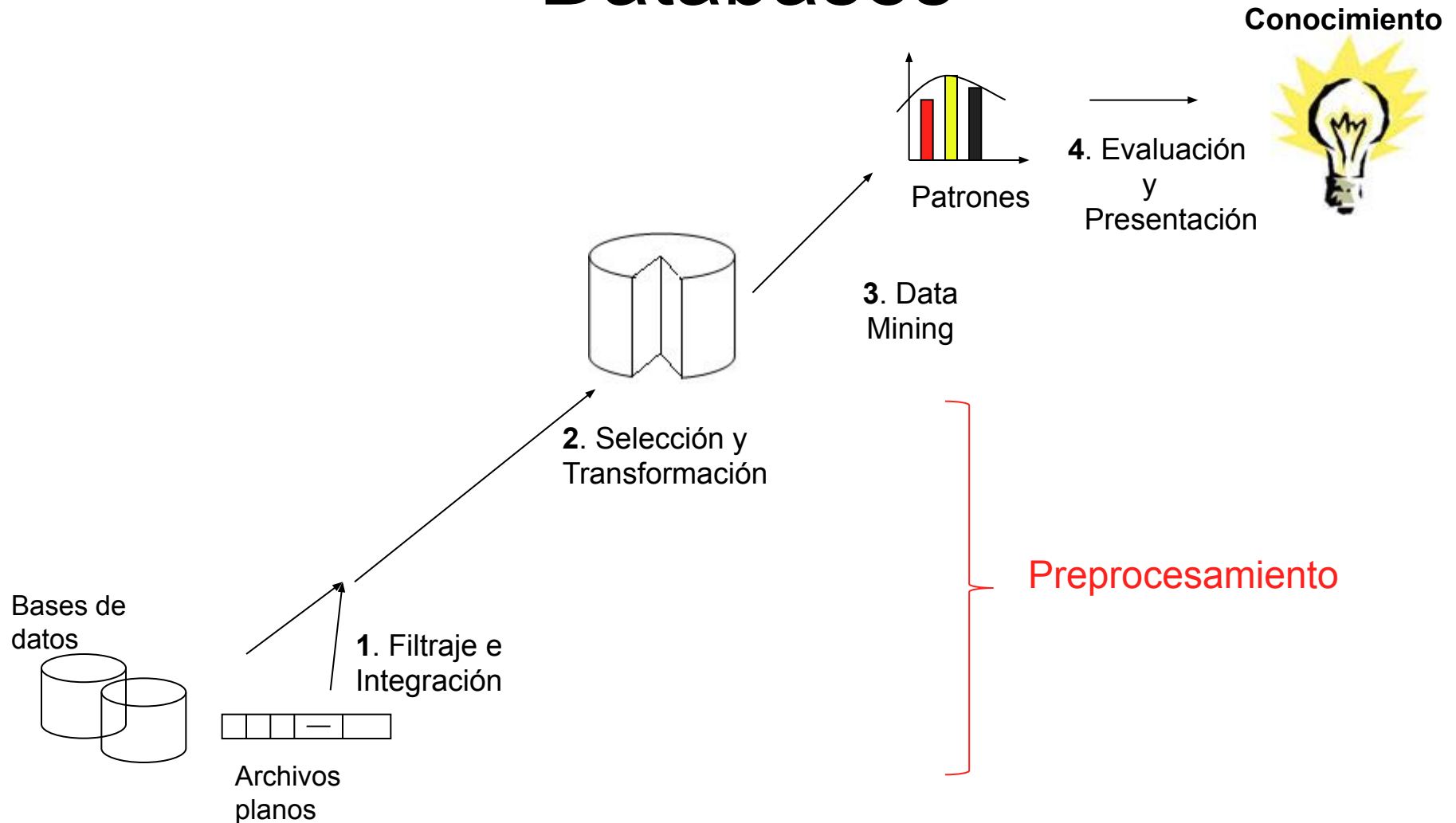
- El Go fue uno de los primeros juegos dominados por estos métodos
- Otro conocido ejemplo es el juego de estrategia StarCraft, donde un bot pudo ganarle a todos los jugadores profesionales contra los que se enfrentó
- Pueden leer un artículo sobre esto [aquí](#).



Cierre

- Durante el curso partimos desde la extracción de datos, hasta la evaluación de modelos.
- Más que aprender sobre matemática o estadística, lo más valioso es que entiendan la noción que hay detrás de todo.
- Lo importante es ser curioso, no quedarse con solo llegar y utilizar un modelo.

Knowledge Discovery in Databases



Habilidades

- Aparte de los contenidos vistos en el curso, espero que hayamos aprendido a:
 - Entender cómo funcionan los modelos de Minería de Datos
 - Poder programar algún modelo visto matemáticamente
 - Comprender y poder explicar el buen y mal desempeño de algún modelo de Minería de Datos
 - Tener la base para poder investigar en áreas más especializadas de Machine Learning
 - Poder modelar problemas y saber como poder aplicar minería de datos sobre ellos.

Aún no se acaba todo

- Ojo que aún queda el proyecto y la última tarea.
- No olviden inscribirse en su horario de presentación.
- Pronto publicaremos el enunciado de la tarea recuperativa.

Nos vemos

- Nos veremos una última vez en las presentaciones de los proyectos.
- Recuerden que cuentan con sus ayudantes para lo que sea.
- Y siempre podrán escribirme después de que dejen de ser mis alumnos, al correo vidominguez@uc.cl
- ¡Mucho éxito!

Me: *uses machine learning*

Machine: *learns*

Me:



Referencias

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6), 734-749.