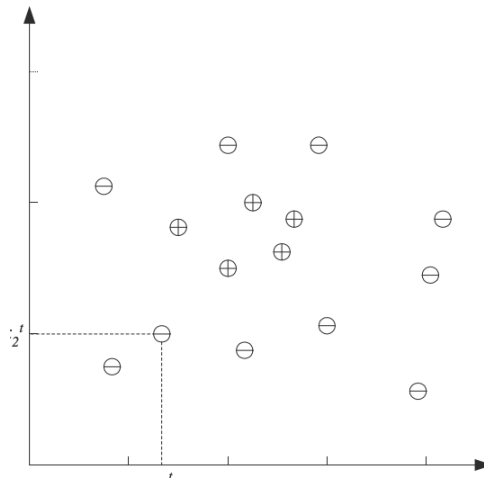


Aprendizaje Supervisado

¿Qué queremos conseguir cuando buscamos una regla de asociación?

Intentamos buscar **propiedades que nos permitan predecir la solución** a un problema de clasificación con la mayor probabilidad posible.

Es decir, el sistema de aprendizaje automático se ajusta a un modelo de datos pasados para poder calcular el riesgo de adoptar una solución al problema planteado y en función de los datos que maneja decidir aceptar o rechazar dicha solución. Por ejemplo: *en un cruce tomar una salida para llegar a un destino; en un sistema de crédito dar o rechazar un préstamo a un cliente, ...*



Para poder tomar estas decisiones existen múltiples características y propiedades algunas pueden tener una incidencia directa en el resultado, otras no. El proceso de clasificación requiere que **definamos correctamente ese conjunto de características y propiedades**, para poder alcanzar el mejor resultado posible en el proceso de aprendizaje. Esta definición suele realizarse mediante el uso de reglas. Supongamos que queremos predecir el riesgo de dar un crédito a alguien; en este caso se utilizarán reglas del estilo

IF income <= θ_1 AND savings > θ_2 THEN low-risk ELSE high-risk

que puede codificarse en la BC de AgentScript en Jason por ejemplo de la siguiente forma:

high_risk(client(C)) :-

```
income(C,X) & // Crédito que solicita el cliente
incomeLimit(Theta1) & // Límite de crédito que puede concederse al cliente
X > Theta1.
```

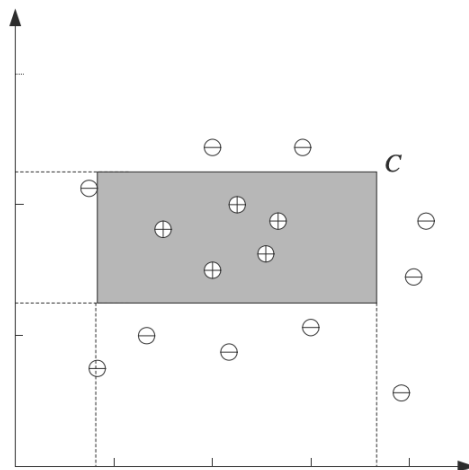
high_risk(client(C)) :-

```
savings(C,Y) & // Ahorro que posee el cliente
savingMin(Theta2) & // Ahorro mínimo que debe tener el cliente
Y <= Theta2.
```

low_risk(client(C)) :-

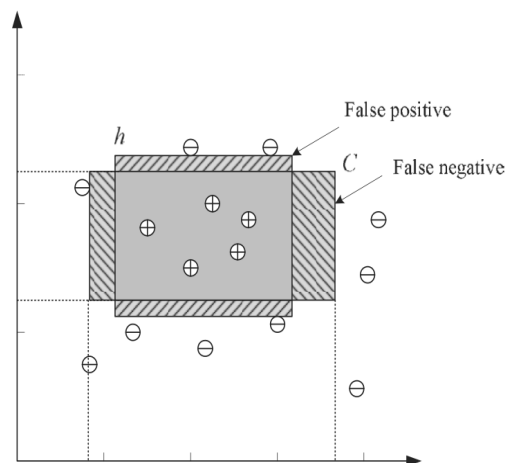
```
income(C,X) &
incomeLimit(Theta1) &
X <= Theta1 &
```

savings(C, Y) &
savingMin(Theta2) &
 $Y > \text{Theta2}$.



¿Qué entendemos por **predicción** en Aprendizaje Automático?

Una vez que tenemos las reglas que clasifican un cliente como de bajo riesgo o de alto riesgo, a partir de un conjunto de datos histórico (conjunto de entrenamiento) Asumiendo que el comportamiento de los clientes es homogéneo y que en el futuro se darán respuestas similares o equivalentes a las pasadas, entonces podemos hacer **predicciones correctas** sobre nuevos clientes.



Dada una nueva aplicación con ciertos ingresos y ahorros, podemos decidir fácilmente si es de bajo riesgo o de alto riesgo.

¿Qué diferencia hay entre los problemas de clasificación y de regresión en Aprendizaje Automático Supervisado?

La diferencia estriba en el rango de la función; en los problemas de clasificación tenemos una función boleana; mientras que en el caso de la regresión tenemos una imagen numérica (real, float, double, ...)

¿Qué se necesita en aprendizaje automático supervisado para **aprender** una clase a partir de una serie de ejemplos?

En el ejemplo anterior se precisa delimitar los valores Theta1 y Theta2 a partir del entrenamiento (casos históricos)

En este caso se puede establecer una relación entre el ahorro y el préstamo solicitado y utilizarla para instanciar adecuadamente Theta1 y Theta2

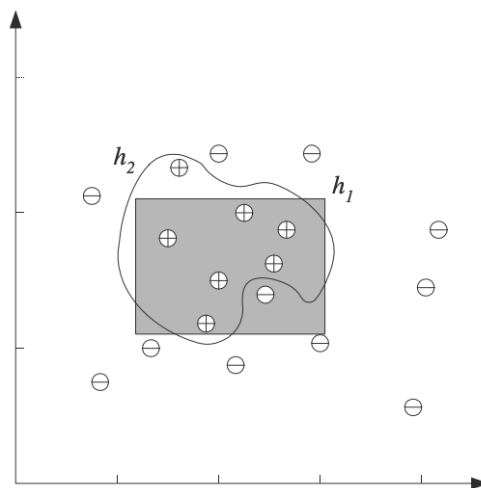
De hecho, suponiendo que los datos históricos están guardados en la BC de Jason mediante creencias del estilo:

positiveCredit(Income, Save).

negativeCredit(Income, Save).

Podríamos utilizar en Jason reglas del estilo:

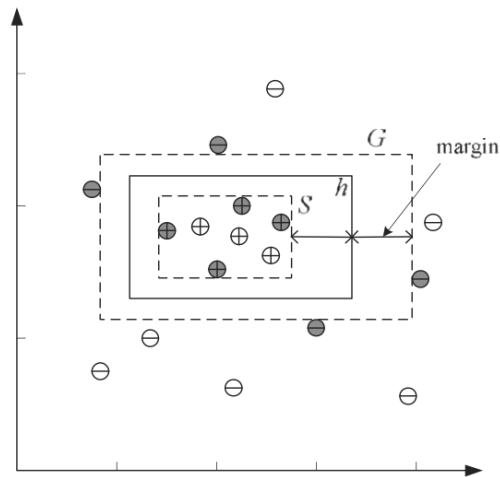
1. incomeLimit(Theta1) :-
 savingMin(Theta2) &
 predict(Theta,Theta1) &
 Theta <= Theta2.
2. incomeLimit(Theta1) :-
 savingMin(Theta2) &
 predict(Theta2,Theta) &
 Theta1 <= Theta.



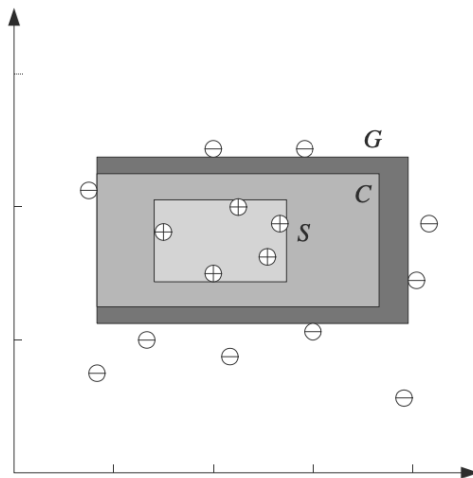
Donde **predict(A,B)** determina a partir de los ejemplos si se verifican las condiciones o no. Aquí varios casos se pueden dar:

1. Que existan prestamos por el limite solicitado en el histórico y que los casos positivos y negativos no presenten conflictos entre ellos (pueden producirse errores en la introducción de casos o existir situaciones que un experto debe aclarar) => aplicar la mejor de las estimaciones.
2. Que existan prestamos por cantidades superiores con ahorros inferiores al que se aporta y podamos estimar el riesgo como una proporción sobre el ahorro del nuevo cliente y los del histórico
3. Que no existan prestamos dados con ahorros inferiores al que proporciona el cliente
4. Que existan situaciones positivas y negativas aplicables => Habrá que ajustar y probablemente preguntar a un experto => incorporar a la BC.

Como se puede comprobar hay un problema con la zona que delimitan los casos históricos positivos y negativos



(Son un numero finito de casos sobre un continuo)



¿Cuales son las **dimensiones** que se precisan en un algoritmo de aprendizaje automático supervisado?

1. **Modelo** que usamos en el aprendizaje => Se obtiene a partir de los casos positivos y negativos.
2. **Función de pérdida** => Permite calcular la diferencia entre la salida deseada, y nuestra aproximación; determina la bondad del modelo utilizado y de la predicción obtenida. Podemos calcularlo como la suma de pérdidas sobre las instancias individuales o utilizar otra distancia que sea más adecuada.
3. **Procedimiento de optimización** => Proporciona el argumento que minimiza el error total (en regresión fundamentalmente)

