# Laboratorio 2: Markov Decision Processes, value iteration y model-free learning.

1. Cargar MDPs del queso y el estocástico. Entender la estructura de la variable *mdp*. Implementar VI y ver que dé la política y los valores correctos. Recomendación: implementar VI que calcula Q(s,a):

Inicializar Q(s,a) a valores arbitrarios (por ej, 0) para todos los estados y acciones.

Ciclar hasta que la estimacion no cambie mucho (delta<epsilon)

Para cada estado s

Para cada accion a

Q(s,a) = R(s,a) + *γ ∑s’∈S T(s, a, s’) V(s’)*

V(s)=max\_a Q(s,a)

1. Implementar Q-learning y correrlo en el queso y el estocástico. Si tu código usa mdp.T o mdp.R, estás haciendo algo mal :-)

