Generalisation and Function Approximation
Lecture 14

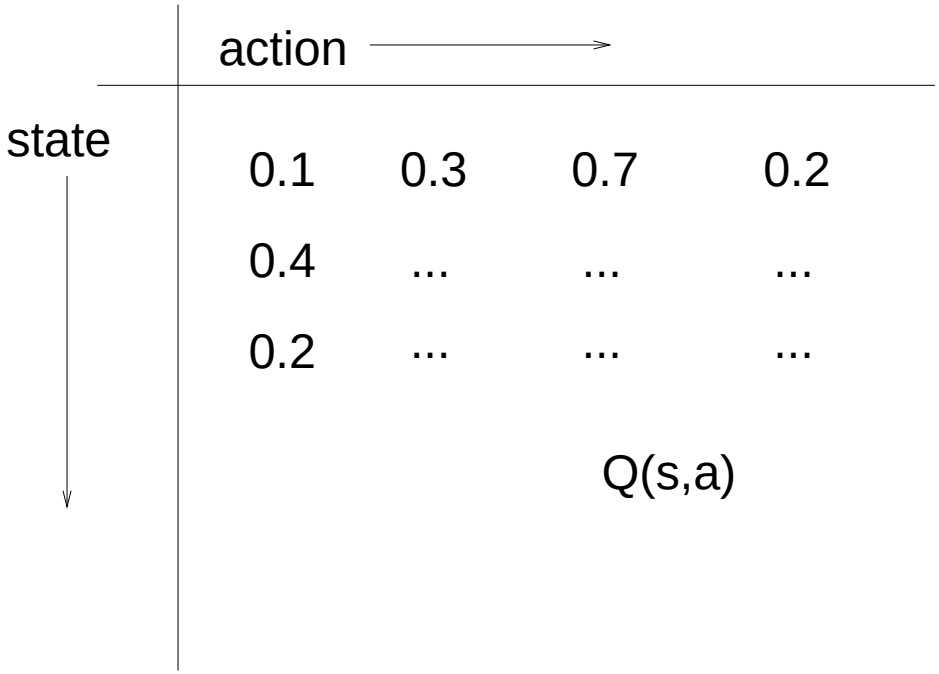
Gillian Hayes

22nd February 2007

School of informatics

Generalisation and Function Approximation

- Tabular value functions



What happens if the size of the state/action space is large?

- Large numbers of states/actions?
- Continuously-valued states/actions?
- Most states never experienced exactly before

  Memory
  Time
  Data

GENERALISATION: how experience with small part of state space is used to produce good behaviour over large part of state space

Methods

Neural networks, decision trees, multivariate regression ...

cf Asterix and Obelix: statistical clustering
Web Crawler: neural network

As long as they can deal with:

- learning while interacting – online
- nonstationarity – policy changes

Combining gradient descent methods with reinforcement learning

May have to use generalisation methods to approximate states, actions, value functions, Q-value functions, policies

Examples of Feature Vectors

$$\vec{\phi_s} = \begin{pmatrix} \text{redness} \\ \text{greenness} \\ \text{roundness} \\ \text{starness} \\ \text{size} \end{pmatrix} = \begin{pmatrix} 25 \\ 3 \\ 2 \\ 15 \\ 25 \end{pmatrix}$$

- "Redness" = say closeness to 11111111000000000000000 (RGB, R=255, G=0, B=0)
- "Roundness" = say distance of points from enclosing circle
- "Starness" = say some combination of number of points, template matching to a star shape, high spatial frequency components of boundary

$$\vec{\phi}_s = \begin{pmatrix} x \\ y \\ \text{heading} \\ \text{batterypower} \end{pmatrix}$$

- Position in x, y coordinates (real numbers)
- Heading in degrees w.r.t. north (real number or quantised)
- Battery power = some real number

Gradient Descent SARSA(λ)

Constant policy $\Rightarrow$ converges like TD(λ)

Combine with policy improvement?

- Discrete action set
  - compute $Q_t(s_t, a)$ for all a possible in s_t
  - find greedy action $a_t^* = \arg max_a \, Q_t(s_t,a)$
  - change estimation policy to
    * greedy (off-policy)
    * soft approximation (on-policy)