

*Cognome e nome dello studente:*

*Matricola:*

*Anno di corso e turno:*

**A.A. 2004-2005 – Tema d'esame del 30 Gennaio 2006**

Descrivere un algoritmo per il Q-learning. Cosa rappresenta la funzione Q? [2]

Cosa si intende per policy evaluation? [1]

Enunciato del teorema del miglioramento della policy [2].

Definire le equazioni di Bellman e quando vengono utilizzate [2].

Cosa si intende per rinforzo? Se chiedete ad un vigile un'indicazione e questi vi risponde di girare a destra al prossimo incrocio, si può considerare un segnale di rinforzo? Perché? [2]

Definire una macchina di Turing [2]

Nelle strategie di ottimizzazione evolutivista, cosa si intende per "covariance adaptation"? Qual è il suo ruolo? Secondo quali equazioni viene modificata la matrice di covarianza e perché? [3]

Come viene definita l'entropia fuzzy? Che cosa rappresenta? La frase "se domani nevierà, metterò le catene sulle ruote dell'auto", può definire un evento regolato dalle leggi della statistica o un evento regolato dalle leggi della fuzziness e perché? [3]

Cosa si intende per modello semi-parametrico? [1]

Apprendimento supervisionato. Secondo quale criterio e quale funzione vengono aggiornati i parametri di una rete neurale? Definite una rete neurale a vostro piacimento e scrivete le equazioni di aggiornamento dei parametri per quella rete [3].

Cos'è un mixture model? Scrivere le equazioni di aggiornamento dei parametri per una mistura a vostro piacimento mediante l'algoritmo di Expectation Maximization [3]

Clustering. Definizione. Scrivere l'algoritmo di clustering mediante mappe di Kohonen [3].

Cosa si intende per switching median filter? Descrivere l'algoritmo visto a lezione [3].

Descrivere un sistema di realtà virtuale [3]

**ESERCIZIO [4]:**

Consideriamo un agente che opera in un ambiente Markoviano stocastico. Supponiamo che l'agente debba stimare il guadagno ottenuto dall'interazione con l'ambiente, considerando che anche il reward fornito dall'ambiente è stocastico. Supponiamo che l'agente osservi otto episodi diversi:

A,0, B,0	B,1	B,1	B,1
B,1	B,1	B,1	B,0

Ciò significa che il primo episodio parte da uno stato A, poi passa a B ricevendo reward pari a zero e lì si ferma ricevendo ancora reward pari a zero. Gli altri sette episodi sono ancora più brevi. Partono dallo stato B e lì terminano con reward pari a uno in sei casi e pari a zero nel settimo caso. Analizzando questi dati, quale è secondo voi la predizione migliore che può fare l'agente della ricompensa a lungo termine (Value function, V), per i due stati? Determinare cioè V(A) e V(B).