|  |
| --- |
|  |
| Laboration 2 |
| Rapport |
|  |
|  |
|  |
|  |

**Innehåll**

[1 Instruktioner - 1 -](#_Toc436247938)

[2 BridgeGrid (Value Iteration) - 1 -](#_Toc436247939)

[3 DiscountGrid (Value Iteration) - 2 -](#_Toc436247940)

[4 BridgeGrid (Q-Learning) - 2 -](#_Toc436247941)

[5 Optimering med den Genetiska Algoritmen - 3 -](#_Toc436247942)

# Instruktioner

Denna rapporten hör till laboration 2 (MDPer, RL och GA). I laborationsdokumnetet ombeds ni fylla i ett antal uppgifter i denna rapporten. När ni lämnar in laborationen skall denna ifyllda rapporten finnas med i arkivfilen.

# BridgeGrid (Value Iteration)

Byt ut värdet för antingen *discount* **eller** *noise* parametern i nedanstående tabell. Med det modifierade värdet skall value iteration agenten kunna följa den genererade policyn för att korsa bron och lämna Bridgeworld brädet via terminaltillståndet +10.00. Förklara också varför parameterinställning leder till önskad policy.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Värde** | **Förklaring** |
| discount | 0.9 | Om man tar bort risken att gå åt fel håll, är det riskfritt att passera bron, varför den större rewarden lockar mer.  (optimal noise: 0.016954852293706043) |
| noise | 0.0 |
|  | |

Uppdateringsformeln för value iteration återges nedan

och är ekvivalent med

där

# DiscountGrid (Value Iteration)

Fyll i de tre parametervärdena *discount*, *noise* och *livingReward* (*reward*) som leder till önskad policy och förklara även varför dessa parameterinställningar leder till önskad policy.

|  |  |  |  |
| --- | --- | --- | --- |
| **Uppgift** | **Parameter** | **Värde** | **Förklaring** |
| a | discount | 0.31 | Här blir det mer värt med en mindre och snabbare belöning |
| noise | 0.0 |
| reward | 0.0 |
|  | |
| b | discount | 0.35 | Farlig nederväg, samt för långt till terminal +10 |
| noise | 0.2 |
| reward | 0.0 |
|  | |
| c | discount | 0.99 | Så länge discount är större än ~0.4 och mindre än 1 så är terminal +10 det bästa målet, vid < 0.4 är det bättre att gå till terminal +1 och vid => 1 är alla ställen lika bra. |
| noise | 0.0 |
| reward | 0.0 |
|  | |
| d | discount | 0.9 | Vid lite brus blir nedervägen farlig, och en discount hög nog; Vid discount > 0.4 är det bättre att gå till terminal +10 |
| noise | 0.2 |
| reward | 0.0 |
|  | |
| e | discount | 0.0 | Vid en reward större än 0 och varken noise eller discount, så är det alltid bättre att bara fortsätta leva |
| noise | 0.0 |
| reward | >0.0 |
|  | |

# BridgeGrid (Q-Learning)

Fyll i de två parametervärdena *epsilon* och *alpha* som leder till den optimala policyn under 50 träningsepisoder. Om den optimala policyn inte kan hittas med någon parameterkombination under 50 träningsepisoder, ange värdet *GÅR EJ* för de båda parametrarna. Motivera också erat svar, dvs varför hittas den optimala policyn med en viss parameterkombination eller varför går det inte att hitta den optimala policyn med någon parameterkombination?

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Värde** | **Förklaring** |
| epsilon | 0.0 |  |
| alpha | 0.0 |
|  | |

# Optimering med den Genetiska Algoritmen

Fyll i de parametrar ni anser vara optimala för att hitta maximat för funktionerna 1-4.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Function | #Gens | PopSize | #Genes | Pcross | Pmut | Selection Method | Ptour | Tournament  Size | #Elite |
| f1 |  |  |  |  |  |  |  |  |  |
| f2 |  |  |  |  |  |  |  |  |  |
| f3 |  |  |  |  |  |  |  |  |  |
| f4 |  |  |  |  |  |  |  |  |  |