Lecture D2 Markov Reward Process2

Baek, Jong min

2021-01-15

차 례

Iterative

```
R = np.array([1.5,1.0]).reshape(2,1)

P = np.array([0.7,0.3,0.5,0.5]).reshape(2,2)
gamma = 0.9
epsilon = 10**-8

v_old = np.zeros(2).reshape(2,1)
v_new = R+gamma*np.dot(P,v_old)

while max(abs(v_new-v_old)> epsilon):
  v_old = v_new
  v_new = R+gamma*np.dot(P,v_old)

print(v_new)
```

```
## [[13.35365845]
##  [12.74390235]]
```

Page 24

```
R = np.array([1.5,1.0]).reshape(2,1)
P = np.array([0.7,0.3,0.5,0.5]).reshape(2,2)
gamma = 0.9
epsilon = 10**-8

v_old = np.zeros(2).reshape(2,1)
v_new = R+gamma*np.dot(P,v_old)
results = v_old.T

results = np.vstack([results,v_new.T])



while np.max(np.abs(v_new-v_old)> epsilon):
  v_old = v_new
  v_new = R+gamma*np.dot(P,v_old)

  results = np.vstack([results,v_new.T])

results=pd.DataFrame(results)
results.columns=['coke','pepsi']
results.head()
```

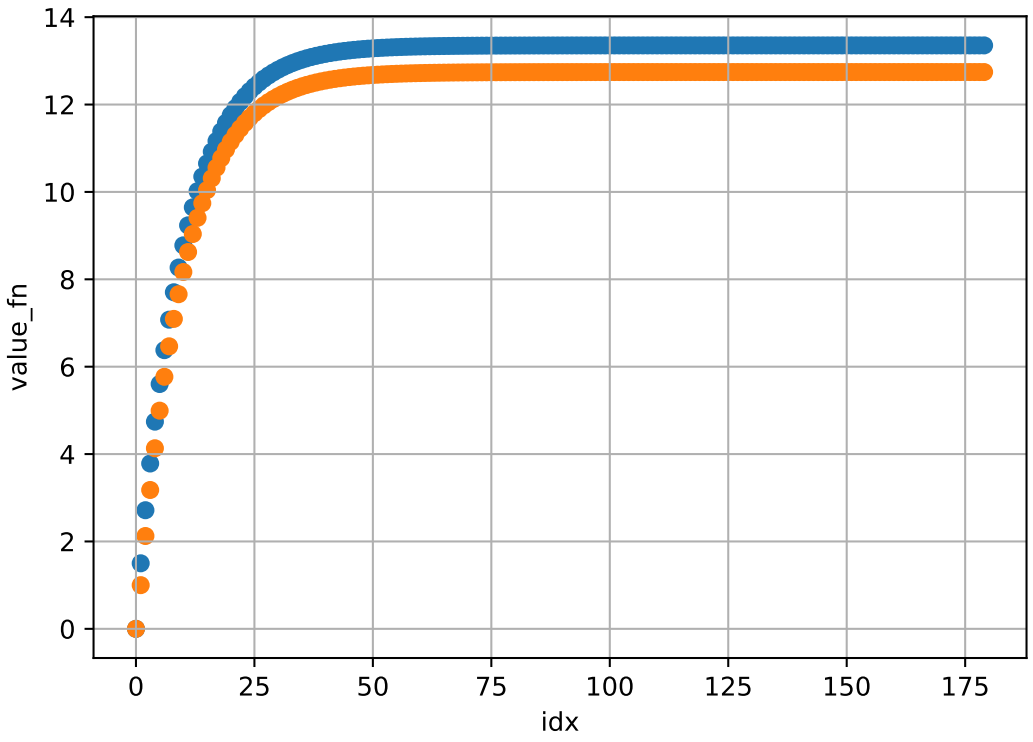
```
##        coke    pepsi
## 0  0.000000  0.00000
## 1  1.500000  1.00000
## 2  2.715000  2.12500
## 3  3.784200  3.17800
## 4  4.742106  4.13299
```

```
results.tail()
```

```
##           coke      pepsi
## 175  13.353658  12.743902
## 176  13.353658  12.743902
## 177  13.353658  12.743902
## 178  13.353658  12.743902
## 179  13.353658  12.743902
```

```
import matplotlib.pyplot as plt
results['idx'] = results.index

plt.scatter(results['idx'],results['coke'])
plt.scatter(results['idx'],results['pepsi'])
plt.grid(True)
plt.xlabel('idx')
plt.ylabel('value_fn')
```

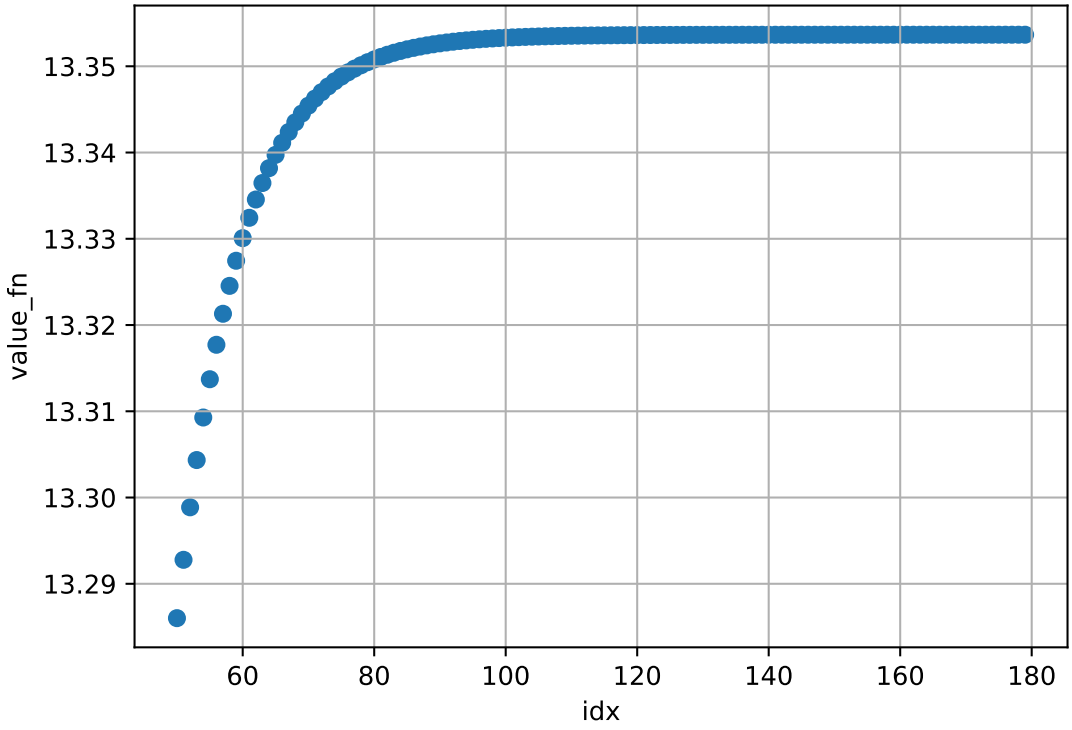


Page 25

```
import matplotlib.pyplot as plt
results['idx'] = results.index

results=results[results['idx']>=50]

plt.scatter(results['idx'],results['coke'])
plt.grid(True)
plt.xlabel('idx')
plt.ylabel('value_fn')
```

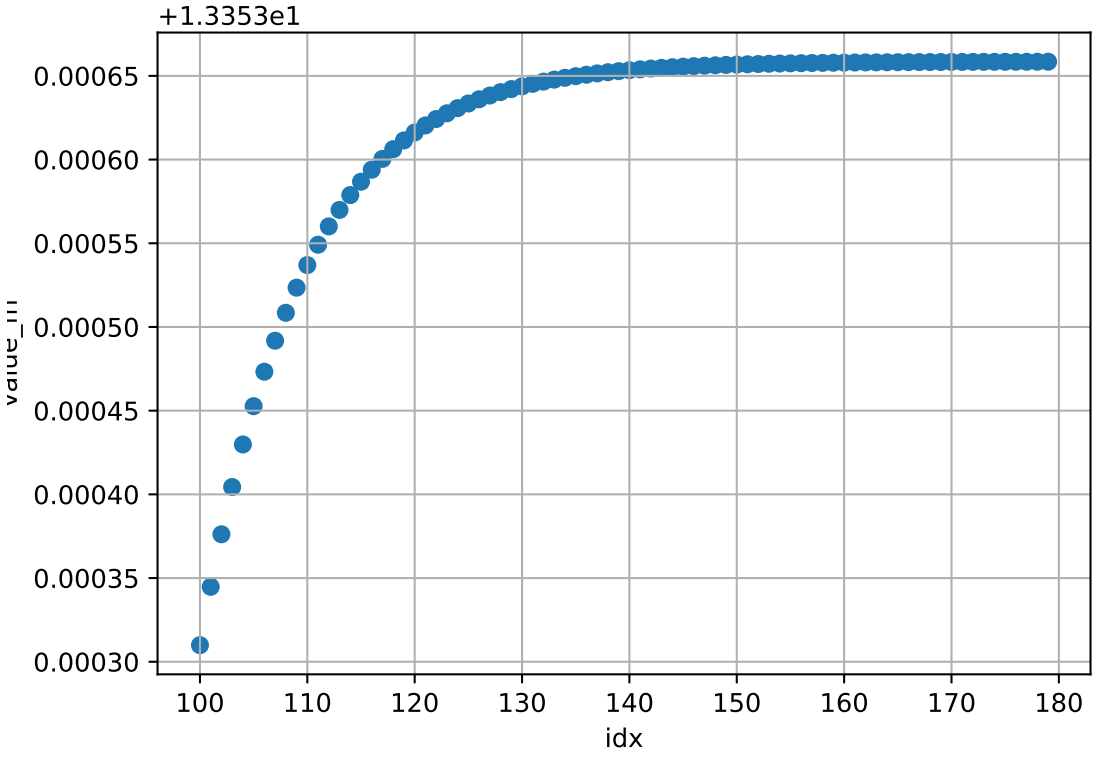


Page 28

```
import matplotlib.pyplot as plt
results['idx'] = results.index

results=results[results['idx']>=100]

plt.scatter(results['idx'],results['coke'])
plt.grid(True)
plt.xlabel('idx')
plt.ylabel('value_fn')
```



D2.Rmd

```
"Hello"
```

```
## [1] "Hello"
```