Problem 1.

Define the fundamental matrix: $H = (I - P + P^*)^{-1}(I - P^*)$

Thus,

$$\begin{aligned} HP^* &= [(I - P + P^*)^{-1}(I - P^*)]P^* \\ &= (I - P + P^*)^{-1}[(I - P^*)P^*] \\ &= (I - P + P^*)^{-1}(P^* - P^*P^*) \end{aligned}$$

According to Lemma 2.1(i): $PP^* = P^*P = P^*P^* = P^*$

$HP^* = (I - P + P^*)^{-1}(P^* - P^*P^*) = (I - P + P^*)^{-1}(P^* - P^*) = 0$

Problem 2.

According to Lemma 2.3: $[I - (P - P^*)]^{-1} = \sum_{k=0}^{\infty}(P - P^*)^k = I + \sum_{k=1}^{\infty}(P^k - P^*)$

$$(I - \alpha P)^{-1} = \sum_{k=0}^{\infty}(\alpha P)^k = I + \sum_{k=1}^{\infty}(\alpha P)^k = I + \sum_{k=1}^{\infty}(\alpha^k P^k)$$

Problem 3.

(a) State space: $X = \{1, 2\}$ – 1-L, 2-H

Control space: $U = \{0, 1\}$ – 0-do not receive catalog, 1-receive catalog

Transition probabilities:

$P\{x_{t+1} = 1 | x_t = 1, u_t = 0\} = 0.5$, $P\{x_{t+1} = 1 | x_t = 1, u_t = 1\} = 0.3$,
$P\{x_{t+1} = 2 | x_t = 1, u_t = 0\} = 0.5$, $P\{x_{t+1} = 2 | x_t = 1, u_t = 1\} = 0.7$,
$P\{x_{t+1} = 1 | x_t = 2, u_t = 0\} = 0.6$, $P\{x_{t+1} = 1 | x_t = 2, u_t = 1\} = 0.2$,
$P\{x_{t+1} = 2 | x_t = 2, u_t = 0\} = 0.4$, $P\{x_{t+1} = 2 | x_t = 2, u_t = 1\} = 0.8$,

Another way to write the transition probabilities is the following:

$$P(1) = \begin{pmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \end{pmatrix} \qquad P(0) = \begin{pmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \end{pmatrix}$$

The cost function is:

$$c_t(x,u) = \begin{cases} -5 & if\ x = 1, u = 1 \\ -10 & if\ x = 1, u = 0 \\ -35 & if\ x = 2, u = 1 \\ -25 & if\ x = 2, u = 0 \end{cases}$$

Dynamic programming equations is:

$$\gamma + h(x) = \min_{u \in U(x)} \left\{ c(x,u) + \sum_{j=1}^{2} P(j|x,u)h(j) \right\},$$

i.e.

$$\begin{cases} \gamma + h(1) = \min \begin{Bmatrix} -10 + 0.5h(1) + 0.5h(2), \\ -5 + 0.3h(1) + 0.7h(2) \end{Bmatrix}, \\ \gamma + h(2) = \min \begin{Bmatrix} -25 + 0.6h(1) + 0.4h(2), \\ -35 + 0.2h(1) + 0.8h(2) \end{Bmatrix} \end{cases}$$

(b) with discount factor $\alpha = 0.9$

$\pi^1 = (0,0)^T$, h^1 is the solution of the system of equations:

$$\begin{cases} h^1(1) = -10 + 0.45h^1(1) + 0.45h^1(2) \\ h^1(2) = -25 + 0.54h^1(1) + 0.36h^1(2) \end{cases}$$

So

$$\begin{cases} h^1(1) = -161.9266 \\ h^1(2) = -175.6881 \end{cases}$$

As

$$\begin{cases} -5 + 0.27h^1(1) + 0.63h^1(2) = -159.4037 > h^1(1) \\ -35 + 0.18h^1(1) + 0.72h^1(2) = -190.6422 < h^1(2) \end{cases}$$

The policy minimize is $\pi^2 = (0,1)^T$, h^2 is the solution of the system of equations:

$$\begin{cases} h^2(1) = -10 + 0.45h^2(1) + 0.45h^2(2) \\ h^2(2) = -35 + 0.18h^2(1) + 0.72h^2(2) \end{cases}$$

So

$$\begin{cases} h^2(1) = -254.1096 \\ h^2(2) = -288.3562 \end{cases}$$

As

$$\begin{cases} -5 + 0.27h^2(1) + 0.63h^2(2) = -255.2740 < h^2(1) \\ -25 + 0.54h^2(1) + 0.36h^2(2) = -266.0274 > h^2(2) \end{cases}$$

The policy minimize is $\pi^3 = (1,1)^T$, h^3 is the solution of the system of equations:

$$\begin{cases} h^3(1) = -5 + 0.27h^3(1) + 0.63h^3(2) \\ h^3(2) = -35 + 0.18h^3(1) + 0.72h^3(2) \end{cases}$$

So

$$\begin{cases} h^2(1) = -257.6923 \\ h^2(2) = -290.6593 \end{cases}$$

As

$$\begin{cases} -10 + 0.45h^3(1) + 0.45h^3(2) = -256.7582 > h^3(1) \\ -25 + 0.54h^3(1) + 0.36h^3(2) = -268.7912 > h^3(2) \end{cases}$$

We have $\pi^* = (1,1)^T$, $h^* = (-256.7582, -268.7912)^T$

(c) $\min h(1) + h(2)$

$$\begin{aligned} s.t. \quad h(1) &\le -10 + 0.45h(1) + 0.45h(2) \\ h(1) &\le -5 + 0.27h(1) + 0.63h(2) \\ h(2) &\le -25 + 0.54h(1) + 0.36h(2) \\ h(2) &\le -35 + 0.18h(1) + 0.72h(2) \end{aligned}$$

```
% v = dne1_LP
function v = dne1_LP
clear();
f=[1;1];

A=[-0.55, 0.45;
    -0.73, 0.63;
    0.54, -0.64;
    0.18, -0.28];

b=[5;10;35;25];

v=linprog(f,A,b);

end
```

```
Optimal solution found.


ans =

 -173.2877
 -200.6849
```

(d) if discount factor is 1, the discounted infinite-horizon problem equivalent to the average reward problem.