Exam 2 - Rajat Sethi

1.) Use $\sum_{i=1}^{n} a_i y_i K(x_i, z) + b = f(z)$

$2.5(1)(2z+1)^2 + 7.\overline{3}(-1)(5z+1)^2 + 4.8\overline{3}(1)(6z+1)^2 + b = f(z)$

$2.5(4z^2+4z+1) + -7.\overline{3}(25z^2+10z+1) + 4.8\overline{3}(36z^2+12z+1) + b = f(z)$

$10z^2 + 10z - 183.\overline{3}z^2 - 73.\overline{3}z + 174z^2 + 58z + b = f(z)$ (All x^0 terms are merged into b)

$f(z) = 0.\overline{6}z^2 - 5.\overline{3}z + b$

Let $z = 2$, $f(z) = 1$, $b = 9$

$b = 9$ for $z = 5$ and 6 too

Discriminate Function:

$f(z) = 0.\overline{666}...z^2 - 5.\overline{333}...z + 9$

$f(3) = -1$

2.)

	a_1	a_2	c	a_1	a_2	c
h_1	10	01	1	11	10	0
h_2	01	11	0	10	01	0

Crossover For h_1 - <1, 3>
Crossover for h_2 - <1, 8>

	a_1	a_2	c	a_1	a_2	c	a_1	a_2	c
h_3	11	11	0	10	01	1	11	10	0
h_4	00	01	0						

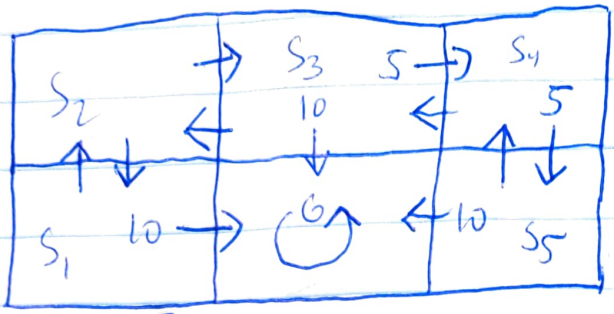
3.) Lazy Learning - Simply storing the training set and not preprocessing or fixing it

Eager Learning - Preprocessing the dataset to ease the burden on future predictions.

Advantages of IBL:
- Works well for complex datasets

Disadvantages of IBL:
- Computationally expensive

4a)



$V^* = r + \gamma \cdot V^*_{n+1}$

$G = 0$

$S_1 = 10 + 0.8 \cdot 0 = 10$

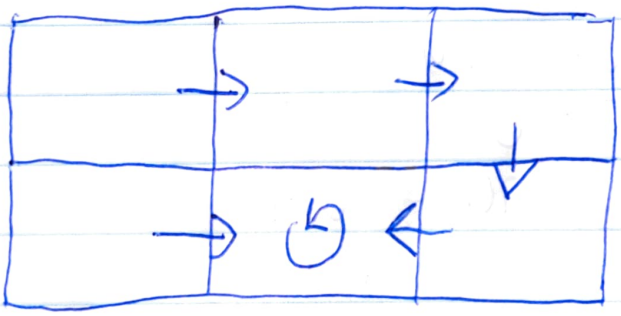
$S_5 = 10 + 0.8 \cdot 0 = 10$

$S_4 = 5 + 0.8 \cdot 10 = 13$

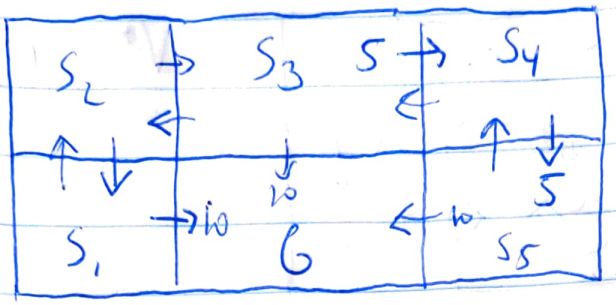
$S_3 = 5 + 0.8 \cdot 13 = 15.4$

$S_2 = 0 + 0.8 \cdot 15.4 = 12.32$

Optimal Policy:



4B.



Episode 1:

$\hat{Q}(S_1, \text{right})$: 10

$\hat{Q}(S_1, \text{up})$: $0 + 0 = 0$

$\hat{Q}(S_2, \text{down})$: $0 + 0.8(10) = 8$

$\hat{Q}(S_2, \text{right})$: $0 + 0 = 0$

$\hat{Q}(S_3, \text{left})$: $0 + 0.8(8) = 6.4$

$\hat{Q}(S_3, \text{down})$: $10 + 0 = 10$

$\hat{Q}(S_3, \text{right})$: $5 + 0 = 5$

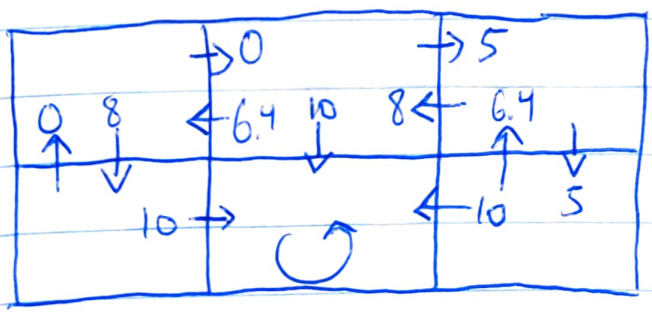
$\hat{Q}(S_4, \text{left})$: $0 + 0.8(10) = 8$

$\hat{Q}(S_4, \text{down})$: $5 + 0 = 5$

$\hat{Q}(S_5, \text{up})$: $0 + 0.8(8) = 6.4$

$\hat{Q}(S_5, \text{left})$: $10 + 0 = 10$

$\hat{Q}$ Values after first episode



5.) $C_2 = (C - (C_1 - \{L\})) \cup \{\sim L\}$

G exists in C_1 but not C
Let $L = G$

$C_1 - \{G\} = (A \vee B) \wedge \sim G$
$C - (C_1 - \{G\}) = (A \vee B) \wedge G$

$(C - (C_1 - \{G\})) \cup \{\sim G\} = \sim G \vee A \vee B$

$\boxed{C_2 = \sim G \quad \text{or} \quad \sim G \vee A \vee B}$

6. $L_1\theta_1 = \neg L_2\theta_2 \therefore L_2 = \neg L_1\theta_1\theta_2^{-1}$

$C = (C_1 - \{L_1\})\theta_1 \cup (C_2 - \{L_2\})\theta_2$

$C - (C_1 - \{L_1\})\theta_1 = (C_2 - \{L_2\})\theta_2$

$(C - (C_1 - \{L_1\})\theta_1)\theta_2^{-1} = C_2 - \{L_2\}$

$\longleftrightarrow$

$C_2 = (C - (C_1 - \{L_1\})\theta_1)\theta_2^{-1} \cup \{L_2\}$

$\boxed{C_2 = (C - (C_1 - \{L_1\})\theta_1)\theta_2^{-1} \cup \{\neg L_1\theta_1\theta_2^{-1}\}}$

7. For every state and action, imagine an infinite number of iterations/episodes.

As the episodes approach infinity, the error between Q-hat and the true Q-value approaches the maximum error.

Specifically, the error will continuously decrease by a factor of γ, the discount rate.

8) Feature Selection Methods

Wrapper - A greedy, heuristic algorithm that searches for features with a local optimum. Combines features until the local optimum declines due to overfitting

Filter - Uses a learning algorithm and statistical measurements to determine the features. Might miss important features.

Embedded - Combines the Wrapper and Filter methods. Less prone to overfitting, reduces computational costs, and gets the useful features.

9.) PCR - Discards eigenvalues that are the smallest. PCR eliminates low variance and prioritizes high variance.

PLS - Shrinks low-variance and inflates high-variance components using least squares

Ridge - Shrinks all components, but shrinks the low-variance components more.

10.) Shrinkage Methods

Ridge Regression - Reduces high variance by adding a little bias. Does so by penalizing the squares of coefficients. Ultimate goal is to shrink coefficients and reduce high variance

LASSO - Similar to Ridge Regression, in that it reduces high variance by adding a little bias. However, instead of squaring the coefficients, it takes the absolute value. Like Ridge, it shrinks coefficients (and can eliminate them) and reduces high variance.