Homework 1

Problem 1

d is the correct answer.
(i) is not learning because the statistical model is the equivalent of pinning the problem of coin classification down mathematically. (ii) is supervised learning because we are given an input of large set of coins, and the correct output(coins are labeled). (iii) is reinforcement learning because it involves an output (moves) and a grade for this output (winning/ losing) that then leads to adjustment of strategy based on the grade.

Problem 2

a is the correct answer.
(i) is not suited for machine learning because determining whether number is prime or not can be done mathematically. Likewise (iii) is not suited because determining the time it would take for a falling object to hit the ground can be modeled by simple physics. (ii) is suited for machine learning because a pattern most certainly exists, fraud in credit card charges cannot be pinned down mathematically, and there is a plethora of data available. (iv) is also suited due to the same reasons that a pattern exists, it cannot be pinned down mathematically, and that there is data on it. Ideally, the lights would be able to "learn" the timing that most matches the desired timing that eases traffic flow of this busy intersection.

Problem 3

d is the correct answer.
Let's call the bag with both black balls bag 1 and the bag with one black and one white ball bag 2. If we know that the first ball drawn is black, then we either picked the first black ball of bag 1 or the second ball of bag 1 or the black ball of bag 2.The only way to get a black ball on the second draw is for the first drawn ball to be from bag 1. There are two ways for this to happen as described above and there are three total ways to get a black ball drawn first so the probability is $\frac{1}{2}$.

Problem 4

b is the correct answer.
We know the probability of drawing a red is 0.55. Thus the probability of NOT drawing a red is 0.45. Because the sample consists of 10 draws, the probability of getting no red marbles in all of the 10 is 0.45^{10} which matches choice b.

Problem 5

c is the correct answer.
The probability that at least one of the samples has v = 0 is equal to the 1- (probability that all samples have v $\neq$ 0).From problem 4, the probability that v = 0 for one sample is 3.405

x 10^{-4}, so the probability that v$\neq$ 0 for one sample is the complement (1- 3.405 x 10^{-4}). Extrapolating this to 1000 samples, the probability that all samples have v $\neq$ 0 is (1 - 3.405 x $10^{-4})^{1000}$ and the probability that at least one of the samples has v = 0 is 1 - (1 - 3.405 x $10^{-4})^{1000}$ = 0.289.

Problem 6

e is the correct answer.
We start by enumerating the 8 possible target functions.

1. $f(101) = 0$, $f(110) = 0$, $f(111) = 0$

2. $f(101) = 0$, $f(110) = 0$, $f(111) = 1$

3. $f(101) = 0$, $f(110) = 1$, $f(111) = 0$

4. $f(101) = 1$, $f(110) = 0$, $f(111) = 0$

5. $f(101) = 1$, $f(110) = 1$, $f(111) = 0$

6. $f(101) = 1$, $f(110) = 0$, $f(111) = 1$

7. $f(101) = 0$, $f(110) = 1$, $f(111) = 1$

8. $f(101) = 1$, $f(110) = 1$, $f(111) = 1$

Now, we calculate the scores of each:

(1): 1 has zero agreeing points, 2 has 1 agreeing point, 3 has one agreeing point, 4 has 1 agreeing point, 5 has two agreeing points, 6 has two agreeing points, 7 has two agreeing points, 8 has three agreeing points so total = 1 * 0 + 1 * 3 + 2 * 3 + 3 * 1 = 12.
(2): By symmetry to (1), also = 12.
(3): 1 has 2 agreeing points, 2 has 3 agreeing points, 3 has 1 agreeing point, 4 has 1 agreeing point, 5 has 0 agreeing points, 6 has 2 agreeing points, 7 has 2 agreeing points, and 8 has 1 agreeing point. Total = 1 * 0 + 3 * 1 + 3 * 2 + 1 * 3 = 12.
(4): 1 has three agreeing points, 2 has 2 agreeing points, 3 has two agreeing points, 4 has two agreeing points, 5 has 1 agreeing point, 6 has 1 agreeing point, 7 has 1 agreeing point, 8 has 0 agreeing points so total = 1 * 0 + 1 * 3 + 3 * 2 + 3 * 1 = 12.

Thus, because all the above hypothesis output 12 as the score and thus are equivalent, e is the correct answer.

Problem 7

b is the correct answer.
See code (attached to end of assignment). I set N = 10, and the number of iterations it takes on average for the PLA to converge was 14.818 which is closest to the answer choice of 15.

Problem 8

c is the correct answer.

See code (attached to end of assignment). I set N = 10 and to estimate the disagreement probability, I counted how many of the 100,000 points (my large set) had $f(x) \neq g(x)$ and the output I got was 0.10902 which is closest to the answer choice of 0.1.

Problem 9

b is the correct answer.

See code (attached to end of assignment). I set N = 100, and the number of iterations it takes on average for the PLA to converge was 164.59 which is closest to the answer choice of 100.

Problem 10

b is the correct answer.

See code (attached to end of assignment). I set N = 100 and to estimate the disagreement probability, I counted how many of the 100,000 points (my large set) had $f(x) \neq g(x)$ and the output I got was 0.0139958 which is closest to the answer choice of 0.01.