Homework I

1. Consider the following set of points: {(-2, -1), (1, 1), (3, 2)}
   a) Find the least square regression line for the given data points.
   b) Plot the given points and the regression line in the same rectangular system of axes.

   [5 Points]

2. The values of y and their corresponding values of y are shown in the table below

x	0	1	2	3	4
y	2	3	5	4	6

   a) Find the least square regression line y = a x + b.
   b) Estimate the value of y when x = 10.

   [5 Points]

3. Sufficient statistics for online linear regression **[15 Points]**

Consider fitting the model $\hat{y} = w_o + w_1 x$ using least squares. Unfortunately we did not keep the original data x_i, y_i, but we do have the following functions (statistics) of the data:

$$\bar{x}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \quad \bar{y}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$C_{xx}^{(n)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad C_{yy}^{(n)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$C_{xy}^{(n)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

a. What are the minimal set of statistic that we need to estimate w_1?
b. What are the minimal set of statistic that we need to estimate w_0?
c. Suppose a new data point x_{n+1} , y_{n+1} arrives, and we want to update our sufficient statistics without looking at the old data which we have not stored. (This is useful for online learning.) Show that we can do this for $\bar{x}$ as follows.

$$\bar{x}^{(n+1)} \triangleq \frac{1}{n+1} \sum_{i=1}^{n+1} x_i = \frac{1}{n+1} \left(n\bar{x}^{(n)} + x_{n+1} \right)$$

$$= \bar{x}^{(n)} + \frac{1}{n+1} \left(x_{n+1} - \bar{x}^{(n)} \right)$$

This has the form: new estimate is old estimate plus correction. We see that the size of correction diminishes over time(i.e. we get more samples). Drive a similar expression to update $\bar{y}$

4. Explain why the size of the hypothesis space in the *EnjoySport* learning task is 973. How would the number of possible instances and possible hypotheses increase with the addition of the attribute *WaterCurrent*, which can take on the values *Light*, *Moderate* or *Strong*? More generally, how does the number of possible instances and hypotheses grow with the addition of a new attribute A that takes on *k* possible values?
[5 Points]

5. Consider the following sequence of positive and negative training examples describing the concept “pairs of people who live in the same house”. Each training example describes an ordered pair of people, with each person described by their sex, hair color (black, brown or blonde), height (tall medium or short), and nationality (US, French, German, Irish, Indian, Japanese or Portuguese). **[15 Points]**

+ <<male brown tall US> <female black short US>

+ <<male brown short French> <female black short US>>

- <<female brown tall German> <female black short Indian>>

+ <<male brown tall Irish> <female brown short Irish>>

Consider a hypothesis space defined over these instances, in which each hypothesis is represented by a pair of 4-tuples, and where each attribute constraints may be a specific value , “*?*” or “*ϑ*”, just as in the *EnjoySport* hypothesis representation. For example, the hypothesis

<<male ? tall ?> < female ? ? Japanese>>

represents the set of all pairs of people where the first is a tall male (of any nationality and hair color), and the second is s Japanese female (of any hair color and height).

a. Provide a hand trace of the Candidate-Elimination algorithm learning from the above training examples and hypothesis language. In particular, show the specific and general boundaries of the version space after it has processed the first training example, then the second training example, etc.

b. How many distinct hypotheses from the given hypothesis space are consistent with the following single positive training example?
<<male black short Portuguese> <female blonde tall Indian>>

c. Assume the learner has encountered only the positive example from part (b), and that it is now allowed to query the trainer by generating any instance and asking the trainer to classify it. Give a specific sequence of queries that assures the learner will converge to the single correct hypothesis, whatever it may be (assuming that the target concept is describable within the given hypothesis language). Give the shortest sequence of queries you can find. How does the length of this sequence relate to your answer to question (b)?

d. Note that this hypothesis language cannot express all concepts that can be defined over instances (i.e., we can define sets of positive and negative examples for which there is no corresponding describable hypothesis). If we were to enrich the language so that it could express all concepts that can be defined over the instance language, then how would your answer to © change?

6. Consider learning a Boolean valued function : $X \rightarrow Y$, where $X = \langle X_1, X_2, \ldots., X_N \rangle$, where *Y* and the X_i are all Boolean valued variables. You decide to consider a hypothesis space *H* where each hypothesis is of the form

   a. $if\left[(X_i = a) \wedge \left(X_j = b\right)\right]\ then\ Y = 1\ else\ Y = 0$

   where $i \neq j$, and where *a* and *b* can be either 0 or 1. Notice each hypothesis constrains exactly two of the features of X.

   How many distinct hypotheses are there in *H*?

[5 Points]