Homework 1

1.

a.

|  |  |  |
| --- | --- | --- |
| X | y | Y=23x/38 + 10/38 |
| -2 | -1 | -36/38 |
| 1 | 1 | 33/38 |
| 3 | 2 | 79/38 |

b.

Chart, line chart

Description automatically generated

2.

M = 9/10

B =11/5

a.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | Y | XY | X^2 | Y = 9x/10 + 11/5 |
| 0 | 2 | 0 | 0 | 11/5 |
| 1 | 3 | 3 | 1 | 31/10 |
| 2 | 5 | 10 | 4 | 4 |
| 3 | 4 | 12 | 9 | 49/10 |
| 4 | 6 | 24 | 16 | 29/5 |

b.

The estimate value of y = 56/5

3.

a.

The hypothesis space cardinality is the number of distinct combinations of the possible values for each attribute in the space. The “?” adds 1 value to all attributes. Since the “Ø” symbol always classifies an instance as negative, hence “Ø” is counted only once.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Attributes | Sky | Temp | Humid | Wind | Water | Forecast |
| Values | 3 | 2 | 2 | 2 | 2 | 2 |
| Values + 1 | 4 | 3 | 3 | 3 | 3 | 3 |

The number of combinations is 4\*3\*3\*3\*3\*3 = 972 + 1(for the “Ø”) = 973

b.

Given the addition of WaterCurrent, there are 2 attributes of 3 values and 5 attributes of 2 values. Hence, the number of combinations/instances = 3^2 \* 2^5 = 288

Given the addition of WaterCurrent, there are 2 attributes of 4 hypotheses and 5 attributes of 3 values of hypotheses. Each attribute has one more hypothesis than the real instance value because “?” is counted. Hence, the number of possible hypotheses is 4^2 \* 2^3 + 1= 3889

c.

Given the addition of the A attribute of k values,

* The new number of instances is current\_instance \* k that current\_instance is the current number of possible instances.
* The new number of possible hypotheses is (current\_hypothesis-1) \* (k + 1) + 1 that current\_hypothesis is the current number of possible hypotheses.

4.

a.

Letter

Description automatically generated with medium confidence

b.

Each hypothesis consistent with the above example can have either the specified value seen above or “?” for each attribute. Given 8 attributes, there are 256 distinct hypotheses.

c.

Given the sequence S1 = *<<male black short Portuguese> <female blonde tall Indian>>*

Then, the query 2 is <<female brown tall Italian><male brown short England>>

Then, S2 = <<(? or male) (? or black) (? or short) (? or Portuguese)> <(? or female) (? or blonde) (? or tall) (? or Indian)>>

Sex attribute has no more value to generalize S

Query 3 = <<done blonde tall German> <done brown short French>>

S3 = <<done (? Or black) (? Or short) (? Or Portuguese)> <done (? Or blonde) (? Or tall) (? Or Indian)>>

Hair color and height attributes have no more value to generalize S.

Query 4 = <<done done done German> <done done done Italian>>

The query and the specific sequence would be generated until the nationality attribute converges.

For each query, the hypothesis length was reduced by half. Only the third query’s length was not reduced because the hair color and height attributes were guaranteed for convergence. This guarantees that the converged sequence has the total number of hypotheses 2^8.

d.

Instead of checking 2 possible values in the pair, I will check every combination of attributes over all possible instances. This leads to more computation but guarantees to cover all possible describable hypotheses.

5.

Given that a and b can be either 0 or 1, Xi and Xj have 2 values. The conditional statement has the and logic; therefore, Xi and Xj are independent from each other. Also, as i not equal to j, Xi and Xj are exclusive. For example, if i=1, then j is among {2,3,4,…N}

The distinct hypotheses are: 2N\*2N-1 = 22N - 1