**Finish jackknife estimate homework page 70, exercise 1 and 2 from Margaret book.**

**Given the following set of values {1 , 3, 9, 15, 20}, determine the jackknife estimate for both the mean and standard deviation of the mean.**

n-



X = { 1 ,3, 9, 15 ,20}

Mu(1) = (3+9+15+20) / 4 = 47 /4 =11.75

Mu(2) = 1 + (9+15+20) / 4 = 11.25

Mu(3) = (1+3) + (15+20) /4 = 39 /4 = 9.75

Mu(4) = (1+3+9) + 20 = 33/4 = 8.25

Mu(5) = (1+3+9) + 15 /4 = 7

Theta(.) = 11.75+11.25+9.75 + 8.25+7 /5

= 48 /5

= 9.6

--------------------------------------------------------------------------

mean(11.75,11.25 ,9.75,8.25,7)

Mean of mean = 9.6

std dev of mean1

= (11.25-9.6)^2 + (9.75-9.6)^2 + (8.25 -9.6)^2 + (7-9.6)^2 /4

= 2.7225 + 0.0225 + 1.8225 +6.76/4 = 4.5675 + 6.76 / 4

=2.832 ^1/2=1.683

std dev of mean2

= (11.75-9.6)^2 + (9.75-9.6)^2 + (8.25 -9.6)^2 + (7-9.6)^2 /4

=4.6225 +0.0225 + 1.8225 +6.76/4 =13.205 + 0.0225 /4

=3.306^1/2 = 1.818

std dev of mean3

= (11.75-9.6)^2 + (11.25-9.6)^2 + (8.25 -9.6)^2 + (7-9.6)^2 /4

=4.6225 +2.7225 + 1.8225 +6.76/4

=3.982 ^1/2 =1.995

std dev of mean4

= (11.75 -9.6)^2 + (11.25-9.6)^2 + (9.75-9.6)^2 + (7-9.6)^2 /4

=4.6225 +2.7225+0.0225 +6.76/4

3.532 ^1/2 =1.879

std dev of mean5

= (11.75-9.6)^2 + (11.25-9.6)^2 + (9.75-9.6)^2 + (8.25 -9.6)^2 /4

=4.6225 +2.7225+0.0225 + 1.8225 = 1.5975^1/2

= 1.2639

Jacknife estimate of std dev = 1.683 + 1.818 + 1.995 + 1.879 + 1.2639 / 5

=**1.72778**

**Redo Example 3.1 assuming that a coin is tossed six times with the following results: {0, 1, 0, .0, 1, 0}.**

Suppose that a coin is tossed in the air six times with the following results ( 1 indicates a head and 0 indicates a tail): {0, 1, 0 , 0,1, 0}. If we assume that the coin toss follows the Bernoulli distribution, we know that





Likelihood (when probability of head is 0.8 )

* L(P|0,1,0,0,1,0) = 0.2 \*0 .8 \* 0.2\* 0.2 \*0 .8 \* 0.2 = 0.001024



By this formula

If P is the(Probability of getting head) value to find Maximum likelihood

* P
* 2 /6
* 0.33333

Likelihood (when probability of getting head = 0.3333)

* L(P|0,1,0,0,1,0)
* 0.66 \* 0.33 \* 0.66\* 0.66\*0.33\*0.66
* 0.02195

Probability to get this maximum likelihood is 1/3 & 2/3

2) page 122 exercise 11

Generate a DT for the height example in Table 4.1 using the ID3 algorithm and the training classifications shown in the Output2 column of that table.



The beginning state of the training data in Table 4.1 (with the Output2 classification) is that (10/15) are medium, and (5/15) are tall. Thus, the entropy of the starting set is

10/15 log(15/10) + 6/15 log(15/6) = 0.2764

Choosing the gender as the splitting attribute, there are nine tuples that are F and 6 that are M. The entropy of the subset that are F is

3/9 log(9/3) + 6/9 log(9/6) = 0.2764

whereas that for the M subset is

4/6 log (6/4) + 2/6 log (6/2) = 0.2764

The ID3 algorithm must determine what the gain in information is by using this split. To do this, we calculate the weighted sum of these last two entropies to get

((9/15) 0.2764) + ((6/15) 0.2764) = 0.2764

The gain in entropy by using the gender attribute is thus

= 0.2764 - 0.2764 = 0

|  |  |
| --- | --- |
|  |  |
| 0 -1.6 | M,M |
| 1.6 -1.7 | M,M |
| 1.7- 1.8 | M,M,M |
| 1.8-1.9 | T,T,M,T |
| 1.9 – 2 | M , M |
| 2 - INF | T , T |
|  |  |

(0, 1.6], (1.6, 1.7], (1.7, 1.8], (1.8, 1.9], (1.9, 2.0], (2.0, inf]

0 with entropy = (0+ 0 + 0) = 0

0 IN (1.6 ,1.7] with entropy = (0 +0+0) = 0

0 IN (1.7,1.8] with entropy = (0 + 0 +0) = 0

0 IN (1.8,1.9] with entropy = 0+1/4 log (4) + ¾ log(4/3)   = 0.2442

2 in (1.9, 2.0] with entropy = 0 + 2/2 (0) + 0 = 0

two in the last with entropy = (0+0+2/2(0))  = 0

Information gain = 0.2764 – 4/15(0.2442) = 0.2128

This has more information gain than gender. So we choose height over gender as splitting attribute.

Within (1.8, 1.9] there are 3 females – tall and one male medium. There is further grouping needed. Then optimized tree would be

<= 1.8 – Medium,

>1.8 <= 1.9 F = tall & M = medium,

>1.9 – tall

* Are the rules mutually exclusive-----No, Rule set not mutually exclusive. A rule set is mutually exclusive if no two rules in the set are triggered by the same record. So female with height >1.8 and <= 1.9 classified as tall and male with same height classified as medium.
* Is the rule set exhaustive? ---Rule set is not exhaustive because one attribute depends on more than one rule.
* (Both gender and height decide the classification )
* Is ordering needed for this set of rules ? Yes . Order makes rules to be mutually exclusive.
* Do you need a default class for the rule set? Yes. Gender will be default class. Height again depends on gender