**Bayesian Concept Learning**

D: Data ( set of example for a concept C)

h: a point Hypothesis about C.

Note: That both p(D|h) D and h can be viewed as function chioas from the …….. of ………….

D: y = {0,1}

D

h and D are contitact if {D (i)= h1 (i) … i }

Bayesian theorem

P (h/D)= P (h/D) p/h

How to choose hypotheses?

Correct the hypotheses?

* Correct on the training net.
* But not overfitting.

**Example:** Learning a real value function.

F: real valued function.

Y= f(x)

D= {(xi, di)}/ di{f(xi) + fi}

I=1,…………, m

ei ~ N (0, ϭi)

hML= any min

**proof,**

hML= any max P (h/D)

= any max P(d1,d2,….dm/h)

=any max=

If we take a log of the prodoce of this

= any max = Log ()

= any max

If iid N(0,)=0

Then,

di iid N(f(xc), )

iid = independent and identically distributed

from this point on we need to know the actual distribution of (di/h).

Notes:

If we use the Hypothesis

H(xi) + ei

Hi=di-h(xi)

hML= any max

hML= any max we can drop this

= any max

= any max

= any max

We can speak about

hMap = hML

P(h) = 1/H.

---------============---------------

The Beta Binomial Model

B-6

P (Xi,di,|h) same as P (di|h,xi)P(xi/h)

……= P(xi|di,h)=P(xi,di/h)/p(h)

…….= P(di,h,xi)/p(h,xi)\*p(xi,h)/p(h) and this same as the formula b4.

P(D/h)=

We going to use

P(D/h)=

P/di=1/h,xi)=h xi

P/di/h, ….) =

(h(xi) (1-h(X))1-di

1-di = can use the {1,0}

-----

Parroly experiment

1-p

When

P^(1-p)^

P(D/h)=

The max likehood

hML=any max

hML= we can drop this, and this formula will refer to the learning predicate

since the di [0,1]

hi [0,1]

--------==================------------------

Most Probable Classification

B-7

H={h1,h2,h3}

Suppose

P (h|D)=0.4

P (h|D)=0.3 the summation of all of these should=1.

P (h|D)=0.3

|  |  |  |
| --- | --- | --- |
|  | - | + |
| h1 | 1 | 0 |
| h2 | 0 | 1 |
| h3 | 0 | 1 |

The P (+|D)

= 1\*0.4

= + 0.3

= + 0.3

=

The

= 0\*0.4

= 1\*0.3

= 1\*0.3

**B-8**

The Gamma Distribution

X. r.v. G (d>0, >0)

t

fGamma (x)= xd-1 e^-ẞx pdf

where,

ᴦ (t) = du (ᴦ (t+1) =

x Gamma () = E (x) = ; Var (x)=

Mode (x)=

f’ (x)= bx/ [ () x e-ẞx - ẞx^ e-ẞx]=0

x e^-ẞx [x-1-ẞx]=x=

E(x)= e^-ẞxdx,,,,=

The beta distribution

X Beta )= 1/ x^-1 (1-x)

ᴦ()ᴦ (ẞ)/ ᴦ (+ẞ) E (x) =

M(x)= (-1)/ (+ẞ-2)

Var (x)= (ẞ)/ (+ẞ)(+ẞ+1)