20CS6037 Machine Learning

MLE, MAP, Bayesian Reasoning - Chapter 3&5 (Lecture 6: 9/11/14)

Lecturer: Anca Ralescu *Scribes: Khaldoon Ashouiliy, Kyungmook Park*

Section 1:
Bayesian Concept Learning

Section 2:
The Beta Binomial Model

Section 3:
Most Probable Classification

Section 4:
The Gamma Distribution

Section 1:
Bayesian Concept Learning

D: Data (set of example for a concept C)
: a point Hypothesis about C.
Note: That both p(D|h) D and can be viewed as functions from the set of instances to {0,1}
C: y $\rightarrow$ {0,1}

$$c(instance) = \begin{cases} 1 & if\ example\ of\ the\ concept\ C \\ 0 & otherwise \end{cases}$$

h and D are consistent if C(i) = h(i) $\forall$ i $\in$ Y

Bayes Theorem
P(h|D) = $\frac{P(h|D)P(h)}{P(D)}$

How to choose hypotheses?

<u>Correct the hypotheses?</u>

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

<u>Example</u> Learning a real value function.

f: real valued function.
<u>Training set</u> D = {$(x_i, d_i)|d_i$=f(x_i) + e_i}
i=1,...,m

$e_i \sim$ N (0, σ_i)

$\Rightarrow h_{ML} = \underset{h}{argmin} \sum_{i=1}^{m}[d_i - h(x_i)]^2$

<u>Proof</u>

$$\begin{aligned} h_{ML} \quad &= \underset{h \in H}{argmin} \\ &= \underset{h \in H}{argmin} \\ &= \underset{h \in H}{argmin} \\ &= \underset{h \in H}{argmin} \\ &= \underset{h \in H}{argmin} \\ &= \underset{h \in H}{argmin} \end{aligned}$$

If iid N(0,σ^2)=0
Then,
di iid N(f(x_c),σ^2)
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) ML= any max $_(i = 1)log[1/2] + e^(- (di.h(xi)2/22)ML = anymax_(i = 1)log[1/22] + log|e^(- (di.h(xi)2/22)wecandropthis = anymax_(i = 1) - di - h(xi)2/22 = anymax - [1/22]_(i = 1)(di.h(xi)2) = anymax_(i = 1)(di.h(xi)2)WecanspeakaboutMap = MLP(h) = 1/H.$

Section 2:
The Beta Binomial Model

X $\sim$ P(X) p:pdf
Y = f(X) What is the distribution of Y?
$P_Y(Y \leq y) = P_Y(f(X) \leq y) = \underline{P}_X(X \leq f^{-1}(y)) = P(f^{-1}(y))$

ex) Y = aX+b $\Rightarrow$ f(x) = aX+b

f^{-1}(y) = $\frac{y-b}{a}$ X = -1 Y = b-a ; X = 1 Y = a+b

P$_Y$(y) = P$_X$ ($\frac{y-b}{a}$)

ex) X $\sim$ u(-1,1)

$$P_X(x) = \begin{cases} 0 & x \leq -1 \\ \frac{1}{2} & -1 < X < 1 \\ 0 & x \geq 1 \end{cases}$$

$$P_X(\frac{y-b}{a}) = \begin{cases} 0 & \frac{y-b}{a} \leq -1 \\ \frac{1}{b-a+a+b} & -1 < \frac{y-b}{a} < 1 \\ 0 & x \geq 1 \end{cases} = \begin{cases} 0 & \frac{y-b}{a} \leq -1 \\ \frac{1}{4} & -1 < \frac{y-b}{a} < 1 \\ 0 & x \geq 1 \end{cases}$$

$$\begin{cases} E(y) = & aE(x) + b \\ Var(y) = & a^2 E(x) \end{cases}$$

For multivariable case : X = (X_1, ..., X$_n$), y=a^{T}x+b
E(y) = y=a^{T}E(x)+b

Section 3: Most Probable Classification

<u>Discrete Case</u> X $\sim$ u(1, . . . , 4)
P(x=i) = $\frac{1}{4}$ i=1,2,3,4

$$Y = \begin{cases} 1 & if\ X\ is\ even \\ 0 & if\ X\ is\ odd \end{cases}$$

<u>P</u>(y = 1) $= P_x$ (x is even)
$\sum P_x(X = k) \ = \ P(X = 2) + P(X = 4) = \frac{2}{4} = \frac{1}{2}$

k $\in$ {1, 2, 3, 4}
k is even

P(y=0) = . . . = $\frac{1}{2}$

<u>Continuous Case</u>
X $\sim$ P_x)(x) :pdf
Y = f(X) $\Rightarrow$ P_Y(y) = ?

Use cdf P_Y(Y$\leq$y) = P_Y(f(X)$\leq$y) = P_X(X$\leq$f^{-1}(y)) = P_X(f^{-1}(y))

Provided that f is invertible
Then P_Y(y) the pdf of Y is obtained by taking the derivative of cdf of Y

P_Y(y)=$\frac{d}{dy}P_Y$(Y$\leq$y)=$\frac{d}{dy}P_X$(f^{-1}(y))=$\frac{d}{dx}P_X$(x)$\cdot\frac{dx}{dy}$
Where x=f^{-1}(y) ignore sign of $\frac{dx}{dy}$
$\Rightarrow$ P_Y(y)=P_X(x)[$\frac{dx}{dy}$] ; For multivariable case J = $(\frac{\sigma y_i}{\sigma x_i})_{i,j}$
[det J]

<u>example</u>

X = (X_1,X_2);
Y = (γ,θ) : $X_1 = \gamma$Cosθ; $X_2 = \gamma$Sinθ

$$J = \begin{pmatrix} \frac{\sigma x_1}{\sigma\gamma} & \frac{\sigma x_1}{\sigma\theta} \\ \frac{\sigma x_2}{\sigma\gamma} & \frac{\sigma x_2}{\sigma\theta} \end{pmatrix} = \begin{pmatrix} Cos\theta & -\gamma Sin\theta \\ Sin\theta & \gamma Cos\theta \end{pmatrix}; det(J) = \gamma Cos^2\theta + \gamma Sin^2\theta = \gamma$$

|det(J)|=$|\gamma|$ P_Y(y)=P_X(x)|det(J)|=$|\gamma| P_X$(x)

Section 4: The Gamma Distribution

X, f(X)

Draw samples from X $x_1, \ldots, x_5$ (observations)
Approximate f(X) by the empirical distribution
of f(X) on $x_1, \ldots, x_5$

$\bar{u}$: A = $\bar{u}\gamma^2 \Rightarrow \bar{u} = \frac{A}{\gamma^2} = Approximate$
A=$4\gamma^2(\frac{1}{5})\sum_{i=1}^{5} f(x_i, y_i) f(x, y) = I(x^2 + y^2 \leq \gamma^2)$