CSCI 561: Homework 1

Deepika Anand

September 20, 2016

Problem 1. (a)(a)

The standard *Beta* distribution gives the probability density of a value x on the interval (0,1):

$$Beta(\alpha, \beta) : \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \tag{1}$$

where B is the beta function

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$$

Replacing $\alpha = 0$ and $\beta = 1$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} dt$$

Hence, the value of B will reduce to α^{-1}.

$$Beta(\alpha, \beta) : x^{\alpha-1} * \alpha \tag{3}$$

Therefore,

$$L(\alpha|X) = \alpha * \prod_{i=1}^{n} (x_i)^{\alpha-1} \tag{4}$$

Taking the log of this function

$\mathrm{LL}' = \log(L(\alpha|X)) = n\log(\alpha) + (\alpha - 1)\sum_{i=1}^{n} \log(x_i)$(5)

Differentiating,

$$\frac{dLL'}{d\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \log(x_i) \tag{6}$$

Equating this to zero, $\alpha_{MLE} = \frac{n}{\sum_{i=1}^{n} log(1/x_i)}$

Problem 1. (a)(b)

N(θ, θ) = $(2\pi\theta)^{-\frac{1}{2}} e^{-\frac{(x_i-\theta)^2}{2\theta}}$
L$(\theta|X) = \prod N(\theta, \theta)$

L$(\theta|X) = (2\pi\theta)^{-\frac{N}{2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2\theta}}$

Log(L$(\theta|X)) = -(N/2) * log(2\pi\theta) - \sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2\theta}$
Now differentiate,
$\frac{dLL}{d\theta}$ = -(N/(2*θ)) $- (N/2) + \frac{\sum x^2}{2*\theta^2}$
So this is an quadratic equation where D = $N^2 - 4 * N * \sum x^2$
Two roots are $\frac{-N+\sqrt{D}}{2*N}$, $\frac{-N-\sqrt{D}}{2*N}$

Problem 1. (b)

f(x) = (1/n) * $\sum \frac{1}{h} * K(\frac{a-X_i}{h})$
Since each x_i is i.i.d. Therefore Expectation value is summation of each of expected of value $E[X] = nE[X_i]$, which is equal to
(n/n) * $\sum \frac{1}{h} * K(\frac{a-X_i}{h})$
which can be written as,
$\frac{1}{h}$ * $\int K(\frac{a-X_i}{h})$
Replacing X_i with t, We get $\frac{1}{h}$ * $\int K(\frac{a-t}{h}) * f(t) * dt$

Now say z = $\frac{x-t}{h}$
=> dz = -dt
Applying Taylor's theorem we get,
E[$\hat{\text{f}}$(x)] = $\int K(z) * f(x - hz)dz$
=$\int K(z)[f(x) - h * z * f'(x) + (h^2)(z^2) * f''(x) + ...]$
Since the variable is continuous therefore, $\int K(z)$ = 1 and instance probability that is $\int z * K(z) = 0$ and $\int (z^2) * K(z) = \sigma^2$
Hence, E[$\hat{\text{f}}$(x)] = f(x) + $\frac{(h^2)*(\sigma^2)f''(x)}{2}$ + higher order terms
E[$\hat{\text{f}}$(x)] - f(x) = $\frac{(h^2)*(\sigma^2)f''(x)}{2}$ + higher order terms

Problem 2. (a)

Given that Y follows Bernoulli distribution
$P(X_j|Y = 1) = p_{j1}^{x_j}(1 - p_{j1})^{1-x_j}$
Similarly, $P(X_j|Y = 0) = p_{j0}^{x_j}(1 - p_{j0})^{1-x_j}$

$P(Y = 1|X) = \frac{(P(X|Y=1)P(Y=1))}{P(X)}$
$P(Y = 1|X) = \frac{(P(X|Y=1)P(Y=1))}{P(X|Y=1)P(Y=1)+P(X|Y=0)P(Y=0)}$
Assume, $P(Y = 1) = \prod$
$P(Y = 0) = 1 - \prod$

$\frac{P(X|Y=0)}{P(X|Y=1)} = \frac{(p_{j0}^{x_j}(1-p_{j0})^{1-x_j})}{(p_{j1}^{x_j}(1-p_{j1})^{1-x_j})}$

$ln(\frac{P(X|Y=0)}{P(X|Y=1)}) = ln\frac{(p_{10}^{(}x_1))*(1-p_{10})^{(}1-x_1)}{(p_{11}^{(}x_1))*(1-p_{11})^{(}1-x_1)} * ......$

Taking prod over all X$_i$

= X $\sum ln\frac{p_{i0}}{p_{i1}} + (1-X) * ln\frac{1-p_{i0}}{1-p_{i1}}$

This can be rearranged to get

$ln\frac{1-p_{i0}}{1-p_{i1}} + X * ln\frac{p_{i0}(1-p_{i1})}{p_{i1}*(1-p_{i0})}$

Now replacing these values in P(Y=1|X) equation

$w_0 = -[ln\frac{\prod}{1-\prod} + \sum ln\frac{1-p_{i1}}{1-p_{i0}}]$

w$^T X = ln\frac{p_{i0}*(1-p_{i1})}{p_{i1}*(1-p_{i0})}X$

Problem 2. (b)

given that X ranges is D-dimensional and Y can belong to any k point

$N(\mu_{jk}, \sigma_{jk}) = \frac{1}{\sqrt{2*\pi*\sigma_{jk}}} * e^{\frac{-(x-_{jk})^2}{\sigma_{jk}}}$

$P(x|Y = y_k) = P(x|Y = y_k) * P(X = X_j|Y = y_k)$

To find the maximum likelihood we need to consider all value of x

$L(\theta|x) = \prod P(x|Y = y_k) * P(X = X_j|Y = y_k)$

Now take log of the statement above $LL(\theta|x) = \sum P(Y = y_i) + \sum\sum logP(X = X_j|Y = y_k)$

Since the distribution is gaussian therefore, log P(X=X$_j$$|Y = y_k) = -\frac{log(2*\pi*\sigma_{jk})}{2} - \frac{(x_j-\mu_{jk})^2}{2*\sigma_{jk}}$

The equation above is for single x value. Therefore because of summation the entire $LL(\theta|x)$ equation will be multiplied by N_k

We will differentiate $LL(\theta|x)$ twice one by μ_{jk} and next time σ_{jk}

To get $\mu_{jk} = \frac{2*((\sum x_j)-(N_k)*\mu_{jk})}{\sigma_{jk}} = 0$

=> $\mu_{jk} = \sum x_{ij}/N_k$, because Y_i can take k values.

Similarly, differentiating wrt σ_{jk}, we get

$\sigma_{jk} = \frac{\sum(x_{ij}-\mu_{jk})^2}{N_k}$

Problem 3. (a)

Xi	Yi	Xi-μ_x	(Xi-μ_x)^2	Yi-μ_y	(Yi - μ_y)^2	N$_x = (Xi - \mu_x)/\sigma_x$	N$_y = (Yi - \mu_y)/\sigma_y$
0	49	-12.769	163.047361	36.693	1346.376249	-0.616383472	1.415079059
-7	32	-19.769	390.813361	19.693	387.814249	-0.954286542	0.759467798
-9	47	-21.769	473.889361	34.693	1203.604249	-1.050830276	1.337948322
29	12	16.231	263.445361	-0.307	0.094249	0.783500676	-0.011839568
49	31	36.231	1312.685361	18.693	349.428249	1.748938019	0.72090243
37	38	24.231	587.141361	25.693	660.130249	1.169675613	0.990860008
8	9	-4.769	22.743361	-3.307	10.936249	-0.230208534	-0.127535673
13	-1	0.231	0.053361	-13.307	177.076249	0.011150801	-0.513189356
-6	-3	-18.769	352.275361	-15.307	234.304249	-0.906014675	-0.590320093
-21	12	-33.769	1140.345361	-0.307	0.094249	-1.630092682	-0.011839568
27	-32	14.231	202.521361	-44.307	1963.110249	0.686956941	-1.708715773
19	-14	6.231	38.825361	-26.307	692.058249	0.300782004	-1.014539144
27	-20	14.231	202.521361	-32.307	1043.742249	0.686956941	-1.245931354

$\mu_x = 12.76923077$
$\mu_y = 12.30769231$
$\sigma_x = 20.71695701$
$\sigma_y = 25.93062738$
Normalised target point(T$_x, T_y$) = $(0.349026608, -0.204688156)$

N$_x - T_x$	(N$_x - T_x$)^2	N$_y - T_y$	(N$_y - T_y$)^2	Euclidean Dist	Manhattan Dist
1.066079059	1.13652456	1.619679059	2.623360254	1.939042241	2.685758118
0.410467798	0.168483813	0.964067798	0.929426719	1.04781226	1.374535596
0.988948322	0.978018784	1.542548322	2.379455327	1.832341156	2.531496645
-0.360839568	0.130205194	0.192760432	0.037156584	0.409098739	0.5536
0.37190243	0.138311417	0.92550243	0.856554747	0.997429779	1.297404859
0.641860008	0.41198427	1.195460008	1.42912463	1.356874681	1.837320015
-0.476535673	0.227086248	0.077064327	0.005938911	0.482726794	0.5536
-0.862189356	0.743370486	-0.308589356	0.095227391	0.915749898	1.170778712
-0.939320093	0.882322236	-0.385720093	0.14877999	1.015432039	1.325040185
-0.360839568	0.130205194	0.192760432	0.037156584	0.409098739	0.5536
-2.057715773	4.234194203	-1.504115773	2.262364259	2.548834726	3.561831546
-1.363539144	1.859238997	-0.809939144	0.656001417	1.585950949	2.173478288
-1.594931354	2.543806023	-1.041331354	1.084370988	1.904777418	2.636262707

For L1 and K=1, nearest to (8,9). Belongs to **Computer Science**.
For L1 and K=5, Top 5 distances (0.5536, C);(0.5536, C);(0.5536, E);(1.17, C);(1.29, E). Belongs to **Computer Science**

For L2 and K=1, nearest to (29, 12).Belongs to **Electrical Engineering**.
For L2 and K=5, Top 5 distances (0.409, E), (0.409, C), (0.915, C),(0.997, E), (1.015, C). Belongs to **computer science**.

Problem 3. (b)

$$p(x) = \sum p(x|Y = c)P(Y = c) \tag{7}$$

$$\sum\left(\frac{k_c}{N_c * V}\right) * \frac{N_c}{N} \tag{8}$$

Since $\sum K_c = K$. Therefore

$$p(x) = \frac{K}{N * V} \tag{9}$$

For the next part,

$$P(Y = c|x) = \frac{p(x|Y = c) * p(Y = c)}{p(x)} \tag{10}$$

$$P(Y = c|x) = \frac{\frac{K_c}{N_c V} * \frac{N_c}{N}}{\frac{K}{N * V}} \tag{11}$$

$$P(Y = c|x) = \frac{K_c}{K} \tag{12}$$

Problem 4. (a)

Information gain is given for attribute outcome Y given X- attribute to split on is given by I(X, Y) = H(Y) - H(Y|X), Y = Accident Rate, H(Y) will remain constant whether X = Weather or X = Traffic.

$p_{high} = 0.73$

$p_{low} = 0.27$

$H(Y|X = Weather) = -p_{high} * (p_{high,sunny} * log(p_{high,sunny}) + p_{high,rainy} * log_{high,rainy}) - p_{low} * (p_{low,sunny} * log(p_{low,sunny}) + p_{low,rainy} * log_{low,rainy})$

$= -0.73 * (0.851 * log0.851 + 0.694 * log0.694) - 0.27 * (0.185 * log0.185 + 0.305 * log0.305)$

H(Y|X=Traffic) = -$p_{high} * (p_{high,heavy} * log(p_{high,heavy}) + p_{high,light} * log(p_{high,light})) - p_{low} * (p_{low,heavy} * log(p_{low,heavy}) + p_{low,light} * log_{low,light})$

$= -0.73 * (1 * log1) - 0.27 * (1 * log0.27)$

$= 0$

H(Y) = -$p_{high} * logp_{high} - p_{low} * p_{low}$

$H(Y) = -0.73 * log0.73 - 0.27 * log0.27 = 0.252$

Since H(Y|X=Weather) >H(Y|X=Traffic), subtracting this from H(Y) which is positive quantity will result in lower information gain from Weather attribute. Hence, **Traffic** attribute will be used since I(X, Y) is maximum in that case

Problem 4. (b)

The tree will remain the same because decision does not depend on the the distribution of values of attributes. It merely depends on the frequency of attributes. So normalizing will not change the graph

Problem 4. (c)

To prove: Gini Index:

$$\sum p_k log(1 - p_k) \tag{13}$$

is greater than Cross entropy

$$-\sum p_k log(p_k) \tag{14}$$

Given that $0 \leq$ p$_k$$\leq 1$
Hence, $0 <$ log p$_k$ < 1
$And, 1 - logp_k > 0$
$Therefore, 1 - p_k - (-logp_{k)}$>0
$=> 1 - p_k > -log(p_k)$
(15)Multiplying both sides with a non negative p_k we get

$$p_k * (1 - p_k) > -p_k * log(p_k) \tag{16}$$

Taking summation on both sides. We get,

$$\sum p_k * (1 - p_k) > \sum -p_k * log(p_k) \tag{17}$$

Problem 5.

- Total of 11 features are present in this data set.
- No, not all features are relevant. For ex: ID field does not convey any meaning information neither helps in decision making.
- There are total 7 types of class 1-7 but not data is present for class 4
- Class 2 is in majority with a total of 73 instances. No, this is not a uniform distribution since there is no symmetry among the values.

No From the results it is clear that Naive classifier is not very accurate or efficient and since naive depends on the frequency of each classification class which can lead to increase in misclassfication. However, kNN is comparatively stable and even if the data is skewed even then the accuracy is approx twice as high of naive bayes.