CS6140: Machine Learning – Homework 2
Name: Vishal Rajpal

Problem 1: Bernoulli Distribution
a)

$$\boldsymbol{p(x \mid \mu) = \left(\frac{1-\mu}{2}\right)^{\frac{(1-x)}{2}} \left(\frac{1+\mu}{2}\right)^{\frac{(1+x)}{2}}}$$

Probability Mass Function = $\sum_{x \in A} \boldsymbol{p}(\boldsymbol{x})$

We need to prove $p_x(-1) + p_x(1) = 1 - \boldsymbol{Equation\ 1}$

$$p_x(-1) = \left(\frac{1-\mu}{2}\right)^{\frac{(1-(-1))}{2}} \left(\frac{1+\mu}{2}\right)^{\frac{(1+(-1))}{2}}$$

$$p_x(-1) = \left(\frac{1-\mu}{2}\right)^{1} \left(\frac{1+\mu}{2}\right)^{0}$$

$$p_x(-1) = \frac{1-\mu}{2}$$

$$p_x(1) = \left(\frac{1-\mu}{2}\right)^{\frac{(1-1)}{2}} \left(\frac{1+\mu}{2}\right)^{\frac{(1+1)}{2}}$$

$$p_x(1) = \left(\frac{1-\mu}{2}\right)^{0} \left(\frac{1+\mu}{2}\right)^{1}$$

$$p_x(1) = \frac{1+\mu}{2}$$

Substituting these values in Equation 1

$$\frac{1-\mu}{2} + \frac{1+\mu}{2} = 1$$

$$\frac{1-\mu+1+\mu}{2} = 1$$

$$\frac{2}{2} = 1$$

Hence Proved

b) Evaluate its mean

$$\boldsymbol{E[X] = \sum_{x=1}^{n} x * p(x)}$$

$$E[X] = -1 * \left(\frac{1-\mu}{2}\right) + 1 * \left(\frac{1+\mu}{2}\right)$$

$$E[X] = \left(\frac{-1+\mu}{2}\right) + \left(\frac{1+\mu}{2}\right)$$

$$E[X] = \frac{-1+\mu+1+\mu}{2}$$

$$E[X] = \frac{2\mu}{2}$$

$$\boldsymbol{E[X] = \mu}$$

c) Evaluate its Variance

$$\boldsymbol{Var(X) = E[X^2] - E[X]^2}$$

$$E[X^2] = (-1)^2 * p(-1|\mu) + 1^2 * p(1|\mu)$$

$$E[X^2] = p(-1|\mu) + p(1|\mu)$$

$$p(-1|\mu) = \left(\frac{1-\mu}{2}\right)^{\frac{1-(-1)}{2}} \left(\frac{1+\mu}{2}\right)^{\frac{1+(-1)}{2}}$$

$$p(-1|\mu) = \left(\frac{1-\mu}{2}\right)$$

$$p(1|\mu) = \left(\frac{1-\mu}{2}\right)^{\frac{1-1}{2}} \left(\frac{1+\mu}{2}\right)^{\frac{1+1}{2}}$$

$$p(1|\mu) = \left(\frac{1+\mu}{2}\right)$$

$$E[X^2] = \left(\frac{1-\mu}{2}\right) + \left(\frac{1+\mu}{2}\right)$$

$$E[X^2] = \left(\frac{1-\mu+1+\mu}{2}\right)$$

$$E[X^2] = \frac{2}{2} = 1$$

Var(X) = 1 – $\boldsymbol{\mu^2}$

Problem 2: Naïve Bayes –I

a) Linear Threshold Function

$$\boldsymbol{f(x) = sgn(w * x - \theta)}$$

$$\boldsymbol{Let\ w = 1\ and\ \theta = m}$$
$$(based\ on\ the\ question\ atleast\ m\ values\ of\ x\ need\ to\ be\ 1)$$

Therefore, this function would be: $\boldsymbol{f(x) = sgn(x - m)}$

Hence, if x is greater than or equal to m this function will return **+(positive label)** else this function will return **–(negative label).**

b) 2-class Naïve Bayes Hypothesis

$$\log\frac{p(y=1)}{p(y=0)} + \sum_{i=1}^{n}\log\frac{1-\mu_i}{1-\chi_i} + \sum_{i=1}^{n}[\log\frac{\mu_i}{1-\mu_i} - \log\frac{\chi_i}{1-\chi_i}]\, x_i > 0$$

$$\boldsymbol{Sample\ Space = 2^8 = 256}$$

$$\boldsymbol{p(y=0) = \binom{8}{0} + \binom{8}{1} + \binom{8}{2}}$$

$$= \frac{8!}{(0! * 8!)} + \frac{8!}{1! * 7!} + \frac{8!}{2! * 6!}$$

$$= 1 + 8 + 28$$

$$p(y=0) = 37$$

$$\boldsymbol{p(y=1) = 256 - p(y=0)}$$

$$= 256 - 37$$

$$= 219$$

Therefore,

$$\log\frac{p(y=1)}{p(y=0)}$$

$$= \log\left(\frac{219}{37}\right)$$

$$\log\frac{p(y=1)}{p(y=0)} = 0.772 - \boldsymbol{Equation\ 1}$$

$$\sum_{i=1}^{n} \log \frac{1-\mu_i}{1-\chi_i}$$

$$\boldsymbol{\mu_i = p(x_i|y = 1)}$$

$$= \frac{\left(\binom{7}{2} + \binom{7}{3} + \binom{7}{4} + \binom{7}{5} + \binom{7}{6} + \binom{7}{7}\right)}{219}$$

$$= \frac{(21 + 35 + 35 + 21 + 7 + 1)}{219}$$

$$\boldsymbol{\mu_i = \frac{120}{219} = 0.547}$$

$$\chi_i = p(x_i = 1\ |y = 0)$$

$$= \frac{\left(\binom{7}{0} + \binom{7}{1}\right)}{219}$$

$$= \frac{1+7}{219}$$

$$\boldsymbol{\chi_i = \frac{8}{37} = 0.216}$$

$$\sum_{i=1}^{n} \log \frac{1-\mu_i}{1-\chi_i}$$

$$= \sum_{i=1}^{n} \log \frac{1-0.547}{1-0.216}$$

$$= \sum_{i=1}^{n} \log \frac{0.453}{0.784}$$

$$= \sum_{i=1}^{n} \log 0.577$$

$$= \sum_{i=1}^{n} -0.238 = -1.91$$

$$\sum_{i=1}^{n} [\log \frac{\mu_i}{1-\mu_i} - \log \frac{\chi_i}{1-\ \chi_i}]\ x_i$$

$$= \sum_{i=1}^{n} \left(\log \left(\frac{0.547}{1-0.547} \right) - \log \left(\frac{0.216}{1-0.216} \right) \right) x_i$$

$$= \sum_{i=1}^{n} \left(\log \left(\frac{0.547}{1-0.547} \right) - \log \left(\frac{0.216}{1-0.216} \right) \right) x_i$$

$$== \sum_{i=1}^{n} (0.081 - (-0.55)\ ) x_i$$

$$\sum_{i=1}^{n} (0.081 - (-0.55)\ ) x_i$$

$$\sum_{i=1}^{n} (0.631) x_i$$

$$0.772 + (-1.91) + \sum_{i=1}^{n} (0.631) x_i$$

$$= \ -1.1 + \sum_{i=1}^{n} (0.631) x_i > 0$$

$$\sum_{i=1}^{n} (0.631) x_i - 1.1 > 0 - in\ the\ form\ of\ w.x - \ \theta > 0$$

c) No it does not learn the target function. Example: if 2 bits in a 8 bit vector are 1 then the result would be

$$1.262 - 1.1 > 0 = 0.162 > 0$$

Hence, this function will return true and this is a case of **false positive**.

Problem 3:

c) Prediction Matrix

Actual/ Predicted	Articles	Corporate	Enron_t_s	Enron_travel_club	Hea_nesa	personal	systems	Tw_commercial_group	Accuracy
Articles	4	2	0	0	0	0	0	0	4/6 = 0.66
Corporate	0	39	1	0	0	0	1	4	39/45 = 0.86
Enron_t_s	0	4	1	0	0	0	0	1	1/6 = 0.16
Enron_travel_club	0	1	0	1	0	0	0	0	1/2 = 0.5
Hea_nesa	0	0	0	0	12	0	0	0	12/12 = 1.0
personal	0	4	1	0	0	22	0	4	22/31 = 0.70
systems	1	5	1	0	0	3	4	3	4/17 = 0.23
Tw_commercial_group	4	2	1	0	0	1	0	143	143/151 = 0.94

d) Interpretation

$$Accuracy = \frac{Sum\ of\ diagonal}{Sum\ of\ every\ row} = \frac{4+39+1+1+12+22+4+143}{6+45+6+2+12+31+17+151} = \frac{226}{270} = 0.837$$