**CS-419m: Practice questions on Probabilistic Classifiers**

1. An insurance company is trying to classify customers as "Risky" ($y = 1$) or "Not-risky" ($y = 2$) based on two attributes of a customer: $x_1$ denoting the number of "incidents" in the past ten years and $x_2$ denoting the type of the vehicle. Assume for each class $y$ that the first attribute follows a Poisson distribution with parameter $\lambda_y$ (Recall that for a Poisson distribution $P(x = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$) and the second attribute follows a multinomial distribution with parameter $\theta_{yr}$ for $r \in (F, T, S)$. The table below denotes a sample training set.

| | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|
| $\mathbf{x}^0$ | 0 | F | 2 |
| $\mathbf{x}^1$ | 3 | T | 1 |
| $\mathbf{x}^2$ | 1 | T | 2 |
| $\mathbf{x}^3$ | 2 | S | 1 |
| $\mathbf{x}^4$ | 2 | F | 2 |
| $\mathbf{x}^5$ | 3 | T | 1 |
| $\mathbf{x}^6$ | 4 | S | 1 |
| $\mathbf{x}^7$ | 5 | S | 1 |

(a) Write the estimate of $\lambda_y$ for each $y$ using the training data above? [Show the derivation for one $y$]

..2

(b) Is the maximum likelihood objective concave in $\lambda_y$? Justify.

..3

(c) Write the estimate of all the $\theta_{yr}$s for $r \in (F, T, S)$ and each $y$

..1

(d) Assume the class labels follow a Bernoulli distribution with parameter $p$ denoting the probability of class 1. What is the maximum likelihood estimate of $p$.

..1

2. Consider a binary classification problem ($y \in \{0, 1\}$) and two dimension data ($d = 2$) where both $x_1$ and $x_2$ are binary. Let $n^y_{00}, n^y_{10}, n^y_{01}, n^y_{11}$ denote the number of training instances in class $y$ where $[x_1\ x_2]$ is [0 0], [1 0], [0 1], [1 1] respectively.

   (a) Suppose we use a naive Bayes classifier where $x_1$ is assumed independent of $x_2$.

      i. What is the maximum likelihood estimate of the two Bernoulli parameters in class $y$ in terms of the counts $n^y_{00}, n^y_{10}, n^y_{01}, n^y_{11}$

      ..1

      ii. What is the maximum likelihood estimate of $\Pr(y)$ in terms of the counts $n^y_{00}, n^y_{10}, n^y_{01}, n^y_{11}$

      ..1

      iii. For what values of $n^y_{00}, n^y_{10}, n^y_{01}, n^y_{11}$ will the naive Bayes classifier incur the maximum percentage training error? Justify. Assume number of instances in both classes is the same.

      ..3

   (b) Now suppose we decide to use a model more powerful than naive Bayes. In each class, we model the probability that $\Pr([x_1\ x_2] = [a\ b]\ |y)$ with a parameter $p^y_{ab}$. Further we follow a LDA like parameter sharing to reduce the number of parameters by requiring that

$p^0_{11} = p^1_{11} = p_{11}$. Clearly, $p^y_{00} + p^y_{10} + p^y_{01} + p_{11} = 1$. Write the log-likelihood function on the training data in terms of the counts $n^y_{ab}$, and parameters $p^y_{ab}$, $\Pr(y = 1), \Pr(y = 0)$.

..2

(c) Derive the maximum likelihood estimates of the $p^y_{00}, p^y_{10}, p^y_{01}$ parameters only in terms of $p_{11}$ and the training counts. [Hint: Use Lagrangian multiplier to push the two constraints $p^y_{00} + p^y_{10} + p^y_{01} + p_{11} = 1$, one for each $y$ to the objective.]

..3

(d) Now derive the maximum likelihood estimate for $p_{11}$.

..1

..1