Gordon Ng CS 505 Homework

1k) $Y$ is boolean

$X = \{X_1, X_2 \ldots X_n\}$

For $X_i$, $P(X_i | Y = y_k)$ is Gaussian/Normal distribution

Every $X_i$ and $X_j$, $i \neq j$ are conditionally independent

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

$$P(Y=1|X) = \frac{1}{1 + \exp\left(\ln\frac{1-\pi}{\pi} + \sum_i \ln\frac{P(X_i|Y=0)}{P(X_i|Y=1)}\right)}$$

We turn the summation to Gaussian form.

$$P(Y=1|X) = \frac{1}{1 + \exp\left(\ln\frac{1-\pi}{\pi} + \sum_i\left(\frac{\mu_{i0} - \mu_{i1}}{\sigma_i^2} X_i + \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2}\right)\right)}$$

Set the weighted sums to $w_0$ and $w_i$

$$w_0 = \ln\frac{1-\pi}{\pi} + \sum_i \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2} \qquad w_i = \frac{\mu_{i0} - \mu_{i1}}{\sigma_i^2}$$

$$P(Y=0|X) = 1 - P(Y=1|X) = \frac{\exp(w_0 + \sum_{i=1}^n w_i X_i)}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)}$$

Conditional Data Likelihood Log Regression parameter

$$W \leftarrow \arg\max_W \prod_l P(Y^l | X^l, W)$$

$$l(W) = \sum_l Y^l = 1 | X^l, W) + (1 - Y^l) \ln P(Y^l = 0 | X^l, W)$$

$$= \sum_l Y^l \left(w_0 + \sum_i^n w_i X_i\right) - \ln\left(1 + \exp\left(w_0 + \sum_i^n w_i X_i\right)\right)$$

No closed form to maximize $l(W)$, so gradient ~~descent~~ ascent

$$\frac{\partial l(W)}{\partial w_i} = \sum_l X_i^l \left(Y^l - \hat{P}(Y^l = 1 | X^l, W)\right)$$

You can plug in $P(Y=0|X)$ to form a Log Regression.

1) 2. $Y$: Sentiment $\in 0,1$
$X_1$: Summer $\in 0,1$
$X_2$: Rowdy $\in 0,1$

$P(\text{Sentiment}=1) = 0.5$
$P(\text{Summer}=1 \mid \text{Sentiment}=1) = 0.8$
$P(\text{Summer}=1 \mid \text{Sentiment}=0) = 0.7$
$P(\text{Rowdy}=1 \mid \text{Sentiment}=1) = 0.4$
$P(\text{Rowdy}=1 \mid \text{Sentiment}=0) = 0.5$

a) decision function $f(x1,x2)=1$
if $P(Y=1 \mid x_1,x_2) > P(Y=0 \mid X_1, X_2)$

$Y=1$ if: $\dfrac{\log P(Y=1)}{\log P(Y=0)} + \dfrac{\log P(X_1=x_1 \mid Y=1)}{\log P(X_1=x_1 \mid Y=0)} + \dfrac{\log P(X_2=x_2 \mid Y=1)}{\log P(X_2=x_2 \mid Y=0)} > 0$ else $Y=0$

b)

$X_1$ = Summer
$X_2$ = Rowdy

| $X_1$ | $X_2$ | $Y$ |
|---|---|---|
| 1 | 0 | 1 |
| 0 | 1 | 0 |
| 0 | 0 | 0 |

$E_e = \sum_{i\in(0,1)} \sum_{j\in(0,1)} \sum_{k\in(0,1)} P(X_1=i\ X_2=j\ Y=k) \times I[k \neq \text{NB-Prediction}(x_1=i, x_2=j)]$

$I[z]$ is an indicator function, where $I[z]=1$ or $I[z]=0$

$E_e = P(X_1=0, Y=1) + P(X_1=1, Y=0) = P(Y=1)P(X_1=0 \mid Y=1) + P(Y=0)P(X_1=1 \mid Y=0)$

~~$E_e = 0.5\times 0.7 + 0.5\times 0.5$~~

For $x_1$ $E_e = 0.5\times 0.2 + 0.5\times 0.7 = 0.1 + 0.35 = 0.45$
For $X_2$ $E_e = 0.5\times 0.6 + 0.5\times 0.5 = 0.6 + 0.25 = 0.85$
For 1 attribute $P(X_1, X_2, Y) = P(Y)P(X_1 \mid Y)P(X_2 \mid Y)$

$E_e = P(X_1=1, X_2=1, Y=0) + P(X_1=1, X_2=0, Y=0) + P(X_1=0, X_2=1, Y=0) + P(X_1=0, X_2=0, Y=1)$

$E_e = 0.5\times 0.7\times 0.5 + 0.5\times 0.7\times 0.5 + 0.5\times 0.3\times 0.5 + 0.5\times 0.2\times 0.6$

$E_e = 0.485$

c) $P(X_1=1, X_2=1, Y=1)$ $\quad P(x_1 \mid Y)\ P(X_2 \mid Y)$

$P(X_1=1 \mid Y=1) = \dfrac{P(Y \mid X_1)\cdot P(X_1)}{P(Y)}$ $\quad P(X_2=1 \mid Y=1) = \dfrac{P(Y \mid X_2)\ P(X_2)}{P(Y)}$

$P(Y \mid X_1)\cdot P(X_1) = 0.5\times 0.8 = .4$ $\quad P(Y \mid X_2)P(X_2) = 0.4\times 0.5 = .2$

$\dfrac{\log(X_1 \mid Y=1)}{\log(X_1 \mid Y=0)} = 0.62562$ $\quad \dfrac{\log(X_2 \mid Y=1)}{\log(X_2 \mid Y=0)} = 1.32192$

$P(X_1) = 0.63936$ $\quad P(X_2) = 0.15129$

$P(X_1, X_2, Y) = 0.5\times 0.63936 \times 0.15129$
$0.04836\cancel{4}872$

1) 2d) The NB assumptions are violated because the attributes are supposed to be conditionally independent, but $X_2$=Winter is not independent.
The ~~gom~~ joint probability should be the same as $P(Y=1, X_1, X_2)$
Winter is just another season, two seasons can't happen at once. So,
0.0483643872.

2e) The expected error rate is higher.

$X_1 = 1 \cap X_2 = 1 \mid Y = 1$
$X_1 = 1 \cap X_2 = 0 \mid Y = 1$
$X_1 = 0 \cap X_2 = 1 \mid Y = 0$
$X_1 = 0 \cap X_2 = 0 \mid Y = 0$

$P(X_1 = 1, X_2 = 1, Y = 0) + P(X_1 = 1, X_2 = 0, Y = 0) + P(X_1 = 0, X_2 = 1, Y = 1) + P(X_1 = 0, X_2 = 0, Y = 1)$

$P(X_1, X_2, Y) = P(Y)\,P(X_1 \mid Y)\,P(X_2 \mid Y)$

$0.5 \times 0.7 \times 0.5 + 0.5 \times 0.5 \times 0.7 + 0.2 \times 0.5 \times 0.4 + 0.2 \times 0.6 \times 0.5$

$0.175 + 0.175 + 0.04 + 0.06 = 0.45 = E_e$

2f) It doesn't improve because we did not provide new information as Winter and Summer are both seasons. Adding more attributes will only make ~~the~~ Expected error go up, as we saw here. (question 2c-2e)

3a) These words will recieve zero probability in maximum likelihood estimates. This zero probability is still there even after the event, so it will mess up your data.

3b) Logistic regression will use parameters of logit/regression functions not estimates of $P(Y)$ and $P(X \mid Y)$.

$w_j := w_j + \eta \frac{\partial}{\partial w_j} l(w)$ The weight is calc using gradient descent and all weights are simultaneously updated. It uses a logit function.