CSE - 575

ASHISH SIROHI

ASSIGNMENT #1

1) <u>Baye's Classifier</u> :

Prior :- Beta$(\beta_H, \beta_T) \Rightarrow \dfrac{\theta^{\beta_H - 1}(1-\theta)^{\beta_T - 1}}{B(\beta_H, \beta_T)}$

$\alpha_H \rightarrow$ Heads

$\alpha_T \rightarrow$ Tails

Likelihood : $\theta^{\alpha_H}(1-\theta)^{\alpha_T}$

We know that

$$P(\theta|D) = \frac{\overset{\text{Prior}}{P(\theta)} \times \overset{\text{Likelihood}}{P(D|\theta)}}{\int_\theta P(\theta)P(D|\theta)\,d\theta} \Big\} \Rightarrow \text{Marginal likelihood} = P(D)$$

Let's say $\int_\theta P(\theta)\,P(D|\theta)\,d\theta = c$

Posterior $\propto \dfrac{1}{c}\,\theta^{\alpha_H}(1-\theta)^{\alpha_T} \cdot \dfrac{\theta^{\beta_H - 1}(1-\theta)^{\beta_T - 1}}{B(\beta_H, \beta_T)}$

$$\propto \frac{1}{c \times B(\beta_H, \beta_T)} \times \theta^{\alpha_H + \beta_H - 1}(1-\theta)^{\alpha_T + \beta_T - 1}$$

$$\propto \frac{1}{c'} \times \theta^{\alpha_H + \beta_H - 1}(1-\theta)^{\alpha_T + \beta_T - 1}$$

$$P(\theta|D) \sim \text{Beta}(\alpha_H + \beta_H, \alpha_T + \beta_T)$$

$$\text{Mean} = \frac{\alpha_H + \beta_H}{\alpha_H + \beta_H + \alpha_T + \beta_T}$$

Posterior: $P(\theta) \times P(D|\theta)$

$$P(\theta|D) = \frac{\theta^{\beta_H - 1}(1-\theta)^{\beta_T - 1}}{B(\beta_H, \beta_T)} \cdot \theta^{\alpha_H}(1-\theta)^{\alpha_T}$$

$$\ln P(\theta|D) = \ln\left[\frac{\theta^{\beta_H - 1}(1-\theta)^{\beta_T - 1}}{B(\beta_H, \beta_T)}\right] + \ln\left[\theta^{\alpha_H}(1-\theta)^{\alpha_T}\right]$$

$$= \ln\frac{1}{B(\beta_H,\beta_T)} + \ln\frac{\theta^{\beta_H-1}}{\cancel{B(\beta_H,\beta_T)}} + \frac{\ln(1-\theta)^{\beta_T-1}}{\cancel{B(\beta_H+\beta_T)}} + \ln\theta^{\alpha_H} + \ln(1-\theta)^{\alpha_T}$$

$$= \frac{\ln}{B(\beta_H,\beta_T)} + (\beta_H - 1)\ln\theta + (\beta_T - 1)\ln(1-\theta) + \alpha_H \ln\theta + \alpha_T\ln(1-\theta)$$

$$\frac{d\ln P(\theta|D)}{d\theta} = 0 + \frac{\beta_H - 1}{\theta} + \frac{\beta_T - 1(-1)}{1-\theta} + \frac{\alpha_H}{\theta} + \frac{\alpha_T}{(1-\theta)}(-1) = 0$$

$$\frac{\beta_H - 1}{\theta} - \frac{(\beta_T - 1)}{1-\theta} + \frac{\alpha_H}{\theta} - \frac{\alpha_T}{1-\theta} = 0$$

$$\frac{\beta_H - 1 + \alpha_H}{\theta} = \frac{\beta_T + \alpha_T - 1}{1-\theta}$$

$$\beta_H + \alpha_H - 1 - \theta\beta_H + \theta - \theta\alpha_H = \theta\beta_T + \theta\alpha_T - \theta$$

$$\alpha_H + \beta_H - 1 - \theta(\alpha_H + \beta_H - 1) = \theta(\alpha_T + \beta_T - 1)$$

$$\hat{\theta}_{MAP} = \frac{\alpha_H + \beta_H - 1}{(\alpha_H + \beta_H - 1) + (\alpha_T + \beta_T - 1)}$$

2 Parameter Expectation →

$x_1, x_2, x_3 \ldots x_N \in R$ (i.i.d samples)

1) We know,

$$\hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$E[\hat{\mu}_{MLE}] = E[\frac{1}{N}\sum_{i=1}^{N} x_i]$$

$$= \frac{1}{N}\sum_{i=1}^{N} E[x_i]$$

$$= \frac{1}{\cancel{N}}(\cancel{N}\mu)$$

$$= \mu \rightarrow$$ It is a true value of μ and not a random variable hence unbiased.

2)

$$\hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu}_{MLE})^2$$

$$E[\hat{\sigma}^2_{MLE}] = E[\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu}_{MLE})^2]$$

$$= \frac{1}{N} E[\sum_{i=1}^{N}(x_i^2 - 2x_i\hat{\mu} + \hat{\mu}^2)]$$

$$= \frac{1}{N} E[\sum_{i=1}^{N} x_i^2 - 2\hat{\mu}\sum_{i=1}^{N} x_i + N\hat{\mu}^2]$$

$$= \frac{1}{N} E[\sum_{i=1}^{N} x_i^2 - 2\hat{\mu}(N\hat{\mu}) + N\hat{\mu}^2]$$

$$= \frac{1}{N} E[\sum_{i=1}^{N} x_i^2 - N\hat{\mu}^2]$$

$$= \frac{1}{N}\sum_{i=1}^{N} E[x_i^2] - \frac{1}{\cancel{N}} \times \cancel{N} E[\hat{\mu}^2]$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{i=1}^{N} E[x_i^2] - E[\hat{\mu}^2]$$

Now, we know that,

$$E[x_i^2] = \sigma^2 + \mu^2 \Rightarrow \sigma^2 = E[x_i^2] - \mu^2 \quad \text{—— (1)}$$

$$E[\hat{\mu}^2] = \frac{\sigma^2}{n} + \mu^2 \quad \text{—— (2)}$$

$$\hat{\sigma}^2_{MLE} = (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) \quad \therefore \text{from (1) \& (2)}$$

$$= \sigma^2 - \frac{\sigma^2}{n} + \cancel{\mu^2} - \cancel{\mu^2}$$

$$\boxed{\hat{\sigma}^2_{MLE} = \frac{(n-1)}{n}\sigma^2}$$

Here we can say that $\hat{\sigma}^2_{MLE} \neq \sigma^2$ hence it is unbaised.

About eq (2) // Additional Work to show eqn (2) //

$$Var(x_1 + x_2 \ldots x_n) = n \cdot \sigma_x^2$$

$$var\left(\frac{x_1 + x_2 \ldots x_n}{n}\right) = \frac{n \cdot \sigma_x^2}{n^2} = \frac{\sigma_x^2}{n} = \sigma_\mu^2$$

$$\frac{\sigma_x^2}{n} = E[\bar{x}^2] - \underbrace{(E[\bar{x}])^2}_{\mu^2}$$

$$E(\bar{x}^2) = \sigma^2 + \mu^2 \quad \text{—— (2)}$$

3 Naive Bayes Classifier I

1) Total No. of Independent Parameters →

2)

$$\text{age} \Rightarrow 2\times(3-1) = 4$$
$$\text{income} \Rightarrow 2\times(3-1) = 4$$
$$\text{student} \Rightarrow 2\times(2-1) = 2$$
$$\text{Credit\_rating} \Rightarrow 2\times(2-1) = 2$$
$$\text{buys\_computer} \Rightarrow (2-1) = 1$$

13 parameters.

Here are they: (MLE Estimations)

$P(\text{age}=\text{youth} \mid Y=\text{yes}) = \frac{2}{9}$

$P(\text{age}=\text{middle} \mid Y=\text{yes}) = 4/9$

$P(\text{income}=\text{high} \mid Y=\text{yes}) = 2/9$

$P(\text{income}=\text{low} \mid Y=\text{yes}) = 3/9$

$P(\text{student}=\text{yes} \mid Y=\text{yes}) = 6/9$

$P(\text{credit\_rating}=\text{fair} \mid Y=\text{yes}) = 6/9$

$P(\text{buys\_computer}) = 9/14$

$P(\text{age}=\text{youth} \mid Y=\text{NO}) = 3/5$

$P(\text{age}=\text{middle} \mid Y=\text{NO}) = 0$

$P(\text{income}=\text{high} \mid Y=\text{NO}) = 2/5$

$P(\text{income}=\text{low} \mid Y=\text{NO}) = 1/5$

$P(\text{student}=\text{yes} \mid Y=\text{NO}) = 1/5$

$P(\text{credit\_rating}=\text{fair} \mid Y=\text{NO}) = 2/5$

3) $x = (\text{youth, medium, yes, fair})$

$$P(Y=\text{yes} \mid x) = \frac{P(x \mid Y=\text{yes})\, P(Y=\text{yes})}{\sum_i P(x \mid y_i)}$$

$$= \frac{P(\text{age}=\text{youth} \mid Y=\text{yes}) \times P(\text{income}=\text{medium} \mid Y=\text{yes}) \times P(\text{student}=\text{yes} \mid Y=\text{yes}) \times P(\text{cr}=\text{fair} \mid Y=\text{yes}) \times P(Y=\text{yes})}{[P(\text{age}=\text{youth} \mid Y=\text{yes}) \times P(\text{income}=\text{med} \mid Y=\text{yes}) \times P(\text{student}=\text{yes} \mid Y=\text{yes}) \times P(\text{cr}=\text{fair} \mid Y=\text{yes}) \times P(Y=\text{yes})] + [P(\text{age}=\text{youth} \mid Y=\text{NO}) \times P(\text{income}=\text{med} \mid Y=\text{NO}) \times P(\text{st}=\text{yes} \mid Y=\text{NO}) \times P(\text{cr}=\text{fair} \mid Y=\text{NO}) \times P(Y=\text{NO})]}$$

$$= \frac{9/14 \times \frac{2}{9} \times \frac{4}{9} \times \frac{6}{9} \times \frac{6}{9}}{\left[\frac{9}{14} \times \frac{2}{9} \times \frac{4}{9} \times \frac{6}{9} \times \frac{6}{9}\right] + \left[\frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{5}{14}\right]}$$

$$= \frac{0.0282}{0.0282 + 0.00685}$$

$$= \frac{0.0282}{0.03505} = 0.8045$$

Since $P(y = yes|x) \geq 0.5$, we will classify it as $y = yes$.

4 Logistic Regression :

$+ : (1,0), (0,-1)$

$- : (0,1), (-1,0)$

$$W^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$P(Y=1 \mid x^1, w^0) = \frac{\exp(w_0 + \sum_i w_i x_i^1)}{1 + \exp(w_0 + \sum_i w_i x_i^1)}$$

$$= \frac{\exp(0+0+0)}{1+\exp(0+0+0)} \Rightarrow \frac{1}{1+1} = 0.5$$

Similarly,

$P(Y=1 \mid x^2, w^0) = 0.5$

$P(Y=1 \mid x^3, w^0) = 0.5$

$P(Y=1 \mid x^4, w^0) = 0.5$

$$w_0^{(1)} = w_0^0 + \eta \cdot \sum_j (y^j - \hat{P}(y=1 \mid x^j, w^{(0)}))$$

$$= w_0^{(0)} + \eta \cdot [(1-0.5) + (1-0.5) + (0-0.5) + (0-0.5)]$$

$$= w_0^{(0)} + \eta \cdot (0.5 + 0.5 - 0.5 - 0.5)$$

$$= 0$$

$$w_1^{(1)} = w_1^{(0)} + \eta \cdot \sum_j x_1^j (y^j - \hat{P}(y=1 \mid x^j, w^{(0)}))$$

$$= 0 + \eta \cdot [1 \times (1-0.5) + 0 \times (1-0.5) + 0 \times (0-0.5) - 1 \times (0-0.5)]$$

$$= 0 + \eta \cdot [0.5 + 0 + 0 + 0.5]$$

$$= \eta$$

$$w_2^{(1)} = w_2^0 + \eta \cdot \sum_j x_2^j \left(y^j - \hat{P}(y=1 \mid x^j, w^{(0)}) \right)$$

$$= 0 + \eta \cdot \left[0 \times (1-0.5) - 1(1-0.5) + 1(0-0.5) + 0(0-0.5) \right]$$

$$= \eta \cdot [0 - 0.5 - 0.5 + 0]$$

$$\Rightarrow -\eta$$

$$w^{(1)} = \begin{bmatrix} 0 \\ \eta \\ -\eta \end{bmatrix} \Bigg\} \quad \text{Here,} \quad \begin{array}{l} w_0^1 = w_0^0 \\ w_1^1 > w_1^0 \\ w_2^1 < w_2^0 \end{array}$$

$$\cancel{w_0^{(2)} = w_0^{(1)} + \eta \cdot \sum_j \left(y^j - \hat{P}(y=1 \mid x^j, w^{(1)}) \right)}$$

$$\cancel{= 0 + \eta \cdot [(1-}$$

$$P(y=1 \mid x^1, w^{(1)}) = \frac{\exp(0 + (1 \times \eta + 0 \times (-\eta)))}{1 + \exp(0 + (1 \times \eta + 0 \times (-\eta)))} = \frac{\exp(\eta)}{1 + \exp(\eta)}$$

$$P(y=1 \mid x^2, w^{(1)}) = \frac{\exp(0 + 0 \times (\eta) - 1(-\eta))}{1 + \exp(0 + 0 \times (\eta) - 1(-\eta))} = \frac{\exp(\eta)}{1 + \exp(\eta)}$$

$$P(y=1 \mid x^3, w^{(1)}) = \frac{\exp(0 + 0 \times (\eta) + 1(-\eta))}{1 + \exp(0 + 0 \times (\eta) + 1(-\eta))} = \frac{\exp(-\eta)}{1 + \exp(-\eta)}$$

$$P(y=1 \mid x^4, w^{(1)}) = \frac{\exp(0 + (-1) \times (\eta) + 0(-\eta))}{1 + \exp(0 + (-1) \times \eta + 0 \times (-\eta))} = \frac{\exp(-\eta)}{1 + \exp(-\eta)}$$

$$w_0^{(2)} = w_0^{(1)} + \eta \sum_j \left(y^j - \hat{P}(y=1 \mid x^j, w^{(1)})\right)$$

$$= 0 + \eta \cdot \left[\left(1 - \frac{\exp(\eta)}{1+\exp(\eta)}\right) + \left(1 - \frac{\exp(\eta)}{1+\exp(\eta)}\right) + \left(0 - \frac{\exp(-\eta)}{1+\exp(-\eta)}\right) \right.$$

$$\left. + \left(0 - \frac{\exp(-\eta)}{1+\exp(-\eta)}\right) \right]$$

$$= \eta \left[\frac{2}{1+\exp(\eta)} - \frac{2}{1+\exp(\eta)} \right]$$

$$= \underline{0}$$

$$w_1^{(2)} = w_1^{(1)} + \eta \sum_j x_1^j \left[y^j - \hat{P}(y=1 \mid x^j, w^{(1)}) \right]$$

$$= \eta + \eta \left[1 \times \left(1 - \frac{\exp(\eta)}{1+\exp(\eta)}\right) + 0 + 0 + (-1) \times \left(0 - \frac{\exp(-\eta)}{1+\exp(-\eta)}\right) \right]$$

$$= \eta + \eta \left[\frac{1}{1+\exp(\eta)} + \frac{\exp(-\eta)}{1+\exp(-\eta)} \right]$$

$$= \eta + \eta \left[\frac{1}{1+\exp(\eta)} + \frac{1}{\exp(\eta)+1} \right]$$

$$= \eta + \eta \left[\frac{2}{1+\exp(\eta)} \right]$$

$$= \eta + \frac{2\eta}{1+\exp(\eta)}$$

$$= \frac{3\eta + \eta \exp(\eta)}{1+\exp(\eta)}$$

$$w_2^{(2)} = w_2^{(1)} + \eta \sum_j x_2^j \left[y^j - \hat{P}(y=1 \mid x^j, w^{(1)}) \right]$$

$$= -\eta + \eta \left[0 + (-1) \times \left(1 - \frac{\exp(\eta)}{1+\exp(\eta)} \right) + 1 \left(0 - \frac{\exp(-\eta)}{1+\exp(-\eta)} \right) + 0 \right]$$

$$= -\eta + \eta \left[\frac{-1}{1+\exp(\eta)} + \frac{\exp(-\eta)}{1+\exp(-\eta)} \right]$$

$$= -\eta + \eta \left[\frac{-1}{1+\exp(\eta)} - \frac{1}{1+\exp(+\eta)} \right]$$

$$= -\eta + \eta \left[\frac{-2}{1+\exp(\eta)} \right]$$

$$= -\eta - \frac{2\eta}{1+\exp(\eta)}$$

$$\Rightarrow \frac{-\eta + \eta \exp(\eta) - 2\eta}{1+\exp(\eta)}$$

$$\Rightarrow \frac{-3\eta + \eta \exp(\eta)}{1+\exp(\eta)}$$

$$w^{(2)} = \begin{bmatrix} 0 \\ \eta \left[\frac{3+\exp(\eta)}{1+\exp(\eta)} \right] \\ -\eta \left[\frac{3+\exp(\eta)}{1+\exp(\eta)} \right] \end{bmatrix}$$

Here, we can observe that,

$w_0^2 = w_0^1 = w_0^0$

$w_1^2 > w_1^1 > w_1^0$

$w_2^2 < w_2^1 < w_2^0$

$$P(Y=1 \mid x^1, w^{(2)}) = \frac{\exp\left(\eta\left[\frac{3+\exp(\eta)}{1+\exp(\eta)}\right]\right)}{1+\exp\left(\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right)}$$

$$P(Y=1 \mid x^2, w^{(2)}) = \frac{\exp\left(\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right)}{1+\exp\left(\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right)}$$

$$P(Y=1 \mid x^3, w^{(2)}) = \frac{\exp\left(-\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right)}{1+\exp\left(-\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right)}$$

$$P(Y=1 \mid x^4, w^{(2)}) = \frac{\exp\left(-\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right)}{1+\exp\left(-\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right)}$$

$$w_0^{(3)} = w_0^{(2)} + \eta \sum_j (y^j - \hat{P}(y=1 \mid x^j, w^{(1)}))$$

$$= 0 + \eta\left[\left(1 - \frac{\exp\left[\frac{(3+\exp(\eta))}{1+\exp(\eta)}\right]}{1+\exp()}\right) + \left(1 - \frac{\exp\left[\frac{\eta(3+\exp(\eta))}{1+\exp(\eta)}\right]}{1+\exp()}\right) + \left(0 + \frac{\exp\left[\frac{\eta(3+\exp(-\eta))}{1+\exp(-\eta)}\right]}{1+\exp()}\right) + \left(0 + \frac{\exp\left[\eta\frac{(3+\exp(-\eta))}{1+\exp(-\eta)}\right]}{1+\exp()}\right)\right]$$

$$= \underline{0}$$

Let's assume

$$C = \eta\left[\frac{3+\exp(\eta)}{1+\exp(\eta)}\right]$$

$$w_1^{(3)} = w_1^{(2)} + \eta \sum_j x_1^j \left[y^j - \hat{P}(y=1 \mid x^j, w^{(2)})\right]$$

$$= C + \eta\left[1\times\left(1-\frac{\exp(C)}{1+\exp(C)}\right) + 0 + 0 - 1\times\left(1-\frac{\exp(-C)}{1+\exp(-C)}\right)\right]$$

$$= C + \eta\left[\frac{1}{1+\exp(C)} + \frac{\exp(-C)}{1+\exp(-C)}\right]$$

$$= C + \eta\left[\frac{1}{1+\exp(C)} + \frac{1}{1+\exp(C)}\right]$$

$$\Rightarrow C + \eta\left[\frac{2}{1+\exp(C)}\right]$$

$$= C + \frac{2\eta}{1+\exp(C)}$$

Replacing value of C

$$w_1^{(3)} = \frac{\eta(3+\exp(\eta))}{1+\exp(\eta)} + \frac{2\eta}{1+\exp\left(\frac{\eta(3+\exp(\eta))}{1+\exp(\eta)}\right)}$$

$$w_{(2)}^{3} = w_2^{(2)} + \eta \sum_j x_2^j \left[y^j - \hat{P}(y=1 \mid x^j, w^{(1)})\right]$$

$$= -C + \eta\left[0 + (-1)\times\left(1-\frac{\exp(+C)}{1+\exp(+C)}\right) + 1\times\left(0-\frac{\exp(-C)}{1+\exp(-C)}\right) + 0\right]$$

$$= -c + \eta\left[\frac{-1}{1+\exp(c)} - \frac{\exp(-c)}{1+\exp(-c)}\right]$$

$$= -c + \eta\left[\frac{-1}{1+\exp(c)} - \frac{1}{1+\exp(c)}\right]$$

$$= -c + \eta\left[\frac{-2}{1+\exp(c)}\right]$$

$$= -c - \frac{2\eta}{1+\exp(c)}$$

Replacing value of c

$$w_2^{(3)} = -\eta\frac{(3+\exp(\eta))}{1+\exp(\eta)} - \frac{2\eta}{1+\exp\left(\frac{\eta(3+\exp(\eta))}{1+\exp(\eta)}\right)}$$

$$w^{(3)} = \begin{bmatrix} 0 \\ \frac{\eta(3+\exp(\eta))}{1+\exp(\eta)} + \frac{2\eta}{1+\exp\left(\frac{\eta(3+\exp(\eta))}{1+\exp(\eta)}\right)} \\ -1\left[\frac{\eta(3+\exp(\eta))}{1+\exp(\eta)} + \frac{2\eta}{1+\exp\left(\frac{\eta(3+\exp(\eta))}{1+\exp(\eta)}\right)}\right] \end{bmatrix}$$

Here, after 3 iterations, we can see that,

w_0 is zero for all iterations.

w_1 is increasing with each iteration and going to $+\infty$

w_2 is decreasing with each iteration and going to $-\infty$

$$w^* = \begin{bmatrix} 0 \\ +\infty \\ -\infty \end{bmatrix}$$

Scanned by CamScanner

Now taking

$$w^0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$P(y=1 \mid x^1, w^0) = \frac{\exp(0+1+0)}{1+\exp(0+1+0)} = \frac{\exp(1)}{1+\exp(1)} = 0.73$$

$$P(y=1 \mid x^2, w^0) = \frac{\exp(0+0+0)}{1+\exp(0)} = \frac{1}{2} = 0.5$$

$$P(y=1 \mid x^3, w^0) = \frac{\exp(0+0+0)}{1+\exp(0)} = 0.5$$

$$P(y=1 \mid x^4, w^0) = \frac{\exp(0-1+0)}{1+\exp(-1)} = \frac{\exp(-1)}{1+\exp(-1)} = 0.27$$

$$w_0^{(1)} = w_0^{(0)} + \eta \cdot \Sigma_j \left(y^j - \hat{P}(y=1 \mid x^j, w^{(0)}) \right)$$

$$= 0 + \eta \cdot [(1-0.73) + (1-0.5) + (0-0.5) + (0-0.27)]$$

$$= \eta [0.27 + 0.5 - 0.5 - 0.27]$$

$$= 0$$

$$w_1^{(1)} = w_1^{(0)} + \eta \cdot \Sigma_j x_1^j \left(y^j - \hat{P}(y=1 \mid x^j, w^{(0)}) \right)$$

$$= 1 + \eta \cdot [1 \times (1-0.73) + 0 + 0 - 1(0-0.27)]$$

$$= 1 + \eta [0.27 + 0.27]$$

$$\Rightarrow 1 + 0.54\eta$$

$$w_2^{(1)} = w_2^{(0)} + \eta \Sigma_j x_2^j \left(y^j - \hat{P}(y=1 \mid x^j, w^{(0)}) \right)$$

$$= 0 + \eta [0 - 1(1-0.5) + 1(0-0.5) + 0]$$

$$= -\eta$$

$$w^1 = \begin{bmatrix} 0 \\ 1+0.54\eta \\ -\eta \end{bmatrix}$$

So, here we can observe that

w_0^1 is same as w_0^0

$w_1^1 > w_1^0$ which is increased by 0.54η

$w_2^1 < w_2^0$

Now,

$$P(y=1 \mid x^1, w^{(1)}) = \frac{\exp(0+1\times(1+0.54\eta)+0)}{1+\exp(1+0.54\eta)} = \frac{\exp(1+0.54\eta)}{1+\exp(1+0.54\eta)}$$

$$P(y=1 \mid x^2, w^{(1)}) = \frac{\exp(0+0+\eta)}{1+\exp(\eta)} = \frac{\exp(\eta)}{1+\exp(\eta)}$$

$$P(y=1 \mid x^3, w^{(1)}) = \frac{\exp(0+0-\eta)}{1+\exp(-\eta)} = \frac{\exp(-\eta)}{1+\exp(-\eta)} \, *$$

$$P(y=1 \mid x^4, w^{(1)}) = \frac{\exp(0+(-1)(1+0.54\eta)+0)}{1+\exp(-1-0.54\eta)} = \frac{\exp(-1-0.54\eta)}{1+\exp(-1-0.54\eta)}$$

$$= \frac{1}{1+\exp(1+0.54\eta)}$$

$$w_0^{(2)} = w_0^{(1)} + \eta \sum_j \left(y^j - \hat{P}(y=1 \mid x^j, w^{(1)})\right)$$

$$= 0 + \eta\left[\left(1 - \frac{\exp(1+0.54\eta)}{1+\exp(1+0.54\eta)}\right) + \left(1 - \frac{\exp(\eta)}{1+\exp(\eta)}\right) + \left(0 - \frac{\exp(-\eta)}{1+\exp(-\eta)}\right)\right.$$

$$\left. + \left(0 - \frac{1}{1+\exp(1+0.54\eta)}\right)\right]$$

$$= \eta\left[\frac{1}{1+\exp(1+0.54\eta)} + \frac{1}{1+\exp(\eta)} - \frac{1}{1+\exp(\eta)} - \frac{1}{1+\exp(1+0.54\eta)}\right]$$

$$= 0$$

$$w_1^{(2)} = w_1^{(1)} + \eta \sum_j x_1^i [y^j - \hat{P}(y=1|x^j, w^{(1)})]$$

$$= 1+0.54\eta + \eta\left[\left(1 - \frac{\exp(1+0.54\eta)}{1+\exp(1+0.54\eta)}\right) - \left(0 - \frac{1}{1+\exp(1+0.54\eta)}\right)\right]$$

$$= 1+0.54\eta + \eta\left[\frac{1}{1+\exp(1+0.54\eta)} + \frac{1}{1+\exp(1+0.54\eta)}\right]$$

$$\Rightarrow 1+0.54\eta + \frac{2\eta}{1+\exp(1+0.54\eta)}$$

$$w_2^{(2)} = w_2^{(2)} + \eta \sum_j x_2^i [y^j - \hat{P}(y=1|x^j; w^{(1)})]$$

$$= -\eta + \eta\left[(-1)\times\left(1 - \frac{\exp(\eta)}{1+\exp(\eta)}\right) + 1\times\left(0 - \frac{1}{1+\exp(\eta)}\right)\right]$$

$$= -\eta + \eta\left[-\frac{1}{1+\exp(\eta)} - \frac{1}{1+\exp(\eta)}\right]$$

$$= -\eta - \frac{2\eta}{1+\exp(\eta)}$$

$$w^2 = \begin{bmatrix} 0 \\ 1+0.54\eta + \frac{2\eta}{1+\exp(1+0.54\eta)} \\ -\eta - \frac{2\eta}{1+\exp(\eta)} \end{bmatrix}$$

Here, we can see that

$w_0^2 = w_0^1 = w_0^0$

$w_1^2 > w_1^1$ since $w_1^2 = w_1^1 +$ (+ve const)

$w_2^2 < w_2^1$ since we are subtracting +ve const from w_2^1

Let's say. (17)

$$C_1 = 1 + 0.54\eta + \frac{2\eta}{1+\exp(1+0.54\eta)}$$

$$c_2 = -\eta - \frac{2\eta}{1+\exp(\eta)}$$

$$P(y=1 \mid x^1, w^{(2)}) = \frac{\exp(c_1)}{1+\exp(c_1)}$$

$$P(y=1 \mid x^2, w^{(2)}) = \frac{\exp(-c_2)}{1+\exp(-c_2)} = \frac{1}{1+\exp(c_2)}$$

$$P(y=1 \mid x^3, w^{(2)}) = \frac{\exp(c_2)}{1+\exp(c_2)}$$

$$P(y=1 \mid x^4, w^{(2)}) = \frac{\exp(-c_1)}{1+\exp(-c_1)} = \frac{1}{1+\exp(c_1)}$$

$$w_0^{(3)} = w_0^{(2)} + \eta \sum_j (y^j - \hat{P}(y=1\ x^j, w^{(1)}))$$

$$= 0 + \eta\left[\left(1 - \frac{\exp(c_1)}{1+\exp(c_1)}\right) + \left(1 - \frac{\cancel{\exp}}{1+\exp(c_2)}\right) + \left(0 - \frac{\exp(c_2)}{1+\exp(c_2)}\right) + \left(0 - \frac{\cancel{\exp}}{1+\exp(c_1)}\right)\right]$$

$$= 0 + \eta\left[\frac{1}{1+\exp(c_1)} + \frac{\exp(c_2)}{1+\exp(c_2)} - \frac{\exp(c_2)}{1+\exp(c_2)} - \frac{1}{1+\exp(c_1)}\right]$$

$$= \underline{0}$$

$$w_1^{(3)} = w_1^{(2)} + \eta \sum_j x_1^j \left[y^j - \hat{P}(y=1 \mid x^j, w^{(2)}) \right]$$

$$= c_1 + \eta \left[1 \times \left(1 - \frac{\exp(c_1)}{1+\exp(c_1)}\right) - 1 \times \left(1 - \frac{\cancel{\exp}\, 1}{1+\exp(c_1)}\right) \right]$$

$$= c_1 + \eta \left[\frac{1}{1+\exp(c_1)} + \frac{1}{1+\exp(c_1)} \right]$$

$$= c_1 + \frac{2\eta}{1+\exp(c_1)}$$

$$w_2^{(3)} = w_2^{(2)} + \eta \sum_j x_1^j \left[y^j - \hat{P}(y=1 \mid x^j, w^{(2)}) \right]$$

$$= c_2 + \eta \left[(-1) \times \left(1 - \frac{\cancel{\exp(c_2)}\, 1}{1+\exp(c_2)}\right) + \left(0 - \frac{\exp(c_2)}{1+\exp(c_2)}\right) \right]$$

$$= c_2 + \eta \left[-\frac{\exp(c_2)}{1+\exp(c_2)} - \frac{\exp(c_2)}{1+\exp(c_2)} \right]$$

$$= c_2 - \left[\frac{2\exp(c_2)}{1+\exp(c_2)} \right] \eta$$

$$= c_2 - \frac{2\eta \exp(c_2)}{1+\exp(c_2)}$$

$$W^3 = \begin{bmatrix} 0 \\ c_1 + \dfrac{2\eta}{1+\exp(c_1)} \\ c_2 - \dfrac{2\eta \exp(c_2)}{1+\exp(c_2)} \end{bmatrix}$$

Here we can see that,

$w_0^3 = w_0^2 = w_0^1$

$w_1^3 > w_1^2 > w_1^1$ Since at each iteration a positive term is being added to previous weights and going to $+\infty$,

$w_2^3 < w_2^2 < w_2^1$ Since at each iteration a positive term is being subtracted from previous weights, and going to $-\infty$,

$$w^* = \begin{bmatrix} 0 \\ +\infty \\ -\infty \end{bmatrix}$$

So for two different w^0, we can see that w^* is same i.e. after enough/large number of iterations in both w_2 & w_3 will converge to $+\infty$ & $-\infty$ respectively

5 i)

$f: X \rightarrow Y$

$X = \langle X_1 \dots X_d \rangle \in \mathbb{R}^d$

$Y = \{0, 1\}$

after Training ↓

$W = \langle W_0, W_1 \dots W_d \rangle$

we are given that,

$$P(X_i \mid Y = k) = \mathcal{N}(\mu_{ik}, \sigma_{ik}) \quad (K = 0, 1,\ i = 1 \dots d)$$

No. of Independent parameters will be :

$$\left.\begin{array}{l} \sigma_i \ \& \ \mu_i \ \text{for } Y=0 \\ \sigma_i \ \& \ \mu_i \ \text{for } Y=1 \end{array}\right\} \text{4 parameter for single feature}$$

for d features, it will be 4d & 1 parameter for $P(Y=1)$ as $P(Y=0)$ can be calculated using $1 - P(Y=1)$

So, total,

$$\boxed{4d + 1}$$

ii) Translation of Logistic weights into equivalent GNB parameters :-

we have,

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

Now, dividing numerator & denominator by $P(Y=1)\,P(X|Y=1)$

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

$$= \frac{1}{1 + \exp\left[\ln\left(\frac{P(Y=0)}{P(Y=1)} \times \frac{P(X|Y=0)}{P(X|Y=1)}\right)\right]}$$

$$= \frac{1}{1 + \exp\left[\left(\ln\frac{1-\theta}{\theta}\right) + \underbrace{\sum_i \ln \frac{P(X_i|Y=0)}{P(X_i|Y=1)}}\right]} \quad \text{——} \quad (1)$$

Now solving the 2nd part

we, know

$$P(X_i|Y=k) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp^{-\frac{(x_i-\mu_{ik})^2}{2\sigma_i^2}}$$

$$= \sum_i \ln\left[\frac{\exp\left(-\frac{(x_i-\mu_{i0})^2}{2\sigma_i^2}\right)}{\exp\left(-\frac{(x_i-\mu_{i1})^2}{2\sigma_i^2}\right)}\right]$$

$$= \sum_i \ln\left[\exp\left(-\frac{(x_i^2 + \mu_{i0}^2 - 2x_i\mu_{i0}) + x_i^2 + \mu_{i1}^2 - 2x_i\mu_{i1}}{2\sigma_i^2}\right)\right]$$

$$= \sum_i \frac{(\mu_{i1}^2 - \mu_{i0}^2) + 2x_i(\mu_{i0} - \mu_{i1})}{2\sigma_i^2}$$

$$= \sum_i \frac{(\mu_{i0} - \mu_{i1})}{\sigma_i^2} x_i + \left(\frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2} \right)$$

Subtituting the above value in Eq ①

$$P(Y=1|X) = \frac{1}{1 + \exp\left[\underbrace{\ln \frac{(1-\theta)}{\theta} + \sum_i \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2}}_{w_0} + \underbrace{\sum_i \frac{(\mu_{i0} - \mu_{i1}) x_i}{\sigma_i^2}}_{w_i} \right]}$$

So from above

$$w_0 = \ln\left(\frac{1-\theta}{\theta}\right) + \sum_i \left(\frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2} \right)$$

$$w_i = \left(\frac{\mu_{i0} - \mu_{i1}}{\sigma_i^2} \right)$$

5 ii) Implementation of Gaussian Naive Bayes and Logistic Regression:

Gaussian Naive Bayes - Pseudo code →

Step-1: Data processing

a) Pre-process the data & divide it into two sets i.e. Training set and Test set based on the five given fractions.

b) Seperate it by labels i.e 0 & 1 in our case

Step-2: Build the Model.

a) To build the model, first we need the Baye's theorem:

posterior = prior * likelihood / marginal (evidence)

b) Since Naive bayes assumes that all the features are independent, it is equivalent to calculate the posterior probability:

$$P(Y=1 \mid x_1 \ldots x_n) = \frac{P(Y=1) * P(x_1 \ldots x_n \mid Y=1)}{P(x_1 \ldots x_n)}$$

Since $x_1, x_2 \ldots x_n$ are independent

$$= \frac{P(Y=1) * P(x_1 \mid Y=1) * \cdots P(x_n \mid Y=1)}{P(x_1 \ldots x_n)}$$

Since $P(x \ldots x_n)$ is constant given the input

$$P(Y \neq 1 \mid x_1 \ldots x_n) \propto P(Y=1) * P(x_1 \mid Y=1) \ldots P(x_n \mid Y=1)$$

c) to calculate the above eqn, we used

Prior = mean(y)

for likelihood, we used Gaussian Model:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi}\,\sigma_y^2} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

where,

$\mu_y \rightarrow$ mean of examples in $y=1$

$\sigma_y^2 \rightarrow$ standard deviation of examples in $y=1$

$x_i \rightarrow$ new test example to predict

Step 3 : Predict

a) We use the test set data and calculate the prior & likelihood for each row and get the posterior for that row by multiplying the prior & likelihood.

b) to get accuracy, we check our predicted label with the actual label.

c) we reapeat the above steps 5 times for each fraction and calculate the average accuracy for each fraction.

Logistic Regression Pseudo code →

Step 1: Process the data

this step is exactly similar to the Naive Bayes classifier's step and I have used same function for both of them

Step 2: Build Model

a) Initialize the coefficients for each attribute/feature randomly.

b) we will use sigmoid funcn as our logistic funcn for our Logistic regression Model i.e

$$\frac{1}{1+e^{-x}}$$

c) Estimate/update the weights/coefficients using stochastic Gradient Descent.

i) Calculate the prediction using current coefficient

ii) Update the coefficients based on error in prediction i.e.

new-coeff
b = b + alpha * (y - prediction) * prediction * (1 - prediction) * x

Learning rate

OR

$$w_i^{(t+1)} \leftarrow w^{(t)} + \eta \frac{\partial (w)}{\partial (w_i)}$$

Learning Rate, $\eta > 0$

d) Repeat the process

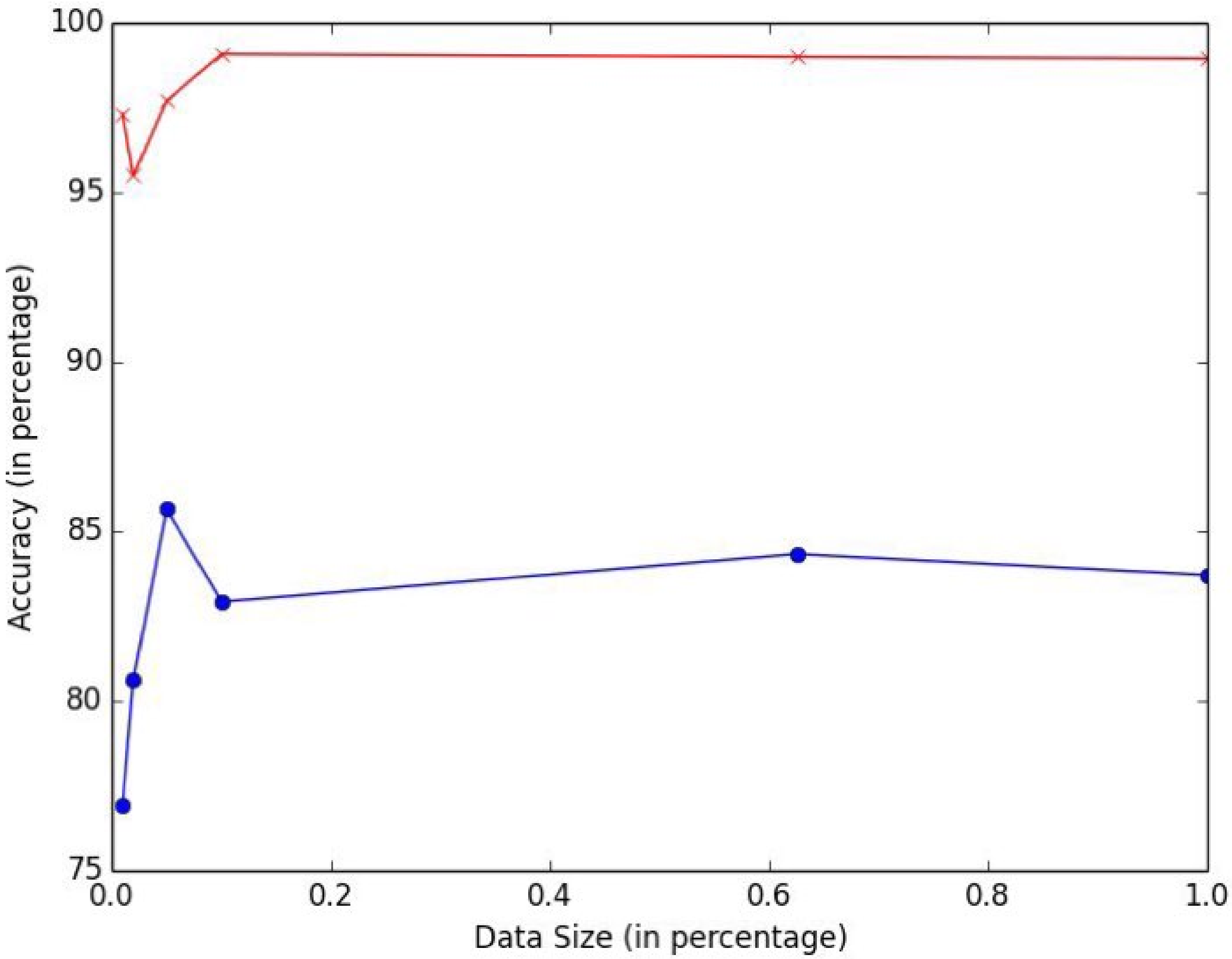
Step 3 : Predict

a) we will use the learned weights to predict the label for test input.

b) Accuracy will be calculated ~~with~~ by comparing the predicted and actual label for some particular row.

c) we repeat the above steps 5 times for each fraction and calculate the average accuracy.

Plot the Learning Curve :-


100
95
90
85
80
75
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy (in percentage)
Data Size (in percentage)

```
Combination: 1


Attribute: 1
Original Training Set's Mean: -1.96649344171
Generated Samples' Mean: -1.86364300074
Difference: 0.102850440973
-----------------
Original Training Set's SD: 1.85284431207
Generated Samples' SD: 1.86267559166
Difference: 0.00983127959427
-----------------
[Attribute: 2
Original Training Set's Mean: -0.567499
Generated Samples' Mean: -0.618954999089
Difference: 0.0514559990894
-----------------
Original Training Set's SD: 5.28183832332
Generated Samples' SD: 5.28684343678
Difference: 0.00500511345608
-----------------
Attribute: 3
Original Training Set's Mean: 1.84226256784
Generated Samples' Mean: 2.03182337991
Difference: 0.189560812072
-----------------
Original Training Set's SD: 5.13805582753
Generated Samples' SD: 4.9860273029
Difference: 0.152028524635
-----------------
Attribute: 4
Original Training Set's Mean: -1.40414835477
Generated Samples' Mean: -1.44380040879
Difference: 0.0396520540194
-----------------
Original Training Set's SD: 2.06651039315
Generated Samples' SD: 2.01938194917
Difference: 0.0471284439827
-----------------
```

```
Combination: 2


Attribute: 1
Original Training Set's Mean: -1.78282045885
Generated Samples' Mean: -1.80718346018
Difference: 0.0243630013244
-----------------
Original Training Set's SD: 1.93454675614
Generated Samples' SD: 1.93281055487
Difference: 0.00173620126134
-----------------
Attribute: 2
Original Training Set's Mean: -0.793434406699
Generated Samples' Mean: -0.742865385263
Difference: 0.0505690214351
-----------------
Original Training Set's SD: 5.35724008266
Generated Samples' SD: 5.68345896539
Difference: 0.326218882732
-----------------
Attribute: 3
Original Training Set's Mean: 1.88618747368
Generated Samples' Mean: 1.45692106266
Difference: 0.429266411028
-----------------
Original Training Set's SD: 5.32950427606
Generated Samples' SD: 5.39534613492
Difference: 0.0658418588561
-----------------
Attribute: 4
Original Training Set's Mean: -1.33667266507
Generated Samples' Mean: -1.3975405891
Difference: 0.0608679240267
-----------------
Original Training Set's SD: 2.11839826059
Generated Samples' SD: 2.14878472877
Difference: 0.0303864681794
-----------------
```

```
Combination: 3


Attribute: 1
Original Training Set's Mean: -1.78282045885
Generated Samples' Mean: -1.73253453385
Difference: 0.0502859250037
-----------------
Original Training Set's SD: 1.93454675614
Generated Samples' SD: 2.02060426588
Difference: 0.0860575097443
-----------------
Attribute: 2
Original Training Set's Mean: -0.793434406699
Generated Samples' Mean: -0.767437365155
Difference: 0.0259970415434
-----------------
Original Training Set's SD: 5.35724008266
Generated Samples' SD: 5.10412861743
Difference: 0.253111465228
-----------------
Attribute: 3
Original Training Set's Mean: 1.88618747368
Generated Samples' Mean: 1.61737288099
Difference: 0.26881459269
-----------------
Original Training Set's SD: 5.32950427606
Generated Samples' SD: 5.69795521831
Difference: 0.368450942248
-----------------
Attribute: 4
Original Training Set's Mean: -1.33667266507
Generated Samples' Mean: -1.33252189683
Difference: 0.00415076824288
-----------------
Original Training Set's SD: 2.11839826059
Generated Samples' SD: 2.17718388418
Difference: 0.0587856235889
-----------------
```

Observation → from above we can see that the differnce b/w the original mean and standard deviation with the mean of standand deviation of 400 generated examples is almost same/difference is close to zero, because the generated examples belong to the same distribution, hence the differnce b/w them is negligible.