JAGDEESH BASAVARAJU
ASU ID: 1213004713

Statistical Machine Learning Assignment #1

1. Given Prior, $P(\theta)$ is $Beta(\beta_H, \beta_T)$, Prove that posterior ~~also~~ follows $Beta(\beta_H + \alpha_H, \beta_T + \alpha_T)$.

Sol^n: Prior, $P(\theta) = Beta(\beta_H, \beta_T)$

$$P(\theta) = \frac{\theta^{\beta_H - 1}(1-\theta)^{\beta_T - 1}}{B(\beta_H, \beta_T)} \quad \text{--- (1)}$$

Posterior $P(\theta/D) \propto P(D/\theta)\, P(\theta)$ --- (2)

Consider $P(D/\theta)\, P(\theta)$

$$P(D/\theta)\,P(\theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T} \cdot \frac{\theta^{\beta_H - 1}(1-\theta)^{\beta_T - 1}}{B(\beta_H, \beta_T)} \quad [\because \text{From (1)}]$$

$$P(D/\theta)\,P(\theta) = \frac{1}{B(\beta_H, \beta_T)}\, \theta^{\alpha_H + \beta_H - 1}(1-\theta)^{\alpha_T + \beta_T - 1} \quad \text{--- (3)}$$

From (1) & (2)

$$P(\theta/D) = \frac{1}{C} P(D/\theta)\, P(\theta)$$

$$P(\theta/D) \propto \theta^{\alpha_H + \beta_H - 1}(1-\theta)^{\alpha_T + \beta_T - 1}$$

$$\therefore P(\theta/D) \sim Beta(\alpha_H + \beta_H, \alpha_T + \beta_T)$$

Mean of posterior = mean of $Beta(\alpha_H + \beta_H, \alpha_T + \beta_T)$

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

Mode of posterior (Map) $= \hat{\theta}_{MAP} = \underset{\theta}{\arg\max}\, P(\theta/D)$ = mode of beta

$$= \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

2) Parameter Estimation

i) Prove that $\hat{\mu}_{MLE}$ is unbiased.

$D = x_1, x_2, \dots, x_N \in R$ and is $N(\mu, \sigma^2)$

$$\therefore P(D/\mu,\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \prod_{i=1}^{N} \exp\left\{\frac{-(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$\ln P(D/\mu,\sigma) = -N\ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}$$

To find $\hat{\mu}_{MLE}$, $\frac{d}{d\mu}\ln(P(D/\mu,\sigma)) = 0$

$$\frac{d}{d\mu}\left[\ln P(D/\mu,\sigma)\right] = \frac{d}{d\mu}\left[-N\ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$= \frac{-1}{2\sigma^2}\sum_{i=1}^{N} 2(x_i-\mu).(-1) = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu) = 0$$

$$\therefore \sum_{i=1}^{N} x_i - \sum_{i=1}^{N}\mu = 0$$

$$\sum_{i=1}^{N} x_i = N\mu$$

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

Expected value of $\hat{\mu}_{MLE} = E(\hat{\mu}_{MLE}) = \frac{1}{N}\sum_{i=1}^{N}E(x_i)$

$$= \frac{1}{N}\sum_{i=1}^{N}\mu = \frac{1}{N}.N\mu$$

$$= \mu$$

Bias $= E(\hat{\mu}_{MLE}) - \mu = \mu - \mu = 0$

$\therefore \hat{\mu}_{MLE}$ is unbiased.

2 ii) Prove that $\hat{\sigma}^2_{MLE}$ is biased.

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

~~Estimate~~

Expected value of $\hat{\sigma}^2_{MLE} = E(\hat{\sigma}^2_{MLE})$

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

$E(x_i - \hat{\mu}_{MLE})^2$ can't be equal to σ^2 as $\hat{\mu}_{MLE}$ is just an estimate and we don't know true value.

$$\therefore \text{Bias} = E(\hat{\sigma}^2_{MLE}) - \sigma^2$$

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

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

~~Bias $\neq 0$~~

~~Hence $\hat{\sigma}^2_{MLE}$ is biased.~~

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

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

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

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

Bias $\neq 0$

Hence $\hat{\sigma}^2_{MLE}$ is biased

Scanned by CamScanner

3) Naive Bayes Classifier

i) 13 independent parameters

Consider age as X_1, income as X_2, student as X_3, credit_rating as X_4 and buys_computer as Y.

1) $P(Y = yes)$
2) $P(X_1 = youth / Y = yes)$
3) $P(X_1 = middle\_aged / Y = yes)$
4) $P(X_2 = high / Y = yes)$
5) $P(X_2 = medium / Y = yes)$
6) $P(X_3 = yes / Y = yes)$
7) $P(X_4 = fair / Y = yes$
8) $P(X_1 = youth / Y = no)$
9) $P(X_1 = middle\_aged / Y = no)$
10) $P(X_2 = high / Y = no)$
11) $P(X_2 = medium / Y = no)$
12) $P(X_3 = yes / Y = no)$
13) $P(X_4 = fair / Y = no)$

Since the class is binary, we need one parameter (1)

Age has 3 distinct values, so need 2 parameters ~~when~~ per class (4)

Income also has 3 distinct values, so need 2 per class (4)

Student has 2 distinct values, so 1 per class (2)

Credit_rating ~~also~~ has 2 distinct values, so 1 per class (2)

Hence totally 13 independent parameters

3 ii) Estimated value of all these parameters

Age : X_1 Income : X_2 Student : X_3 Credit_rating : X_4

Buys_Computer : Y

1) $P(Y = yes) = 9/14$ $\therefore P(Y = no) = 5/14$

2) $P(X_1 = youth \mid Y = yes) = 2/9$

3) $P(X_1 = middle\_aged \mid Y = yes) = 4/9$ $\therefore P(X_1 = senior \mid Y = yes) = 3/9$

4) $P(X_2 = high \mid Y = yes) = 2/9$

5) $P(X_2 = medium \mid Y = yes) = 4/9$ $\therefore P(X_2 = low \mid Y = yes) = 3/9$

6) $P(X_3 = yes \mid Y = yes) = 6/9$ $\therefore P(X_3 = no \mid Y = yes) = 3/9$

7) $P(X_4 = fair \mid Y = yes) = 6/9$ $\therefore P(X_4 = excellent \mid Y = yes) = 3/9$

8) $P(X_1 = youth \mid Y = \cancel{yes}\ no) = 3/5$

9) $P(X_1 = middle\_aged \mid Y = no) = 0/5 = 0$ $\therefore P(X_1 = senior \mid Y = no) = \frac{2}{5}$

10) $P(X_2 = high \mid Y = no) = 2/5$

11) $P(X_2 = medium \mid Y = no) = 2/5$ $\therefore P(X_2 = low \mid Y = no) = 1/5$

12) $P(X_3 = yes \mid Y = no) = 1/5$ $\therefore P(X_3 = no \mid Y = no) = 4/5$

13) $P(X_4 = fair \mid Y = no) = 2/5$ $\therefore P(X_4 = excellent \mid Y = no) = 3/5$

3 iii) $X = (\text{youth}, \text{medium}, \text{yes}, \text{fair})$, find $P(y = \text{yes} \mid x)$

$$P(Y=1 / X_1 \ldots X_n) = \frac{P(Y=1)\, P(X_1 \ldots X_n / Y=1)}{P(Y=1)\, P(X_1 \ldots X_n / Y=1) + P(Y=0)\, P(X_1 \ldots X_n / Y=0)}$$

$$= \frac{P(Y=1)\, P(X_1/Y=1) \ldots P(X_n/Y=1)}{P(Y=1)\, P(X_1/Y=1) \ldots P(X_n/Y=1) + P(Y=0)\, P(X_1/Y=0) \ldots P(X_n/Y=0)}$$

$$\therefore P(Y=\text{yes}/X) = \frac{P(Y=\text{yes})\, P(X_1=\text{youth}/Y=\text{yes})\, P(X_2=\text{medium}/Y=\text{yes})\, P(X_3=\text{yes}/Y=\text{yes}) \cdot P(X_4=\text{fair}/Y=\text{yes})}{\left[P(Y=\text{yes})\, P(X_1=\text{youth}/Y=\text{yes})\, P(X_2=\text{medium}/Y=\text{yes})\, P(X_3=\text{yes}/Y=\text{yes})\, P(X_4=\text{fair}/Y=\text{yes}) + P(Y=\text{no})\, P(X_1=\text{youth}/Y=\text{no})\, P(X_2=\text{medium}/Y=\text{no})\, P(X_3=\text{yes}/Y=\text{no})\, P(X_4=\text{fair}/Y=\text{no})\right]}$$

$$\therefore P(Y=\text{yes}/X) = \frac{\left(\frac{9}{14}\right)\left(\frac{2}{9}\right)\left(\frac{4}{9}\right)\left(\frac{6}{9}\right)\left(\frac{6}{9}\right)}{\left[\left(\frac{9}{14}\right)\left(\frac{2}{9}\right)\left(\frac{4}{9}\right)\left(\frac{6}{9}\right)\left(\frac{6}{9}\right) + \left(\frac{5}{14}\right)\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)\left(\frac{1}{5}\right)\left(\frac{2}{5}\right)\right]}$$

$$= \frac{0.0282}{[0.0282 + 0.006857]} = \frac{0.0282}{0.035057}$$

$P(Y=\text{yes}/X) = 0.8044$

Since its greater than 0.5, the classifier has predicted $y = \text{yes}$ for this person.

i.e buys_computer = yes

4. Logistic Regression

$X_1 = (1, 0)$ $X_2 = (0, -1)$ $X_3 = (0, 1)$ $X_4 = (-1, 0)$

$y_1 = 1$ $y_2 = 1$ $y_3 = 0$ $y_4 = 0$

i) Initial weight vector, $w^{(0)} = (0, 0, 0)'$

<u>Iteration 1:</u> $\hat{P}(Y^j = 1 / X^j, w^{(t)}) = \dfrac{\exp(w_0^{(t)} + \sum_{i=1}^{2} w_i^{(t)} x_i)}{1 + \exp(w_0^{(t)} + \sum_{i=1}^{2} w_i^{(t)} x_i)}$

$j = 1, 2, 3, 4$

$$\hat{P}(Y^1 = 1 / X^1, w^{(0)}) = \frac{\exp(0 + 0\times1 + 0\times0)}{1 + \exp(0 + 0\times1 + 0\times0)} = \frac{1}{2} = 0.5$$

$$\hat{P}(Y^2 = 1 | X^2, w^{(0)}) = \frac{\exp(0 + 0\times0 + 0\times(-1))}{1 + \exp(0 + 0\times0 + 0\times(-1))} = \frac{1}{2} = 0.5$$

$$\hat{P}(Y^3 = 1 / X^3, w^{(0)}) = \frac{\exp(0 + 0\times0 + 0\times1)}{1 + \exp(0 + 0\times0 + 0\times1)} = \frac{1}{2} = 0.5$$

$$\hat{P}(Y^4 = 1 | X^4, w^{(0)}) = \frac{\exp(0 + 0\times(-1) + 0\times0)}{1 + \exp(0 + 0\times(-1) + 0\times0)} = \frac{1}{2} = 0.5$$

$$w_0^{(t+1)} = w_0^{(t)} + \eta \sum_j [y^j - \hat{P}(Y^j = 1 | X^j, w^{(t)})]$$

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

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

$$w_i^{(t+1)} = w_i^{(t)} + \eta \sum_j x_i^j [y^j - \hat{P}(Y^j = 1 / X^j, w^{(t)})]$$

$i = 1, 2$

$$\therefore w_1^{(1)} = w_1^{(0)} + \eta[1(1-0.5) + 0(1-0.5) + 0(0-0.5) + (-1)(0-0.5)]$$

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

$$\therefore w_2^{(1)} = w_2^{(0)} + \eta[0(1-0.5) + (-1)(1-0.5) + 1(0-0.5) + 0(0-0.5)]$$

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

$\therefore w^{(1)} = (0, \eta, -\eta)'$

<u>Iteration 2 :-</u>

$$\hat{P}(Y^1=1/X^1, w^{(1)}) = \frac{\exp(0+\eta(1)+(-\eta)(0))}{1+\exp(0+\eta(1)+(-\eta)(0))} = \frac{\exp(\eta)}{1+\exp(\eta)}$$

$$\hat{P}(Y^2=1/X^2, w^{(1)}) = \frac{\exp(0+\eta(0)+(-\eta)(-1))}{1+\exp(0+\eta(0)+(-\eta)(-1))} = \frac{\exp(\eta)}{1+\exp(\eta)}$$

$$\hat{P}(Y^3=1/X^3, w^{(1)}) = \frac{\exp(0+\eta(0)+(-\eta)(1))}{1+\exp(0+\eta(0)+(-\eta)(1))} = \frac{\exp(-\eta)}{1+\exp(-\eta)} = \frac{1}{1+\exp(\eta)}$$

$$\hat{P}(Y^4=1/X^4, w^{(1)}) = \frac{\exp(0+\eta(-1)+(-\eta)(0))}{1+\exp(0+\eta(-1)+(-\eta)(0))} = \frac{\exp(-\eta)}{1+\exp(-\eta)} = \frac{1}{1+\exp(\eta)}$$

$$w_0^{(2)} = w_0^{(1)} + \eta\left[\left(1-\frac{\exp(\eta)}{1+\exp(\eta)}\right)+\left(1-\frac{\exp(\eta)}{1+\exp(\eta)}\right)+\left(0-\frac{1}{1+\exp(\eta)}\right)+\left(0-\frac{1}{1+\exp(\eta)}\right)\right]$$

$$= 0+\eta\left[\frac{2}{1+\exp(\eta)}-\frac{2}{1+\exp(\eta)}\right] = 0$$

$$w_1^{(2)} = w_1^{(1)} + \eta\left[1\left(\frac{1}{1+\exp(\eta)}\right)+0\left(\frac{1}{1+\exp(\eta)}\right)+0\left(\frac{-1}{1+\exp(\eta)}\right)+(-1)\left(\frac{-1}{1+\exp(\eta)}\right)\right]$$

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

$$w_2^{(2)} = w_2^{(1)} + \eta\left[0\left(\frac{1}{1+\exp(\eta)}\right)+(-1)\left(\frac{1}{1+\exp(\eta)}\right)+1\left(\frac{-1}{1+\exp(\eta)}\right)+0\left(\frac{-1}{1+\exp(\eta)}\right)\right]$$

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

$$\therefore w^{(2)} = \left(0, \eta+\frac{2\eta}{1+\exp(\eta)}, -\eta-\frac{2\eta}{1+\exp(\eta)}\right)'$$

As we continue the number of iterations like this, we get a final w, i.e $w^{(*)} = (0, \infty, -\infty)'$

Hence $w^{(*)} = (0, \infty, -\infty)'$ when $w^{(0)} = (0,0,0)'$

4 ii) Initial weight vector, $w^{(0)} = (0,1,0)'$

Iteration 1:

$$\hat{P}(Y^1=1 / X^1, w^{(0)}) = \frac{\exp(0+1(1)+0(0))}{1+\exp(0+1(1)+0(0))} = \frac{\exp(1)}{1+\exp(1)} = \frac{e}{1+e}$$

$$\hat{P}(Y^2=1 / X^2, w^{(0)}) = \frac{\exp(0+1(0)+0(-1))}{1+\exp(0+1(0)+0(-1))} = \frac{1}{2} = 0.5$$

$$\hat{P}(Y^3=1 / X^3, w^{(0)}) = \frac{\exp(0+1(0)+0(1))}{1+\exp(0+1(0)+0(-1))} = \frac{1}{2} = 0.5$$

$$\hat{P}(Y^4=1 / X^4, w^{(0)}) = \frac{\exp(0+1(-1)+0(0))}{1+\exp(0+1(-1)+0(0))} = \frac{e^{-1}}{1+e^{-1}} = \frac{1}{1+e}$$

$$\therefore w_0^{(1)} = w_0^{(0)} + \eta\left[\left(1-\frac{e}{1+e}\right)+(1-0.5)+(0-0.5)+\left(0-\frac{1}{1+e}\right)\right]$$

$$= 0+\eta\left[\left(\frac{1}{1+e}\right)+(0.5)+(-0.5)+\left(\frac{-1}{1+e}\right)\right] = 0+\eta(0) = 0$$

$$\therefore w_1^{(1)} = w_1^{(0)} + \eta\left[1\left(\frac{1}{1+e}\right)+0(0.5)+0(-0.5)+(-1)\left(\frac{-1}{1+e}\right)\right]$$

$$= 1+\eta\left[\frac{2}{1+e}\right] = 1+\frac{2\eta}{1+e}$$

$$\therefore w_2^{(1)} = w_2^{(0)} + \eta\left[0\left(\frac{1}{1+e}\right)+(-1)(0.5)+1(-0.5)+0\left(\frac{-1}{1+e}\right)\right]$$

$$= 0+\eta\left[-0.5-0.5\right] = -\eta$$

$$\therefore w^{(1)} = \left(0,\ 1+\frac{2\eta}{1+e},\ -\eta\right)'$$

Iteration 2:

$$\hat{P}(Y^1=1 / X^1, w^{(1)}) = \frac{\exp\left(0+\left(1+\frac{2\eta}{1+e}\right)(1)+(-\eta)(0)\right)}{1+\exp\left(0+\left(1+\frac{2\eta}{1+e}\right)(1)+(-\eta)(0)\right)}$$

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

$$\hat{P}(Y^2=1 / X^2_1, w^{(1)}) = \frac{\exp\left(0 + \left(1+\frac{2\eta}{1+e}\right)(0) + (-\eta)(-1)\right)}{1+\exp\left(0+\left(1+\frac{2\eta}{1+e}\right)(0) + (-\eta)(-1)\right)}$$

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

$$\hat{P}(Y^3=1 / X^3_1, w^{(1)}) = \frac{\exp\left(0 + \left(1+\frac{2\eta}{1+e}\right)(0) + (-\eta)(1)\right)}{1+\exp\left(0+\left(1+\frac{2\eta}{1+e}\right)(0) + (-\eta)(1)\right)}$$

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

$$\hat{p}(Y^4=1 | X^4_1, w^{(1)}) = \frac{\exp\left(0 + \left(1+\frac{2\eta}{1+e}\right)(-1) + (-\eta)(0)\right)}{1+\exp\left(0+\left(1+\frac{2\eta}{1+e}\right)(-1) + (-\eta)(0)\right)}$$

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

$$\therefore w_0^{(2)} = w_0^{(1)} + \eta\left[\left(1 - \frac{\exp\left(1+\frac{2\eta}{1+e}\right)}{1+\exp\left(1+\frac{2\eta}{1+e}\right)}\right) + \left(1 - \frac{\exp(\eta)}{1+\exp(\eta)}\right)\right.$$

$$\left. + \left(0 - \frac{1}{1+\exp(\eta)}\right) + \left(0 - \frac{1}{1+\exp\left(1+\frac{2\eta}{1+e}\right)}\right)\right]$$

$$= 0 + \eta\left[\frac{1}{1+\exp\left(1+\frac{2\eta}{1+e}\right)} + \frac{1}{1+\exp(\eta)} - \frac{1}{1+\exp(\eta)} - \frac{1}{1+\exp\left(1+\frac{2\eta}{1+e}\right)}\right]$$

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

Scanned by CamScanner

$$\therefore\ w_1^{(2)} = w_1^{(1)} + \eta\left[1\cdot\left(\frac{1}{1+\exp\left(1+\frac{2\eta}{1+e}\right)}\right) + 0(\ ) + 0(\ ) + (-1)\left(\frac{-1}{1+\exp\left(1+\frac{2\eta}{1+e}\right)}\right)\right]$$

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

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

$$w_2^{(2)} = w_2^{(1)} + \eta\left[0(\ ) + (-1)\left(\frac{1}{1+\exp(\eta)}\right) + 1\left(\frac{-1}{1+\exp(\eta)}\right) + 0(\ )\right]$$

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

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

$$\therefore\ w^{(2)} = \left(0,\ 1+\frac{2\eta}{1+e} + \frac{2\eta}{1+\exp\left(1+\frac{2\eta}{1+e}\right)},\ -\left[\eta+\frac{2\eta}{1+\exp(\eta)}\right]\right)^{T}$$

As we continue the number of iterations, the pattern continues and we get a final w.

i.e $w^{(*)} = (0, \infty, -\infty)'$

Hence the final vector, $w^{(*)}$ will be the same for the two different initial values.

5) Naive Bayes Classifier and Logistic Regression

i) Guassian Naive Bayes & Logistic Regression

- Number of independent parameters in GNB?

Sol^n:- $4d+1$, d being the number of features

1 of $P(Y=1)$

$$d\begin{cases}\mu_{1,0}\\ \mu_{2,0}\\ \mu_{3,0}\\ \vdots\\ \mu_{d,0}\end{cases} \quad d\begin{cases}\mu_{1,1}\\ \mu_{2,1}\\ \mu_{3,1}\\ \vdots\\ \mu_{d,1}\end{cases} \quad d\begin{cases}\sigma_{1,0}\\ \sigma_{2,0}\\ \sigma_{3,0}\\ \vdots\\ \sigma_{d,0}\end{cases} \quad d\begin{cases}\sigma_{1,1}\\ \sigma_{2,1}\\ \sigma_{3,1}\\ \vdots\\ \sigma_{d,1}\end{cases}$$

Hence $4d+1$ independent parameters

- No, 'w' cannot be translated into the parameters of an equivalent GNB classifier without any extra assumption.

Extra assumption: Variance is the same across all the classes.

i.e $\sigma_{iK} = \sigma_i$ $\forall K$ (no. of classes)

With this assumption, w can be translated.

Explanation:

From Gaussian Naive Bayes,

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

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

Assuming $P(Y=1) = \theta$, $P(Y=0) = 1-\theta$ & $\sigma_{ik} = \sigma_i$,
i.e variance same across classes

we get,

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

This is of the form

$$P(Y=1/X) = \frac{1}{1 + \exp\left(w_0 + \sum_i w_i X_i\right)}$$

which is the form of Logistic Regression. ~~Hence~~ Hence parameters of logistic regression w, can be translated into equivalent Gaussian Naive Bayes Classifier.

Question 5.2: Implementation of Gaussian Naïve Bayes and Logistic Regression

i) Pseudocode

Load X(Samples), y(class) from the dataset

For fraction in fractions (0.01, 0.02, 0.05, 0.1, 0.625, 1)

Divide data into 3 folds making one of the folds test and other 2 folds training set

We will end up with 3 different combinations of train, test data

In each such combination, randomly pick a fraction of train data

Repeat the above step 5 times so that we will have 5 sets of data for each combination

Train the model for each of the set and learn the parameters

Predict the class for the test data from the learnt parameters and calculate the accuracy

Calculate the mean of accuracies which will be the accuracy for the fraction

Plot accuracy vs the fraction size curve

Gaussian Naïve Bayes

Function **Train**(X_train, y_train):

P(y=1) = number of positive samples / total number of samples

Find $Mean_{i,1}$ for training data of class y=1

Find $Mean_{i,0}$ for training data of class y=0

Find $Variance_{i,1}$ for training data of class y=1

Find $Variance_{i,0}$ for training data of class y=0

Learnt_parameters = P(y=1), $Mean_{i,1}$, $Mean_{i,0}$, $Variance_{i,1}$, $Variance_{i,0}$

Return Learnt_parameters

Function **Predict**(X_test, y_test, Learnt_parameters):

P(X/Y=1) = (1/sqrt(2*pi* $Variance_{i,1}$)) exp(-((X_test - $Mean_{i,1}$)^2)/(2* $Variance_{i,1}$))

H_pos = P(Y=1) * $\prod_{I}$ P(X_i/Y=1)

P(X/Y=0) = (1/sqrt(2*pi* $Variance_{i,0}$)) exp(-((X_test - $Mean_{i,0}$)^2)/(2* $Variance_{i,0}$))

H_zero = P(Y=0) * $\prod_{I}$ P(X_i/Y=0)

If H_pos > H-zero then

y_pred = 1.0

else

y_pred = 0.0

Accuracy = (# y_pred == y_test) / (# y_pred)

Return Accuracy

Logistic Regression

Function **train**(X_train, y_train, learning_rate)

W_0=0, W=(0, 0, 0, 0)

For 500 iterations

Z = $w_0 + \sum_i w_i$X_train$_i$

P(Y=1/X_train,w) = exp(z)/(1+exp(z)) = 1/(1+exp(-z)) = sigmoid(-z) = a

w_0 = w_0 + learning_rate * ($\sum_j$ (y_train^j – P(Y=1/X_train^j,w)))

w = w + learning_rate * ($\sum_j X^j$(y_train^j – P(Y=1/X_train^j,w)))

return w, w_0

Function **predict**(X_test, y_test, w, w_0)

Z = $w_0 + \sum_i w_i$X_test$_i$

P(Y=1/X_test,w) = exp(z)/(1+exp(z)) = 1/(1+exp(-z)) = sigmoid(-z) = a

If P(Y=1/X_test,w) > 0.5 then

y_pred = 1.0

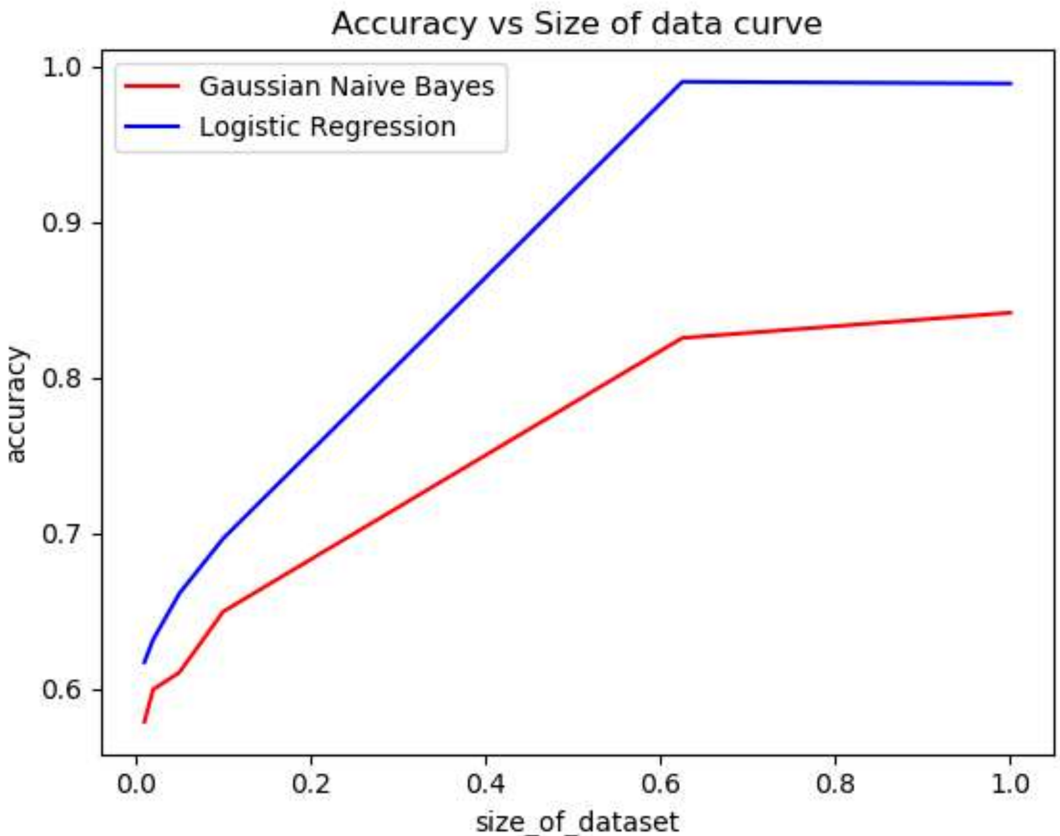
else

y_pred = 0.0

Accuracy = (# y_pred == y_test) / (# y_pred)

Return Accuracy

ii) Learning curve



iii) Generate examples and compare their mean and variance with the training data

Fold 1

Accuracy = 0.849015317287

Mean of training set: [-1.84753379 -0.89508644 2.03368838 -1.26910129]

Mean of generated set: [-1.76017452 -0.99775458 2.19185291 -1.19954756]

Percentage of Mean Deviated: [4.72842612 10.28991948 7.21601956 5.48054968]

Variance of training set: [3.52720065 29.00592331 26.85858549 4.40942191]

Variance of generated set: [3.92013342 28.76514662 26.13434515 4.76906046]

Percentage of Variance Deviated: [10.02345405 0.83009489 2.6964947 7.54107763]

Fold 2

Accuracy = 0.875273522976

Mean of training set: [-1.80954275 -0.92629223 2.0706694 -1.2504281]

Mean of generated set: [-1.811013 -0.9795895 2.1839032 -1.20256079]

Percentage of Mean Deviated: [0.08118422 5.44077653 5.18492754 3.82807361]

Variance of training set: [3.54048344 28.86707855 27.99184211 4.24441466]

Variance of generated set: [3.61466085 30.28615443 27.9742008 4.2960772]

Percentage of Variance Deviated: [2.05212644 4.68555982 0.06302304 1.20255159]

Fold 3

Accuracy = 0.80306345733

Mean of training set: [-1.94775075 -1.16378884 2.34564159 -1.21900678]

Mean of generated set: [-1.99693031 -0.95619555 2.63533007 -1.11657451]

Percentage of Mean Deviated: [2.46275768 17.83771089 10.99249318 8.40292865]

Variance of training set: [3.52197547 29.57897479 28.0784832 4.18090146]

Variance of generated set: [3.44216542 26.90525506 29.04871296 4.52075785]

Percentage of Variance Deviated: [2.26605923 9.0392576 3.34000947 7.5176862]