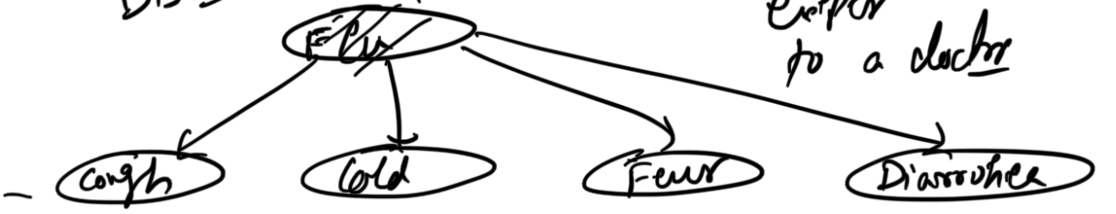
COL774
Machine Learning
Sep 8, 2021

Last Class:- GDA
(Generative) Discriminative
$\hookrightarrow \Rightarrow \{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

Naive Bayes: I (Generative Model)

Disease $Y \in \{0,1\}$

Automated helper to a doctor



} symptoms

X_1 X_2 X_3 X_4 → Features

$\in \{0,1\}$ N|Y $\quad \{0,1\}$ N|Y $\quad \{1, 2, 3\}$ Low/Medium/High $\quad \{0,1\}$ N|Y

$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

$\hookrightarrow x \qquad P(y^{\in\{0,1\}} \mid x)$

$x \in \{0, 1\}$
$x \in \{1 \dots L\}$
$x \in R$

$y \in \{0, 1\}$
$\in \{1 \dots r\}$ $\quad | \quad y \in R$

⇒ Generative Model: I

$\hookrightarrow y \sim \text{Bernoulli}(\phi)$

$\left[x \mid y \begin{cases} 0 \\ 1 \end{cases} N(\mu, \Sigma) \begin{cases} \mu_0, \Sigma_0 \\ \mu_1, \Sigma_1 \end{cases} \right.$ ← [For GDA]

Naive Bayes's

$$x_j \perp x_k \mid y \qquad \forall j, k$$

$$\Rightarrow \quad x_1 \perp x_n \mid y$$

independent — conditional independence

$$\left.P(x_1 \ldots x_n \mid y) = \prod_{j=1}^{n} P(x_j \mid y)\right]$$

$\forall y \in \{0, 1\}$

Naive Bayes Assumption

↳ Features/Attributes are independent of each other given the c

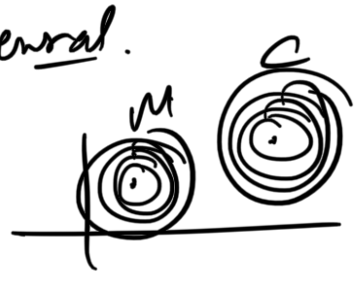
Cough $\perp$ Cold | Flu ↳ 0 ↳ 1

Cough $\not\perp$ Cold?

↳ No:- in general.

Under Naive Bayes Assumption

$$P(x \mid y) = \prod_{j=1}^{n} P(x_j \mid y) \Big\} \quad P(y) = \text{Bernoulli } (\phi)$$



$$P(x, y) = P(y)\, P(x \mid y)$$

$$\hookrightarrow \left[P(y) \prod_{j=1}^{n} P(x_j \mid y) \right]$$

$$P(y \mid x)$$

Effectively we can estimate the parameters corresponding to each attribute independently

$$\frac{P(x|y)\,P(y)}{P(x)} \}$$

$$= \frac{\prod_{j=1}^{n} P(x_j|y)\,P(y)}{\sum_{y}\left[\prod_{j=1}^{n} P(x_j|y)\,P(y)\right]} \Big\} \text{ Normalizing constant}$$

Inference

$$\Big\{ \arg\max_{y} P(y|x) = \arg\max_{y} \frac{P(x|y)\,P(y)}{\cancel{P(x)}}$$

$$= \arg\max_{y} P(x|y)\,P(y)$$

Generative Model :- I

$$P(x,y) = P(y) \prod_{j=1}^{n} P(x_j|y)$$

$y \sim$ Bernoulli (ϕ)

$x|y \sim$ Dependent fun of x

$x_j|y \rightarrow \forall j$ } Naive Bayes Assumption

$x_j|y=0 \sim$ Multinoulli $(\theta_{j|y=0})$ $x_j \in$ Discrete Set $\{1 \ldots L\}$

$x_j|y=1 \sim$ Multinoulli $(\theta_{j|y=1})$

parameters of multinoulli

$$\theta_{j|y=0} = (\theta_{j1|y=0} \ldots \theta_{jL|y=0}) \quad \sum_{l=1}^{L} \theta_{jl|y=0} = 1$$

$$\theta_{j|y=1} = (\theta_{j1|y=1}, \ldots \theta_{jL|y=1}) \quad \sum^{L} \theta_{jl|y=1} = 1$$

Log-Likelihood

Discrete Case:

$$\underset{\theta}{\operatorname{argmax}}\ LL(\theta)$$

class prior

$$\Theta = \left(\phi, \{\theta_{j\ell|y=0}\}_{\ell=1}^{L}, \{\theta_{j\ell|y=1}\}_{\ell=1}^{L}\right) \quad \forall j \quad \forall j$$

params $\simeq O(nL)$ → # of class

$$LL(\theta) = \log \prod_{i=1}^{m} P(x^{(i)}, y^{(i)}; \theta)$$

$$= \sum_{i=1}^{m} \log P(x^{(i)}, y^{(i)}; \theta)$$

$$= \sum_{i=1}^{m} \left[\log P(y^{(i)}; \phi) + \log P(x^{(i)} | y^{(i)}; \theta)\right]$$

Naive Bayes Assumption

$$= \sum_{i=1}^{m} \left[\log P(y^{(i)}; \phi) + \log \prod_{j=1}^{n} P(x_j^{(i)} | y^{(i)}; \theta)\right]$$

$$= \sum_{i=1}^{m} 1\{y^{(i)}=0\} \log P(y^{(i)}=0; \phi) + 1\{y^{(i)}=1\} \log P(y^{(i)}=1; \phi) \quad T1$$

$$+ \sum^{m} \sum^{n} \left[\log P(x_j^{(i)} | y^{(i)}; \theta)\right] \quad T2$$

Aside:

$x \in \mathbb{R}^n$

Naive Bayes Assumption

$\forall j$

$$x_j | y=1 \sim N(\mu_{j1}, \sigma_{j1}^2)$$

Connection with GDA

$$\prod_{j=1}^{n} P(x_j | y=1) \Leftrightarrow \text{HW}$$

$$\sim N(\mu_1, \Sigma_1)$$

$$\mu_1 = (\mu_{11} \ldots \mu_{n1})$$

$$\Sigma_1 = \mathrm{Diag}(\sigma_{11}^2 \ldots \sigma_{n1}^2)$$

Gaussian Naive Bayes

$$i=1 \quad j=1$$

$$T2:- \sum_{i=1}^{m}\left[\sum_{j=1}^{n}\left[1\{y^{(i)}=0\} \underbrace{\log P(x_j^{(i)}|y^{(i)}=0;\theta)} + 1\{y^{(i)}=1\}\log P(x_j^{(i)}|y^{(i)}=1;\theta)\right]\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} 1\{y^{(i)}=0\} \sum_{l=1}^{L} 1\{x_j^{(i)}=l\} \log \underbrace{P(x_j^{(i)}=l|y^{(i)}=0;\theta)}_{\theta_{jl|y=0}}$$

$$+ \sum_{i=1}^{m}\sum_{j=1}^{n} 1\{y^{(i)}=1\} \sum_{l=1}^{L} 1\{x_j^{(i)}=l\} \log \underbrace{P(x_j^{(i)}=l|y^{(i)}=1;\theta)}_{\theta_{jl|y=1}}$$

$$\Downarrow$$

$$\nabla_\theta LL(\theta) \in \left(\phi, \{\theta_{jl|y=0}\}_{l=1}^{L} \ \forall j, \{\theta_{jl|y=1}\}_{l=1}^{L} \ \forall j\right)$$

$= 0$ Analytical

Note:- not all parameter are independent HW/Exam

$\{\theta_{jl|y=0}\}_{l=1}^{L}$ are not independent

$\{\theta_{jl|y=1}\}_{l=1}^{L}$

$$\nabla_\phi LL(\theta) \quad \nabla_{\{\theta_{jl|y=0}\}_{l=1}^{L} \forall j} LL(\theta) \quad \nabla_{\{\theta_{jl|y=1}\}_{l=1}^{L} \forall j} LL(\theta) = 0$$

$$\sum^{m} 1\{y^{(i)}=1\} \checkmark$$

$$\Rightarrow \quad \phi = \frac{\sum_{i=1} 1\{y^{(i)} = 1\}}{m}$$

$$\theta_{jl|y=0} = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=0\} \, 1\{x_j^{(i)}=l\}}{\sum_{i=1}^{m} 1\{y^{(i)}=0\}}$$

$$\theta_{jl|y=1} = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=1\} \, 1\{x_j^{(i)}=l\}}{\sum_{i=1}^{m} 1\{y^{(i)}=1\}}$$