CS3780/5780
Naive Bayes

- Recap: MLE : model $P(x, y)$ with P_θ where $\theta \in \Theta$

$$\theta_{MLE} = \underset{\theta \in \Theta}{\text{argmax}} \; P_\theta(D)$$

model θ that maximizes likelihood of da

MAP: $$\hat{\theta}_{MAP} = \underset{\theta \in \Theta}{\text{argmax}} \; P(\theta \mid D) = \underbrace{P(D \mid \theta)}_{\text{likelihood}} \; \underbrace{P(\theta)}_{\text{Prior}}$$

Eg

Y	X = Favorite dish
Adult	Soup
child	Mac N Cheese
child	Mac N Cheese
Adult	Tacos
Adult	Soup
Child	Tacos
Adult	Soup
Adult	Mac N Cheese
child	Mac N Cheese

3

4

$X = \{$Soup, Mac N cheese, Tacos$\}$

Estimate

P (Y= "child" | X =" Mac or cheese")? $= 3/4$

$$\boxed{\text{Estimate } P(Y = y \mid X = x)\ ? \quad \frac{\sum_{i=1}^{n} \mathbb{1}\{x_i = x\} \, \mathbb{1}\{y_i = y\}}{\sum_{i=1}^{n} \mathbb{1}\{x_i = x\}}}$$

What is the issue with this?

exact matches are rare with many features

Eg

Y	X[1] = Favorite dish	X[2] = # words known	X[3] = Favorite Movie	X[4] = Hours of sleep
Adult	Soup	20000	Godfather	8
child	Mac N Cheese	200	Frozen	11
child	Mac N Cheese	400	Frozen	12
Adult	Tacos	17000	La la land	6
Adult	Soup	15000	Godfather	5
Child	Tacos	1000	Eternals	10
Adult	Soup	21000	Avengers	10
Adult	Mac N Cheese	11000	Avengers	8
child	Mac N Cheese	700	Avengers	11

$\hat{P}(y = $ "Adult" $\mid X = ($"soup", 20000, "Avengers", 8$))$?

Naive Bayes Model

Assumption: $P(X=\vec{x} \mid Y=y) = \prod_{\alpha=1}^{d} P(X[\alpha]=x[\alpha] \mid Y=y)$

"Given it's a child, favorite dish, # words known, hours of sleep. are all independent"

why is this useful?

$$\boxed{\begin{aligned} P(Y=y \mid X=x) &= \frac{\underline{P(X=x \mid Y=y)\, P(Y=y)}}{P(X=x)} \\ &= \frac{\prod_{\alpha=1}^{d} P(X[\alpha]=x[\alpha] \mid Y=y)\, P(Y=y)}{P(X=x)} \\ &= \frac{\prod_{\alpha=1}^{d} P(X[\alpha]=x[\alpha] \mid Y=y)\, P(Y=y)}{\sum_{c\in\mathcal{Y}} \prod_{\alpha=1}^{d} P(X[\alpha]=x[\alpha] \mid Y=c)\, P(Y=c)} \end{aligned}}$$

$P(Y=y)$ and $\forall \alpha$, $P(X[\alpha]=x[\alpha] \mid Y=y)$ are easy to estimate

$$\boxed{\begin{array}{l}\text{Eg. estimate} \\ \hat{P}(Y=\text{"Adult"} \mid X=(\text{"soup"}, 20000, \text{"Avengers"}, 8))\ ?\end{array}}$$

$$\begin{aligned} h(x) &= \underset{y\in\mathcal{Y}}{\operatorname{argmax}}\ \hat{P}(Y=y \mid X=x) \\ &= \underset{y\in\mathcal{Y}}{\operatorname{argmax}}\ \prod_{\alpha} P(X[\alpha]=x[\alpha] \mid Y=y)\, P(Y=y) \\ &= \underset{y\in\mathcal{Y}}{\operatorname{argmax}}\ \sum_{\alpha} \log(P(X[\alpha]=x[\alpha] \mid Y=y)) \\ &\qquad + \log P(Y=y) \end{aligned}$$

When $X_i[\alpha]$ are counts

Eg $X_i[\alpha] = j$ means α^{th} word in the dictionary occurs j times in the document i

x is an m word document: $X[\alpha] \in \{0, 1, \ldots, m\}$ $\sum_{\alpha=1}^{d} X[\alpha] = m$

Multinomial Distribution

$$P(X = x \mid m, Y = y) = \frac{m!}{X[1]!\, X[2]! \times \cdots \times X[d]!} \prod_{i=1}^{d} (\theta_{\alpha,y})^{X[\alpha]}$$

MLE estimate:

$$\hat{\theta}_{\alpha,y} = \frac{\sum_{i=1}^{n} 1\{y_i = y\} X_i[\alpha]}{\sum_{i=1}^{n} 1(y_i = y)\, m_i}$$

m_i : # words in document i

$$h(x) = \underset{y \in \mathcal{Y}}{\operatorname{argmax}}\ \hat{P}(Y = y) \prod_{\alpha=1}^{d} \hat{\theta}_{\alpha,y}^{X[\alpha]}$$

$X[\alpha]$'s are Continuous variables.: Gaussian distribution conditioned on Y

$$p(X[\alpha] = x \mid Y = y) = \frac{1}{\sqrt{2\pi}\,\sigma_y^{2}} \exp\left(-\frac{(x - \mu_y)^2}{2\sigma_y^2}\right)$$

Parameter estimation:

$$\hat{\mu}_y[\alpha] = \frac{\sum_{i=1}^{n} 1(y_i = y)\, x_i[\alpha]}{\sum_{i=1}^{n} 1(y_i = y)}$$

$$\hat{\sigma}_y^2 = \frac{\sum_{i=1}^{n} 1(y_i = y)\,(x_i[\alpha] - \hat{\mu}_y[\alpha])^2}{\sum_{i=1}^{n} 1(y_i = y)}$$

1. For both multinomial case and Gaussian case (with variance between class per feature fixed) classification boundary is linear.

2. For Gaussian case

$$P(Y = y \mid x) = \frac{1}{1 + \exp(-y(w^T x + b))}$$

logistic link function.

Y	X[1] = Favorite dish	X[2] = # words known	X[3] = Favorite Movie	X[4] = Hours of sleep
Adult	Soup	20000	Godfather	8
child	Mac N Cheese	200	Frozen	11
child	Mac N Cheese	400	Frozen	12
Adult	Tacos	17000	La la land	6
Adult	Soup	15000	Godfather	5
Child	Tacos	1000	Eternals	10
Adult	Soup	21000	Avengers	10
Adult	Mac N Cheese	11000	Avengers	8
child	Mac N Cheese	700	Avengers	11

$$\hat{P}(y = \text{"Adult"} \mid x = (\text{"soup"}, 20000, \text{"Avengers"}, 8)) \ ?$$

$$= \frac{\prod_{\alpha=1}^{d} P(x[\alpha] = x[\alpha] \mid y = y)\ P(y=y)}{\sum_{c\in y} \prod_{\alpha=1}^{d} P(x[\alpha] = x[\alpha] \mid y = c)\ P(y=c)}$$

$\hat{P}(y = \text{"Adult"}) = \frac{5}{9}$ $\qquad \frac{5 \text{ adults}}{9 \text{ people}}$ $\qquad P(y = \text{child})$

$\hat{P}(x[1] = \text{"soup"} \mid y = \text{Adult}) = 3/5$ $\qquad \frac{3 \text{ soups for adults}}{5 \text{ adults}}$

$\hat{P}(x[2] = 20000 \mid y = \text{Adult}) = 1/5$

$\hat{P}(x[3] = \text{"Avengers"} \mid y = \text{Adult}) = 2/5$

$\hat{P}(x[4] = 8 \mid y = \text{Adult}) = 2/5$

$$\hat{P}(x[1] = \text{"soup"} \mid y = \text{child}) = 0/4$$

$$\hat{P}(y = \text{"Adult"} \mid x = (\text{"soup"}, 20000, \text{"Avengers"}, 8))$$

$$= \frac{\prod_{\alpha=1}^{d} P(x[\alpha] = x[\alpha] \mid y = y)\ P(y = y)}{\sum_{c \in y} \prod_{\alpha=1}^{d} P(x[\alpha] = x[\alpha] \mid y = c)\ P(y = c)}$$

$$= \frac{\left(\frac{3}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{2}{5}\right) \times \frac{5}{9}}{\left(\frac{3}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{2}{5}\right) \times \frac{5}{9} + 0 \times \frac{4}{9}}$$

$$= 1$$