Navie Bayes Assumption

: Assumes all feature values are independent give the label.

$P(X=x|Y=y) : \prod_{i=1}^{d} P([x]_i = x_i | Y=y)$ Way2

Note: use Bayes classifier $h(x) = \arg\max_y P_\theta(y=y|X=x)$ which

$P_\theta(y=y|X=x)$ have 2 optional

① $P(Y=y|X=\vec{x}) = P(Y=y|[x]_1=x_1, [x]_2=x_2, \ldots, [x]_d=x_d)$ way1

② $P(y=y|X=x) = \frac{P(X=x|Y=y)\cdot P(Y=y)}{P(X=x)}$ way2

Bayes Rule

Naive bayes use this!

way1

MLE: $P_\theta(Y=y|X=x) = \frac{P_\theta(Y=y \wedge X=x)}{P_\theta(X=x)}$

$= \frac{\sum_{i=1}^{n} I(Y_i=y \wedge X_i=x)}{\sum_{i=1}^{n} I(X_i=x)}$

problem!

when $d \gg 0$ and $n \to +\infty$

(d: dimension)

$\Rightarrow P_\theta(Y=y \wedge X=x)\cdot\frac{1}{n} = 0$

$\Rightarrow P_\theta(X=x)\cdot\frac{1}{n} = 0$

So: $P_\theta(Y=y|X=x) = \frac{0}{0}$ undifind

way2

MLE: Using Bayes classifier

$h(x) = \arg\max_y P(X=x|Y=y)\cdot P(Y=y)$

$P_\theta(Y=y) = \sum_{i=1}^{n} \frac{I(Y_i=y)}{n}$

$P_\theta(X=x|Y=y) = ?$ Can't estimate (use naive Bayes assumptions) $\Rightarrow P([\vec{x}]_\alpha = \vec{x}^\alpha | Y=y)$

Naive Bayes assumptions

: Assumes all feature values are independent give the label.

$P(X=x|Y=y) : \prod_{i=1}^{d} P([x]_i = x_i | Y=y)$ Way2

Naive Bayes classifier

$$h(x) = \arg\max_y \prod_{\alpha=1}^{d} P([\vec{x}]_\alpha = \vec{x}^\alpha | Y=y)\, P(Y=y)$$

Estimate $P([\vec{X}]_\alpha \mid Y)$ 3 cases

cases 1) Categorical Naive Bayes Classifier

MLE : $\frac{\sum_{i=1}^{n} I(y_i = y) \cdot I(x_i^\alpha = j)}{\sum_{i=1}^{n} I(y_i = y)}$ eg: {single, widowed, married}

cases 2) Multinomial feature : Multinomial Naive Bayes classifier

eg. text data : "An ant is animal"

$$\frac{P(Y=\text{spam}) \times \prod_{\alpha=1}^{d} P([X]_\alpha = x^\alpha \mid Y=\text{spam})}{P(Y=\text{ham}) \times \prod_{\alpha=1}^{d} P([X]_\alpha = x^\alpha \mid Y=\text{ham})} \propto \frac{\left(\sum_{\alpha=1}^{d} x^\alpha \log_e([\theta_{\text{spam}}]_\alpha)\right) + \log_e(P(Y=\text{spam}))}{\left(\sum_{\alpha=1}^{d} x^\alpha \log_e([\theta_{\text{ham}}]_\alpha)\right) + \log_e(P(Y=\text{ham}))}$$

which $[\theta_y]_\alpha = \frac{\sum_{i=1}^{n} I(y_i=y) \cdot x_i^\alpha}{\sum_{i=1}^{n} I(y_i=y)\left(\sum_{b=1}^{d} x_i^b\right)}$

→ จน. ครั้งแต่ละ word ปรากฏอยู่ใน y

→ จน. ของ word ทั้งหมดที่อยู่ใน y

cases3) continous feature : Gvassian Naive Bayes classifier

$[X]_\alpha \in \mathbb{R}$

$$P(y \mid x) = \frac{1}{1+e^{-y(\vec{w}^T \vec{x})}} \; ; \; y = \{-1, 1\}$$