## Anatomy of Naïve Bayes Classifier

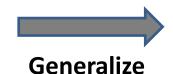
## Two categories: $\theta_1$ and $\theta_2$

$$score(d) = \log \frac{p(\theta_1 \mid d)}{p(\theta_2 \mid d)} = \log \frac{p(\theta_1) \prod_{w \in V} p(w \mid \theta_1)^{c(w,d)}}{p(\theta_2) \prod_{w \in V} p(w \mid \theta_2)^{c(w,d)}}$$

$$= \log \frac{p(\theta_1)}{p(\theta_2)} + \sum_{w \in V} c(w,d) \log \frac{p(w \mid \theta_1)}{p(w \mid \theta_2)}$$
 Weight on each word (feature)  $\beta_i$  doesn't depend on d! Sum over all words Feature value:  $x = c(w,d)$ 

Sum over all words (features  $\{x_i\}$ )

Feature value: x<sub>i</sub>=c(w,d)



$$d = (x_1, x_2, ..., x_M), x_i \in \mathcal{Y}$$

$$d = (x_1, x_2, ..., x_M), \ x_i \in \Re$$
 
$$score(d) = \beta_0 + \sum_{i=1}^{M} x_i \beta_i \quad \beta_i \in \Re$$
 = Logistic Regression!