Activity Recognition 5
Classification

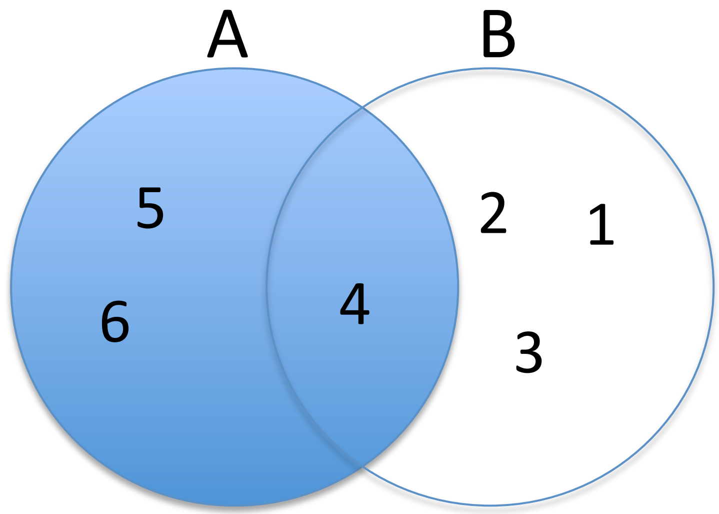
Enterprise Computing
Minho Shin
2012. 10

Conditional Probability

- Conditional probability

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

- A: >= 4
- B: even (2, 4, 6)
- P(A) = ½
- P(B) = ½
- P(A,B) = 1/3
- P(A|B) = 2/3



Bayesian Theorem

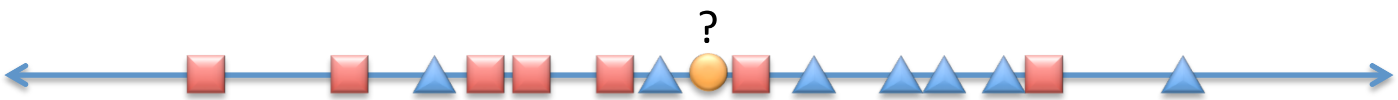
- Thomas Bayes (1701-1761)
- Bayesian Theorem

$$P(A|B) = \frac{P(A,B)}{P(B)} \qquad P(A,B) = P(A|B)P(B)$$

$$\begin{aligned} P(S|E) &= \frac{P(S,E)}{P(E)} \\ &= \frac{P(S)P(E|S)}{P(E)} \end{aligned}$$

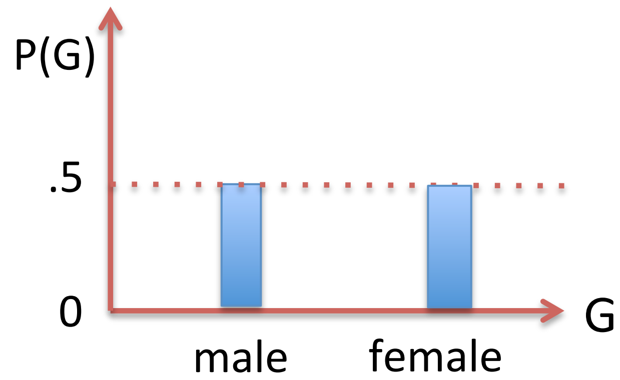
Example: Gender classification

- G: Gender {male, female}, as statement
- H: height, as evidence
- Want to know:
  - P(G|H): Guess gender given evidence of height
  - P(G=m|H=165cm) = ?
- Classification
  - Given feature set {height}, classify gender



Prior probability

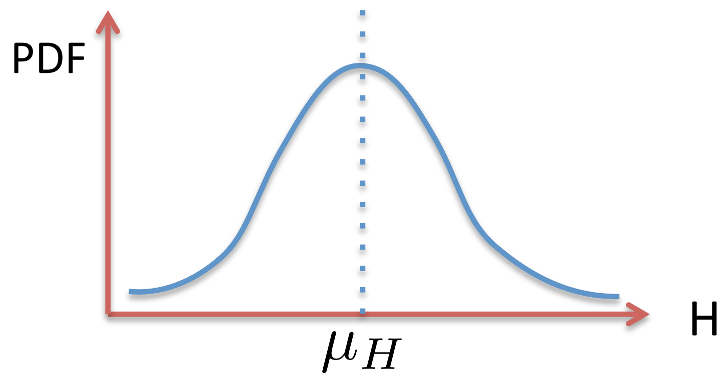
- P(G): prob. of gender (gender distribution)
- P(G=m) =
- P(G=f) =



- Probability of gender before any evidence is given
  - *Prior probability*

Evidence

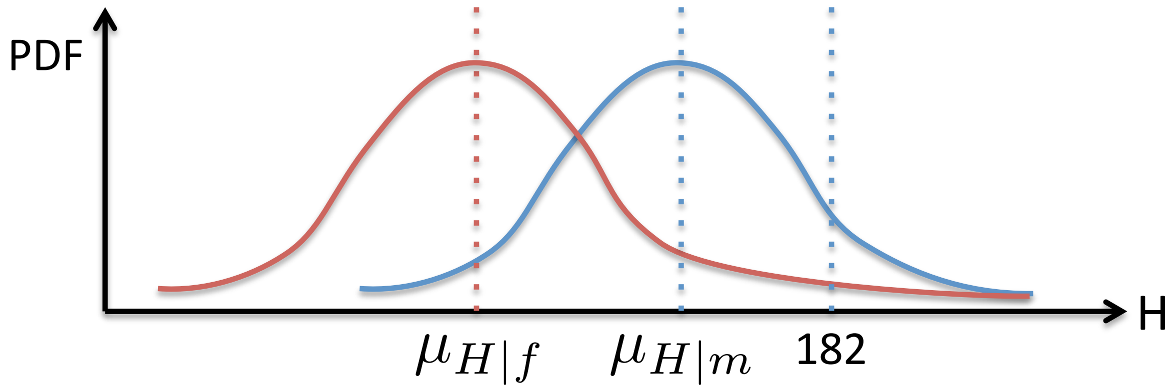
- P(H): Probability of height (Height distribution)
- P(H=165)=?



- Probability of the evidence
  - *evidence probability*

Likelihood

- P(H|G): prob. of height given gender G
- P(H=182 | G=f)? and P(H=182 | G=m)?



- How much likely to observe an evidence when the gender was *g*?
  - *Likelihood*

Bayesian Theorem

- Probability model:

$$P(G|H) = \frac{P(G)P(H|G)}{P(H)}$$

$$\text{Posterior} \quad = \quad \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}$$

Naïve Bayesian Classificaiton

- Given evidences, we want to choose gender *g* that maximizes

$$
\begin{aligned}
P(G = g|H) &= \frac{P(G = g)P(H|G = g)}{P(H)} \\
&\propto P(G = g)P(H|G = g)
\end{aligned}
$$

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

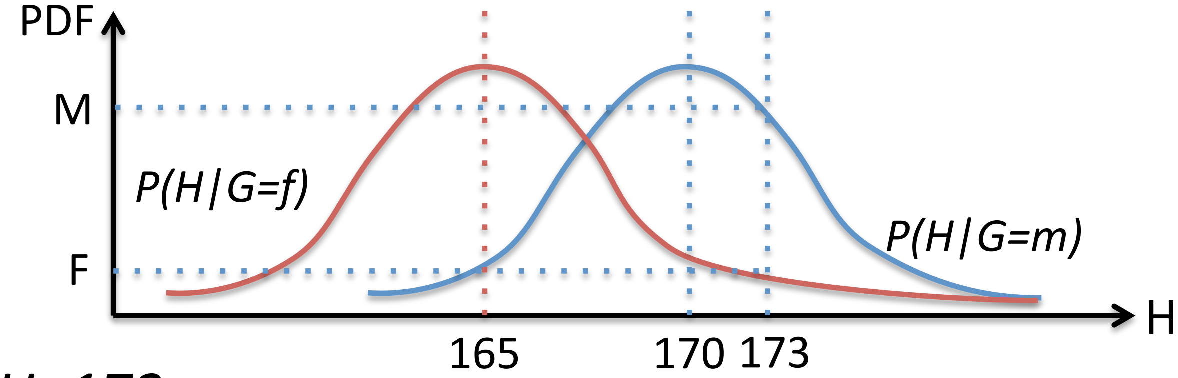
- Maximum A Posteriori (MAP) classification

Example Naïve Bayesian Class'ion

- Goal: Find gender *g* maximizing posterior

$$\text{Posterior} = P(G|H) \propto P(G)P(H|G)$$

- P(G=m) = P(G=f) = 0.5
- P(H|G=m), P(H|G=f) is given by (obtained from training)



- Classification when *H=173:*
  - *Posterior of male = 0.5 * M*
  - *Posterior of female = 0.5 * F*
  - *Posterior of male > female, therefore, it's a male!*

Naïve Bayesian w/ multiple features

- G: Gender {male, female}, as classes
- H: height, as an evidence
- W: weight, as an evidence
- F: foot size, as an evidence
- Classification
  - Given feature set {height, weight, foot-size}, predict the gender

Naïve Bayesian w/ multiple features

- Posterior: probability of gender given height, weight, and foot size

$$P(G|H,W,F)$$

- By Bayesian theorem,

$$\begin{aligned} P(G|H,W,F) &\propto P(G)P(H,W,F|G) \\ &\propto P(G)P(H|G)P(W,F|G,H) \\ &\propto P(G)P(H|G)P(W|G,H)P(F|G,H,W) \\ &\propto P(G)P(H|G)P(W|G)P(F|G) \end{aligned}$$

- *Feature independence assumption!*

Naïve Bayesian Assumption

- Features are independent
- Not always true, and mostly not true
- But this simplification works well in many cases
  - Defeats curse of dimensionality
  - What matters is the relative comparison between posteriors of classes, and feature independence simplification keeps the comparison

Spam Filter

$$\text{Posterior} \quad = \quad \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}$$

$$P(C|E) = \frac{P(C)P(E|C)}{P(E)} \qquad C = S \text{ or } \sim S$$

$$E \text{ is non-spam if } P(\sim S|E) > P(S|E)$$

$$E \text{ is spam if } P(\sim S|E) < P(S|E)$$

Prior

- P(S): probability of an email being spam

$$P(S) = \frac{\text{number of spam emails}}{\text{number of all emails}}$$

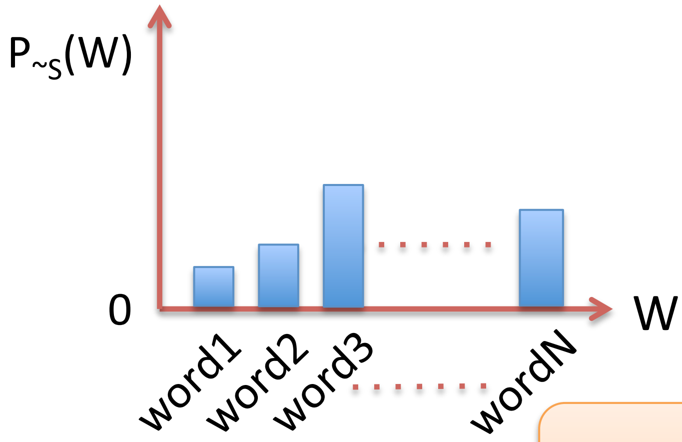
- P(~S): probability of an email being non-spam

$$P(\sim S) = \frac{\text{number of non-spam emails}}{\text{number of all emails}}$$

Spam Filter: Likelihood Model

Likelihood of E given non-spam

E



Independence Assumption

$P(E|\sim S) = P_{\sim S}(\text{"Subj" and "Replica" and ... "XXX"})$

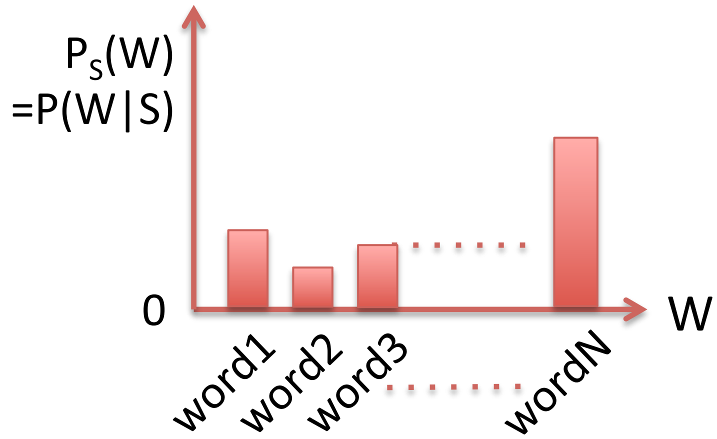
$= P_{\sim S}(\text{"Subj"}) \times P_{\sim S}(\text{"Replica"}) \times \ldots \times P_{\sim S}(\text{"XXX"})$

$E = \{\text{"Subj", "Replica", ..., "XXX"}\}$

$$P(E = \{w_1, \ldots, w_n\}|S = \text{non-spam}) = P_{\sim S}(w_1, \ldots, w_n) = \prod_i P_{\sim S}(w_i)$$

Likelihood of E given spam

E



$P(E|S) = P_S(\text{"Subj" and "Replica" and ... "XXX"})$

$= P_S(\text{"Subj"}) \times P_s(\text{"Replica"}) \times \ldots \times P_s(\text{"XXX"})$

$$P(E = \{w_1, \ldots, w_n\} | S = \text{spam}) = P_S(w_1, \ldots, w_n)$$
$$= \prod_i P_S(w_i)$$

Classification

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

$$\begin{aligned}
\text{Posterior}_{\text{spam}} &= P(S|E) \\
&\propto P(S)P(E|S) \\
&= P(S) \prod_{w \in E} P_S(w)
\end{aligned}$$

$$\begin{aligned}
\text{Posterior}_{\text{non-spam}} &= P(\sim S|E) \\
&\propto P(\sim S)P(E|\sim S) \\
&= P(\sim S) \prod_{w \in E} P_{\sim S}(w)
\end{aligned}$$