Activity Recognition 4
Classification

Mobile Computing

Minho Shin

2012. 10

REVIEW: FEATURE SELECTION/ EXTRACTION, CLASSIFIERS

Feature selection

- Discarding features that are little helpful for classification
- But, finding the subset is exponentially expensive

Feature Extraction

- Idea: another data representation can be constructed in a subspace (less dimension) while keeping discriminative capability
- Example algorithms
  - PCA (Principle Component Analysis): transform features into small number of uncorrelated variables
  - ICA (Independent Component Analysis)

Type of Classifiers

- Supervised & Unsupervised
- Single-frame & Sequential
  - Single-frame: Each frame is classified regardless of previous frames
  - Sequential: Each frame is classified in consideration of previous frames

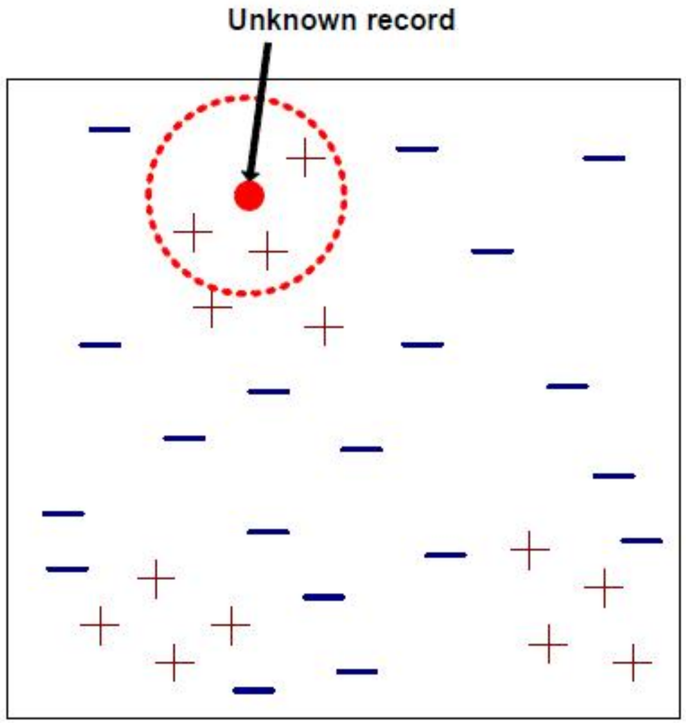
Type of Classifiers

- Template matching
  - Based on similarity between data and
- Binary classifier
  - binary decision tree from root to leaves refining classification
- Geometric: Construct decision boundaries
  - k-NN/ Nearest Mean (NM), Support Vector Machine (SVM)
- Probabilistic
  - feature vector $\boldsymbol{x}$ is classified to class C_{i*} if class-conditional PDF $p(\boldsymbol{x}|C_i)$ is maximized for $i=1, \ldots, C$
  - naïve Bayesian, Logistic, Parzen, Gaussian Mixture Model (GMM)

k-NN

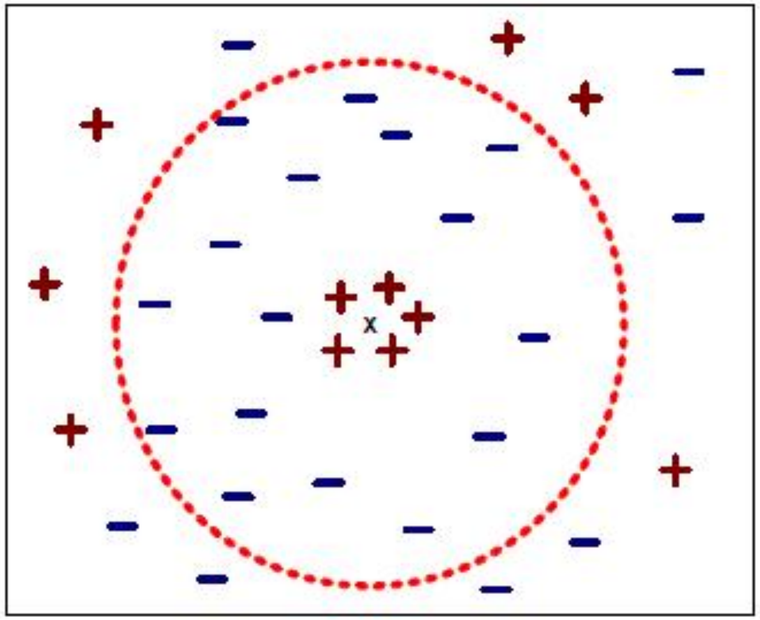
- *"If it walks like a duck, quacks like a duck, and looks like a duck, then it's probably a duck"*

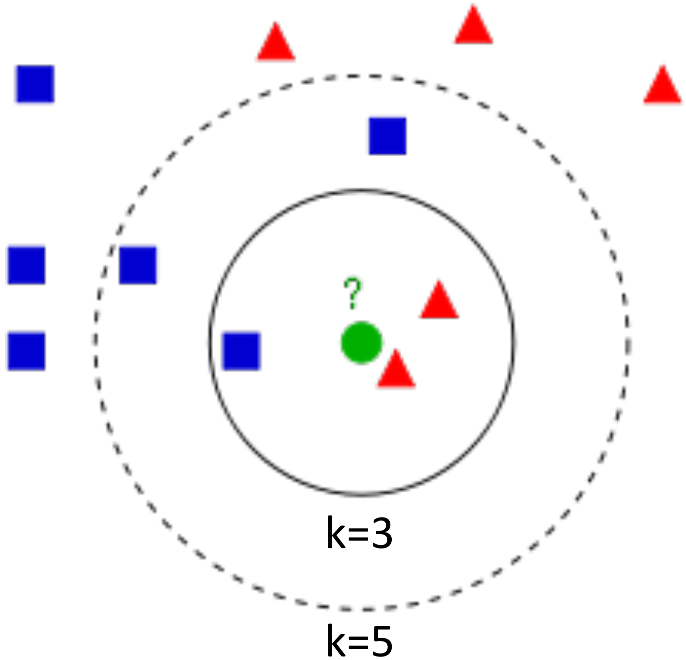
  - Plot each training data in space
  - Given test data, compute the k nearest training data
  - Test data is classified to the same class of majority of k-NN



kNN: Choosing K

- Not too small, not too large

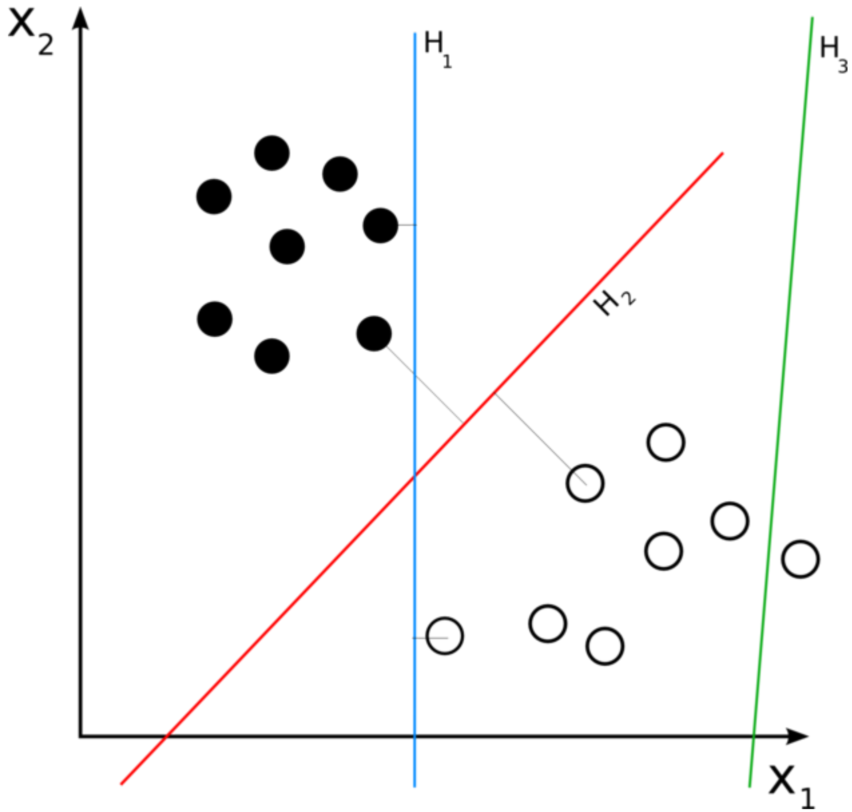




SVM(Support Vector Machine)

- Binary classification

- Find a hyperplane that separates two classes
- (evenly) maximize the distance to the nearest training data of any class (margin)



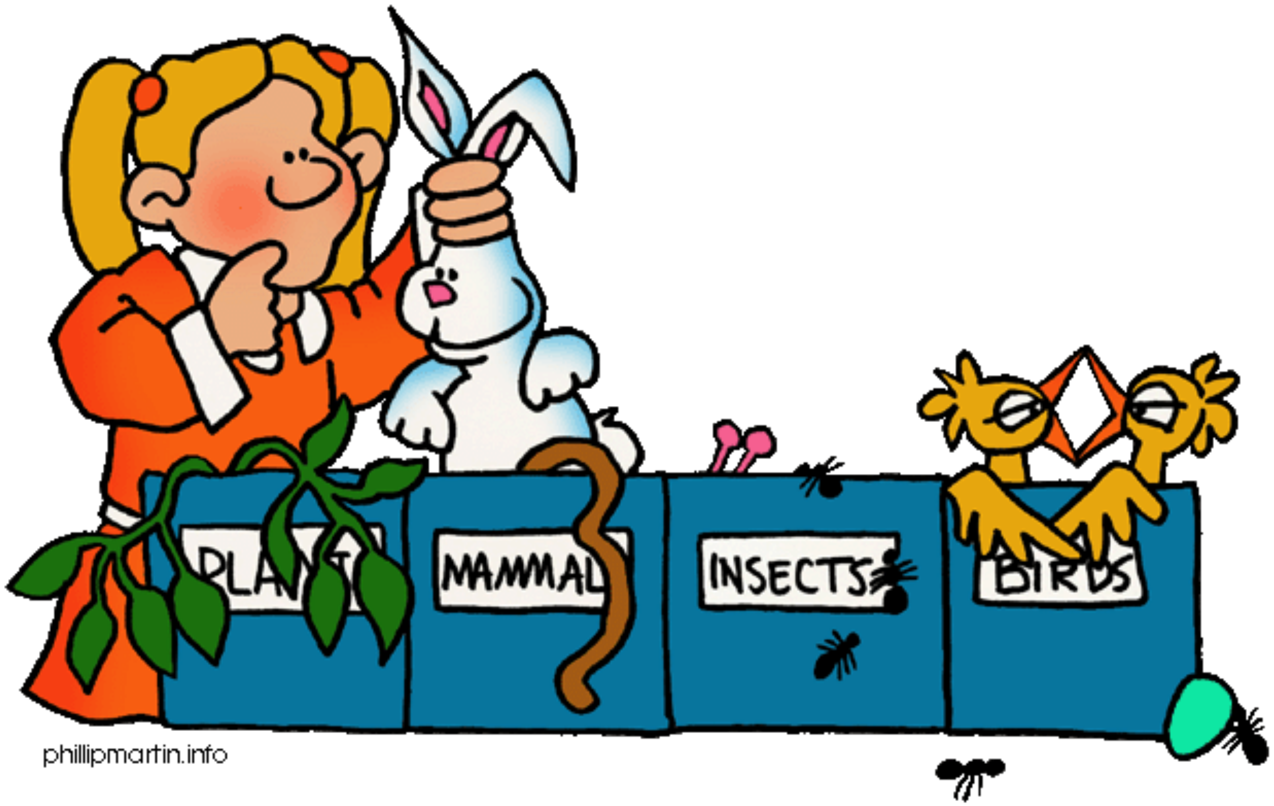
Machine Learning

Machine Learning

- Design and development of algorithms that allow computers to evolve behaviors based on empirical data
- A computer program is said to 'learn' if its performance in a task improves with experience

Classification

- Assign a given piece of data into one of given categories



Example: Spam Filter

- Determine if an email is a spam or non-spam

Example: Spam Filter

- Determine if an email is a spam or non-spam

Classification of spam

- Observe a set of emails labeled with “spam” or “non-spam”
- Build a model for words used in each class of emails
- Given a new email, analyze words in it and determine whether it is a spam or not

Probabilistic Classification

- Use *Statistical Inference* to find the best class for a given instance
- Output probabilities of the instance being a member of each class

- Naïve Bayesian Classifier
- Logistic
- Gaussian Mixture Model classifier

Probability: two schools

- *Physical probability by Frequentists*
  - Objective
  - Assigned to an event of random experiments
  - Relative frequency of an event occurring in a repeating experiment (in the long run)
  - *What is the probability of a dice showing six?*

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

Probability: two schools

- Evidential probability by Bayesians
  - Subjective probability, Epistemic probability
  - Assigned to any statement
  - Subjective plausibility, Degree of belief
  - How much the statement is supported by available evidence
  - *How probably that a suspect is guilty based on the evidences presented?*

$$P(S|E) = \frac{P(S)P(E|S)}{P(E)}$$

Bayesian Theorem

- Thomas Bayes (1701-1761)
- Conditional probability

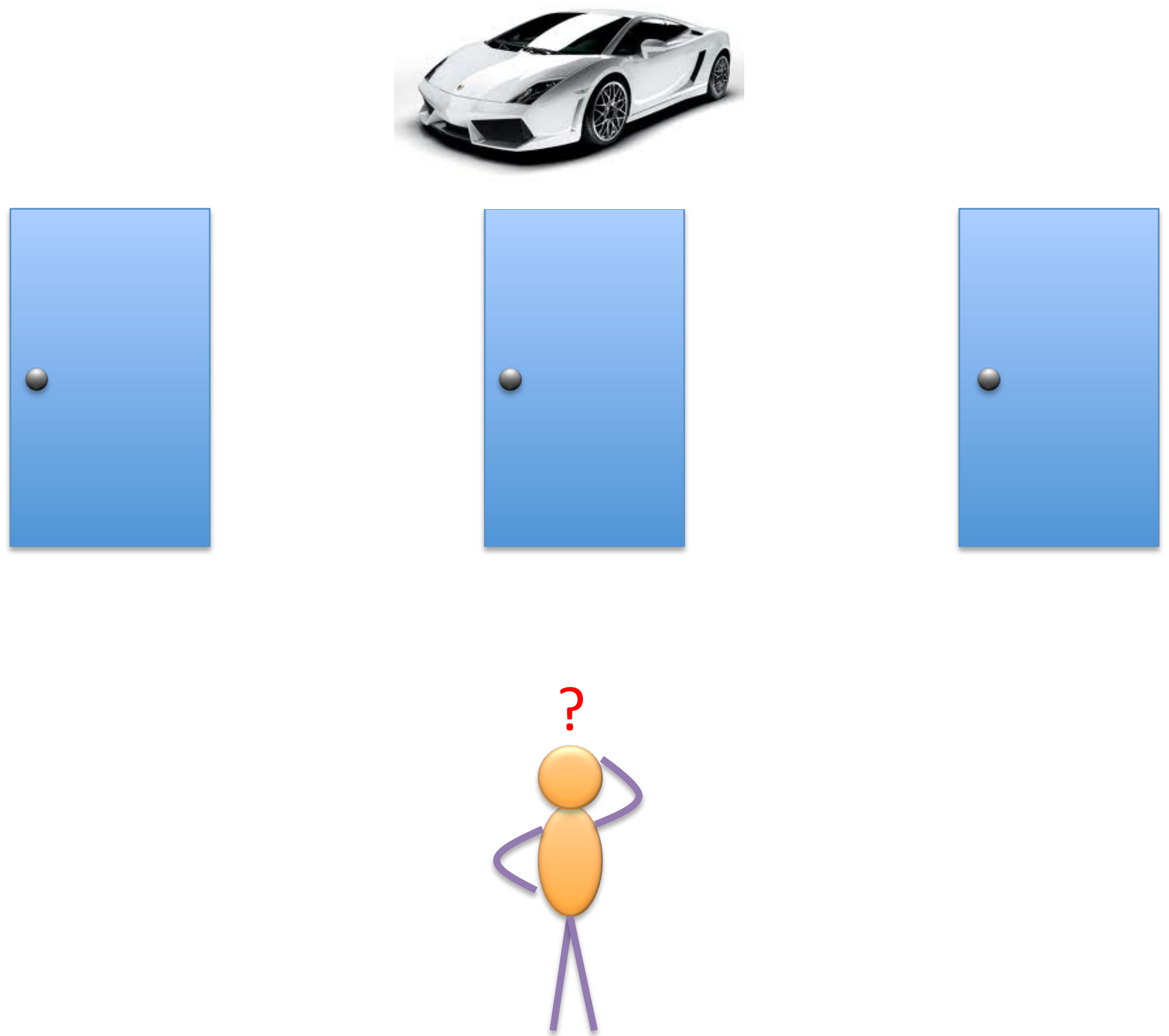
$$P(A|B) = \frac{P(A,B)}{P(B)} \qquad P(A,B) = P(A|B)P(B)$$

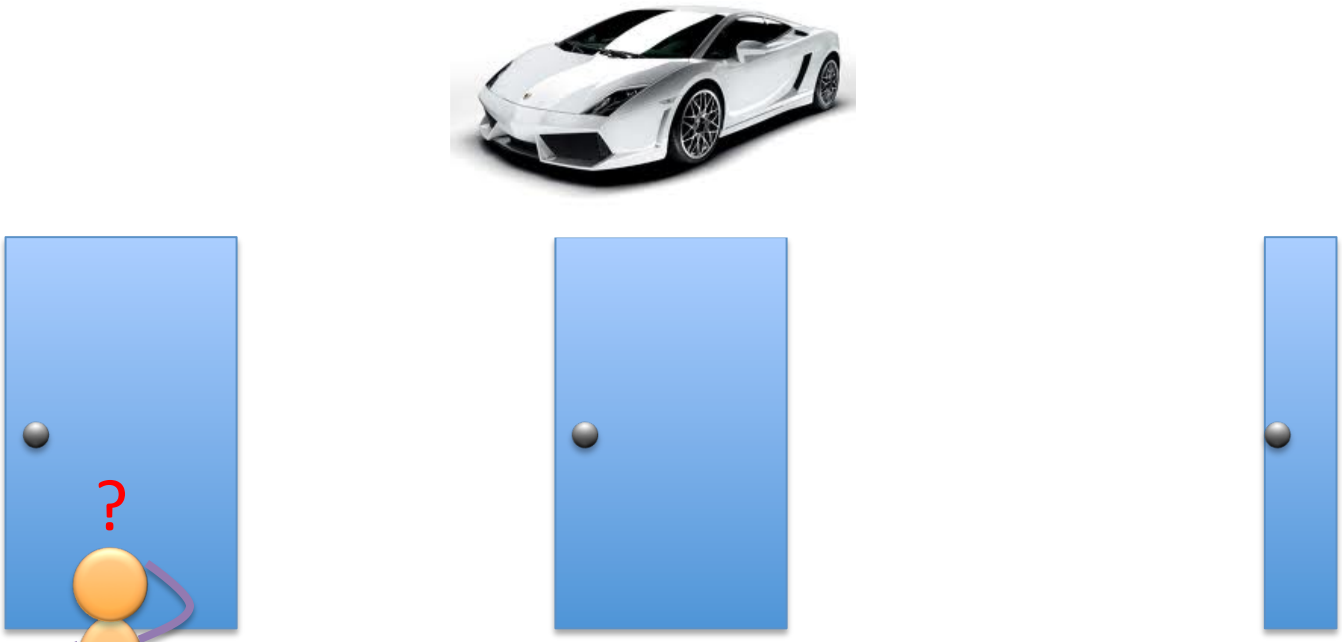
- Therefore

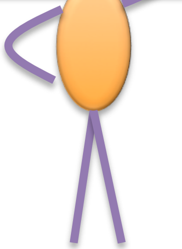
$$\begin{aligned} P(S|E) &= \frac{P(S,E)}{P(E)} \\ &= \frac{P(S)P(E|S)}{P(E)} \end{aligned}$$

Monty Hall Paradox

- You are standing in front of three doors closed.
- A luxury car is behind one of the doors
- You pick a door that you think the car is behind
- Then, the host opens up a door other than the door you chose, and that doesn't have car behind
- Now you are asked to decide: Would you change your choice of the door?
- What will be better choice? Move? Stay?





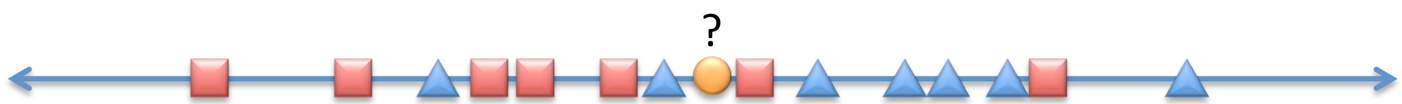


- $s=1$, $c=2$, $h=3$
- $P(c=1|h=3) = P(c=1)P(h=3|c=1) \,/\, P(h=3)$
  $= (1/3)*(1/2) \,/\, P(h=3)$
- $P(c=2|h=3) = P(c=2)P(h=3|c=2) \,/\, P(h=3)$
  $= (1/3)*(1) \,/\, P(h=3)$
- $P(c=2|h=3) > P(c=1|h=3)$

- So, you better change your choice!

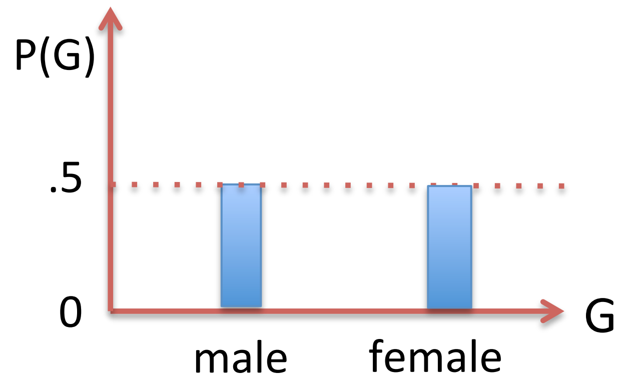
Example: Gender classification

- 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

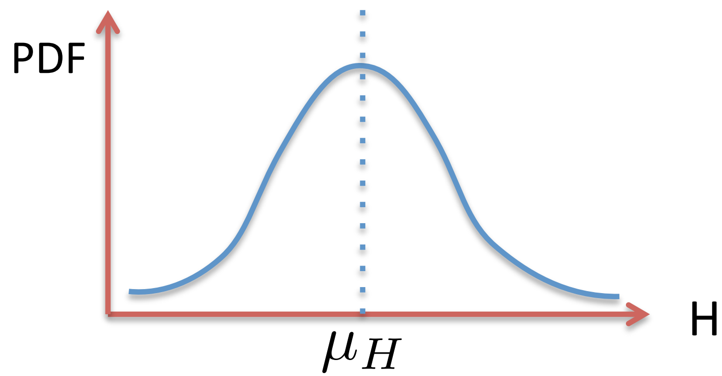
- P(G): prob. of gender (gender distribution)
- P(G=m) =
- P(G=f) =



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

Evidence

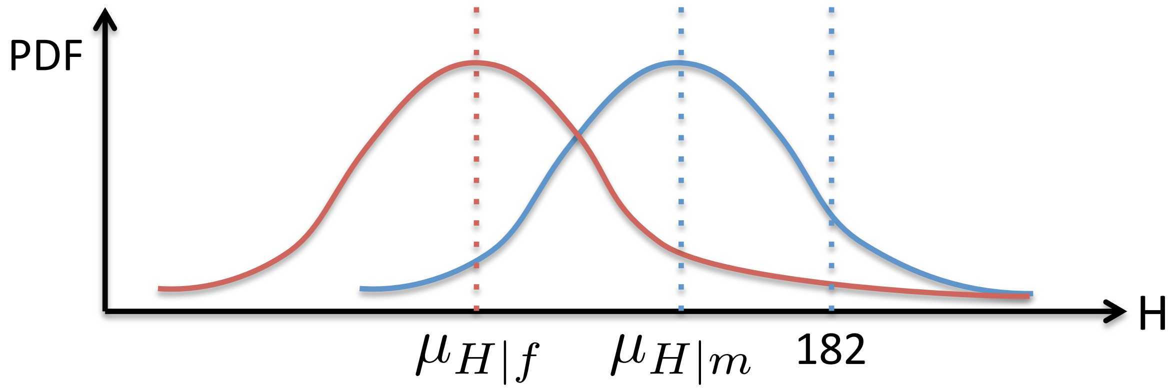
- P(H): Probability of height (Height distribution)
- P(H=165)=?



- 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

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

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

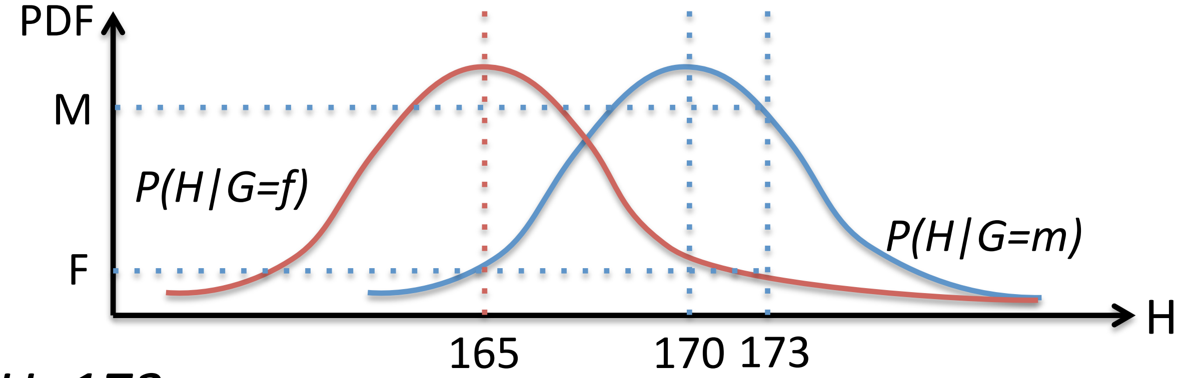
- Maximum A Posteriori (MAP) classification

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

Naïve Bayesian Spam Filter

- S: class of a document {spam, email}
- D: feature of a document $\{w_1, w_2, \ldots, w_n\}$, a set of words appearing in the document
- Classify a document of $\{w_i\}$ into spam or email
- Posterior:

$$P(S|D) \propto P(S)P(D|S)$$

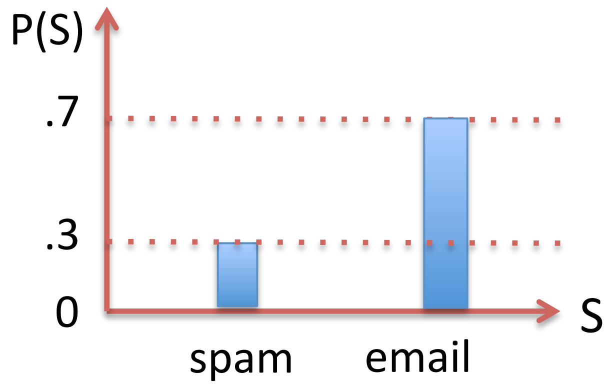
- Prior: P(S) is

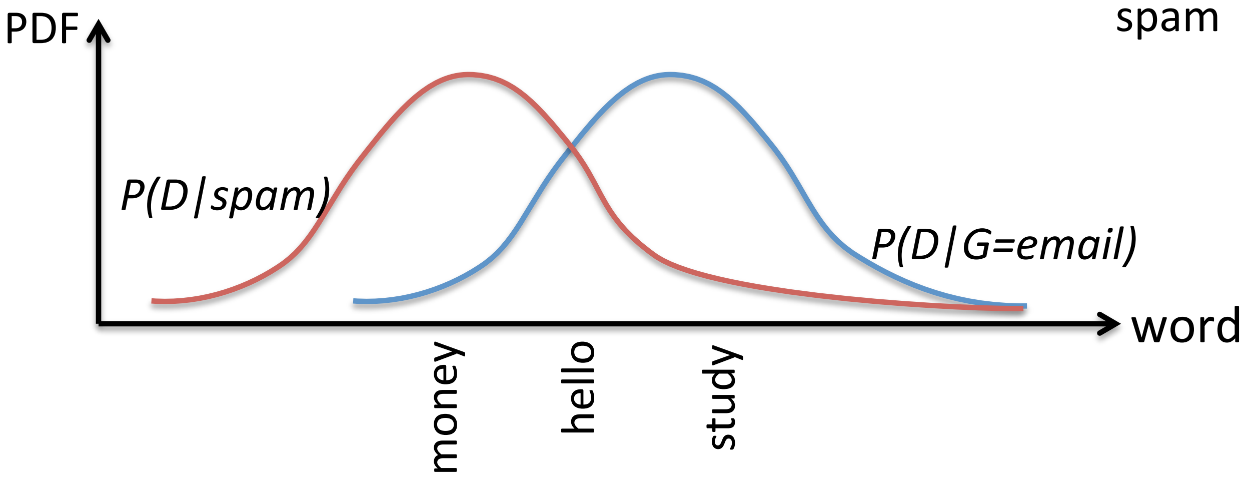
Naïve Bayesian Spam Filter

- Posterior:

$$P(S|D) \propto P(S)P(D|S)$$

- Prior: P(S) is
- Likelihood:
  - Distribution of words





Naïve Bayesian Spam Filter

- Likelihood P(D|S)

$$
\begin{aligned}
P(D = \{w_1, \ldots, w_n\}|S = \text{spam}) &= P(w_1, ..., w_n|S = \text{spam}) \\
&= \prod_i P(w_i|S = spam)
\end{aligned}
$$

- multiplication of prob. of each word showing in a spam email

$$
\begin{aligned}
P(D = \{w_1, \ldots, w_n\}|S = \text{email}) &= P(w_1, ..., w_n|S = \text{email}) \\
&= \prod_i P(w_i|S = email)
\end{aligned}
$$

- multiplication of prob. of each word showing in a non-spam email