Bayes Classifiers

PROF XIAOHUI XIE
SPRING 2019

CS 273P Machine Learning and Data Mining

Slides courtesy of Alex Ihler

Machine Learning

Bayes Classifiers

Naive Bayes Classifiers

Bayes Error

Gaussian Bayes Classifiers

A basic classifier

- Training data $D=\{x^{(i)},y^{(i)}\}$, Classifier $f(x ; D)$
  - Discrete feature vector x
  - $f(x ; D)$ is a contingency table
- Ex: credit rating prediction (bad/good)
  - X_1 = income (low/med/high)
  - How can we make the most # of correct predictions?

Features	# bad	# good
X=0	42	15
X=1	338	287
X=2	3	5

A basic classifier

- Training data $D=\{x^{(i)},y^{(i)}\}$, Classifier $f(x ; D)$
  - Discrete feature vector x
  - $f(x ; D)$ is a contingency table
- Ex: credit rating prediction (bad/good)
  - X_1 = income (low/med/high)
  - How can we make the most # of correct predictions?
  - Predict more likely outcome
    for each possible observation

Features	# bad	# good
X=0	42	15
X=1	338	287
X=2	3	5

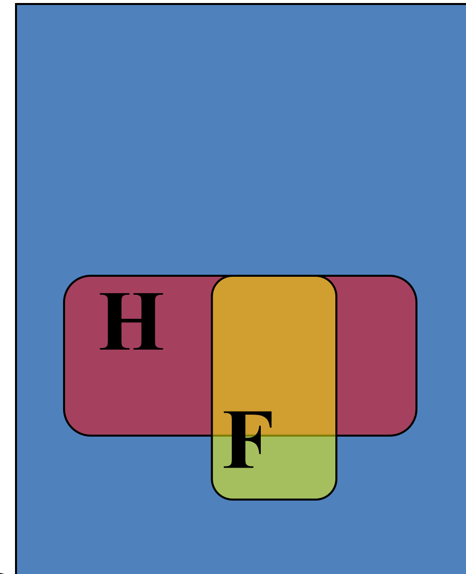
A basic classifier

- Training data $D=\{x^{(i)},y^{(i)}\}$, Classifier $f(x ; D)$
  - Discrete feature vector x
  - f(x ; D) is a contingency table
- Ex: credit rating prediction (bad/good)
  - X_1 = income (low/med/high)
  - How can we make the most # of correct predictions?
  - Predict more likely outcome
    for each possible observation
  - Can normalize into probability:
    p(y=good | X=c)
  - How to generalize?

Features	# bad	# good
X=0	.7368	.2632
X=1	.5408	.4592
X=2	.3750	.6250

Bayes rule

- Two events: headache, flu
- p(H) = 1/10
- p(F) = 1/40
- p(H|F) = 1/2



- You wake up with a headache – what is the chance that you have the flu?

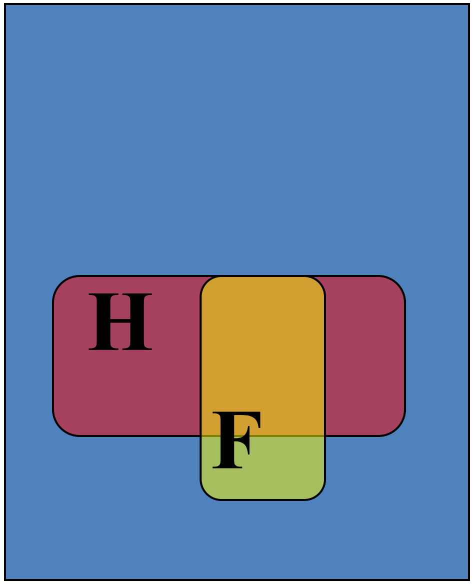
Example from Andrew Moore's slides

Bayes rule

- Two events: headache, flu
- p(H) = 1/10
- p(F) = 1/40
- p(H|F) = 1/2

- P(H & F) = ?

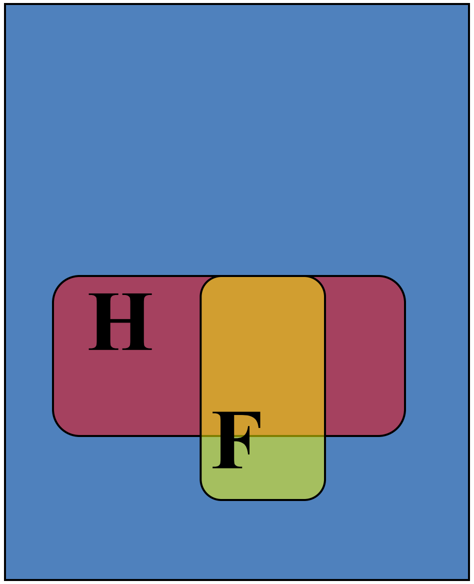
- P(F|H) = ?



Example from Andrew Moore's slides

Bayes rule

- Two events: headache, flu
- p(H) = 1/10
- p(F) = 1/40
- p(H|F) = 1/2



- P(H & F) = p(F) p(H|F)

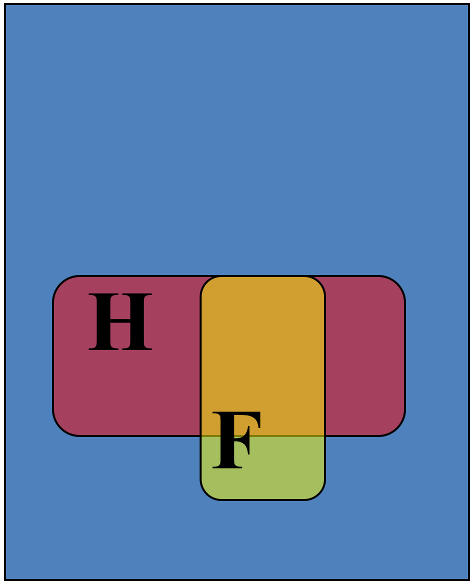
  = (1/2) * (1/40) = 1/80
- P(F|H) = ?

Example from Andrew Moore's slides

Bayes rule

- Two events: headache, flu
- $p(H) = 1/10$
- $p(F) = 1/40$
- $p(H|F) = 1/2$

- $P(H \& F) = p(F)\, p(H|F)$
  $= (1/2) * (1/40) = 1/80$
- $P(F|H) = p(H \& F) / p(H)$
  $= (1/80) / (1/10) = 1/8$



Example from Andrew Moore's slides

Classification and probability

- Suppose we want to model the data

- Prior probability of each class, p(y)
  - E.g., fraction of applicants that have good credit
- Distribution of features given the class, p(x | y=c)
  - How likely are we to see "x" in users with good credit?

- Joint distribution

$$p(y|x)p(x) = p(x,y) = p(x|y)p(y)$$

- Bayes Rule:

$$\Rightarrow \quad p(y|x) = p(x|y)p(y)/p(x)$$

$$= \frac{p(x|y)p(y)}{\sum_c p(x|y=c)p(y=c)}$$

(Use the rule of total probability to calculate the denominator!)

Bayes classifiers

- Learn "class conditional" models
  - Estimate a probability model for each class
- Training data, D
  - Split by class, $D_c = \{ x^{(j)} : y^{(j)} = c \}$
- Estimate $p(x \mid y=c)$ using D_c
- For a discrete x, this recalculates the same table…

Features	# bad	# good
X=0	42	15
X=1	338	287
X=2	3	5

$\Rightarrow$

p(x | y=0)	p(x | y=1)
42 / 383	15 / 307
338 / 383	287 / 307
3 / 383	5 / 307

$\Rightarrow$

p(y=0|x)	p(y=1|x)
.7368	.2632
.5408	.4592
.3750	.6250

p(y)	383/690	307/690