Data Mining & Machine Learning

CS37300
Purdue University

Sep 15, 2023

Today's topics

- Generative probabilistic classification
  - Naïve Bayes classifier

Generative Models

Generative Models

- Model x given y: this approach is called **generative model**
  - Model the **class-conditional** probabilities: $P(x|y)$
  - And the **class-prior** probability: $P(y)$

Generative Models

- Model x given y: this approach is called **generative model**
  - Model the **class-conditional** probabilities: $P(x|y)$
  - And the **class-prior** probability: $P(y)$

- Once we estimate $P(x|y)$ and $P(y)$ from data
- Use Bayes rule to solve for $P(y|x)$ for predictions on test points x

- Classify test points x with label **$\mathrm{argmax}_y P(y|x)$**
- called the **maximum a posteriori** assignment (MAP)

Bayes rule for probabilistic classifier

$$P(y \mid \underline{x}) = \frac{P(\underline{x}, y)}{P(\underline{x})} = \frac{P(\underline{x} \mid y)P(y)}{P(\underline{x})}$$

$$= \frac{P(\underline{x} \mid y)P(y)}{P(\underline{x} \mid y=1)P(y=1) + P(\underline{x} \mid y=-1)P(y=-1)}$$

$$\propto P(\underline{x} \mid y)P(y)$$

Bayes rule for probabilistic classifier

$$P(y \mid \underline{x}) = \frac{P(\underline{x}, y)}{P(\underline{x})} = \frac{P(\underline{x} \mid y)P(y)}{P(\underline{x})}$$

Bayes rule

$$= \frac{P(\underline{x} \mid y)P(y)}{P(\underline{x} \mid y=1)P(y=1) + P(\underline{x} \mid y=-1)P(y=-1)}$$

$$\propto P(\underline{x} \mid y)P(y)$$

Bayes rule for probabilistic classifier

$$P(y \mid \underline{x}) = \frac{P(\underline{x}, y)}{P(\underline{x})} = \frac{P(\underline{x} \mid y)P(y)}{P(\underline{x})}$$

Bayes rule

$$= \frac{P(\underline{x} \mid y)P(y)}{P(\underline{x} \mid y=1)P(y=1) + P(\underline{x} \mid y=-1)P(y=-1)}$$

$$\propto P(\underline{x} \mid y)P(y)$$

Denominator is not important for Classification

$$\underset{y}{\operatorname{argmax}} \frac{P(x|y)P(y)}{P(x)} = \underset{y}{\operatorname{argmax}}\, P(x|y)P(y)$$

Naïve Bayes classifier

- Simple generative model
- Based on the assumption that attributes in the feature vector are **conditionally independent** given the label
- For feature vector $\underline{x} = [x_1, \ldots, x_d]^\top$

$$P(\underline{x}|y) = \prod_{j=1}^{d} P(x_j|y)$$

$$P(y \mid \underline{x}) \propto P(\underline{x} \mid y) P(y)$$

$$\propto \left(\prod_{j=1}^{d} P(x_j \mid y) \right) P(y)$$

Naïve Bayes classifier

- Simple generative model
- Based on the assumption that attributes in the feature vector are **conditionally independent** given the label
- For feature vector $\underline{x} = [x_1, \dots, x_d]^\top$

$$P(\underline{x}|y) = \prod_{j=1}^{d} P(x_j|y)$$

$$\boxed{P(y \mid \underline{x}) \propto P(\underline{x} \mid y) P(y)}$$

Bayes rule

$$\propto \left(\prod_{j=1}^{d} P(x_j \mid y) \right) P(y)$$

Naïve Bayes classifier

- Simple generative model
- Based on the assumption that attributes in the feature vector are **conditionally independent** given the label
- For feature vector $\underline{x} = [x_1, \dots, x_d]^\top$

$$P(\underline{x}|y) = \prod_{j=1}^{d} P(x_j|y)$$

$$P(y \mid \underline{x}) \propto P(\underline{x} \mid y)P(y)$$

Bayes rule

$$\propto \left(\prod_{j=1}^{d} P(x_j \mid y) \right) P(y)$$

Naïve assumption

Naïve Bayes classifier

$$
\begin{aligned}
P(BC|A,I,S,CR) &= \frac{P(A,I,S,CR|BC)P(BC)}{P(A,I,S,CR)} \\
&= \frac{P(A|BC)P(I|BC)P(S|BC)P(CR|BC)P(BC)}{P(A,I,S,CR)} \\
&\propto P(A|BC)P(I|BC)P(S|BC)P(CR|BC)P(BC)
\end{aligned}
$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Naïve Bayes classifier

$$P(BC|A,I,S,CR) = \frac{P(A,I,S,CR|BC)P(BC)}{P(A,I,S,CR)}$$
$$= \frac{P(A|BC)P(I|BC)P(S|BC)P(CR|BC)P(BC)}{P(A,I,S,CR)}$$
$$\propto P(A|BC)P(I|BC)P(S|BC)P(CR|BC)P(BC)$$

parameters = conditionals + prior

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

```
Conditionals: P(A|BC)
              P(I|BC)
              P(S|BC)
              P(CR|BC)
Prior:        P(BC)
```

Naïve Bayes classifier: Discrete vs Continuous

- If the attributes are discrete, model $P(x_j|y)$ as a **Multinomial** distribution
  - Each attribute x_j can have values in $\{1,\ldots,k\}$
  - For each possible y value and each attribute x_j we have k parameters:

    $P(x_j = a \mid y = b)$
- $P(y=1)$ is also a parameter

Naïve Bayes classifier: Discrete vs Continuous

- If the attributes are discrete, model $P(x_j|y)$ as a **Multinomial** distribution
  - Each attribute x_j can have values in {1,…,k}
  - For each possible y value and each attribute x_j we have k parameters:

    $P(x_j = a \mid y = b)$
- $P(y=1)$ is also a parameter
- Question: If y is binary, how many parameters are there?

Naïve Bayes classifier: Discrete vs Continuous

- If the attributes are discrete, model $P(x_j|y)$ as a **Multinomial** distribution
  - Each attribute x_j can have values in {1,…,k}
  - For each possible y value and each attribute x_j we have k parameters:

    $P(x_j = a \mid y = b)$
- P(y=1) is also a parameter
- Question: If y is binary, how many parameters are there?
- 2dk + 1 (or 2d(k-1)+1 if we're clever)

Naïve Bayes classifier: Discrete vs Continuous

- If the attributes are discrete, model $P(x_j|y)$ as a **Multinomial** distribution
  - Each attribute x_j can have values in {1,…,k}
  - For each possible y value and each attribute x_j we have k parameters:

    $P(x_j = a \mid y = b)$
- P(y=1) is also a parameter
- Question: If y is binary, how many parameters are there?
- 2dk + 1 (or 2d(k-1)+1 if we're clever)
- If the attributes are real-valued, typically model $P(x_j|y)$ as **Normal** distribution

Learning the Naïve Bayes Classifier

- Suppose we have a dataset of n samples $D = \{(\underline{\boldsymbol{x}}_1, y_1), (\underline{\boldsymbol{x}}_2, y_2) \ldots (\underline{\boldsymbol{x}}_n, y_n)\}$
- Estimate P(x|y)P(y) using maximum likelihood estimation:

$\mathrm{P}(\mathrm{x}|\mathrm{y}; \hat{\theta}) P(y; \hat{\theta})$ where

$$\hat{\theta} = \operatorname*{argmax}_{\theta} L(D; \theta)$$

$$L(D; \theta) = \prod_{i=1}^{n} \left(\left(\prod_{j=1}^{d} P(x_{ij} | y_i; \theta) \right) P(y_i; \theta) \right)$$

Naïve assumption (conditional independence)

Naïve Bayes Maximum Likelihood Estimator

- We want to maximize this, jointly over the set of all parameters, to find $\hat{\theta}$

$$L(D;\theta) = \left(P(Y=1)^{\sum_{i=1}^n y_i} P(Y=0)^{n-\sum_{i=1}^n y_i}\right) \prod_{j=1}^{d} \left(\prod_{i:y_i=0} P(X_j = x_{ij} | Y=0)\right) \left(\prod_{i:y_i=1} P(X_j = x_{ij} | Y=1)\right)$$

- **Solution:**

$$\hat{P}(y=1) = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Estimate of P(Y=1)

$$\hat{P}(x_j = a | y = b) = \frac{1}{n_b} \sum_{i:y_i=b} \mathbb{I}[x_{ij} = a]$$

Estimate of $P(X_j = a \mid Y = b)$

$$n_b = |\{i : y_i = b\}|$$

Computing conditionals from training examples

		X	
	Low	Medium	High
Yes	10	13	17
No	2	13	0

Y

$$P(X = \text{Low} \mid Y = \text{Yes}) = \frac{10}{(10 + 13 + 17)}$$

$$P(Y = \text{No}) = \frac{(2 + 13)}{(2 + 13 + 10 + 13 + 17)}$$

Naïve Bayes classifier: learning

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- Estimate prior P(BC) and conditional probability distributions P(A | BC), P(I | BC), P(S | BC), P(CR | BC) independently with maximum likelihood estimation

Naïve Bayes classifier: learning

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- Estimate prior P(BC) and conditional probability distributions P(A | BC), P(I | BC), P(S | BC), P(CR | BC) independently with maximum likelihood estimation

P(BC)

BC	θ
yes	$9/14$
no	$5/14$

Naïve Bayes classifier: learning

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- Estimate prior P(BC) and conditional probability distributions P(A | BC), P(I | BC), P(S | BC), P(CR | BC) independently with maximum likelihood estimation

P(BC)

BC	θ
yes	$9/14$
no	$5/14$

P(A I BC)

BC	A	θ
yes	$<= 30$	$2/9$
	$31..40$	$4/9$
	> 40	$3/9$
no	$<= 30$	$3/5$
	$31..40$	$0/5$
	> 40	$2/5$

Naïve Bayes classifier: learning

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- Estimate prior P(BC) and conditional probability distributions P(A | BC), P(I | BC), P(S | BC), P(CR | BC) independently with maximum likelihood estimation

P(BC)

BC	θ
yes	$9/14$
no	$5/14$

P(A I BC)

BC	A	θ
yes	$<= 30$	$2/9$
	31..40	$4/9$
	> 40	$3/9$
no	$<= 30$	$3/5$
	31..40	$0/5$
	> 40	$2/5$

P(I I BC)

BC	I	θ
yes	high	$2/9$
	med	$4/9$
	low	$3/9$
no	high	$2/5$
	med	$2/5$
	low	$1/5$

Naïve Bayes classifier: learning

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- Estimate prior P(BC) and conditional probability distributions P(A | BC), P(I | BC), P(S | BC), P(CR | BC) independently with maximum likelihood estimation

P(BC)

BC	θ
yes	$9/14$
no	$5/14$

P(A I BC)

BC	A	θ
yes	$<= 30$	$2/9$
	$31..40$	$4/9$
	> 40	$3/9$
no	$<= 30$	$3/5$
	$31..40$	$0/5$
	> 40	$2/5$

P(I I BC)

BC	I	θ
yes	high	$2/9$
	med	$4/9$
	low	$3/9$
no	high	$2/5$
	med	$2/5$
	low	$1/5$

P(S I BC)

BC	S	θ
yes	yes	$6/9$
	no	$3/9$
no	yes	$1/5$
	no	$4/5$

Naïve Bayes classifier: learning

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- Estimate prior P(BC) and conditional probability distributions P(A | BC), P(I | BC), P(S | BC), P(CR | BC) independently with maximum likelihood estimation

P(BC)

BC	θ
yes	$9/14$
no	$5/14$

P(A I BC)

BC	A	θ
yes	$<= 30$	$2/9$
	$31..40$	$4/9$
	> 40	$3/9$
no	$<= 30$	$3/5$
	$31..40$	$0/5$
	> 40	$2/5$

P(I I BC)

BC	I	θ
yes	high	$2/9$
	med	$4/9$
	low	$3/9$
no	high	$2/5$
	med	$2/5$
	low	$1/5$

P(S I BC)

BC	S	θ
yes	yes	$6/9$
	no	$3/9$
no	yes	$1/5$
	no	$4/5$

P(CR I BC)

BC	CR	θ
yes	exc	$3/9$
	fair	$6/9$
no	exc	$4/5$
	fair	$1/5$

Naïve Bayes classifier: prediction

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- What is the probability that a new person will buy a computer?

Naïve Bayes classifier: prediction

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no
31..40	**high**	**no**	**excellent**	**?**

- What is the probability that a new person will buy a computer?

Naïve Bayes classifier: prediction

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no
31..40	**high**	**no**	**excellent**	**?**

- What is the probability that a new person will buy a computer?

$$P(BC = yes|A = 31..40, I = high, S = no, CR = exc)$$
$$\propto P(A = 31..40|BC = yes)P(I = high|BC = yes)$$
$$P(S = no|BC = yes)P(CR = exc|BC = yes)P(BC = yes)$$

Naïve Bayes classifier: prediction

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no
31..40	**high**	**no**	**excellent**	**?**

- What is the probability that a new person will buy a computer?

$$P(BC = yes | A = 31..40, I = high, S = no, CR = exc)$$
$$\propto P(A = 31..40 | BC = yes) P(I = high | BC = yes)$$
$$P(S = no | BC = yes) P(CR = exc | BC = yes) P(BC = yes)$$

P(BC)

BC	θ
yes	$9/14$
no	$5/14$

P(A I BC)

BC	A	θ
yes	$<= 30$	$2/9$
	$31..40$	$4/9$
	> 40	$3/9$
no	$<= 30$	$3/5$
	$31..40$	$0/5$
	> 40	$2/5$

P(I I BC)

BC	I	θ
yes	high	$2/9$
	med	$4/9$
	low	$3/9$
no	high	$2/5$
	med	$2/5$
	low	$1/5$

P(S I BC)

BC	S	θ
yes	yes	$6/9$
	no	$3/9$
no	yes	$1/5$
	no	$4/5$

P(CR I BC)

BC	CR	θ
yes	exc	$3/9$
	fair	$6/9$
no	exc	$4/5$
	fair	$1/5$

Naïve Bayes classifier: prediction

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no
31..40	**high**	**no**	**excellent**	**?**

- What is the probability that a new person will buy a computer?

$$P(BC = yes|A = 31..40, I = high, S = no, CR = exc)$$
$$\propto P(A = 31..40|BC = yes)P(I = high|BC = yes)$$
$$P(S = no|BC = yes)P(CR = exc|BC = yes)P(BC = yes)$$

P(BC)

BC	θ
yes	9/14
no	5/14

P(A I BC)

BC	A	θ
	$<= 30$	2/9
yes	31..40	4/9
	> 40	3/9
	$<= 30$	3/5
no	31..40	0/5
	> 40	2/5

P(I I BC)

BC	I	θ
	high	2/9
yes	med	4/9
	low	3/9
	high	2/5
no	med	2/5
	low	1/5

P(S I BC)

BC	S	θ
yes	yes	6/9
	no	3/9
no	yes	1/5
	no	4/5

P(CR I BC)

BC	CR	θ
yes	exc	3/9
	fair	6/9
no	exc	4/5
	fair	1/5

Naïve Bayes classifier: prediction

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no
31..40	**high**	**no**	**excellent**	**?**

- What is the probability that a new person will buy a computer?

$$P(BC = yes|A = 31..40, I = high, S = no, CR = exc)$$
$$\propto P(A = 31..40|BC = yes)P(I = high|BC = yes)$$
$$P(S = no|BC = yes)P(CR = exc|BC = yes)P(BC = yes)$$
$$\propto \frac{4}{9} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14}$$

P(BC)

BC	θ
yes	9/14
no	5/14

P(A I BC)

BC	A	θ
	<= 30	2/9
yes	31..40	4/9
	> 40	3/9
	<= 30	3/5
no	31..40	0/5
	> 40	2/5

P(I I BC)

BC	I	θ
	high	2/9
yes	med	4/9
	low	3/9
	high	2/5
no	med	2/5
	low	1/5

P(S I BC)

BC	S	θ
yes	yes	6/9
	no	3/9
no	yes	1/5
	no	4/5

P(CR I BC)

BC	CR	θ
yes	exc	3/9
	fair	6/9
no	exc	4/5
	fair	1/5

Smoothing: Laplace correction

Smoothing: Laplace correction

- Zero counts are a problem
- If an attribute value does not occur in training example, we assign **zero** probability to that value

Smoothing: Laplace correction

- Zero counts are a problem
- If an attribute value does not occur in training example, we assign **zero** probability to that value
- How does that affect the value $P(x|y)P(y)$?
  - It equals 0 !!! (for both y values)
- Adjust for zero counts by “smoothing” probability estimates
- It also helps compensate for having small data set size

Smoothing: Laplace correction

X

Y

	Low	Medium	High
Yes	10	13	17
No	2	13	0

Smoothing: Laplace correction

X

Y		Low	Medium	High
	Yes	10	13	17
	No	2	13	0

P(X = High | Y = No) =

Smoothing: Laplace correction

X

Y		Low	Medium	High
	Yes	10	13	17
	No	2	13	0

$$P(\, X = \text{High} \mid Y = \text{No}\,) = \frac{0}{(2+13+0)}$$

Smoothing: Laplace correction

X

Y		Low	Medium	High
	Yes	10	13	17
	No	2	13	0

$$P(X = \text{High} \mid Y = \text{No}) = \frac{0 \quad +1}{(2+13+0)+3}$$

Smoothing: Laplace correction

X

Y		Low	Medium	High
	Yes	10	13	17
	No	2	13	0

$$P(X = \text{High} \mid Y = \text{No}) = \frac{0 + 1}{(2 + 13 + 0) + 3}$$

Adds uniform prior

Smoothing: Laplace correction

X

Y

	Low	Medium	High
Yes	10	13	17
No	2	13	0

$$P(X = High \mid Y = No) = \frac{0 + 1}{(2 + 13 + 0) + 3}$$

Adds uniform prior

Laplace correction
Numerator: ***add 1***
Denominator: ***add k***,
where k=number of possible values of X

Naïve Bayes classifier: Discrete vs Continuous

- If the attributes are real-valued, typically model $P(x_j|y)$ as a **Normal** distribution
  - Each attribute x_j is conditionally Normal
  - For each possible y value and each attribute x_j we have 2 parameters: $\mu_{(j,y)}$, $\sigma_{(j,y)}$
  - $P(x_j \mid y = b)$ is density for $N(\mu_{(j,b)}, \sigma^2_{(j,b)})$
  - Recall:

$$P(x_j|y=b) = \frac{1}{\sigma_{(j,b)}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_j-\mu_{(j,b)}}{\sigma_{(j,b)}}\right)^2}$$

- As before, denote by **p** the parameter for $P(y=1)$

Naïve Bayes Maximum Likelihood Estimator

- We want to maximize this, jointly over the set of all parameters, to find $\hat{\theta}$

$$L(D;\theta) = \left(p^{\sum_{i=1}^{n} y_i} (1-p)^{n-\sum_{i=1}^{n} y_i}\right) \prod_{j=1}^{d} \prod_{b=0}^{1} \left(\prod_{i:y_i=b} \frac{1}{\sigma_{(j,b)}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{ij}-\mu_{(j,b)}}{\sigma_{(j,b)}}\right)^2} \right)$$

- **Solution:**

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Estimate of P(Y=1)

$$\hat{\mu}_{(j,b)} = \frac{1}{n_b}\sum_{i:y_i=b} x_{ij} \quad ,$$

$$\hat{\sigma}^2_{(j,b)} = \frac{1}{n_b}\sum_{i:y_i=b} \left(x_{ij} - \hat{\mu}_{(j,b)}\right)^2$$

Estimate of $P(X_j \mid Y = b)$

$$n_b = |\{i : y_i = b\}|$$

Naïve Bayes classifier

- Simplifying (naive) assumption:
  attributes are conditionally independent given the class
- Strengths:
  - Easy to implement
  - Often performs well even when assumption is violated
- Weaknesses:
  - Dependencies among variables aren't being modeled