# Naive Bayes

Boston University CS 506 - Lance Galletti

# Conditional Probability

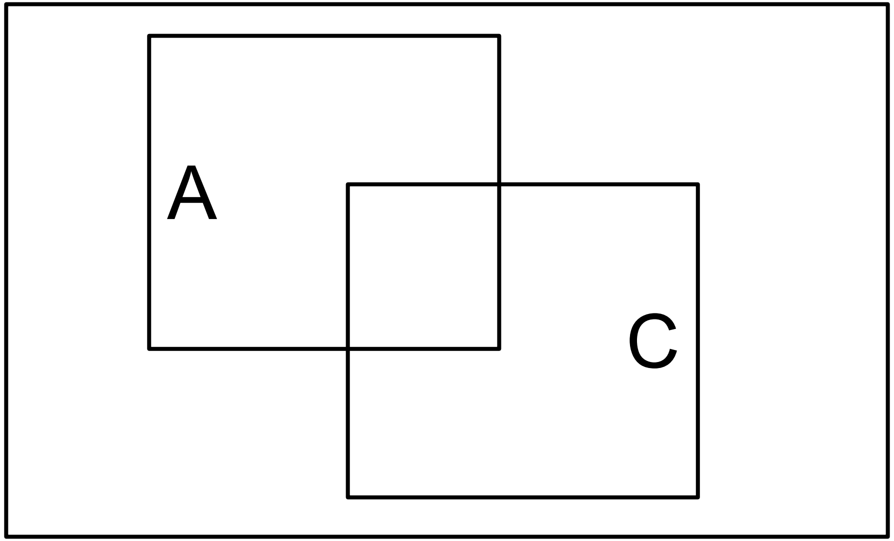
Recall

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$

# Conditional Probability

Recall

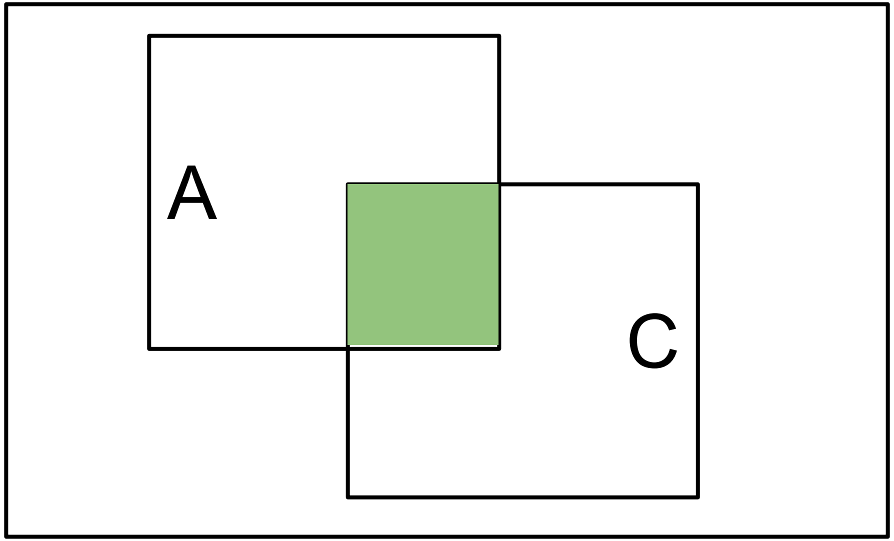
$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$



# Conditional Probability

Recall

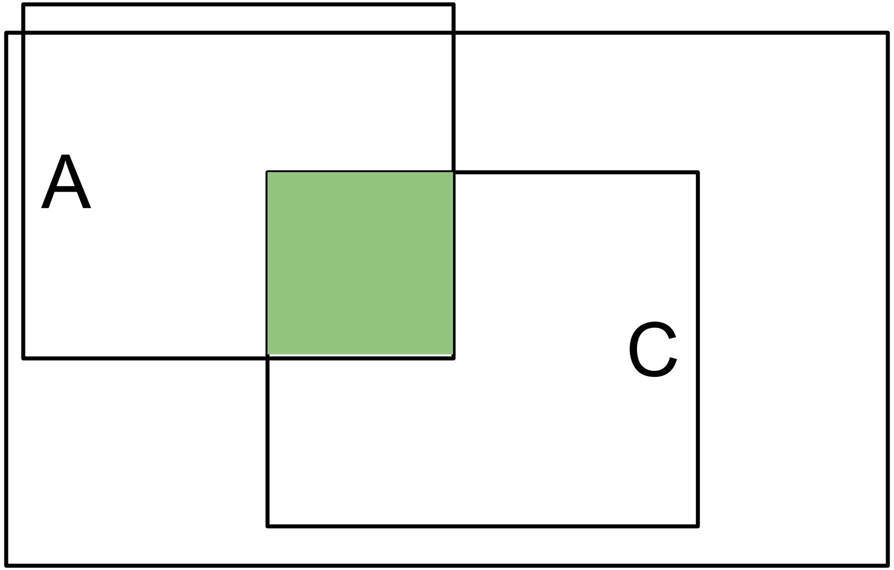
$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$



# Conditional Probability

Recall

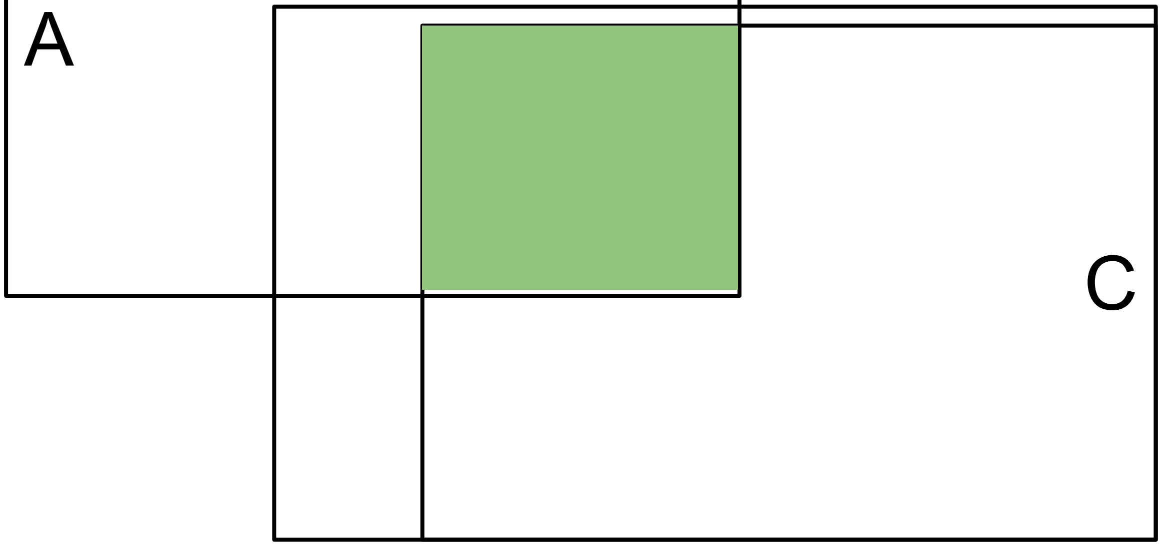
$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$



# Conditional Probability

Recall

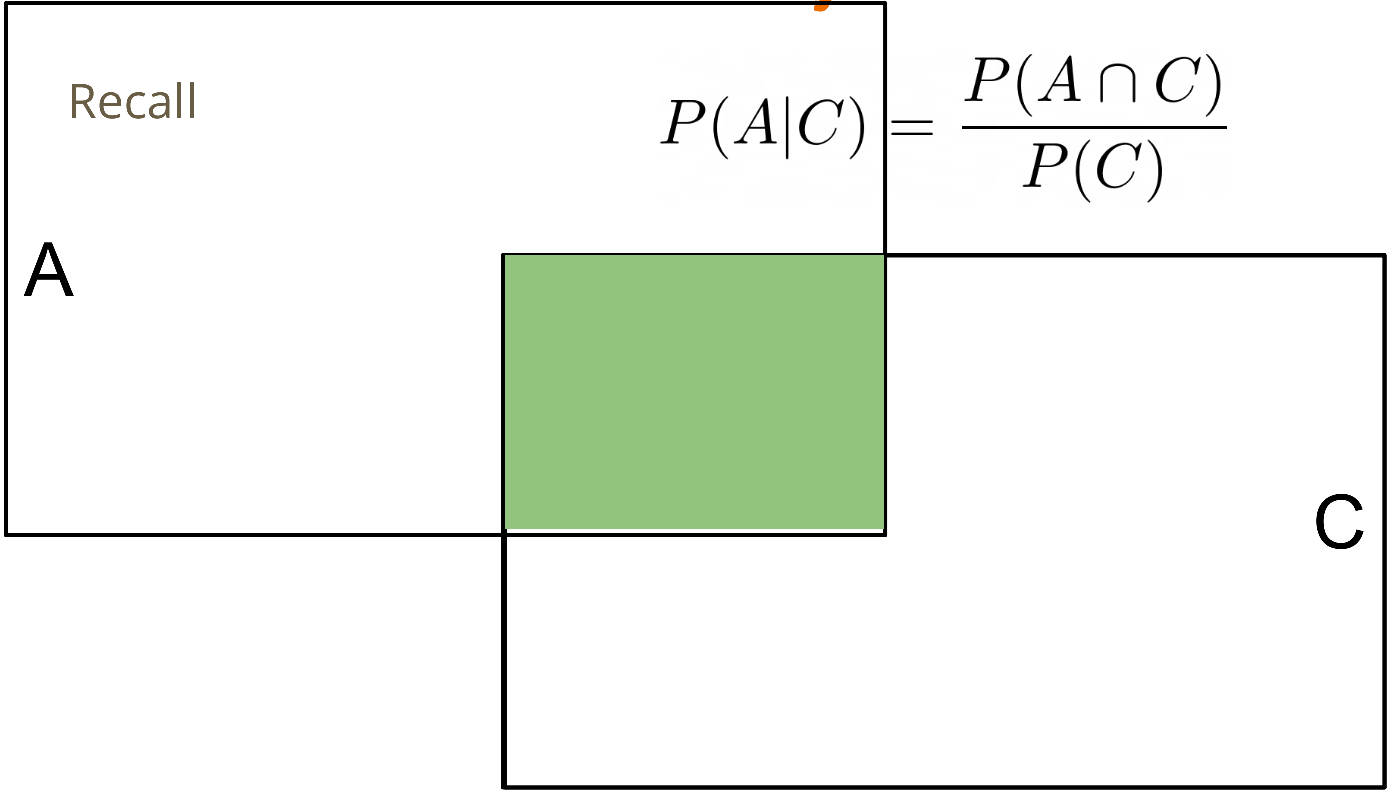
$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$



# Conditional Probability

Recall

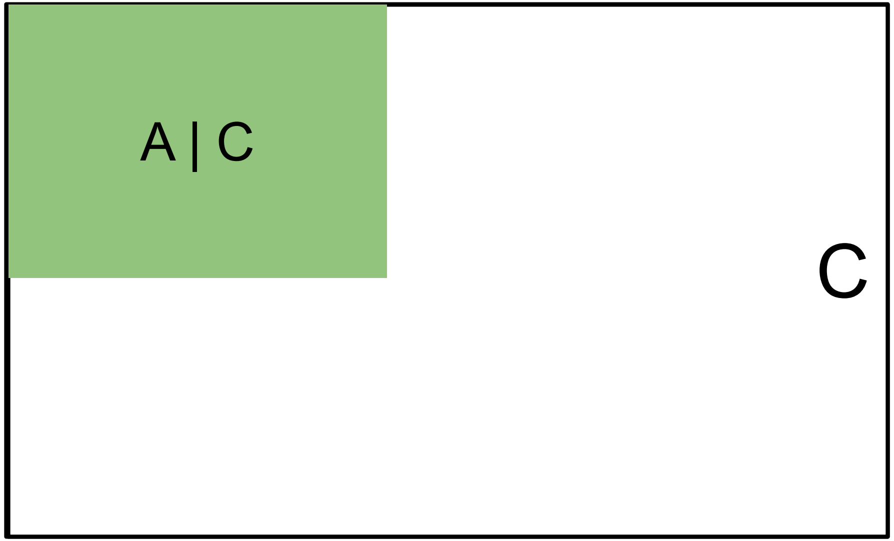
$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$



# Conditional Probability

Recall

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$



# Bayes Theorem

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)}$$

# Bayes Theorem

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)}$$

$$\frac{P(A \cap C)}{P(A)}$$

# Bayes Theorem

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)}$$

$$\frac{P(A \cap C)}{\text{P(A)}}$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$

# Example

Given:

- Meningitis causes a stiff neck 50% of the time
- Prior probability of any patient having meningitis is 1/50,000
- Prior probability of any patient having a stiff neck is 1/20

If a patient has a stiff neck, what is the probability that they have meningitis?

# Example

Given:

- Meningitis causes a stiff neck 50% of the time
- Prior probability of any patient having meningitis is 1/50,000
- Prior probability of any patient having a stiff neck is 1/20

If a patient has a stiff neck, what is the probability that they have meningitis?

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = \frac{.5 \cdot 1/50,000}{1/20} = .0002$$

# Bayesian Classifiers

Given x = some attribute values

Predict the class C that maximizes P(C | some attribute values)

# Bayesian Classifiers

Predict the class C that maximizes P(C | some attributes)

Example: **binary class {yes, no}**

To classify unseen record **(marital status = “married”, income = 100k)**

# Bayesian Classifiers

Predict the class C that maximizes P(C | some attributes)

Example: binary class {yes, no}

To classify unseen record **(marital status = "married", income = 100k)**

1. Compute **P(yes | marital status = "married" and income = 100k)**
2. Compute **P(no | marital status = "married" and income = 100k)**
3. Compare and predict the class that has the highest prob given the attribute values

# Bayesian Classifiers

How do we estimate P(C | some attributes) from the data?

$$P(C|A_1 \cap A_2 \cap \cdots \cap A_n)$$

# Bayesian Classifiers

How do we estimate P(C | some attributes) from the data?

$$P(C|A_1 \cap A_2 \cap \cdots \cap A_n) = \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n|C)P(C)}{P(A_1 \cap A_2 \cap \cdots \cap A_n)}$$

# Bayesian Classifiers

How do we estimate P(C | some attributes) from the data?

$$P(C|A_1 \cap A_2 \cap \cdots \cap A_n) = \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n|C)P(C)}{P(A_1 \cap A_2 \cap \cdots \cap A_n)}$$

Does not depend on C

# Bayesian Classifiers

How do we estimate P(C | some attributes) from the data?

$$P(C|A_1 \cap A_2 \cap \cdots \cap A_n) = \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n|C)P(C)}{P(A_1 \cap A_2 \cap \cdots \cap A_n)}$$

Maximizing $P(C|A_1A_2...A_n)$ is
equivalent to maximizing

# Bayesian Classifiers

How do we estimate P(C | some attributes) from the data?

$$P(C|A_1 \cap A_2 \cap \cdots \cap A_n) = \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n|C)P(C)}{P(A_1 \cap A_2 \cap \cdots \cap A_n)}$$

Maximizing $P(C|A_1A_2 \ldots A_n)$ is equivalent to maximizing the numerator on the right

# Bayesian Classifiers

So how to we estimate $P(A_1 A_2 \ldots A_n | C)P(C)$ from the data?

# Bayesian Classifiers

So how to we estimate $P(A_1 A_2 \ldots A_n | C)P(C)$ from the data?

$P(C)$ is easy we can just count how many instances of each class we have

But $P(A_1 A_2 \ldots A_n | C)$ is tricky because it requires knowing the **joint distribution** of the attributes...

# Bayesian Classifiers

Can we make some assumptions about the attributes in order to simplify the problem?

# Bayesian Classifiers

Assume that $A_1A_2 \ldots A_n$ are independent!

Then

$P(A_1A_2 \ldots A_n | C) = P(A_1 | C)\ P(A_2 | C) \ldots P(A_n | C)$

Can we estimate $P(A_j | C)$ from the data?

# Bayesian Classifiers

Assume that $A_1 A_2 \ldots A_n$ are independent!

Then

$P(A_1 A_2 \ldots A_n | C) = P(A_1 | C)\ P(A_2 | C) \ldots P(A_n | C)$

Can we estimate $P(A_j | C)$ from the data?

Yes! Just count the occurrence of $A_j$ for that class.

# Example

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

$P(C = Yes) = 3/10$

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

$P(\text{Marital Status} = \text{“Single”} \mid C = \text{Yes})$

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

P(Marital Status = “Single” | C = Yes)

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| No | Divorced | 90k | Yes |
| No | Single | 85k | Yes |
| No | Single | 90k | Yes |

P(Marital Status = “Single” | C = Yes)
= 2/3

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

P(Marital Status = “Married” | C = No)

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

P(Marital Status = “Married” | C = No)

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Married | 75k | No |

P(Marital Status = “Married” | C = No)
= 4/7

# Worksheet a) and b)

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

$$P(\text{Income} = 120k \mid C = \text{No})$$

# Continuous Attributes

- Binning / 2-way or multi-way split
  - Create new attribute for each bin
  - Issue is that these attributes are no longer independent
- Pdf estimation
  - Assume attribute follows a particular distribution (example: normal)
  - Use data to estimate the parameters of the distribution

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

Assume normal distribution

$P(\text{Income} = 120\text{k} \mid C = \text{No})$

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

Assume normal distribution

$P(\text{Income} = 120k \mid C = \text{No})$

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Married | 75k | No |

Assume normal distribution

P(Income = 120k | C = No)

Sample mean = 110

Sample variance = 2975

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Married | 75k | No |

Assume normal distribution

P(Income = 120k | C = No)

Sample mean = 110

Sample variance = 2975

$$P(Income = 120|No) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120-110)^2}{2(2975)}} = .0072$$

# Putting it all together

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

Test Record:

X = (Refund = No, Married, Income = 120k)

- P( X | No ) =

- P( X | Yes ) =

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

Test Record:
X = (Refund = No, Married, Income = 120k)

- P( X | No ) = P( Refund = No | No) P(Married | No) P(Income=120k | No) = 4/7 * 4/7 * .0072 = .0024

- P( X | Yes ) =

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

Test Record:
X = (Refund = No, Married, Income = 120k)

- P( X | No ) = P( Refund = No | No) P(Married | No) P(Income=120k | No) = 4/7 * 4/7 * .0072 = .0024

- P( X | Yes ) = P( Refund = No | Yes) P(Married | Yes) P(Income=120k |Yes) = 1 * 0 * 1.2 * $10^{-9}$ = 0

| Refund | Marital Status | Income | Class |
|---|---|---|---|
| Yes | Single | 125k | No |
| No | Married | 100k | No |
| No | Single | 70k | No |
| Yes | Married | 120k | No |
| No | Divorced | 90k | Yes |
| No | Married | 60k | No |
| Yes | Divorced | 220k | No |
| No | Single | 85k | Yes |
| No | Married | 75k | No |
| No | Single | 90k | Yes |

Test Record:
X = (Refund = No, Married, Income = 120k)

- P( X | No ) = P( Refund = No | No) P(Married | No) P(Income=120k | No) = 4/7 * 4/7 * .0072 = .0024

- P( X | Yes ) = P( Refund = No | Yes) P(Married | Yes) P(Income=120k |Yes) = 1 * 0 * 1.2 * $10^{-9}$ = 0

Since P(X|No)P(No) > P(X|Yes)P(Yes)
=> predict No

# Limitation

If one of the conditional probabilities is zero, the entire expression becomes zero...

Original estimate of $P(A_i \mid C) = N_{ic} / N_c$

Laplace estimate : $P(A_i \mid C) = (N_{ic} + 1) / (N_c + \text{constant})$

m-estimate : $P(A_i \mid C) = (N_{ic} + mp) / (N_c + m)$

p = prior probability
m = parameter