Naive Bayesian Classifier Example, *m*-estimate of probability

Relevant Readings: Section 6.9.1

CS495 - Machine Learning, Fall 2009

Training data for function *playTennis* [Table 3.2, Mitchell]

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it ($v =$ Yes/No)

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it ($v = \text{Yes}/\text{No}$)
  - $v = \operatorname{argmax}_{b \in \{\text{Yes}, \text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it (v = Yes/No)
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it ($v =$ Yes/No)
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try $b =$ Yes and $b =$ No and see which comes out higher

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it (v = Yes/No)
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try b = Yes and b = No and see which comes out higher
  - We can estimate each term using the data, for example:

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it (v = Yes/No)
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try b = Yes and b = No and see which comes out higher
  - We can estimate each term using the data, for example:
    - $\Pr(\text{Yes}) = 9/14$

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it (v = Yes/No)
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try b = Yes and b = No and see which comes out higher
  - We can estimate each term using the data, for example:
    - $\Pr(\text{Yes}) = 9/14$
    - $\Pr(\text{No}) = 5/14$

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it (v = Yes/No)
  - $v = \operatorname{argmax}_{b \in \{\text{Yes}, \text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \operatorname{argmax}_{b \in \{\text{Yes}, \text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try b = Yes and b = No and see which comes out higher
  - We can estimate each term using the data, for example:
    - $\Pr(\text{Yes}) = 9/14$
    - $\Pr(\text{No}) = 5/14$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{Yes}) = 2/9$

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it (v = Yes/No)
  - $v = \mathrm{argmax}_{b \in \{\mathsf{Yes},\mathsf{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \mathrm{argmax}_{b \in \{\mathsf{Yes},\mathsf{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try b = Yes and b = No and see which comes out higher
  - We can estimate each term using the data, for example:
    - $\Pr(\text{Yes}) = 9/14$
    - $\Pr(\text{No}) = 5/14$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{Yes}) = 2/9$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{No}) = 3/5$

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it (v = Yes/No)
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try b = Yes and b = No and see which comes out higher
  - We can estimate each term using the data, for example:
    - $\Pr(\text{Yes}) = 9/14$
    - $\Pr(\text{No}) = 5/14$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{Yes}) = 2/9$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{No}) = 3/5$
    - We end up with $\frac{9}{14} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} = .0053$ for Yes and $\frac{5}{14} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} = .0206$ for No.

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it ($v =$ Yes/No)
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \operatorname{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $$\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$$
    $$\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$$
    $$\cdot \Pr(\text{Humidity} = \text{High} \mid b)$$
    $$\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$$
  - So just try $b =$ Yes and $b =$ No and see which comes out higher
  - We can estimate each term using the data, for example:
    - $\Pr(\text{Yes}) = 9/14$
    - $\Pr(\text{No}) = 5/14$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{Yes}) = 2/9$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{No}) = 3/5$
    - We end up with $\frac{9}{14} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} = .0053$ for Yes and $\frac{5}{14} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} = .0206$ for No.
  - We thus predict that No is the output

Classifying a new instance

- Consider the instance: (Sunny, Cool, High, Strong)
  - We will use Naive Bayes to classify it ($v =$ Yes/No)
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b) \prod_i \Pr(a_i \mid b)$
  - $v = \mathrm{argmax}_{b \in \{\text{Yes},\text{No}\}} \Pr(b)$
    $\cdot \Pr(\text{Outlook} = \text{Sunny} \mid b)$
    $\cdot \Pr(\text{Temperature} = \text{Cool} \mid b)$
    $\cdot \Pr(\text{Humidity} = \text{High} \mid b)$
    $\cdot \Pr(\text{Wind} = \text{Strong} \mid b)$
  - So just try $b =$ Yes and $b =$ No and see which comes out higher
  - We can estimate each term using the data, for example:
    - $\Pr(\text{Yes}) = 9/14$
    - $\Pr(\text{No}) = 5/14$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{Yes}) = 2/9$
    - $\Pr(\text{Outlook} = \text{Sunny} \mid \text{No}) = 3/5$
    - We end up with $\frac{9}{14} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} = .0053$ for Yes and $\frac{5}{14} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} = .0206$ for No.
  - We thus predict that No is the output

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$
- To avoid this, we fix the following numbers p and m beforehand:

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$
- To avoid this, we fix the following numbers p and m beforehand:
  - A nonzero prior estimate p for $\Pr(A \mid B)$, and

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$
- To avoid this, we fix the following numbers p and m beforehand:
  - A nonzero prior estimate p for $\Pr(A \mid B)$, and
  - A number m that says how confident we are of our prior estimate p, as measured in number of samples

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$
- To avoid this, we fix the following numbers p and m beforehand:
  - A nonzero prior estimate p for $\Pr(A \mid B)$, and
  - A number m that says how confident we are of our prior estimate p, as measured in number of samples
- Then instead of using $\frac{n_c}{n}$ for the estimate, use $\frac{n_c + mp}{n+m}$

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$
- To avoid this, we fix the following numbers p and m beforehand:
  - A nonzero prior estimate p for $\Pr(A \mid B)$, and
  - A number m that says how confident we are of our prior estimate p, as measured in number of samples
- Then instead of using $\frac{n_c}{n}$ for the estimate, use $\frac{n_c + mp}{n+m}$
- Just think of this as adding a bunch of samples to start the whole process

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$
- To avoid this, we fix the following numbers p and m beforehand:
  - A nonzero prior estimate p for $\Pr(A \mid B)$, and
  - A number m that says how confident we are of our prior estimate p, as measured in number of samples
- Then instead of using $\frac{n_c}{n}$ for the estimate, use $\frac{n_c + mp}{n+m}$
- Just think of this as adding a bunch of samples to start the whole process
- If we don't have any knowledge of p, assume the attribute is uniformly distributed over all possible values

m-estimate of probability

- Note that we estimated conditional probabilities $\Pr(A \mid B)$ by $\frac{n_c}{n}$ where n_c is the number of times $A \wedge B$ happened and n is the number of times B happened in the training data
- This can cause trouble if $n_c = 0$
- To avoid this, we fix the following numbers p and m beforehand:
  - A nonzero prior estimate p for $\Pr(A \mid B)$, and
  - A number m that says how confident we are of our prior estimate p, as measured in number of samples
- Then instead of using $\frac{n_c}{n}$ for the estimate, use $\frac{n_c + mp}{n + m}$
- Just think of this as adding a bunch of samples to start the whole process
- If we don't have any knowledge of p, assume the attribute is uniformly distributed over all possible values