CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019
• Office hours **TODAY** 1-3pm

• **Midterm 1 Feb 27** (in lab)
  – Study guide posted
  – You may use a 1-page (front and back) “cheat-sheet”, but no other resources
  – Advice for studying: make cheat sheet, go through notes/slides, redo practice problems, read book for conceptual ideas, make sure study guide makes sense

• **Lab 4 due March 8** (Friday before spring break)
  – Please start soon, but don’t neglect midterm studying
Outline for February 22

• Recap Naïve Bayes for discrete features
  – Finish Handout 4

• Examples of Bayes Rule
  – Handout 5

• Naïve Bayes for continuous features
Outline for February 22

• Recap Naïve Bayes for discrete features
  – Finish Handout 4

• Examples of Bayes Rule
  – Handout 5

• Naïve Bayes for continuous features
Handout 4

\[ \Theta_1 = \frac{4}{9}, \quad \Theta_2 = \frac{5}{9} \]

<table>
<thead>
<tr>
<th>( f_1 )</th>
<th>( y = 1 )</th>
<th>( y = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos</td>
<td>( \frac{2}{5} )</td>
<td>( \frac{2}{3} )</td>
</tr>
<tr>
<td>neg</td>
<td>( \frac{3}{5} )</td>
<td>( \frac{1}{3} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( f_2 )</th>
<th>( y = 1 )</th>
<th>( y = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>neg</td>
<td>( \frac{4}{5} )</td>
<td>( \frac{1}{2} )</td>
</tr>
</tbody>
</table>

LaPlace

\[ o + 1 = 2 \]

\[ 1 + f_2 = 2 \]

\[ \hat{X}_{\text{test}} = [\text{neg}, \text{pos}] \]

\[ p(y = 1 | \hat{X}_{\text{test}}) \propto \Theta_1 \cdot \Theta_1 \cdot \text{neg, neg} \cdot \Theta_2 \cdot \text{pos, 1} \]

\[ p(y = k) \prod_{j=1}^{p} p(x_j | y = k) \]

\[ = \frac{4}{5} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} \]

\[ = \frac{4}{75} \]
\[ P(y = 2 | \hat{x}_{\text{test}}) \propto \Theta_2 \cdot \Theta_1, \text{neg}_2 \cdot \Theta_2, \text{pos}_2 \]

\[ = \frac{5}{9} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{54} \]

\[ \hat{y} = 2 \rightarrow p(y = 2 | \hat{x}_{\text{test}}) = \frac{\frac{5}{54}}{\frac{4}{75} + \frac{5}{54}} \approx 63\% \]
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

\[ p(y = k | \mathbf{x}) = \frac{p(y = k)p(\mathbf{x} | y = k)}{p(\mathbf{x})} \]

• What do each of these 4 terms mean?
• Why is the LHS a natural quantity to estimate? When have we seen it before?
• (open-ended) how could we extend Naïve Bayes to continuous features?
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

\[ p(y = k | x) = \frac{p(y = k)p(x | y = k)}{p(x)} \]

• **Evidence**: this is the data (features) we actually observe, which we think will help us predict the outcome we’re interested in
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

\[
p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}
\]

• **Prior**: without seeing any evidence (data), what is our prior believe about each outcome (intuition: what is the outcome in the population as a whole?)
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

\[ p(y = k|x) = \frac{p(y = k)p(x|y = k)}{p(x)} \]

• **Posterior**: this is the quantity we are actually interested in. *Given* the evidence, what is the probability of the outcome?
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

\[ p(y = k | \mathbf{x}) = \frac{p(y = k)p(\mathbf{x} | y = k)}{p(\mathbf{x})} \]

• **Likelihood**: given an outcome, what is the probability of observing this set of features?
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

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

• Why is the LHS a natural quantity to estimate? When have we seen it before?
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

\[ p(y = k | \mathbf{x}) = \frac{p(y = k) p(\mathbf{x} | y = k)}{p(\mathbf{x})} \]

• Why is the LHS a natural quantity to estimate? When have we seen it before?

We are given a new data point and want to classify/predict what class it belongs to. Probability is important for uncertainty! Seen before: logistic regression, even k-nearest neighbors.
Outline for February 22

• Recap Naïve Bayes for discrete features
  – Finish Handout 4

• Examples of Bayes Rule
  – Handout 5

• Naïve Bayes for continuous features
Example (email)

\[ P(\text{spam} | \text{words}) = \frac{P(\text{spam}) P(\text{words} | \text{spam})}{P(\text{words})} \]

\[ P(\text{words}) = \sum_{\text{vals in spam status}} P(\text{spam status, words}) \]

\[ = P(\text{spam, words}) + P(\overline{\text{spam, words}}) \]
\[
P(s, \text{words}) + P(\overline{s}, \text{words})
\]

\[
P(\text{spam})P(\text{words} | \text{spam}) + P(\overline{\text{spam}})P(\text{words} | \overline{\text{spam}})
\]

\[
\{ \text{normalizer} \}
\]

\[
P(\text{spam} | \text{words}) + P(\overline{\text{spam}} | \text{words}) = 1
\]
Handout 5

\[ P(\text{disease}) = P(D) = \frac{1}{100} \quad \text{Prior} \]
\[ P(\text{healthy}) = P(H) = \frac{99}{100} \]

Test with 90% accuracy.

\[ P(\text{pos} | D) = \frac{9}{10} \]
\[ P(\text{neg} | H) = \frac{9}{10} \]
Q: \[ P(D/\text{pos}) = \frac{p(D)p(\text{pos}|D)}{p(\text{pos})} \]

Expand

\[
= \frac{P(D)P(\text{pos}|D)}{P(D)P(\text{pos}|D) + P(H)P(\text{pos}|H)}
\]

\[
= \frac{\frac{1}{100} \cdot \frac{9}{10}}{\frac{1}{100} \cdot \frac{9}{10} + \frac{99}{100} \cdot \frac{1}{10}}
\]

\[
= \frac{9}{9 + 99} = \frac{9}{108} = \frac{1}{12} \approx 8.33\%.
\]
Outline for February 22

• Recap Naïve Bayes for discrete features
  – Finish Handout 4

• Examples of Bayes Rule
  – Handout 5

• Naïve Bayes for continuous features
Informal Quiz: discuss with a partner

• Identify the evidence, prior, posterior, and likelihood in the equation below

\[ p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})} \]

• (open-ended) how could we extend Naïve Bayes to continuous features?
Discriminative vs. Generative

- **Regression**: discriminative model $\rightarrow$ finds decision boundary
- **Naïve Bayes**: generative model $\rightarrow$ estimates probability distributions

Figure: Ameet Soni
Example with $K=2$, $p=1$

**FIGURE 4.4.** Left: Two one-dimensional normal density functions are shown. The dashed vertical line represents the Bayes decision boundary. Right: 20 observations were drawn from each of the two classes, and are shown as histograms.
FIGURE 4.6. An example with three classes. The observations from each class are drawn from a multivariate Gaussian distribution with $p = 2$, with a class-specific mean vector and a common covariance matrix. Left: Ellipses that contain 95% of the probability for each of the three classes are shown. The dashed lines are the Bayes decision boundaries. Right: 20 observations were generated from each class, and the corresponding LDA decision boundaries are indicated using solid black lines. The Bayes decision boundaries are once again shown as dashed lines.