Probabilistic ML models
Part 1

Recommended text:
Pattern Recognition and Machine Learning
By Christopher Bishop
http://research.microsoft.com/~cmbishop/PRML/

These slides: mostly Ch 1&2
Other relevant Chs: 8 (available online), 9

Contents

• Part 1
  – Introduction to probability theory
  – Baye’s theorem
  – Likelihood functions in curve fitting
  – Probability distributions

• Part 2
  – Graphical models
  – Mixture Models

• Part 3
  – An example of some of my current work using probabilistic models for biological data
Introduction to Probability

- Experiments, Outcomes, Events and Sample Spaces
- What is probability?
- Basic Rules of Probability
- Probabilities of Compound Events

MT2004

Olivier GIMENEZ

Telephone: 01334 461827
E-mail: olivier@mcs.st-and.ac.uk
Website: http://www.creem.st-and.ac.uk/olivier/OGimenez.html
1. Probability

Probability is a branch of mathematics
Deals with quantifying and modelling the uncertainty relating to random experiments
But what's a random experiment?
Any process with a number of possible outcomes, but the occurrence of any outcome is not known in advance
Think of *tossing a coin (head/tail) or rolling a die (1,...,6)* e.g.

1.1 Definitions

The **sample space** $\Omega$ is the set of all possible outcomes

A **sample point** $\omega$ is a possible outcome i.e. a point in $\Omega$

An **event**, e.g. $A$, is a set of possible outcomes satisfying a given condition (so $A$ is a set of sample points)

The **null event** $\emptyset$ contains no sample point - **impossible event**
1.1 Definitions

Example 1: Toss a coin
\(\Omega = \{\text{head, tail}\} = \{H, T\}\)
Let A be event 'a head is obtained', then A = \{H\}

Example 1': Toss 2 coins
\(\Omega = \{TT, TH, HT, HH\}\)
Let B be event 'at least one H is obtained', then B = \{TH, HT, HH\}

Example 2: Roll a die and note the number shown
\(\Omega = \{1, 2, 3, 4, 5, 6\}\)
Let A be event 'the number shown is even', then A = \{2, 4, 6\}
Let B be event 'the number shown is \(\leq 4\)', then B = \{1, 2, 3, 4\}

*When studying a random experiment, it's important to spend time to define both the sample space and the events*
1.1 Definitions

Let A and B be any two events in sample space \( \Omega \), then:

\[ A \cup B = \text{union of } A \text{ and } B = \text{the set of all outcomes in } A \text{ or } B \]
(i.e. only A, only B in both A and B)

Example 2 (cont’d): Roll a die

A event 'the number shown is even', i.e. \( A = \{2,4,6\} \)
B event 'the number shown is \( \leq 4 \)', i.e. \( B = \{1,2,3,4\} \)
\[ A \cup B = \{2,4,6\} \cup \{1,2,3,4\} = \{1,2,3,4,6\} \]

---

1.1 Definitions

Let A and B be any two events in sample space \( \Omega \), then:

\[ A \cap B = \text{intersection of } A \text{ and } B = \text{the set of all outcomes in } A \text{ and } B \]

--- joint probability --- denoted as \( P(A,B) \) and verbalized as 'probability of A and B'.

Example 2 (cont’d): Roll a die

A event 'the number shown is even', i.e. \( A = \{2,4,6\} \)
B event 'the number shown is \( \leq 4 \)', i.e. \( B = \{1,2,3,4\} \)
\[ A \cap B = \{2,4,6\} \cap \{1,2,3,4\} = \{2,4\} \]
1.1 Definitions

Let A be any event in sample space Ω, then:

\[ A^c = \text{the set of points that do not occur in } A, \text{ but do occur in } \Omega \]

Example 2 (cont'd): Roll a die

A event 'the number shown is even', i.e. \( A = \{2,4,6\} \)

\( A^c = \{1,3,5\} = \text{event 'number shown is odd'} \)
1.1 Definitions

Let $A$ and $B$ be any two events in sample space $\Omega$, then:

A and $B$ are **mutually exclusive** or **disjoint** if $A \cap B = \emptyset$

![Venn diagram of mutually exclusive events]

1.2 Axioms of Probability

Let $\Omega$ be a sample space. A probability $Pr$ is a function which assigns a real number $Pr(A)$ to each event $A$ such that:

1) $0 \leq Pr(A) \leq 1$ for any event $A \subseteq \Omega$

2) $Pr(\Omega) = 1$ (honesty condition)

3) If $A_1, \ldots, A_n$ are a finite sequence of disjoint events,

$$Pr(A_1 \cup A_2 \ldots \cup A_n) = \sum Pr(A_i)$$

i.e. the probability that either of the events $A_i$ happens is the sum of the probabilities that each happens
1.2 Axioms of Probability

Example 3:

Let $\Omega$ be \{1,2,...,n\}. Define $#A$ to be the number of elements of $A$, e.g. if $A=\{1,2\}$ then $#A = 2$.

The function $#$ is called the \textit{cardinality function}.

Let $Pr(A) = #A/n$, the number of elements in $A$, divided by the total number of elements in $\Omega$.

Here, the function $Pr$ is called \textit{the uniform probability distribution on} $\Omega$.

Let us see why this function satisfies the axioms of probability.

\begin{itemize}
  \item \textbf{1. The number of elements in any subset $A$ of $\Omega$ is at least zero ($#A = 0$), and at most $n$ ($#A = n$), so $0/n \leq Pr(A) \leq n/n$.}
  \item \textbf{2. $Pr(\Omega) = #\Omega/n = n/n = 1$.}
  \item \textbf{3. If $A$ and $B$ are disjoint, then the number of elements in the union $A \cup B$ is the number of elements in $A$ plus the number of elements in $B$, i.e. $#(A \cup B) = #A + #B$. Therefore,}
  \begin{align*}
    P(A \cup B) &= #(A \cup B)/n \\
    &= (#A + #B)/n \\
    &= #A/n + #B/n \\
    &= Pr(A) + Pr(B).
  \end{align*}
\end{itemize}
1.3 Conditional Probability

Let $A$ and $B$ be any two events in sample space $\Omega$, such that $\Pr(B) > 0$ (i.e. $\Pr(B) \neq 0$).

Then the conditional probability of $A$, given that event $B$ has already occurred is denoted by $\Pr(A|B)$ and is defined by:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Example 4:

Suppose that we randomly choose a family from the set of all families with 2 children. Then, $\Omega = \{(g,g),(g,b),(b,g),(b,b)\}$ (b=boy, g=girl) where we assume that each event is equally likely.

Given the family has a boy, what is the probability both children are boys?
1.3 Conditional Probability

Example 4 (cont'd):

Let A be event 'both children are boys' and B event 'family has a boy'; we have to calculate \( \Pr(A|B) \).

\[ A = \{(b,b)\}, \quad B=\{(g,b),(b,g),(b,b)\} \quad \text{and} \quad A \cap B = \{(b,b)\}. \]

Since each outcome is equally likely, we can use the uniform probability distribution on \( \Omega \), then

\[ \Pr(B) = \frac{3}{4} \quad \text{and} \quad \Pr(A \cap B) = \frac{1}{4} \quad \text{and therefore} \]

\[ \Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \]

\[ = \frac{1/4}{3/4} \]

\[ = \frac{1}{3} \]

1.4 Multiplication Rule

Let A and B be any two events in sample space \( \Omega \).

Then, by rearranging the definition of conditional probability,

\[ \Pr(A \cap B) = \Pr(A|B) \Pr(B) \]

This can be extended to any \( n \) events. Let A, B and C be events, then,

\[ \Pr(A \cap B \cap C) = \Pr(A | B \cap C) \Pr(B \cap C) \]

\[ = \Pr(A | B \cap C) \Pr(B | C) \Pr(C) \]

And so on to any number \( n \) of events \( A_1, A_2, \ldots, A_n \)

\[ \Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = \Pr(A_1) \Pr(A_2|A_1) \Pr(A_3|A_1 \cap A_2) \ldots \Pr(A_n|A_1 \cap A_2 \ldots \cap A_{n-1}) \]
1.5 Law of Total Probability

A set of disjoint (or mutually exclusive) events $A_1, A_2, ..., A_n$ on sample space $\Omega$ such that $\Omega = \bigcup A_n$, are said to be a partition of $\Omega$.

Given a partition $A_1, A_2, ..., A_n$ on sample space $\Omega$, the Law of Total Probability states that:

$$Pr(B) = \sum Pr(B \mid A_i) Pr(A_i)$$

1.6 Bayes Theorem

Let $A_1, A_2, ..., A_n$ be a partition of $\Omega$ with $Pr(A_i) > 0$ for $i = 1, ..., n$. Let $B$ be an event, such that $Pr(B) > 0$.

Then, the Bayes theorem states that:

$$Pr(A_i \mid B) = \frac{Pr(B \mid A_i) Pr(A_i)}{Pr(B)} = \frac{Pr(B \mid A_i) Pr(A_i)}{\sum_{i=1}^{n} Pr(B \mid A_i) Pr(A_i)}$$

Established by British cleric Thomas Bayes in his 1764 posthumously published masterwork, "An Essay Toward Solving a Problem in the Doctrine of Chances". See module MT4531.
1.6 Bayes Theorem

Example 5: Drivers in the age-range 18-21 can be classified into 4 categories for car insurance purposes:

<table>
<thead>
<tr>
<th>Category</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of population in this category</td>
<td>20</td>
<td>40</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>Pr(no accidents in a year)</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

What is the probability that a randomly chosen driver came from category 3, given that he had no accident in the year?

Using Bayes Theorem, we have:

\[
P(B_3|A) = \frac{Pr(A|B_3)Pr(B_3)}{Pr(A)}
\]

Recall the data:

<table>
<thead>
<tr>
<th>Category</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of population in this category</td>
<td>20</td>
<td>40</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>Pr(no accidents in a year)</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

So, from the data, we know that \(Pr(B_3) = 0.25\) and \(Pr(A|B_3) = 0.4\)
1.6 Bayes Theorem

Example 5 (cont'd): What is the probability that a randomly chosen driver came from category 3, given that he had no accidents in the year?

Using the Law of Total Probability, we have:

\[ \Pr(A) = \sum \Pr(A \mid B_i) \Pr(B_i) \]

\[ = 0.8 \times 0.2 + 0.6 \times 0.4 + 0.4 \times 0.25 + 0.2 \times 0.15 \]

\[ = 0.53 \]

So, substituting into the general equation, we get:

\[ \Pr(B_3 \mid A) = \frac{\Pr(A \mid B_3) \Pr(B_3)}{\Pr(A)} = \frac{0.25 \times 0.4}{0.53} = 0.1887 \]

1.7 Independence

Let A and B be any two events in sample space \( \Omega \).

Then, A and B are said to be independent if,

\[ \Pr(A \cap B) = \Pr(A) \Pr(B) \]

This implies that

\[ \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \Pr(A) \]

That is, knowing whether or not B occurs gives no information about the occurrence of A.

Example of disjoint events: tossing a coin
Example of dependent events: extracting cards from a deck of cards
1.7 Independence

Note: If events A and B are disjoint, this does NOT imply that A and B are independent.

Suppose that we toss a coin. Let A event 'obtain a head' and B 'obtain a tail', then A and B are disjoint (you can't have H and T at the same time).

Now, we have \( \Pr(A) = \Pr(B) = 1/2 \).

However the probability that a head is obtained, given that we have obtained a tail with the coin toss is 0, i.e. \( \Pr(A|B) = 0 \).

Thus, we have A and B disjoint but A and B are not independent since \( \Pr(A|B) \neq \Pr(A) \).

Expectations

The expectation of \( f(x) \) over a discrete distribution is

\[
E[f] = \sum_x p(x)f(x)
\]

For continuous variables, the probability density is

\[
E[f] = \int p(x)f(x)dx
\]

The conditional expectation with respect to a conditional distribution is

\[
E_x[f|y] = \sum_x p(x|y)f(x)
\]

where \( E[f|y] \) denotes the expectation of \( f(x) \) given \( y \).
Variance and Covariance

The variance of a variable $x$ is

$$\text{var}(x) = \text{E}[x^2] - \text{E}[x]^2$$

For two random variables $x$ and $y$, the covariance is

$$\text{cov}[x,y] = \text{E}[xy] - \text{E}[x]\text{E}[y]$$

Covariance expresses the extent to which the two variables $x$ and $y$ vary together.

Question: What happens if $x$ and $y$ are independent?

Think of coin tossing ...

Curve fitting

Given: inputs $X = [x_1, x_2, \ldots x_N]$ and outputs $t = [t_1, t_2, \ldots t_N]$.

Assumption: Given $x$, the corresponding value of $t$ has a Gaussian distribution with a mean of $y(x, w)$.

So, $p(t|x, w, \beta) = N(t|y(x, w), \beta^{-1})$ where $\beta^{-1}$ is the precision, the variance.

Use training data $(x, t)$ and maximum likelihood estimate to find $w$ and $\beta$. 
Likelihood function

The likelihood function is

\[ p(t|x, w, \beta) = \prod_{i=1}^{n} N(t_i|y(x_i, w), \beta^{-1}) \]

Easier to maximize the log likelihood ...

\[ \ln p(t|x, w, \beta) = \sum_{i=1}^{n} \ln \frac{1}{\sqrt{2\beta^{-1}}} \exp -\frac{1}{2\beta^{-1}}(t - y(x_i, w))^2 \]

\[ = -\frac{\beta}{2} \sum_{i=1}^{n} (t - y(x_i, w))^2 + \frac{n}{2} (\ln \beta - \ln 2\pi) \]

Maximizing the log likelihood is equivalent to minimizing the sum-of-squares error function.

A more Bayesian approach

Introduce prior distributions over \( w \) like so:

\[ p(w|\alpha) = N(w|0, \alpha^{-1} I) = \left( \frac{\alpha}{2\pi} \right)^{(M+1)/2} \exp \left\{ -\frac{\alpha}{2} w^T w \right\} \]

\( w \) is an \( M+1 \) vector for a polynomial of order \( M+1 \)

Hyperparameter \( \alpha \) controls the distribution of the model parameter \( w \)

Using Baye’s theorem, the posterior distribution of \( w \) is proportional to the product of the prior distribution and the likelihood function

\[ p(w|x, t, \alpha, \beta) \propto p(t|x, w, \beta)p(w|\alpha) \]
The predictive distribution resulting from a Bayesian treatment of polynomial curve fitting using an M=0 polynomial, with the fixed parameters $\alpha=0.0005$ and $\beta=11.1$ (corresponding to the known noise variance), in which the red curve denotes the mean of the predictive distribution and the red region corresponds to $\pm 1$ standard deviation around the mean.

Classwork/Homework 4

From Bishop’s book (Q6 and Q10 and Q11)

6. Show that the covariance of two independent variables $x$ and $y$ is zero.

10. Suppose that two variables $x$ and $y$ are statistically independent. Show that the mean and variance of their sum satisfies

$$E[x+y] = E[x] + E[y]$$
$$\text{var}(x+y) = \text{var}(x) + \text{var}(y)$$

11. By setting the derivatives of

$$\ln p(x|\mu, \sigma^2) = \ln \prod_{i=1}^{n} N(x_i|\mu, \sigma^2)$$

with respect to $\mu$ and $\sigma^2$ equal to zero, find the maximum likelihood estimate for $\mu$ and $\sigma^2$. 

\[\text{Beta} (\mu | a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1}\]
Classwork/Homework 4

From Bishop's book - Ch1 Q 17 ... for you math geeks only!!

17. The gamma function is defined by

\[ \Gamma(x) = \int_0^\infty \mu^{x-1} e^{-\mu} \, d\mu \]

Using integration by parts, prove the relation \( \Gamma(x+1) = x \Gamma(x) \). Show also that \( \Gamma(1) = 1 \) and hence that \( \Gamma(x+1) = x! \) when \( x \) is an integer.

Probability Distributions (ch 2)

- **Binary** variables
  - Bernoulli, Binomial, Beta distributions

- **Multinomial** variables
  - The Dirichlet distribution

- The Gaussian distribution

- Non parametric methods, frequentist approaches
**Binary variables**

Eg. Coin tossing - Prob. of getting heads is $\mu$

Bernoulli distribution: $\text{Bern}(x|\mu) = \mu^x(1 - \mu)^{1-x}$

Expectation and variance of Bernoulli:

$$E[x] = \mu, \quad \text{var}[x] = \mu(1 - \mu)$$

Tossing a coin multiple times … the distribution of the number $m$ of heads given $N$ tosses is

$$\text{Bin}(m|N, \mu) = \binom{N}{m}\mu^m(1 - \mu)^{N-m} \text{ where } \binom{N}{m} = \frac{N!}{(N-m)!m!}$$

This is the Binomial distribution.

---

**The beta distribution**

The maximum likelihood for $\mu$ can be determined by experimental observation … $\mu = m/N$

Bayesian treatment: introduce a prior distribution over the $\mu$ parameter

$$\text{Beta}(\mu|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1 - \mu)^{b-1}$$

where $\Gamma(x)$ is the gamma function defined earlier …

$$\Gamma(x) = \int_0^\infty \mu^{x-1}e^{-\mu}d\mu$$

This is the Beta distribution and $a$ and $b$ are referred to as the hyperparameters.
**Multinomial variables**

Now consider variables that can take one of \( K \) possible values. Denote the probability of getting the \( k \)th state for an experiment by \( \mu_k \).

The distribution of \( x \) is given by

\[
p(x|\mu) = \prod_{k=1}^{K} \mu_k^{x_k}
\]

Consider a series of \( N \) experiments. What is the outcome now? (Earlier in binary situation, we had \([0,1,0,0,1,1,\ldots]\) ... what now?)

The likelihood function for a dataset \( D \) of \( N \) experiments is

\[
p(D|\mu) = \prod_{k=1}^{K} \mu_k^{m_k}
\]

where \( m_k \) is a count of the number of \( k \)th states.

**The Dirichlet distribution**

Let’s use priors again!!! Now introduce a family of prior distributions for the parameters \( \mu_k \) as follows:

\[
p(\mu|\alpha) \propto \prod_{k=1}^{K} \mu_k^{\alpha_k-1}
\]

where \( 0 \leq \mu_k \leq 1, \sum_k \mu_k = 1 \)

\[\alpha = [\alpha_1, \ldots, \alpha_K]\]

The normalized form of this is the **Dirichlet distribution**:

\[
\text{Dir}(\mu|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} \quad \alpha_0 = \sum_{k=1}^{K} \alpha_k
\]
The Gaussian Probability Distribution Function

- The Gaussian probability distribution is perhaps the most used distribution in all of science. Sometimes it is called the “bell shaped curve” or normal distribution.
- Unlike the binomial and Poisson distribution, the Gaussian is a continuous distribution:
  \[ p(y) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \]
- The integral for arbitrary \( a \) and \( b \) cannot be evaluated analytically. The value of the integral has to be looked up in a table.
- The total area under the curve is normalized to one by the \( \sqrt{2\pi} \) factor.
  \[ P(-\infty < y < \infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy = 1 \]
- Only 5% of area outside 2\( \sigma \) from the mean.
- 95% of area within 2\( \sigma \).
- It is very unlikely (< 0.3%) that a measurement taken at random from a Gaussian pdf will be more than \( \pm 3 \sigma \) from the true mean of the distribution.

Karl Friedrich Gauss 1777-1855
For a binomial distribution:

- The mean number of heads is given by $\mu = Np = 5000$.
- The standard deviation is $\sigma = \sqrt{Np(1-p)} = 50$.

The probability to be within $\mu \pm \sigma$ for this binomial distribution is:

For a Gaussian distribution:

Both distributions give about the same probability!

Relationship between Gaussian and Binomial distribution

- The Gaussian distribution can be derived from the binomial (or Poisson) assuming:
  - $p$ is finite
  - $N$ is very large
  - we have a continuous variable rather than a discrete variable
- An example illustrating the small difference between the two distributions under the above conditions:
  - Consider tossing a coin 10,000 times.
    - $p(\text{head}) = 0.5$
    - $N = 10,000$

For a binomial distribution:

mean number of heads $= \mu = Np = 5000$
standard deviation $\sigma = \sqrt{Np(1-p)} = 50$

The probability to be within $\mu \pm \sigma$ for this binomial distribution is:

For a Gaussian distribution:

$P(\mu - \sigma < y < \mu + \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \int_a^b e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = 0.68$

Both distributions give about the same probability!

Why is the Gaussian pdf so applicable? ⇒ Central Limit Theorem

A crude statement of the Central Limit Theorem:

Things that are the result of the addition of lots of small effects tend to become Gaussian.

A more exact statement:

Let $Y_1, Y_2, \ldots, Y_n$ be an infinite sequence of independent random variables each with the same probability distribution.

Suppose that the mean ($\mu$) and variance ($\sigma^2$) of this distribution are both finite.

For any numbers $a$ and $b$:

$$\lim_{n \to \infty} P\left[ a < \frac{Y_1 + Y_2 + \ldots + Y_n - n\mu}{\sigma \sqrt{n}} < b \right] = \frac{1}{\sqrt{2\pi} a} \int_a^b e^{-\frac{1}{2} y^2} dy$$

The C.L.T. tells us that under a wide range of circumstances the probability distribution that describes the sum of random variables tends towards a Gaussian distribution as the number of terms in the sum $\to \infty$.

How close to $\infty$ does $n$ have to be??

Alternatively we can write the CLT in a different form:

$$\lim_{n \to \infty} P\left[ a < \frac{\sum Y - \mu}{\sigma \sqrt{n}} < b \right] = \lim_{n \to \infty} P\left[ a < \frac{\bar{Y} - \mu}{\sigma^2} < b \right] = \frac{1}{\sqrt{2\pi} a} \int_a^b e^{-\frac{1}{2} y^2} dy$$

Actually, the $Y$'s can be from different pdf's!
Nonparametric methods

So far the probability distributions were governed by a small number of parameters that can be induced from the data. This is known as the parametric approach to density modeling.

What about nonparametric approaches?
Requires assumptions about the form of distribution
- histogram method
  Build a histogram of the data to get a density estimation
- kernel functions
  Use a kernel function to obtain density estimates using hypercubes
- nearest neighbor methods
  Useful for local density estimation

More Classwork/Homework 4

From Bishop’s book, Ch 2

2.1 Verify that the Bernoulli distribution $\text{Bern}(x|\mu) = \mu \cdot (1 - \mu)$ satisfies the following properties:

$E[x] = \mu$, $\text{var}[x] = \mu \cdot (1 - \mu)$, $\sum_{x=0}^{1} p(x|\mu) = 1$

Some hypothesis-test questions using the Gaussian distribution.

Defective Products: A large mail-order company has placed an order for 5,000 electric can openers with a supplier on the condition that no more than 2% of the devices will be defective. To check the shipment, the company tests a random sample of 400 of the can openers and finds that 11 (sample proportion $= 11/400 = 0.0275$) are defective. Does this provide sufficient evidence to indicate that the proportion of defective can openers in the shipment exceeds 2%? Test using $\alpha = 0.05$ (significance level).
Contaminated Soil Environmental Science & Technology (Oct. 1993) reported on a study of contaminate soil in The Netherlands. Seventy-two (n=72) 400-gram soil specimens were sampled, dried, and analyzed for the contaminant cyanide. The cyanide concentration [in milligrams per kilogram (mg/kg) of soil] of each soil specimen was determined using an infrared microscopic method. The sample resulted in a mean cyanide level of \( \bar{x} = 84 \text{mg/kg} \) and a standard deviation of \( s = 80 \text{mg/kg} \).

Test the hypothesis that the true mean cyanide level in soil in The Netherlands exceeds 100 mg/kg. Use \( \alpha = .10 \).

Would you reach the same conclusion as in part a using \( \alpha = .05 \)? Using \( \alpha = .01 \)?