Remember linear regression techniques?

Let's take a few moments and consider the evaluation of the minimization criterion used to obtain the “best fit” line.
The best fit line is a line which minimizes the difference between the estimated and actual values of $y$. 

$y = mx + b$
\( \hat{y}_i \) is the estimate of \( y_i \).

\[ \hat{y}_i = mx_i + b \]

We want to minimize these differences for all \( y_i \).
.. and the best way to do this is to minimize the sum of the squares of these departures. Mathematically the sum of the squares of the departures or differences is

$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$$\sum_{i=1}^{N} (y_i - mx_i - b)^2$$

Let the sum of these differences = D. How can we minimize D?
Remember, when you want to find the minimum of something you compute its derivative (its tangents) and set the derivative equal to 0, i.e., find a tangent to the curve whose slope is zero.

Where is the minimum of the function $y = (x - a)^2$?
Given

\[ D = \sum_{i=1}^{N} (y_i - mx_i - b)^2 \]

there are two ways we could minimize this expression - one with respect to the slope \( m \) - and the other with respect to the intercept \( b \).

We’ll work through a little of this on the board.
Differentiation w.r.t. \( m \) and \( b \) yield

\[
\frac{\partial D}{\partial m} \rightarrow b \sum x_i + m \sum x_i^2 = \sum x_i y_i
\]

\[
\frac{\partial D}{\partial b} \rightarrow nb + m \sum x_i = \sum y_i
\]

The result yields two linear equations in two unknowns that can be combined to evaluate the slope \( m \) and intercept \( b \) of the best fit line.
The end result -

The intercept

\[ b = \bar{y} - mx \]

The slope

\[ m = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n(n-1)s^2} \]

Where \( s^2 \) = variance of \( x \)
It also turns out that

\[ m = \frac{\text{covariance}_{xy}}{\text{variance}_x} \]

Where the covariance between \( x \) and \( y \) is

\[ s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \]

or

\[ m = \frac{s_{xy}}{s_x^2} \]
224 spectral measurements per pixel
Think of the AVIRIS data set as consisting of 224 images of the same area each recorded at different wavelengths extending from the 400 to 2500 nanometer (0.4 to 2.5 micron) range.
The electromagnetic spectrum
Cuprite mining district in Nevada
Spectral characteristics of different minerals serve as fingerprints that can be used to identify the mineral from which light has been reflected.
Color composite photo of the Cuprite Nevada area derived from three Landsat TM bands.

For more information about these remote sensing methods, visit http://speclab.cr.usgs.gov/map.intro.html
Another image derived from a combination of Landsat bands. Mineral distributions are suggested by the color distribution but are not uniquely associated with individual mineral types.
This image is compiled from the hyperspectral AVIRIS data. The different colors are associated with the distribution of specific minerals across the surface.
To determine what mineral a specific spectral signature represents, one must make comparisons to know mineral spectra. See the USGS spectral library at http://speclab.cr.usgs.gov/spectral-lib.html
From the AVIRIS image

From the USGS spectral library.
Some AVIRIS data extracted for a class exercise

Individual pixels defined by a given value of reflectance

Region 1
Region 2
Region 3
Region 4

Band 20 - 0.57 microns
~ TM band 2
Band 100 - 1.28 microns
Band 200 - 2.27 microns

~ TM band 7
<table>
<thead>
<tr>
<th>Band</th>
<th>Wavelength (nm)</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>450 - 520</td>
<td>Blue-green. Maximum penetration of water</td>
</tr>
<tr>
<td>2</td>
<td>520 - 600</td>
<td>Green. Matches green reflection peak of vegetation</td>
</tr>
<tr>
<td>3</td>
<td>630 - 690</td>
<td>Red. Matches a chlorophyll absorption band important to distinguish vegetation types</td>
</tr>
<tr>
<td>4</td>
<td>760 - 900</td>
<td>Reflected IR. Useful for determining biomass content and mapping shorelines</td>
</tr>
<tr>
<td>5</td>
<td>1550 - 1750</td>
<td>Reflected IR. Indicates moisture content of soil and vegetation</td>
</tr>
<tr>
<td>6</td>
<td>10400 - 12500</td>
<td>Thermal IR. Nighttime thermal mapping and soil moisture.</td>
</tr>
<tr>
<td>7</td>
<td>2080 - 2350</td>
<td>Reflected IR. Mineral absorption band associated with hydrothermally altered rocks.</td>
</tr>
</tbody>
</table>

The data sets provided for this exercise include AVIRIS bands 16 and 18. AVIRIS bands 16, 18 and 20 correspond to 517.6nm, 537.33nm and 557.07nm. These bands span Landsat TM Green band 2 (520-600 above). Also included in the data sets available for this exercise is AVIRIS band 52 recorded at 846.28nm. This band lies within Landsat TM band 4. Note also that AVIRIS band 200 recorded at a 2268.4nm wavelength lies within band 7 of the Landsat TM data.
<table>
<thead>
<tr>
<th>Wavelength</th>
<th>Landsat TMBand #</th>
<th>AVIRIS Band #</th>
<th>Data set name</th>
</tr>
</thead>
<tbody>
<tr>
<td>517.6</td>
<td>2</td>
<td>16</td>
<td>16(1 through 4).dat</td>
</tr>
<tr>
<td>537.33</td>
<td>2</td>
<td>18</td>
<td>18(1 through 4).dat</td>
</tr>
<tr>
<td>557.07</td>
<td>2</td>
<td>20</td>
<td>20(1 through 4).dat</td>
</tr>
<tr>
<td>846.28</td>
<td>4</td>
<td>52</td>
<td>52(1 through 4).dat</td>
</tr>
<tr>
<td>1284.2</td>
<td>NA</td>
<td>100</td>
<td>100(1 through 4).dat</td>
</tr>
<tr>
<td>2268.4</td>
<td>7</td>
<td>200</td>
<td>200(1 through 4).dat</td>
</tr>
</tbody>
</table>
Agjourn to computer lab