Lab: Learning, Prediction and Optimization

Last time, we saw that if we think about learning as building an association between stimulus and a desired response, then learning is function building.

Additionally, there is usually some trade-off between accuracy of the function and its ability to generalize- these are usually at odds with each other. For example, the best generalization will be from a function that has a constant (or zero) derivative. This will mean that, if the stimulus is changed slightly, the response will change accordingly (by the constant, or not at all). In this case, the function will be a line through the data.

1. The Line of Best Fit

In this case, we are interested in constructing the equation of a line, $y = mx + b$, that will best approximate the data. We now have to define what we mean by "best".

In what follows, "best" will mean that the line gets as close as possible to all of the $y-$values. That is, we are given a set of points, $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ and the error for any given m and b will be measured as the sum of the squares of the difference between the $y-$values on the line, and the actual $y-$values. Note that, for a given x_i, the $y-$value on the line is $mx_i + b$, and the desired $y-$value is y_i. Thus, the error is given by:

$$\mathrm{E} = (mx_1 + b - y_1)^2 + (mx_2 + b - y_2)^2 + \ldots + (mx_n + b - y_n)^2$$

or, using Σ notation,

$$\mathrm{E} = \sum_{k=1}^{n} (mx_k + b - y_k)^2$$

Note that in this equation, the x_k and y_k are already known. That means that the function `Error` is a function of m and b.

1.1. Pre Lab Questions.

(1) Is there a maximum value for the Error function?
(2) To find the minimum value:
   (a) Compute $\dfrac{\partial E}{\partial m}$ and $\dfrac{\partial E}{\partial b}$.
   (b) Set each of those equal to zero. We're looking for two equations in m and b- try to write your equations in the form:

$$m \cdot A + b \cdot B = C$$
$$m \cdot V + b \cdot W = Z$$

   Note that we can solve this for m and b (by substitution if necessary), and we've solved the problem!

1.2. Line of Best Fit and Maple.

Maple has a built-in method for determining the best fitting line using our previous error measure (the sum of squares error). Here is an example (try it!). The first line defines the data points, the second set of commands constructs the line of best fit, and the third set of commands plots the results.

```
t:=[[-1,0],[2,-1],[3,0],[5,1]];

with(CurveFitting):
g:=LeastSquares(t,x);

with(plots):
A:=pointplot(t,symbol=diamond,symbolsize=18):
B:=plot(g,x=-1..5,color=black):
display({A,B});
```

2. Curves of Best Fit

Suppose our data does not look linear. Suppose, for example, that the data looks like it could be:

$$y = A\mathrm{e}^{kx}$$

Can we use our line of best fit to find A, k?

2.1. Pre Lab Questions, Continued.

(3) Take the natural log of both sides of the model equation. Use the rules of logarithms to show that $\ln(y)$ has a linear relationship with x.

(4) If you put $(x_i, \ln(y_i))$, and Maple gives you the model:

$$g = 3 + 2x$$

then what are the values of A and k?

2.2. Linear Models. We can use other model functions. For example, if I assume that the function that describes the data is of the form:

$$y = ax^2 + bx + c$$

then we can construct the error function, and it will depend only on a, b, c. We can look at the partial derivatives, set them to zero, and get three equations in three unknowns.

Similarly, if we assume that the model equation is of the form $y = c_1 f_1(x) + c_2 f_2(x) + \ldots + c_n f_n(x)$, then the error depends only on $c_1, \ldots, c_n$, and the partial derivatives will give us n linear equations to solve. Fortunately, Maple will give the solution. This is why these are called *linear models*, even though the functions $f_i(x)$ may be highly nonlinear.

Here is a sample, where I am assuming that the model equation is $y = ax^2 + b\cos(x)$. I need to find a, b so that the error is minimized.

```
t:=[[-1,0],[2,-1],[3,0],[5,1]];
with(CurveFitting):
g:=evalf(LeastSquares(t,x,curve=a*x^2+b*cos(x)));
with(plots):
A:=pointplot(t,symbol=diamond,symbolsize=18):
B:=plot(g,x=-1..5,color=black):
display({A,B});
```

2.3. Pre Lab Questions, Continued.

(5) Find the best fitting cubic function, $y = ax^3 + bx^2 + cx + d$ for the given data in the example, and plot the results (as in the example).

3. The Lab:

Find historical data on the population of the United States between the years 1790 and 1990. Type this data into Maple, and find the line, quartic and cubic best fits. Find the exponential function best fit. Use all 4 approximations to estimate the population of the United States for the year 2000, and give an analysis of your results.