CogSci 109: Lecture 9

Tuesday Oct 19, 2006
More basic fits, correlation (goodness of fit), nonlinear fits

Announcements

- Chancellor's challenge tomorrow!
- Midterm - Thursday?

A little about the midterm

- Is Thursday better than Tuesday?
- Review session will be scheduled
- Practice midterm will be posted with answers
- Matlab list of commands will be posted
- Answers to homework 0.2 will be posted, but you won't have to do a lot of calculation on the midterm
- Part mult choice, part short answer
- Comprehension more than computation

Another example in matlab

- consider (1,2) (3,4) (2, 3.5)

```
x=[1 3 2]
y=[2 4 3.5]
plot(x,y,'*')
```

- 1m +b =2
- 3m +b =4

- 2m +b = 3.5

Example continued

```
A=[1 1; 3 1; 2 1]
y= [ 2; 4; 3.5]
```

- if we use the m=1, b=1 solution to the first two it doesn't fit the third
- e.g. 3 equations and 2 unknowns
- This is what is known as an overconstrained problem. People commonly like to find the solution that minimizes the mean square error

Example continued

- This means we want to find the solution that minimizes

  \sum_{(x,y) pairs} (y-mx-b)^2
- Matlab again solves this with

```
mb=A\y
hold on
newA=[0 1; 5 1]
plot([0 5],newA*mb)
```

Correlation coefficient

- <board>
- Matlab function - returns a matrix of coefficients
-

r=corrcoef(x,y)

r=corrcoef(X)

[r,p]=corrcoef(X)

$$r(i,j) = \frac{C(i,j)}{\sqrt{C(i,i)C(j,j)}}$$

C=covariance matrix of X (or x,y)

Correlation as a measure of fit

- Closer to 0 worse fit
- Closer to 1 better fit

But what if the data isn't linear?

- Nonlinear least squares
  - Polynomial fit
  - Exponential fit