Class 26 in-class problems, 18.05, Spring 2022

Board questions

Problem 1. Make it fit
We are given bivariate data: $(1, 3)$, $(2, 1)$, $(4, 4)$.

(a) Do (simple) linear regression to find the best fitting line.

(i) Give the model for simple linear regression.

(ii) Write down the formula for the total squared error.

(iii) Use calculus to find the parameters that minimize the total squared error.

(b) Do linear regression to find the best fitting parabola. (Really just set this up and get as far as needing to solve equations to find the coefficients.)

(c) Find the best fitting exponential $y = \mathrm{e}^{ax+b}$. (As before, set up the equations but don't solve them.)

Hint: take $\ln(y)$ and do simple linear regression.

(d) For data $(x_1, y_1), \ldots, (x_n, y_n)$. Set up the linear regression to find the best fitting cubic. Don't try to take derivatives or actually find the formulas for the coefficients.

Problem 2. Using the formulas plus some theory
Bivariate data: $(1, 3)$, $(2, 1)$, $(4, 4)$

(a) Calculate the sample means for x and y.

(b) Use the formulas to find a best-fit line in the xy-plane.

$$\hat{a} = \frac{s_{xy}}{s_{xx}} \qquad \hat{b} = \overline{y} - \hat{a}\overline{x}$$

$$s_{xy} = \frac{1}{n-1}\sum(x_i - \overline{x})(y_i - \overline{y}) \qquad s_{xx} = \frac{1}{n-1}\sum(x_i - \overline{x})^2.$$

(c) Show the point $(\overline{x}, \overline{y})$ is always on the fitted line.

(d) (For fun later!) Under the assumption $E_i \sim \mathrm{N}(0, \sigma^2)$ show that the least squares method is equivalent to finding the MLE for the parameters (a, b).

Hint: $f(y_i \mid x_i, a, b) \sim \mathrm{N}(ax_i + b, \sigma^2)$.