Math 1512: Honors Calculus II Fall 2010

Homework for the lecture on Oct 28

1. Find the least squares regression line for the data points

$$(0,10),(2,6),(3,7),(4,6),(5,3),(8,1).$$

2. Let

$$A=\begin{pmatrix}2&1\\4&2\\1&1\end{pmatrix},\ b=\begin{bmatrix}1\\0\\0\end{bmatrix}.$$

   (a) Find the orthogonal projection of b onto the range of A.

   (b) Find the best approximate (least squares) solution to the overdetermined system of equations $Ax=b$.

3. The *mean* $\overline{x}$ and *standard deviation* σ_x of a collection $x_1,\ldots,x_n$ of real numbers are given by the formulas

$$\overline{x}=\frac{\sum_{i=1}^n x_i}{n},\ \sigma=\sqrt{\frac{\sum_{i=1}^n(x_i-\overline{x})^2}{n}}.$$

The *Pearson correlation coefficient* of a collection of $n\geq 2$ distinct data points $(x_1,y_1),\ldots,(x_n,y_n)\in\mathbf{R}^2$ is defined to be

$$r=\frac{\sum_{i=1}^n(x_i-\overline{x})(y_i-\overline{y})}{n\sigma_x\sigma_y},$$

where σ_x (resp. σ_y) denotes the standard deviation of $x_1,\ldots,x_n$ (resp. $y_1,\ldots,y_n$) and $\overline{x}$ (resp. $\overline{y}$) denotes the mean of $x_1,\ldots,x_n$ (resp. $y_1,\ldots,y_n$). (The quantity $\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})(y_i-\overline{y})$ is called the *covariance*, so the Pearson correlation coefficient is just the covariance divided by the product of the standard deviations.)

   (a) Prove that $-1\leq r\leq 1$, and that the data points lie on a straight line if and only if $r=\pm 1$. [**Hint:** Use the Cauchy-Schwartz inequality.]

   (b) Calculate the Pearson correlation coefficient for the data points in Problem 1.