HW: Mean and Variance of a Projection

We will suppose we have p points in $\mathbb{R}^2$, and store the points in a $p \times 2$ matrix A. We'll also recall that the formula for the projection of $\mathbf{x}$ onto $\mathbf{u}$

$$\mathrm{Proj}_u(\mathbf{x}) = \frac{\mathbf{x} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$$

And, if $\|\mathbf{u}\|1$, then we define the **scalar projection** of $\mathbf{x}$ onto $\mathbf{u}$ as: $\mathbf{x} \cdot \mathbf{u}$.

- If we project the data to the $x-$axis, the data of interest is the scalar projection, or in this case, just the values in the first column. We can then compute the mean and variance of that, as usual.
- Similarly, if we project the data to the $y-$axis, the data of interest is the scalar projection, or in this case, just the data in the second column of A. Again, we can then compute the mean and variance of that data.
- Now suppose we project the data to the vector $\mathbf{u}$. We will look at the mean and variance of the *scalar projection*, $\mathbf{x} \cdot \mathbf{u}$.

For example, suppose we have three points:

x	1	2	−1
y	0	1	1

$$\mathbf{u} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Then the new scalar values from the projection are:

$$\frac{1+0}{\sqrt{2}}, \qquad \frac{2+1}{\sqrt{2}}, \qquad \frac{-1+1}{\sqrt{2}}$$

from which we can compute a mean and variance. Note that in matrix form, we could have written:

$$A\mathbf{u} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$$

We want to investigate the relationship of the mean and variance to the projection. To do that, let's set our notation to be the following (assume $\|\mathbf{u}\| = 1$):

$$A = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_p & y_p \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad A\mathbf{u} = \begin{bmatrix} x_1u_1 + y_1u_2 \\ x_2u_1 + y_2u_2 \\ x_3u_1 + y_3u_2 \\ \vdots \\ x_pu_1 + y_pu_2 \end{bmatrix}$$

Further, define the mean of the first column of A be denoted by $\bar{x}$, and mean of the second by $\bar{y}$.

Exercises

1. Show that the mean of the (scalar) projection is the (scalar) projection of the mean.

2. In linear algebra, if $\mathbf{b}$ is a column vector, then what operation(s) should I perform in order to sum the squared elements of $\mathbf{b}$?

3. If the first and second columns of A have been mean-subtracted, show that the variance of the projection is given by
$$\mathbf{u}^T(A^TA)\mathbf{u}$$

4. Show that you're correct by writing a short script in Maple using 100 randomly placed points in the plane, and the same vector $\mathbf{u}$ as in our example above. That is, compute the mean and variance of the projection by first performing the projection, then show that your answers correspond to the computations in exercises 1 and 3 above.